Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk
w Poznaniu

mgr inż. Małgorzata Marcinkowska-Swojak

„Opracowanie i zastosowanie nowej metody do genotypowania powszechnego polimorfizmu liczby kopii w genomie człowieka”

Praca doktorska
wykonana pod kierunkiem
dr hab. Piotra Kozłowskiego, prof. IChB PAN
w Instytucie Chemii Bioorganicznej PAN w Poznaniu

Poznań 2013

Niniejsza praca doktorska była w całości finansowana z grantu
Ministerstwa Nauki i Szkolnictwa Wyższego Nr N N302-278937.

W trakcie realizacji pracy doktorskiej, autorka dwukrotnie była stypendystką w ramach projektu pt. „Wsparcie stypendialne dla doktorantów na kierunkach uznanych za strategiczne z punktu widzenia rozwoju Wielkopolski”, Poddziałanie 8.2.2 Programu Operacyjnego Kapitał Ludzki finansowanego ze środków Europejskiego Funduszu Społecznego.

Serdecznie dziękuję mojemu Promotorowi,
Panu dr hab. Piotrowi Kozłowskiemu, prof. IChB PAN
za wprowadzenie w interesującą tematykę,
cenne wskazówki i dyskusje w trakcie realizacji pracy
oraz za wsparcie i wyrozumiałość.

Pragnę również podziękować wszystkim Pracownikom
Zakładu Biomedycyny Molekularnej IChB PAN oraz
Europejskiego Centrum Bioinformatyki i Genomiki
za życzliwość i miłą atmosferę pracy.

Dziękuję również mojej Rodzinie i Przyjaciołom
za nieustannie okazywane wsparcie.

Niniejsza rozprawa doktorska składa się z następujących części:

STRESZCZENIE

Tytuł: „Opracowanie i zastosowanie nowej metody do genotypowania powszechnego polimorfizmu liczby kopii w genomie człowieka”

Zmienność liczby kopii w genomie człowieka jest w ostatnich latach intensywnie badanym zjawiskiem. Warianty liczby kopii (CNV) definiowane są jako segmenty DNA (około 1kpz-1Mpz długości), które wykazują zmienną liczbę kopii w porównywanych genomach. CNV przyjmują formę delecji, duplikacji, wielokrotnych duplikacji lub bardziej złożonych rearanżacji. Powszechne CNV obejmują około 10% ludzkiego genomu, zawierając setki genów, sekwencji regulatorowych i innych funkcjonalnych elementów genomu. Chociaż większość CNV ma prawdopodobnie neutralny charakter, odkrywanych jest coraz więcej CNV wpływających na ludzki fenotyp, w tym zdrowie człowieka. Dotychczas opracowano wiele metod służących do identyfikacji i analizy CNV zarówno w skali całego genomu, jak i pojedynczych CNV, jednakże wciąż istnieje potrzeba opracowania precyzyjnej i niedrogiej metody, pozwalającej na jednoznaczne genotypowanie wybranych CNV w dużej liczbie próbek.

Z tego względu opracowaliśmy nową metodę genotypowania CNV, która wykorzystuje podstawowe założenia metody zależnej od ligacji multipleksowej amplifikacji sond (MLPA). Jednak, w porównaniu z oryginalną wersją metody MLPA, w której wykorzystuje się długie sondy generowane w specjalnie przygotowanych wektorach, nasza strategia wykorzystuje krótkie oligonukleotydowe sondy, które można otrzymać na drodze chemicznej syntezy. Pozwala to zaprojektować sondy i przygotować testy MLPA dla praktycznie dowolnie wybranego miejsca w genomie.

Modelem badawczym dla opracowanej metody były regiony CNV, obejmujące geny ludzkich mikroRNA (CNV-miRNA), które zidentyfikowaliśmy i scharakteryzowaliśmy z wykorzystaniem narzędzi bioinformatycznych. Dla wybranych CNV-miRNA zaprojektowaliśmy testy MLPA. Opracowane testy pozwoliły eksperymentalnie zidentyfikować oraz scharakteryzować wybrane CNV-miRNA pod kątem zmienności liczby kopii w trzech populacjach ludzkich. Przeprowadzona analiza jakości wykazała dużą powtarzalność i rzetelność genotypów przypisanych z wykorzystaniem opracowanej metody.

Proponowaną metodę wykorzystaliśmy również do analiz wielo-allelicznych CNV związanych z powszechnymi chorobami człowieka, a także do połączonej analizy zmienności liczby kopii i małych mutacji w genie *EGFR*.

Zaproponowana przez nas metoda genotypowania CNV obejmuje projektowanie i generowanie sond oraz testów MLPA, optymalizację i wykonanie reakcji MLPA, a także analizę i interpretację uzyskanych wyników. Metoda ta pozwala na opracowanie testów do analizy dowolnie wybranego regionu w genomie oraz na genotypowanie zarówno prostych, jak i złożonych CNV. Relatywnie niski koszt sprawia, że metoda ta jest atrakcyjna do genotypowania poszczególnych CNV w dużej liczbie próbek, często wymaganej w badaniach genetycznych.

SUMMARY

Title: „The development and applications of the new method for genotyping of common copy number polymorphism in the human genome”

Copy number variation in the human genome has become well recognized in recent years. Copy number variants (CNVs) are genomic regions (roughly 1kb-1Mb in length) that show variable number of copies in compared genomes. CNVs include deletions, duplications multiple duplications or more complex rearrangements. Common CNVs account for approximately 10% of human genome, overlapping hundreds of genes, regulatory sequences, and other functional genetic elements. Although the majority of CNVs are probably neutral, increasing numbers of CNVs are being associated with various human phenotypes, including diseases. Several methods, both genome-wide and locus-specific, have been developed for CNVs identification and analysis. However, there is still a need for inexpensive method allowing discrete (with integer resolution) genotyping of selected CNVs in large number of samples.

We have developed a new method for CNV genotyping that takes advantage of the general principles of the multiplex ligation-dependent probe amplification method (MLPA). However, in comparison to standard MLPA, instead of long MLPA probes generated in special vectors, our strategy uses short oligonucleotide probes which can be generated through chemical synthesis. It allows easy custom design and generation of assays for almost any genomic region of interest.

As a model for testing our method, we employed the CNV regions overlapping with miRNA genes (CNV-miRNAs). All CNV-miRNAs in human genome were identified and validated with the use of computational analysis of different genomic data. For selected CNV-miRNAs we developed MLPA assays. With the use of developed assays, we experimentally identified 8 CNV-miRNAs which copy number polymorphism was characterized in three distinct human populations. Extensive quality control analysis demonstrated high reproducibility and reliability of the genotypes determined with the use of our method.

We have successfully used our method also for the analyses of multi-allelic CNVs involved in common human diseases and for parallel copy number and small mutation analysis in *EGFR* gene.

The proposed strategy includes the design and generation of MLPA probes and assays, optimization and implementation of MLPA reactions and the analysis and interpretation of the obtained results. The strategy allows assays designing for almost any genomic region of interest and discrete genotyping of both bi-allelic and multi-allelic CNVs. The relatively low per-genotype cost makes this method attractive for the genotyping of individual CNV in large number of samples, allowing it to be applied in genotype-phenotype association studies.

OPIS PUBLIKACJI ZAWARTYCH

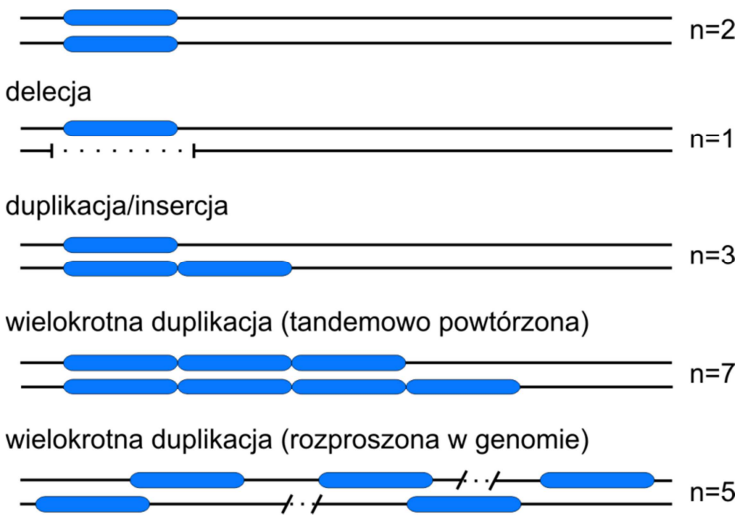
W ROZPRAWIE DOKTORSKIEJ

„Opracowanie i zastosowanie nowej metody
do genotypowania powszechnego polimorfizmu liczby kopii
w genomie człowieka”

Genom człowieka, którego sekwencję w zasadniczej części poznano w 2001 roku, obejmuje blisko 3 miliardy par zasad (Lander i wsp. 2001; Venter i wsp. 2001). Ich charakterystyczny układ stanowi informację genetyczną wspólną dla genomów wszystkich ludzi. Mimo to, porównanie genomów reprezentujących różne ludzkie populacje, jak również bezpośrednie porównanie genomów nawet blisko spokrewnionych osób, ujawnia istnienie szeregu różnic. Różnice te zwane są polimorfizmem genetycznym, który w znacznym stopniu odpowiada za zróżnicowanie w obrębie naszej populacji. Polimorfizm genetyczny może modyfikować większość cech fenotypowych, takich jak wygląd zewnętrzny czy poziom markerów biochemicznych. Polimorfizm może również wpływać na stan zdrowia człowieka, determinując występowanie chorób, modyfikując ich ryzyko, zróżnicowanie objawów, przebieg oraz reakcje na stosowane terapie.

Do niedawna sądzono, że główną przyczyną genetycznej zmienności w ludzkiej populacji są małe polimorfizmy, obejmujące od jednego do kilku nukleotydów. Wśród nich występują niewielkie insercje, delecje lub inwersje, jednak najpowszechniejszą formą takiego polimorfizmu są substytucje pojedynczych nukleotydów, SNP (ang. *single nucleotide polymorphism*). Szacuje się, że w genomie człowieka występuje około 10 milionów SNP o częstości >5% (Frazer i wsp. 2007). Ze względu na powszechność SNP podjęto wiele projektów, mających na celu zarówno dokładne scharakteryzowanie tego polimorfizmu w genomie człowieka (np. International HapMap Project czy 1000 Genomes Project), jak również identyfikację jego związku z różnymi, powszechnie występującymi chorobami lub ich fenotypami składowymi. W wyniku badań asocjacji zidentyfikowano setki SNP, z których część związana jest z takimi chorobami jak: cukrzyca (Doria i wsp. 2008), astma (Weiss i wsp. 2004), choroby krążenia (Romeo i wsp. 2007), czy rak płuc, piersi i prostaty (Easton i wsp. 2007; Wang i wsp. 2008; Gudmundsson i wsp. 2009).

Innym typem polimorfizmu genetycznego są duże zmiany strukturalne, określane mianem zmienności liczby kopii (ang. *copy number variation*). Chociaż ten rodzaj zmienności genetycznej znany był już od dawna, głównie jako mutacje uszkadzające geny związane z chorobami człowieka, w ostatnim czasie wykazano, że zmienność liczby kopii występuje również w formie powszechnych polimorfizmów (Iafrate i wsp. 2004; Sebat i wsp. 2004). Poszczególne regiony genomu, w których występuje zmienność liczby kopii określane są mianem CNV (ang. *copy number variant*) lub analogicznie do SNP, CNP (ang. *copy number polymorphism*). CNV definiowane są jako segmenty DNA o wielkości od 1kpz do nawet kilku Mpz, w których zaobserwowano relatywne zwiększenie (duplikacje/amplifikacje) lub zmniejszenie (delecje) liczby kopii w porównywanych genomach (Rycina 1).



Rycina 1. *Najczęściej występujące typy polimorfizmu CNV. Niebieski element reprezentuje polimorficzny region o zmiennej liczbie kopii. Z prawej strony podana jest obserwowana liczba kopii danego regionu.*

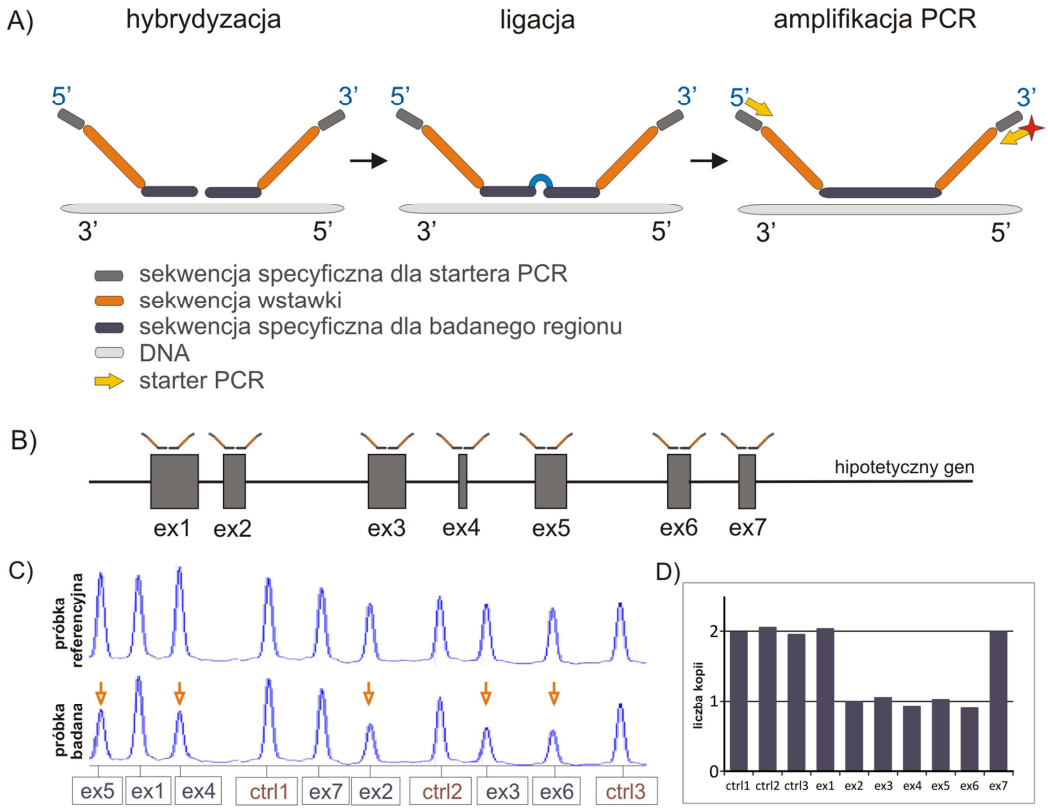
Dotychczas, dzięki zastosowaniu takich metod jak: porównawcza hybrydyzacja genomowa do macierzy (aCGH) (Conrad i wsp. 2010), mikromacierze SNP (Redon i wsp. 2006), analiza błędów dziedziczenia markerów SNP (McCarroll i wsp. 2006), czy wprowadzona w ostatnich latach technologia masowego sekwencjonowania (Conrad i wsp. 2010), w genomie człowieka zidentyfikowano tysiące CNV. Szacuje się, że częste (>1%) CNV stanowią około 10% ludzkiego genomu, obejmując setki ważnych funkcjonalnie elementów genomu, m.in. geny kodujące białka, czy sekwencje regulatorowe. CNV, które obejmują geny, mogą, choć nie muszą, zmieniać liczbę funkcjonalnych kopii tych genów, a tym samym wpływać na ich ekspresję, wyrażoną zarówno jako ilość powstającego transkryptu, jak również ilość funkcjonalnego białka (tzw. efekt dawki). Takie CNV mogą znacząco modyfikować ludzki fenotyp, w tym wpływać na ryzyko występowania lub przebieg różnych chorób. Wśród przykładów wpływu zmienności liczby kopii na fenotyp człowieka na uwagę zasługują: CNV genu *AMY1* wpływający na wydajność hydrolizy skrobi (Perry i wsp. 2007), CNV genu *UGT2B17*, który związany jest z występowaniem osteoporozy (Yang i wsp. 2008), CNV genu *CCL3L1*, który wpływa na podatność na infekcje wirusem HIV (Gonzalez i wsp. 2005), CNV genów beta-defensyn, który modyfikuje ryzyko wystąpienia łuszczycy (Hollox i wsp. 2008) oraz CNV genu *CYP2D6*, który wpływa na szybkość metabolizmu leków (Ingelman-Sundberg 2005). Mimo identyfikacji licznych związków CNV z fenotypem człowieka, badania tego typu zmienności genetycznej są wciąż znacznie utrudnione przez brak odpowiednich metod, umożliwiających jednoznaczne i precyzyjne określenie liczby kopii (genotypowanie) poszczególnych CNV.

Dotychczas do genotypowania CNV stosowano różne metody molekularne (opisane niedawno w (Cantsilieris i wsp. 2012)), m.in. FISH (ang. *fluorescence in situ hybridization*), hybrydyzację Southerna (ang. *Southern blotting*), qPCR (ang. *quantitative polymerase chain reaction*), PRT (ang. *paralog ratio test*), MLPA (ang. *multiplex ligation-dependent probe amplification*) oraz MAPH (ang. *multiplex amplification and probe hybridization*). Z wymienionych metod najpopularniejszą i najczęściej stosowaną do genotypowania CNV jest qPCR. Metoda ta jednak w większości przypadków nie pozwala na jednoznaczne określenie faktycznej liczby kopii danego CNV w badanej próbce. Zamiast tego bezwzględna wartość relatywnego sygnału qPCR używana jest jako odpowiednik (ang. *proxy*) liczby kopii (Hosono i wsp. 2009). Takie podejście znacząco utrudnia analizy CNV (m.in. wnioskowanie o allelach, czy analiza dziedziczenia mendlowskiego i nierównowagi sprzężeń) oraz obniża siłę statystyczną analiz asocjacji CNV (Fernandez-Jimenez i wsp. 2011; Fode i wsp. 2011). Inną metodą, często stosowaną do genotypowania CNV, jest wspomniany już PRT. Metoda ta polega na porównaniu intensywności sygnałów równolegle amplifikowanych regionów CNV oraz niepolimorficznych paralogów tych regionów (Armour i wsp. 2007). Chociaż PRT umożliwia jednoznaczne określenie liczby kopii danego CNV, testy PRT można zastosować jedynie dla nielicznych CNV, zawierających odpowiednie sekwencje paralogów. Ograniczenia obecnie stosowanych metod oraz potrzeba opracowania bardziej precyzyjnej metody do genotypowania CNV, były wielokrotnie podkreślane w literaturze (McCarroll i Altshuler 2007; Cantsilieris i wsp. 2012).

W związku z powyższym, w ramach niniejszej pracy doktorskiej, podjęta została próba opracowania uniwersalnej, precyzyjnej i stosunkowo niedrogiej metody genotypowania CNV w dużej liczbie próbek. Cel ten realizowany był w następujących etapach: (i) wybór modelu badawczego dla prowadzonych badań, (ii) opracowanie i optymalizacja metody genotypowania CNV oraz (iii) wykorzystanie opracowanej metody do genotypowania CNV obejmujących geny mikroRNA, jak również innych CNV oraz mutacji w genomie człowieka.

Proponowana przez nas metoda do genotypowania CNV w genomie człowieka opiera się na wspomnianej wyżej metodzie zależnej od ligacji multipleksowej amplifikacji sond (MLPA). Metoda MLPA, opisana po raz pierwszy w 2002 roku (Schouten i wsp. 2002), oryginalnie została opracowana i jest z powodzeniem stosowana do detekcji dużych mutacji. Z wykorzystaniem tej metody wykrytych zostało tysiące mutacji w licznych genach związanych z chorobami człowieka (Schouten i wsp. 2002; Aretz i wsp. 2005; Bunyan i wsp. 2007; Kozlowski i wsp. 2007; De Luca i wsp. 2007). W skrócie, MLPA jest multipleksową

metodą, wykorzystującą do 45 sond specyficznych dla różnych miejsc w genomie (Rycina 2). Każda sonda MLPA składa się z dwóch pół-sond, które hybrydyzują do ściśle przylegających do siebie sekwencji docelowych. Tylko pół-sondy, które prawidłowo rozpoznają sekwencję docelową, podlegają w kolejnych etapach ligacji i amplifikacji w multipleksowej reakcji PCR. Następnie produkt PCR rozdzielany jest przy pomocy elektroforezy kapilarnej (ang. *capillary electrophoresis*, CE). Wynikiem takiego rozdziału jest specyficzny układ pików, reprezentujących poszczególne sondy, których intensywność proporcjonalna jest do liczby kopii sekwencji docelowej występującej w genomie.



Rycina 2. *Schemat metody MLPA i analizy wyników. A) Kolejne etapy reakcji MLPA. Poszczególne sekwencje stanowiące sondy MLPA zostały zaznaczone odpowiednimi kolorami. B) Mapa hipotetycznego genu z zaznaczonymi eksonami i pozycjami sond MLPA. C) Przykładowe elektroforegramy próbki referencyjnej i badanej. Obniżone sygnały zaznaczono pomarańczową strzałką. D) Wykres słupkowy przedstawia stosunek intensywności sygnału poszczególnych sond w próbce badanej i referencyjnej. Przedstawiony przykład reprezentuje heterozygotyczną delecję pięciu kolejnych eksonów (2-6) (na podstawie Marcinkowska i wsp. 2010).*

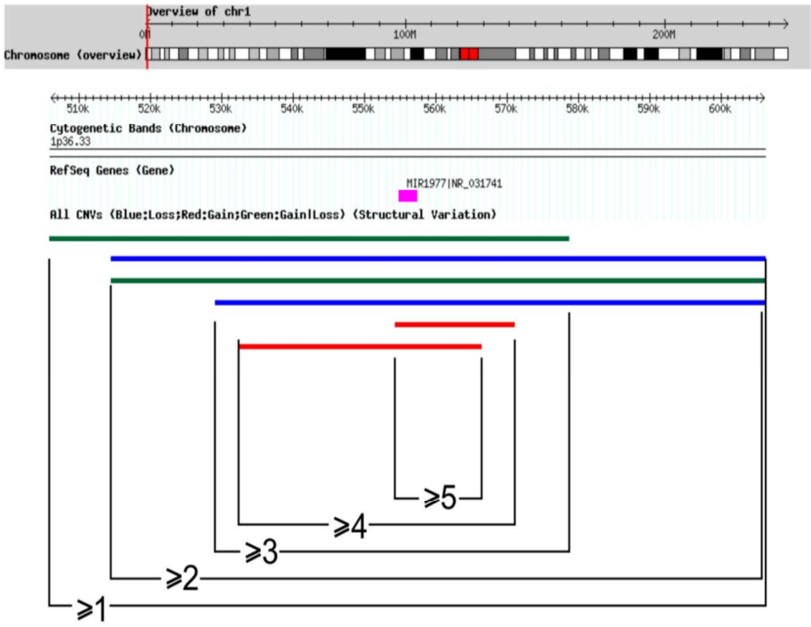
Zasadniczym ograniczeniem oryginalnej wersji MLPA jest złożony, a tym samym czasochłonny i kosztochłonny proces generowania długich sond MLPA w specjalnie przygotowanych wektorach M13. W praktyce ogranicza to zastosowania tej metody wyłącznie do genów, dla których dostępne są gotowe, komercyjne zestawy (firma MRC-Holland). Zastosowana w proponowanej metodzie strategia generowania sond MLPA,

pozwala obejść kłopotliwe stosowanie długich pół-sond, poprzez wykorzystanie krótkich oligonukleotydowych pół-sond, które w łatwy sposób można otrzymać na drodze chemicznej syntezy. Ogólny zarys strategii projektowania i generowania krótkich pół-sond został opracowany już wcześniej (Kozlowski i wsp. 2007). Strategia ta umożliwia zaprojektowanie sond MLPA do dowolnie wybranego miejsca w genomie, co znacznie poszerza zastosowania tej metody.

Poniżej przedstawiam skrótowe omówienie publikacji, które stanowią wynik uzyskany w trakcie realizacji niniejszej pracy doktorskiej. Dla odróżnienia referencje odnoszące się do tych publikacji zostały podkreślone.

Modelem badawczym dla analiz wykonywanych w ramach opracowywania nowej metody do genotypowania CNV były regiony CNV, które obejmowały geny ludzkich mikroRNA. Dla uproszczenia nazwaliśmy je CNV-miRNA i jako takie zaczynają funkcjonować w literaturze (Wu i wsp. 2012; Vaishnavi i wsp. 2013). CNV-miRNA zostały zidentyfikowane z wykorzystaniem narzędzi bioinformatycznych (Marcinkowska i wsp. 2011), na podstawie porównania genomowej lokalizacji genów miRNA, zdeponowanych w bazie miRBase (www.mirbase.org) z genomową lokalizacją regionów CNV z grup: (i) CNV zdeponowanych w Database of Genomic Variants (DGV, http://projects.tcag.ca/variation) oraz (ii) wysoko-polimorficznych CNV zidentyfikowanych w dwóch niezależnych pracach (McCarroll i wsp. 2008; Conrad i wsp. 2010), wykorzystujących mikromacierze dedykowane wykrywaniu CNV (Rycina 3). W toku analizy bioinformatycznej zidentyfikowaliśmy 221 CNV-miRNA, obejmujących delecje, duplikacje i wielokrotne duplikacje. Zidentyfikowaliśmy również 38 miRNA leżących w regionach chromosomowych zaangażowanych w mikrodelecyjne/mikroduplikacyjne syndromy, m.in. w syndrom DiGeorge'a (DECYPHER v5.0). Zidentyfikowane CNV-miRNA scharakteryzowaliśmy pod względem szeregu parametrów opisujących ich polimorficzność, a także pod kątem konserwatywności oraz ekspresji miRNA. Analiza współwystępowania genów miRNA i CNV wykazała, iż geny miRNA rzadziej występowały w regionach objętych przez wysoko-polimorficzne CNV niż by to wynikało z ich losowego rozkładu. Sugeruje to, że CNV podlegają negatywnej selekcji w regionach występowania genów miRNA, co potwierdza zachowawczy charakter tych ostatnich. Zależność tę potwierdziliśmy również poprzez analizę częstości SNP, która w sekwencjach prekursorów miRNA była istotnie niższa (3,7 SNP/1000pz) niż w całym genomie (4,8 SNP/1000pz). W toku tej analizy

zidentyfikowaliśmy 229 SNP zlokalizowanych w sekwencjach ludzkich pre-miRNA. Na podstawie przeprowadzonych analiz zaproponowaliśmy także kilka mechanizmów, w jaki sposób CNV może wpływać na funkcje genów miRNA, w tym na poziom funkcjonalnych kopii sekwencji kodujących pre-miRNA oraz poziom ekspresji miRNA. Jako, że sekwencje pre-miRNA są krótkie i niepodzielone na eksony, mechanizmy wpływu CNV na funkcje tych genów mogą być odmienne niż te dla genów kodujących białka.



Rycina 3. *Identyfikacja genów miRNA objętych zmiennością liczby kopii. Zrzut z ekranu z bazy Database of Genomic Variants (DGV) przedstawia mapę fragmentu chromosomu 1, na którym znajduje się jeden ze zidentyfikowanych CNV-miRNA. W panelu „RefSeq Genes" zaznaczona jest lokalizacja genu mir-1977. Panel „All CNVs" przedstawia różne CNV występujące w tym regionie (delecje, insercje lub bardziej złożone rearanżacje zaznaczono odpowiednimi kolorami). Jako czynnik weryfikujący polimorfizm poszczególnych genów miRNA, przyjmowaliśmy między innymi liczbę CNV zdeponowanych w DGV, obejmujących dany region. Jako minimalne regiony polimorficzne przyjmowaliśmy regiony genomu objęte przez co najmniej 5 CNV zgłoszonych do DGV przez różnych autorów.*

Spośród CNV-miRNA, które na podstawie przeprowadzonych analiz bioinformatycznych zaklasyfikowaliśmy do grupy o najlepiej udokumentowanym polimorfizmie liczby kopii (ang. *top-validated*) (Marcinkowska i wsp. 2011), wybraliśmy 17, które poddaliśmy analizie eksperymentalnej. Te CNV-miRNA posłużyły nam do opracowania metody do genotypowania CNV. Wybrane CNV-miRNA reprezentowały zarówno unikatowe regiony genomu, jak również regiony segmentowo zduplikowane. Dla każdego z wybranych CNV-miRNA zaprojektowaliśmy po dwie sondy MLPA, dostosowując ich sekwencję docelową do typu obejmowanego regionu. Z wykorzystaniem zaprojektowanych sond, opracowaliśmy dwa multipleksowe testy MLPA, które obok sond testujących poszczególne

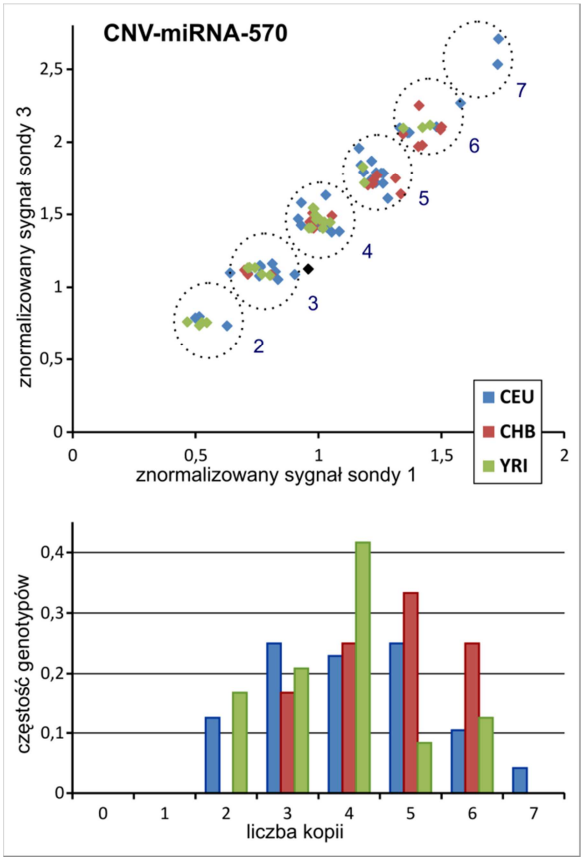
CNV-miRNA, zawierały również pięć sond kontrolnych. Opracowane testy wykorzystaliśmy do wykonania reakcji MLPA na próbkach HapMap pochodzących z trzech populacji ludzkich: europejskiej, azjatyckiej i afrykańskiej (Marcinkowska-Swojak i wsp. 2013).

W analizie wyników MLPA wykorzystaliśmy odmienną od standardowej procedurę przypisywania liczby kopii poszczególnym CNV w poszczególnych próbkach. Zwykle, po znormalizowaniu sygnałów pochodzących z sond badanych względem średniego sygnału sond kontrolnych, sygnał próbki badanej porównywany jest do sygnału pochodzącego z próbki/próbek referencyjnych o znanym genotypie liczby kopii (Rycina 2). Takie podejście jest niepraktyczne w przypadku multipleksowego genotypowania polimorficznych CNV, gdyż znalezienie odpowiedniej próbki referencyjnej, bez wcześniejszej wiedzy o posiadanej przez nią kombinacji genotypów, jest praktycznie niemożliwe. Z tego względu zaproponowaliśmy alternatywny system przypisywania genotypów, w którym znormalizowany sygnał dwóch sond testujących dany CNV prezentowany jest na wykresie dwuwymiarowym (Marcinkowska-Swojak i wsp. 2013). Ponieważ sygnał MLPA jest proporcjonalny do liczby kopii, sygnały pochodzące z wielu próbek tworzą na wykresie wyraźnie oddzielone grupy, odpowiadające poszczególnym genotypom liczby kopii.

Genotypowanie z zastosowaniem opisanej wyżej metody pozwoliło na jednoznaczne przypisanie genotypów analizowanym CNV-miRNA w badanych próbkach oraz na potwierdzenie zmienności liczby kopii w ośmiu z 17 analizowanych regionów. Trzy CNV-miRNA sklasyfikowaliśmy jako dwu-alleliczne, zaś pozostałych pięć jako wielo-alleliczne CNV (Rycina 4). Dla większości polimorficznych CNV-miRNA rozkład genotypów oraz częstość alleli różniły się znacząco pomiędzy badanymi populacjami. Może to świadczyć o tym, iż są to polimorfizmy funkcjonalne, podlegające zróżnicowanej presji selekcyjnej w różnych populacjach. W czasie przeprowadzanych badań zidentyfikowaliśmy również wcześniej nie notowaną w bazach danych insercję typu AluY, która znajdowała się w obrębie sekwencji docelowej jednej z zaprojektowanych sond (Marcinkowska-Swojak i wsp. 2013).

Przeprowadzona analiza CNV-miRNA pozwoliła na eksperymentalną identyfikację ośmiu polimorficznych genów miRNA, których liczba kopii w analizowanych próbkach wahała się od 0 do 9. Dostępna literatura wskazuje, iż większość z tych polimorficznych miRNA zaangażowana jest w regulację genów i procesów związanych m.in. z nowotworami (Wulfken i wsp. 2011; Wang i wsp. 2012), metabolizmem leków (Tili i wsp. 2010), czy apoptozą (Sudbery i wsp. 2010). Ciekawym przykładem tych miRNA jest hsa-mir-383, którego obniżenie ekspresji obserwowano w azoospermii, prowadzącej do męskiej

niepłodności (Lian i wsp. 2010; Lian i wsp. 2009). Obniżona ekspresja tego miRNA może wynikać, przynajmniej częściowo, z wykrytego przez nas polimorfizmu.



Rycina 4. *Przykładowy wynik genotypowania wielo-allelicznego CNV-miRNA uzyskany z wykorzystaniem opracowanej metody. Górny panel przedstawia wykres dwuwymiarowy prezentujący znormalizowany sygnał dwóch sond zaprojektowanych dla CNV-miRNA-570. Sygnały pochodzące od poszczególnych próbek grupują się w klastry, odpowiadające kolejnym genotypom liczby kopii (2-7). Kolorami zaznaczono próbki pochodzące z trzech populacji ludzkich: europejskiej (CEU), azjatyckiej (CHB) i afrykańskiej (YRI). Dolny panel przedstawia częstość poszczególnych genotypów liczby kopii CNV-miRNA-570 w badanych populacjach.*

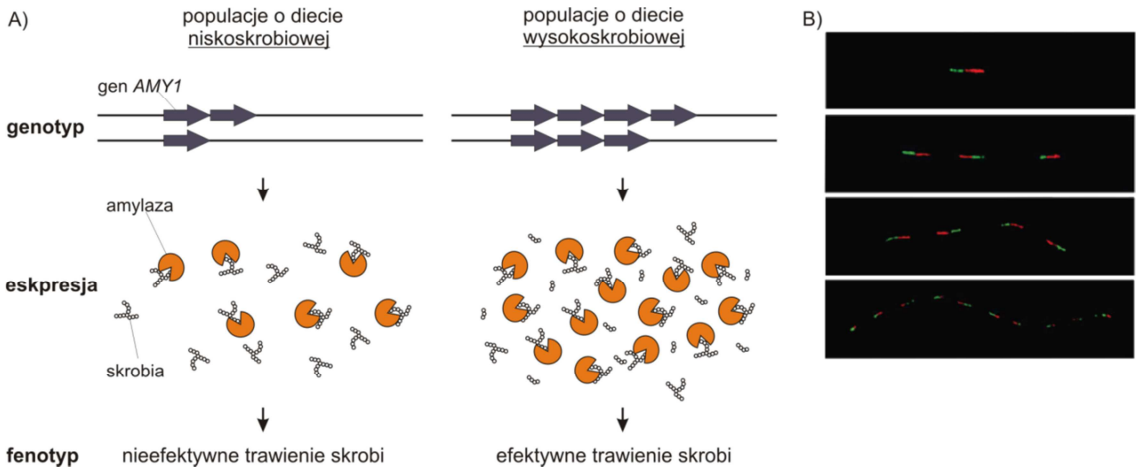
Ponieważ proponowana przez nas metoda jest nowa, a wśród dostępnych metod brak „złotego standardu", do którego moglibyśmy uzyskane wyniki porównać, przeprowadziliśmy bardzo restrykcyjną analizę walidacyjną, stosując szereg technicznych, statystycznych, bioinformatycznych i genetycznych kryteriów (Marcinkowska-Swojak i wsp. 2013). W trakcie weryfikacji wyników obserwowaliśmy wysoką korelację sygnałów pochodzących od sond zaprojektowanych dla danego CNV-miRNA, bardzo wysoką powtarzalność między poszczególnymi eksperymentami, zgodność naszych wyników z wynikami uzyskanymi we wcześniejszych badaniach (McCarroll i wsp. 2008; Conrad i wsp. 2010), a także zgodność przypisanych przez nas genotypów z prawami dziedziczenia Mendla i prawem Hardy-Weinberga. Rezultaty przeprowadzonej weryfikacji wskazują na dużą powtarzalność oraz wiarygodność i rzetelność uzyskanych przez nas wyników.

Skuteczność zaproponowanej metody została potwierdzona również przez zastosowanie jej do analizy wielo-allelicznych CNV, związanych z powszechnie występującymi chorobami człowieka: (i) CNV genu *UGT2B17*, którego wysoka liczba kopii predysponuje do występowania osteoporozy (Yang i wsp. 2008), (ii) CNV genu *CCL3L1*, którego wysoka liczba kopii ma działanie ochronne przeciwko zakażeniom wirusem HIV (Gonzalez i wsp. 2005) oraz (iii) CNV obejmujący grupę genów beta-defensyn, których wysoka liczba kopii związana jest ze zwiększonym ryzykiem wystąpienia łuszczycy (Hollox i wsp. 2008). Przeprowadzone testy pozwoliły praktycznie bezbłędnie określić genotypy wyżej wymienionych wariantów w analizowanych próbkach z trzech populacji. Wyniki analiz zostały opisane w publikacji (Marcinkowska-Swojak i wsp., *under review*), która nie wchodzi w skład niniejszej rozprawy doktorskiej.

Zdobyte doświadczenie w projektowaniu sond posłużyło nam również do przygotowania szczegółowego protokołu, opisującego kolejne kroki metody genotypowania z wykorzystaniem krótkich sond MLPA (Marcinkowska i wsp. 2010). Strategia ta obejmuje: (i) wybór odpowiednich sekwencji docelowych, (ii) projektowanie i generowanie sond oraz testów MLPA, (iii) wykonanie reakcji MLPA oraz (iv) analizę i interpretację wyników. Protokół ten przedstawiliśmy na przykładzie testu do analizy genu *EGFR* w próbkach pochodzących z nowotworów. Opracowany test umożliwiał jednoczesną analizę amplifikacji genu *EGFR*, powszechnie występujących w różnych typach nowotworów (Murray i wsp. 2008) oraz analizę małych mutacji. Występowanie małych mutacji w genie *EGFR* (m.in. T790M w eksonie 20 czy L858R w eksonie 21) jest jednym z czynników warunkujących oporność lub podatność na terapię przeciwnowotworową z użyciem inhibitorów kinaz tyrozynowych (Paez i wsp. 2004; Kobayashi i wsp. 2005). Badania z wykorzystaniem przygotowanego testu pozwoliły na identyfikację szeregu mutacji w genie *EGFR*, zarówno małych mutacji, jak i amplifikacji całego genu, sięgających nawet 12 kopii.

Ogólną charakterystykę zjawiska zmienności liczby kopii zawarliśmy w publikacji przeglądowej, która opisuje strukturę CNV, mechanizmy ich powstawania, metody identyfikacji i analizy oraz liczne przykłady związku CNV z ekspresją genów i ich wpływu na fenotyp człowieka (Marcinkowska i Kozłowski, 2011). Jednym z opisanych przykładów związku CNV z fenotypem człowieka, jest CNV obejmujący gen *AMY1*, którego liczba kopii koreluje z poziomem kodowanego przez ten gen enzymu, amylazy ślinowej i jest silnie zróżnicowana pomiędzy populacjami różniącymi się pod względem stosowanej diety (Perry i wsp. 2007) (Rycina 5). Populacje, których podstawę diety tradycyjnie stanowią produkty

zawierające duże ilości skrobi (np. rolnicze populacje europejskie, których dieta bogata jest w wysokoskrobiowe korzenie i bulwy), posiadają wyższą liczbę kopii genu *AMY1*, a co za tym idzie wytwarzają więcej amylazy, co zwiększa wydajność hydrolizy wielocukrów, a tym samym ułatwia trawienie dostępnych pokarmów. Analogicznie, populacje, w diecie których udział skrobi jest nieznaczny (np. populacje północne, których podstawą żywienia są zwierzęta hodowlane i ryby), posiadają niższą liczbę kopii genu *AMY1*, gdyż ich układ trawienny nie wymaga zwiększonej ilości amylazy. Niniejsza praca przeglądowa została wyróżniona Nagrodą im. Bolesława Skarżyńskiego w Konkursie na najlepszy artykuł opublikowany w kwartalniku „Postępy Biochemii" w 2011 roku.



Rycina 5. *Przykład wpływu zmienności liczby kopii na fenotyp człowieka. A) CNV obejmujący gen AMY1 modyfikuje funkcjonalną liczbę kopii tego genu, a tym samym wpływa na poziom kodowanej przez ten gen amylazy ślinowej. Wyższy poziom amylazy ślinowej umożliwia bardziej efektywne trawienie skrobi, szczególnie ważne dla populacji, których dieta tradycyjnie wzbogacona jest w ten wielocukier. B) Analiza liczby kopii genu AMY1 z wykorzystaniem metody FISH. Poszczególne panele przedstawiają przykłady alleli z różną liczbą kopii genu AMY1. Czerwona i zielona sonda obejmują przylegające do siebie regiony genu AMY1 (Perry i wsp. 2007).*

Podsumowując, wszystkie prace przedstawione w niniejszej rozprawie doktorskiej dotyczą zagadnienia zmienności liczby kopii w genomie człowieka oraz opisują kolejne kroki analizy bioinformatycznej oraz eksperymentalnej, zmierzające do zaproponowania metody, która umożliwi precyzyjne i jednoznaczne genotypowanie CNV. Poszczególne etapy analiz opisane w publikacjach stanowiących niniejszą rozprawę doktorską obejmowały: (i) bioinformatyczną identyfikację regionów CNV obejmujących geny ludzkich mikroRNA (CNV-miRNA), (ii) zaprojektowanie i wygenerowanie testów do analizy wybranych CNV-miRNA, (iii) eksperymentalną identyfikację i charakterystykę polimorficznych CNV-miRNA, oraz (iv) opracowanie szczegółowego protokołu zaproponowanej metody,

obejmującego zarówno projektowanie testów MLPA, jak również analizę i interpretację uzyskanych wyników.

Opracowana przez nas metoda pozwala na projektowanie sond oraz testów MLPA, umożliwiających analizę zmienności liczby kopii oraz detekcję małych mutacji w dowolnie wybranym genie lub regionie w genomie człowieka, a tym samym na uniezależnienie się od komercyjnie dostępnych testów MLPA. Stosunkowo wysoka przepustowość, łatwość projektowania testów, wysoka powtarzalność wyników, uniwersalność i elastyczność w wyborze regionu genomu, jak również relatywnie niski koszt (zależny od skali prowadzonych eksperymentów) opracowanej metody, to zalety, które sprawiają, iż jest ona atrakcyjna do genotypowania CNV w dużej liczbie próbek, często niezbędnego w różnych badaniach genetycznych, w tym analizach asocjacji. Większość proponowanych przez nas rozwiązań może być zastosowana do analizy zmienności liczby kopii nie tylko w genomie człowieka, ale również w genomach innych gatunków zwierząt lub roślin.

Bibliografia

Aretz S, Stienen D, Uhlhaas S, Loff S, Back W, Pagenstecher C, McLeod DR, Graham GE, Mangold E, Santer R, Propping P, Friedl W. 2005. High proportion of large genomic STK11 deletions in Peutz-Jeghers syndrome. Human Mutation 26: 513–9.

Armour JAL, Palla R, Zeeuwen PLJM, Heijer M den, Schalkwijk J, Hollox EJ. 2007. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Research 35: e19.

Bunyan DJ, Skinner AC, Ashton EJ, Sillibourne J, Brown T, Collins AL, Cross NCP, Harvey JF, Robinson DO. 2007. Simultaneous MLPA-based multiplex point mutation and deletion analysis of the dystrophin gene. Molecular Biotechnology 35: 135–40.

Cantsilieris S, Baird PN, White SJ. 2012. Molecular methods for genotyping complex copy number polymorphisms. Genomics. [*epub ahead of print*]

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, i wsp. 2010. Origins and functional impact of copy number variation in the human genome. Nature 464: 704–12.

Doria A, Patti M-E, Kahn CR. 2008. The emerging genetic architecture of type 2 diabetes. Cell Metabolism 8: 186–200.

Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, i wsp. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087–93.

Fernandez-Jimenez N, Castellanos-Rubio A, Plaza-Izurieta L, Gutierrez G, Irastorza I, Castaño L, Vitoria JC, Bilbao JR. 2011. Accuracy in copy number calling by qPCR and PRT: a matter of DNA. PloS ONE 6: e28910.

Fode P, Jespersgaard C, Hardwick RJ, Bogle H, Theisen M, Dodoo D, Lenicek M, Vitek L, Vieira A, Freitas J, Andersen PS, Hollox EJ. 2011. Determination of beta-defensin genomic copy number in different populations: a comparison of three methods. PloS ONE 6: e16768.

Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, i wsp. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851–61.

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, i wsp. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307: 1434–40.

Gudmundsson J, Sulem P, Gudbjartsson DF, Blondal T, Gylfason A, Agnarsson BA, Benediktsdottir KR, Magnusdottir DN, Orlygsdottir G, Jakobsdottir M, Stacey SN, Sigurdsson A, i wsp. 2009. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. Nature Genetics 41: 1122–6.

Hollox EJ, Huffmeier U, Zeeuwen PLJM, Palla R, Lascorz J, Rodijk-Olthuis D, Kerkhof PCM van de, Traupe H, Jongh G de, Heijer M den, Reis A, Armour JAL, i wsp. 2008. Psoriasis is associated with increased beta-defensin genomic copy number. Nature Genetics 40: 23–5.

Hosono N, Kato M, Kiyotani K, Mushiroda T, Takata S, Sato H, Amitani H, Tsuchiya Y, Yamazaki K, Tsunoda T, Zembutsu H, Nakamura Y, i wsp. 2009. CYP2D6 genotyping for functional-gene dosage analysis by allele copy number detection. Clinical Chemistry 55: 1546–54.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. Nature Genetics 36: 949–51.

Ingelman-Sundberg M. 2005. Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. The Pharmacogenomics Journal 5: 6–13.

Kobayashi S, Boggon TJ, Dayaram T, Jänne PA, Kocher O, Meyerson M, Johnson BE, Eck MJ, Tenen DG, Halmos B. 2005. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. The New England Journal of Medicine 352: 786–92.

Kozlowski P, Roberts P, Dabora S, Franz D, Bissler J, Northrup H, Au KS, Lazarus R, Domanska-Pakiela D, Kotulska K, Jozwiak S, Kwiatkowski DJ. 2007. Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. Human Genetics 121: 389–400.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, i wsp. 2001. Initial sequencing and analysis of the human genome. Nature 409: 860–921.

Lian J, Tian H, Liu L, Zhang X-S, Li W-Q, Deng Y-M, Yao G-D, Yin M-M, Sun F. 2010. Downregulation of microRNA-383 is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation by targeting IRF1. Cell Death & Disease 1: e94.

Lian J, Zhang X, Tian H, Liang N, Wang Y, Liang C, Li X, Sun F. 2009. Altered microRNA expression in patients with non-obstructive azoospermia. Reproductive Biology and Endocrinology 7: 13.

Luca A De, Bottillo I, Dasdia MC, Morella A, Lanari V, Bernardini L, Divona L, Giustini S, Sinibaldi L, Novelli A, Torrente I, Schirinzi A, i wsp. 2007. Deletions of NF1 gene and exons detected by multiplex ligation-dependent probe amplification. Journal of Medical Genetics 44: 800–8.

Marcinkowska-Swojak M, Klonowska K, Figlerowicz M, Kozlowski P. An MLPA-based approach for discrete genotyping of disease-related multi-allelic CNVs. *under review*.

Marcinkowska M, Kozłowski P. 2011. Wpływ polimorfizmu liczby kopii na zmienność fenotypową człowieka. Postepy Biochemii 57: 240–8.

Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozlowski P. 2011. Copy number variation of microRNA genes in the human genome. BMC Genomics 12: 183.

Marcinkowska M, Wong KK, Kwiatkowski DJ, Kozlowski P. 2010. Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. TheScientificWorldJournal 10: 2003–18.

Marcinkowska-Swojak M, Uszczynska B, Figlerowicz M, Kozlowski P. 2013. An MLPA-Based Strategy for Discrete CNV Genotyping: CNV-miRNAs as an Example. Human Mutation 34: 763–73.

McCarroll SA, Altshuler DM. 2007. Copy-number variation and association studies of human disease. Nature Genetics 39: S37–42.

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM. 2006. Common deletion polymorphisms in the human genome. Nature Genetics 38: 86–92.

McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, Bakker PIW de, Maller JB, Kirby A, Elliott AL, Parkin M, i wsp. 2008. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nature Genetics 40: 1166–74.

Murray S, Dahabreh IJ, Linardou H, Manoloukos M, Bafaloukos D, Kosmidis P. 2008. Somatic mutations of the tyrosine kinase domain of epidermal growth factor receptor and tyrosine kinase inhibitor response to TKIs in non-small cell lung cancer: an analytical database. Journal of Thoracic Oncology 3: 832–9.

Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, i wsp. 2004. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304: 1497–500.

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, i wsp. 2007. Diet and the evolution of human amylase gene copy number variation. Nature Genetics 39: 1256–60.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, i wsp. 2006. Global variation in copy number in the human genome. Nature 444: 444–54.

Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC. 2007. Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nature Genetics 39: 513–6.

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Research 30: e57.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, i wsp. 2004. Large-scale copy number polymorphism in the human genome. Science 305: 525–8.

Sudbery I, Enright AJ, Fraser AG, Dunham I. 2010. Systematic analysis of off-target effects in an RNAi screen reveals microRNAs affecting sensitivity to TRAIL-induced apoptosis. BMC Genomics 11: 175.

Tili E, Michaille J-J, Adair B, Alder H, Limagne E, Taccioli C, Ferracin M, Delmas D, Latruffe N, Croce CM. 2010. Resveratrol decreases the levels of miR-155 by upregulating miR-663, a microRNA targeting JunB and JunD. Carcinogenesis 31: 1561–6.

Vaishnavi V, Manikandan M, Tiwary BK, Munirajan AK. 2013. Insights on the functional impact of microRNAs present in autism-associated copy number variants. PloS one 8: e56781.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, i wsp. 2001. The sequence of the human genome. Science 291: 1304–51.

Wang W, Sun J, Li F, Li R, Gu Y, Liu C, Yang P, Zhu M, Chen L, Tian W, Zhou H, Mao Y, i wsp. 2012. A frequent somatic mutation in CD274 3'-UTR leads to protein over-expression in gastric cancer by disrupting miR-570 binding. Human Mutation 33: 480–4.

Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Spitz MR, Eisen T, i wsp. 2008. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nature Genetics 40: 1407–9.

Weiss ST, Raby BA. 2004. Asthma genetics 2003. Human Molecular Genetics 13 Spec No: R83–9.

Wu X, Zhang D, Li G. 2012. Insights into the regulation of human CNV-miRNAs from the view of their target genes. BMC genomics 13: 707.

Wulfken LM, Moritz R, Ohlmann C, Holdenrieder S, Jung V, Becker F, Herrmann E, Walgenbach-Brünagel G, Ruecker A von, Müller SC, Ellinger J. 2011. MicroRNAs in renal cell carcinoma: diagnostic implications of serum miR-1233 levels. PloS ONE 6: e25787.

Yang TL, Chen XD, Guo Y, Lei SF, Wang JT, Zhou Q, Pan F, Chen Y, Zhang ZX, Dong SS, Xu XH, Yan H, i wsp. 2008. Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. American Journal of Human Genetics 83: 663–74.

PUBLIKACJE WCHODZĄCE

W SKŁAD ROZPRAWY DOKTORSKIEJ

WRAZ Z MATERIAŁAMI UZUPEŁNIAJĄCYMI

1

Marcinkowska M, Wong K-K, Kwiatkowski DJ, Kozlowski P

„Design and Generation of MLPA Probe Sets for Combined Copy Number and Small-Mutation Analysis of Human Genes: EGFR as an Example”

TheScientificWorldJOURNAL 2010, 10:2003-2018

Peer-Reviewed Protocol
TheScientificWorldJOURNAL (2010) 10, 2003–2018
ISSN 1537-744X; DOI 10.1100/tsw.2010.195

TheScientificWorldJOURNAL
www.thescientificworld.com

Design and Generation of MLPA Probe Sets for Combined Copy Number and Small-Mutation Analysis of Human Genes: *EGFR* as an Example

Malgorzata Marcinkowska[1], Kwok-Kin Wong[2,3], David J. Kwiatkowski[4], and Piotr Kozlowski[1,*]

[1]Laboratory of Cancer Genetics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland; [2]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA; [3]Ludwig Center at Dana-Farber/Harvard Cancer Center, Boston, MA; [4]Division of Translational Medicine, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

E-mail: marcinkm@man.poznan.pl; kwong1@partners.org; dk@rics.bwh.harvard.edu; kozlowp@yahoo.com

Received July 26, 2010; Revised September 6, 2010; Accepted September 23, 2010; Published October 12, 2010

Multiplex ligation-dependent probe amplification (MLPA) is a multiplex copy number analysis method that is routinely used to identify large mutations in many clinical and research labs. One of the most important drawbacks of the standard MLPA setup is a complicated, and therefore expensive, procedure of generating long MLPA probes. This drawback substantially limits the applicability of MLPA to those genomic regions for which ready-to-use commercial kits are available. Here we present a simple protocol for designing MLPA probe sets that are composed entirely of short oligonucleotide half-probes generated through chemical synthesis. As an example, we present the design and generation of an MLPA assay for parallel copy number and small-mutation analysis of the *EGFR* gene.

KEYWORDS: multiplex ligation-dependent probe amplification, MLPA, copy number variation, CNV, EGFR, large deletion, amplification, mutation detection

INTRODUCTION

Copy number variation (CNV) in the human genome has become well recognized in recent years. CNVs are heritable and somatic losses and gains of DNA segments that range in size from <1 kb to >1 Mb, and may include entire genes or even multiple genes[1,2]. The physiological effects of CNVs are a subject of continuing investigation, and range from neutral to phenotype-modifying to disease-causing mutations. Polymorphic CNVs account for about 10% of the human genome, overlapping hundreds of genes. Genomic deletion mutations occurring in genes that cause Mendelian disorders are a special subcategory of germline CNVs, and account for up to 70% of all mutations seen in some genes (e.g., *BRCA1*, *DMD*, *TSC2*, *STK11*)[3,4,5,6,7]. In addition, it is well known that CNV is widespread throughout the typical cancer genome and very likely contributes to cancer pathogenesis as much as point mutations[2,8]. A

*Corresponding author.
©2010 with author.
Published by TheScientificWorld; www.thescientificworld.com

number of methods have been developed to assess CNV at the genome-wide level. Array comparative genomic hybridization, high-density single nucleotide polymorphism (SNP) arrays (reviewed in [9]), and, more recently, second-generation sequencing[10] are widely used for CNV identification, and major improvements (regarding the precision of CNV genotyping and breakpoint mapping) to these methods have recently been achieved[11,12,13]. However, the major laboratory tool for the analysis of CNV mutations over small genomic regions, particularly for clinical diagnostic laboratories, is multiplex ligation-dependent probe amplification (MLPA) (reviewed in [14,15]).

MLPA is a method first described by Schouten et al.[16] 8 years ago as a multiplex assay utilizing up to 45 probes specific for different genomic locations (often exons in a gene of interest). Each probe is composed of two sister half-probes (a 5' half-probe and a 3' half-probe). The first step of the MLPA procedure is hybridization, during which the sister half-probes hybridize to adjacent target sequences in the input genomic DNA. In the next step, ligation of sister half-probes is performed under stringent conditions, and then the ligation products are amplified by polymerase chain reaction (PCR) using fluorescently tagged universal primers to sequences incorporated in the sister half-probes (Fig. 1A). The PCR products are separated by capillary electrophoresis (CE) (Fig. 1C), and the signal from each probe is normalized against a control probe signal and is compared to a corresponding normalized signal observed in a set of reference samples (Fig. 1D).

Originally, MLPA was designed as a copy number analysis tool, and it has been successfully used in the testing and identification of hundreds of large mutations in numerous disease-related genes, including *DMD*, *BRCA1*, *NF1*, *STK11*, and *TSC2*. Further modifications of the MLPA protocol broadened its range of applications. The additional applications of MLPA are SNP genotyping[16], methylation status determination[17], copy number analysis in segmentally duplicated regions[18,19], expression profiling[20], mouse transgene genotyping[21], analysis of DNaseI hypersensitive sites[22], determination of the effectiveness of conditional allele conversion[23], and strand-specific expression analysis (Mykowska et al., submitted for publication).

The main disadvantage of the standard MLPA setup is a complicated and time-consuming (and therefore expensive) process of probe design and generation. This is due to the necessity for creating long 3' half-probes (~100–400 nt). Usually this is done by cloning 3' half-probes in specially prepared M13 vectors, enabling insertion of arbitrary numbers of nucleotides into those probes[16]. In practice, this disadvantage seriously limits the applicability of MLPA to novel genes or sets of genes for which ready-to-use commercial kits are not available.

This M13-based method of probe generation can be avoided by designing MLPA probe sets composed entirely of oligonucleotide probes that can be generated through chemical synthesis. Although several successful applications of fully synthetic MLPA probe sets have been reported (e.g., [24,25,26,27,28]), the vast majority of MLPA applications are still restricted to genes for which it is possible to use commercially available kits (MRC-Holland, http://www.mlpa.com).

Here we describe a protocol for the simple design and generation of MLPA assays that utilize exclusively synthetic probes. Critical modifications applied in our strategy are (1) a shortest probe length of 90 nt; (2) separation of subsequent probes by 3 and 4 nt for probes shorter and longer than 120 nt, respectively; (3) placing stuffer sequences into both 5' and 3' half-probes, making them of approximately equal length; and (4) restricting the longest probe/half-probe lengths to 200/100 nt, respectively. This leads to a capacity for analysis of 31 probes at once; longer oligonucleotide synthesis is also possible, expanding the capacity of this approach. A further increase of multiplexing capacity can be achieved by the use of two-color (or multiple-color) labeling on two distinct pairs of universal primers that enable a simultaneous CE analysis of two sets of MLPA products[24]. The strategy described here can be applied to any genomic region(s) of interest. We have used this strategy to generate over 10 different MLPA assays (examples are shown in Fig. 2). Published applications include the identification of large mutations in *TSC1*, *TSC2*[6], and *PKD1*[18] genes; analysis of loss of heterozygosity in cancer samples[23]; genotyping of several mouse transgenes[21]; and strand-specific expression analysis (Mykowska et al., submitted for publication).

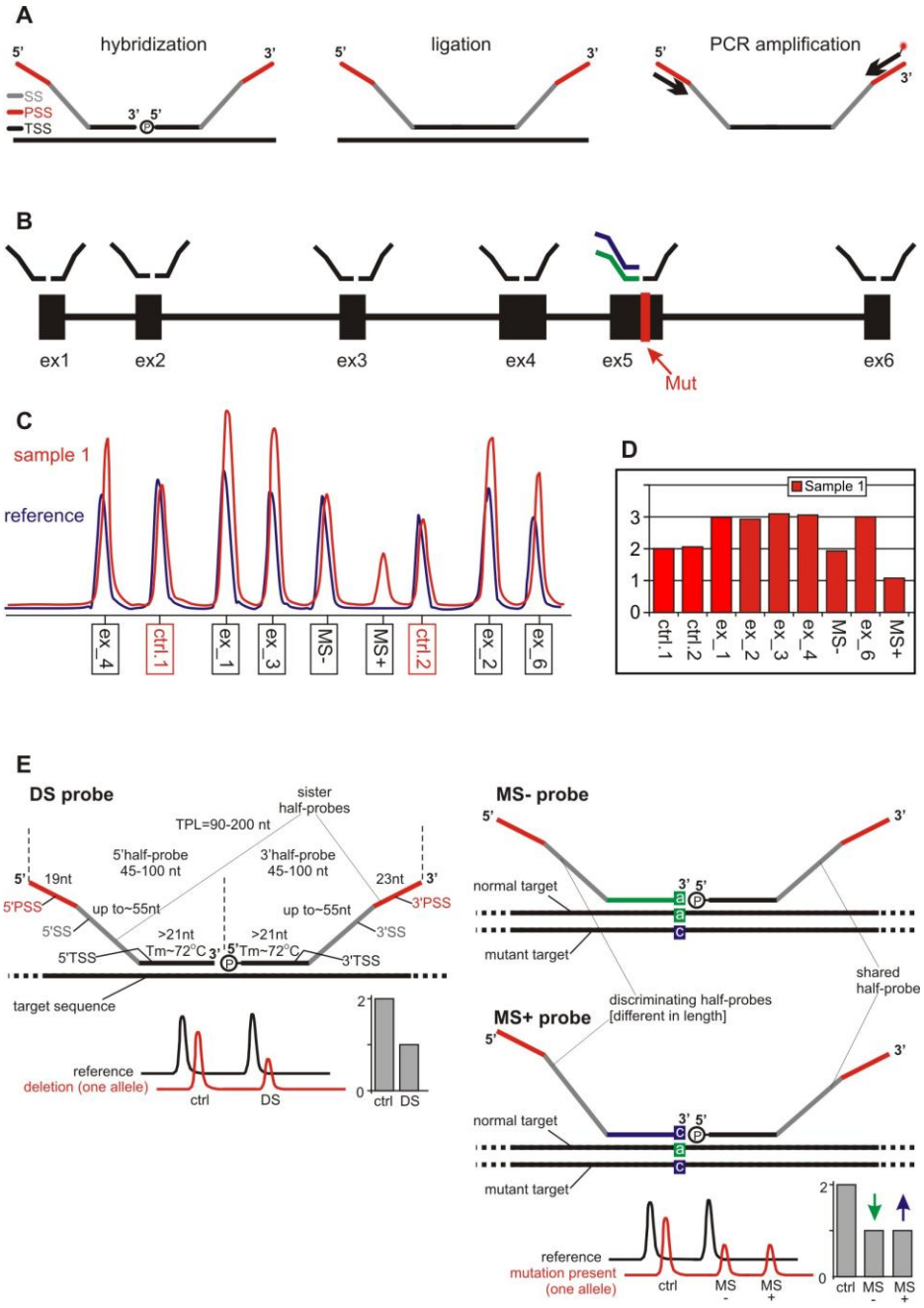


FIGURE 1. The principle of MLPA analysis for simultaneous identification of CNV and small mutations. (A) Three subsequent steps in the MLPA reaction (from left to right): hybridization of sister half-probes to the target sequence, ligation of correctly hybridized probes, and PCR amplification of ligated probes with universal primers. Primer-specific sequences (PSSs), stuffer sequences (SSs), and target-specific sequences (TSSs) are indicated in red, gray, and black, respectively. (B) Structure of a hypothetical model gene with the locations of MLPA probes (above). The probe located in exon 5 has two alternative 5' half-probes: one (MS-) specific for normal (green) and the other (MS+) specific for mutant (blue) sequence. The alternative 5' half-probes are different in length. (C) Overlapped hypothetical electropherograms of subject (red) and reference (blue) samples. Probe IDs are indicated below the electropherogram. (D) Bar graph showing relative copy number values calculated for each probe. Increased signal from all exonic probes (ex_1 to ex_6) indicates entire gene duplication. Relatively low signal from probe MS- located in exon 5 indicates the presence of a small mutation that is additionally confirmed by the appearance of a signal from the mutation-specific (MS+) probe. (E) Characteristics of the three types of MLPA probes; (left-hand side) copy number–sensitive (DS) probe, (above, right-hand side) small-size mutation–sensitive, negative (MS-) probe, and (below, right-hand side) small-size mutation–sensitive, positive (MS+) probe. In each upper panel, a schematic representation of an MLPA probe hybridized to its target sequence is shown. PSSs, SSs, and TSSs are indicated and marked as in panel A. TSSs specific for normal and mutant sequences are indicated in green and blue, respectively. In panels DS and MS (below), a schematic electropherogram of the analyzed (red line) and reference (black line) sample is shown. The results of copy number analysis presented in the form of a bar plot are shown below on the right-hand side.

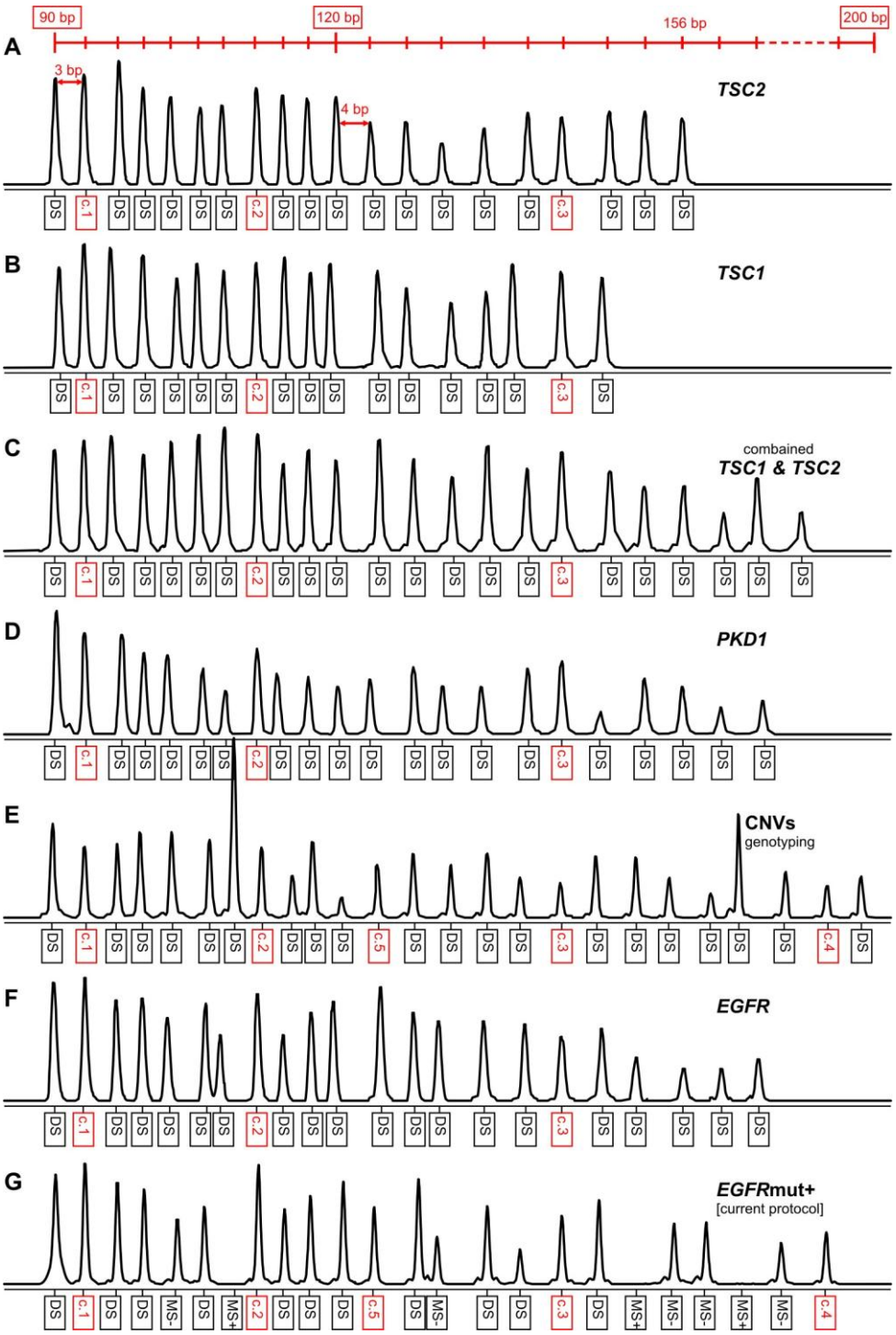


FIGURE 2. Examples of MLPA probe sets designed according to the described protocol. Electropherogram profiles representing a normal DNA sample analyzed with different MLPA probe sets (the signal of each probe [except panel E] represents two target sequence copies). (Top) Schematic representation of an MLPA probe set layout. Probe sets for large-mutation analysis in (A) *TSC2*, (B) *TSC1*, (C) both *TSC1* and *TSC2*, and (D) *PKD1*. (E) Probe set for genotyping several polymorphic CNVs at different sites in the genome. (F) Probe set for CNV analysis of *EGFR*. (G) Probe set for combined copy number and small-mutation analysis of *EGFR* (the assay described in this article). The types of the MLPA probes are indicated under the electropherograms. Control probes are indicated in red.

As an example, we present here the design of an MLPA probe set (assay) for the combined copy number and small-mutation analysis of the *EGFR* gene. *EGFR* is a well-known tumor proto-oncogene frequently mutated in various types of cancer. Oncogenic variants activating *EGFR* can be both copy number (*EGFR* amplification and vIII deletion) and small-size mutations (substitutions, in-frame deletions, and in-frame insertions)[29]. The status of *EGFR* mutations is an important factor modifying the effectiveness of tyrosine kinase inhibitor (TKI) treatment (reviewed in [30]). Lung cancers with certain *EGFR* mutations (e.g., L858R and exon 19 in-frame deletions) are sensitive to TKI treatment[31,32], whereas the occurrence of the secondary mutation T790M causes resistance to TKI[33,34].

The proposed MLPA setup allows for copy number or combined copy number and small-mutation analysis of up to ~30 genomic locations (probes) with a per-sample cost of ~$3 plus a starting cost (probe synthesis) of about $3000 (once synthesized, the number of probes obtained is sufficient for hundreds of thousands of analyses).

MATERIALS

1. Reagents
   A. MLPA reactions
      (i) Genomic DNA sample: 20–50 ng/µl (3 µl per assay)
      (ii) Probe mix: composed of self-designed synthetic probes. Synthesis parameters: synthesis scale, 100 nmol; purification: IE HPLC; modification, 5’ phosphorylation (only 3’ half-probes) (IDT-DNA)
      (iii) MLPA reagent kit (includes all reagents except probe mix): SALSA MLPA Reagents (MRC-Holland EK1, EK5, EK20, or EK50)
      (iv) Deionized water (resistance <18 MΩ cm)
   B. Sample preparation and CE analysis
      (i) HiDi formamide (Applied Biosystems Cat. No. 4311320)
      (ii) CE polymer: ABI POP7 (Applied Biosystems)
      (iii) DNA size standard: Gene Scan LIZ-600 (Applied Biosystems)
2. Equipment and consumables
   A. 96-well plates: Certified Thin Wall 96 × 0.2 ml PCR Plates (Starlab)
   B. PCR thermocycler: GeneAmp PCR System 9700 (Applied Biosystems) or PTC-200 Thermo Cycler (MJ Research)
   C. Capillary electrophoresis: CE analysis can be performed on any standard multicapillary DNA analyzer (e.g., ABI-Prism 3130XL, 3100, 1700 [Applied Biosystems], CEQ-2000, 8000, 8800 [Beckman])

PROCEDURE

1. General MLPA design

A. Probe set layout

The MLPA assay can be composed of up to 31 probes, with a total probe length (TPL) ranging from 90 to 200 nt (half-probe length [HPL] ranging from 45 to 100 nt). The (EGFRmut+) MLPA probe set presented in this protocol was composed of 24 probes with TPL ranging from 90 to 172 nt. The difference between the lengths of the probes (spacing) was 3 and 4 nt for probes shorter and longer than 120 nt, respectively (Fig. 2).

COMMENT: The proposed spacing of the probe lengths ensures proper separation of PCR products during CE. Smaller differences in length can cause the adjacent peaks to overlap, making interpretation difficult. Larger spacing intervals can be used, but this reduces the capacity of an MLPA assay.

Most probes in the set are used to investigate CNV in the genomic region(s) of interest. Probes should be evenly distributed over the investigated region. If the region of interest contains a gene, probes can be preferentially located in exons.

COMMENT: The lengths of the probes do not have to correspond to the order of their genomic locations. Mixing up the lengths of the probes allows CNV to be distinguished from artifacts related to the size of the probes. True CNVs often affect probes that are located in adjacent positions in the genome, whereas length-dependent artifacts affect probes of similar lengths. The most common length-dependent artifact is a gradual increase or decrease of relative signal intensity corresponding to the probe length.

Each probe set should contain at least a few control probes (in the EGFRmut+ probe set, five control probes specific for locations in different chromosomes were used). The control probes should be chosen from outside the genomic region of interest, ideally from different chromosomes, and not subject to CNV in the general population. The Database of Genomic Variations (DGV - http://projects.tcag.ca/variation/) and other resources[11,12,35] can be used to avoid known CNV regions. Alternatively, the control probes proposed here (Supplementary Table 1) can be used. If an MLPA assay is intended to analyze somatic variation in cancer samples, the control probes should not be located within or close to known cancer-variable regions (e.g., proto-oncogenes and tumor suppressors)[36]. Recently published results of genome-wide somatic CNV analysis across numerous cancer samples[2] can be used to avoid regions highly variable in cancers.

B. Probe layout

Each probe is composed of two sister half-probes (a 5' half-probe and a 3' half-probe) of equal length (Fig. 1). In the case of probes with an odd TPL, the length of sister half-probes can differ by 1 nt. Each half-probe consists of a target-specific sequence (TSS), a universal primer-specific sequence (PSS), and a stuffer sequence (SS) that allows the TPL to be modulated. The 3' half-probe is phosphorylated at its 5' end to enable ligation of sister half-probes. The 5' half-probe is composed of (from the 5' end): 5' PSS (19 nt), 5' SS (variable length), and 5' TSS (≥21 nt). The 3' half-probe is composed of (from the 5' end): 5' phosphate, 3' TSS (≥21 nt), 3' SS (variable length), and 3' PSS (23 nt) (Fig. 1E).

MLPA probe design depends on the purpose of the probe. The subsequent steps of the protocol describe the design of three types of MLPA probes that were used in the EGFRmut+ assay: (1) dosage (copy number) sensitive (DS); (2) small-size mutation sensitive, negative (MS-); and (3) small-size mutation sensitive, positive (MS+) probes (Fig. 1E).

2. Design of DS probes: general MLPA probe design

The basic MLPA probe is a DS probe (Fig. 1E). The signal intensity of a DS probe corresponds to the dosage (copy number) of the target sequence.

A. Selection of TSSs

The TSSs specifically recognize analyzed sequences and are thus the most critical part of MLPA probes. The design of TSSs depends on the purpose of the probe.

(i) Select a genomic region or a gene of interest, retrieve genomic sequence, and paste it into a word-processing program (e.g., MS Word) (Supplementary File 1).
(ii) In the sequence of interest, mark exons or other sequences on which you are going to focus.
(iii) Label/mask the sequence of interest with (1) repeat/low-complexity regions, (2) positions of SNPs, (3) segmental duplication, and (4) polymorphic CNV regions.

COMMENT: There are many alternative resources that can be used for sequence selection (e.g., UCSC Genome Browser [UCSC GB]: http://genome.ucsc.edu/; Ensembl: http://www.ensembl.org; NCBI: http://www.ncbi.nlm.nih.gov; SNPper: http://snpper.chip.org) and labeling/masking (UCSC DB; DGV; RepeatMasker: http://www.repeatmasker.org/; dbSNP database: http://www.ncbi.nlm.nih.gov/snp/; HapMap: www.hapmap.org/).

(iv) [This step is a convenient alternative to steps (i–iii).] The sequence of interest marked with all the above-mentioned genetic features can be retrieved from UCSC GB in a few steps: (1) select the sequence of interest, (2) select "DNA" on the upper toolbar, (3) select "extend case/color options", (4) fill in the "Extended DNA Case/Color Options" table as shown in Fig. 3, (5) press "submit", and (6) copy the sequence to a word-processing program (Supplementary File 1).
(v) In the region for which the probe will be designed (e.g., exons), select candidate target sequences (about 100 nt long) free of polymorphisms and repetitive elements. Extremely high GC-content sequences should be avoided.
(vi) In the candidate target region, select directly adjacent 5' and 3' TSSs. Each TSS should be at least 21 nt long. Sister TSSs should be of similar lengths, and mononucleotide tracts of C or G (≥3 nt) should be avoided at or close to the ligation point. The melting temperature (Tm) of each TSS should be as close as possible to 71°C. To calculate the annealing temperature, use RaW-Probe v.0.15B. RaW-Probe is freely available at the MRC-Holland webpage (http://www.mlpa.com).

CAUTION: Avoid shortening the target sequence below 21 nt. If, due to high GC content, it is not possible to find any sequence with the optimal Tm value (71°C), it is better to select a sequence with a higher Tm rather than to shorten the sequence below 21 nt.

(vii) BLAST the selected TSSs against the appropriate reference sequence (here, the human genome) to verify that they are unique in the human genome. Use the algorithm BLASTN with the following parameters: no filtering, no repeat masking, and E (expectancy) = 1. We recommend using the BLAST program available at the NCBI webpage (http://blast.ncbi.nlm.nih.gov).
(viii) Mark selected TSSs in different colors (a different color for the 5' TSS [e.g., yellow] and the 3' TSS [e.g., green]) (Supplementary File 1) and paste them into the appropriate positions in the "Probe Set Assembly Table" (Supplementary Table 1).

COMMENT: The "Probe Set Assembly Table" serves as a tool for combining segments of half-probes into oligonucleotide sequences of the desired length. Each row represents one probe. Predefined probe lengths are indicated in the last column. Each row is divided into two

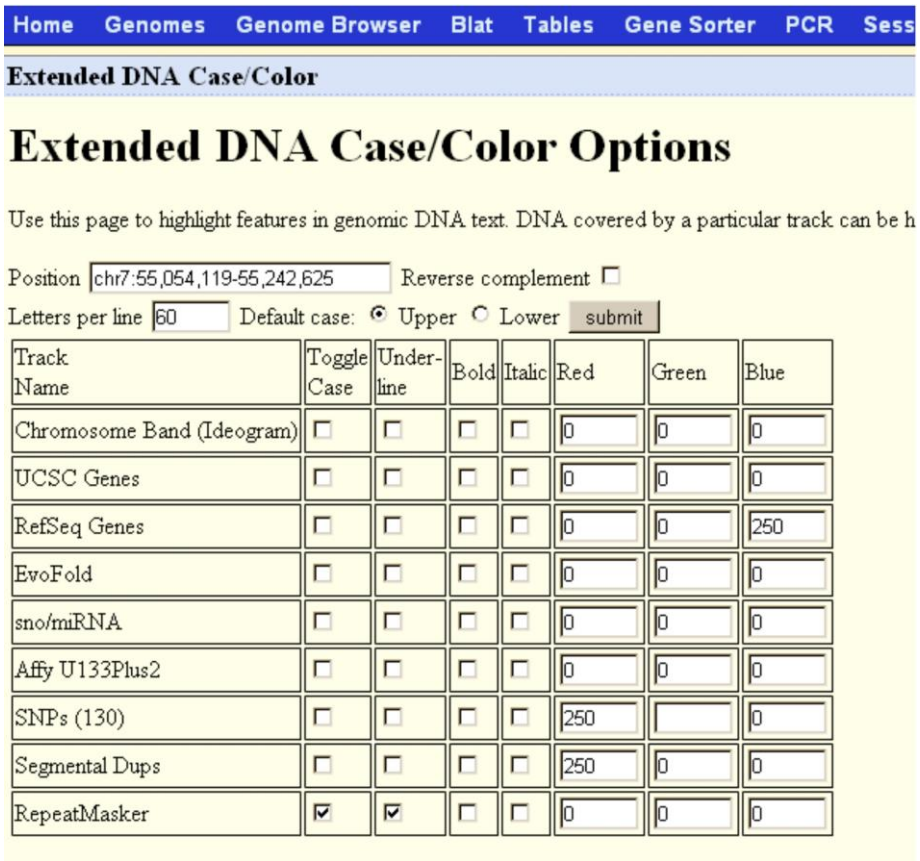


FIGURE 3. Screenshot of completed "Extended DNA Case/Color Options" table for labeling/masking DNA sequences retrieved from UCSC GB.

sections: a 5' half-probe (yellow panels) and a 3' half-probe (green panels). Each half-probe section includes columns with the sequences and lengths of the probe segments (from 5'): 5' PSS, 5' SS, 5' TSS, 3' TSS, 3' SS, and 3' PSS. The sequences and lengths of the PSSs as well as the sequences of the control probes can be pasted into the "Probe Set Assembly Table" prior to the start of a probe set designing project.

(ix) Use a strategy similar to that presented above (i–viii) to design control probes. Control probes should be located in genomic regions expected to be free of CNV in the intended experiments. Alternatively, control probes included in the EGFRmut+ set can be used as controls for any MLPA set. These control probes were already tested in several MLPA assays (Fig. 2).

B. Addition of PSSs

The PSSs correspond to a pair of universal primers included in all commercially available MLPA reagent kits (MRC-Holland). They enable multiplex amplification of all MLPA probes.

(i) Paste the 5' PSS (GGGTTCCCTAAGGGTTGGA) and 3' PSS (TCTAGATTGGATCTTGCTGGCGC) into the appropriate positions of the "Probe Set Assembly Table" (Supplementary Table 1).

C. Addition of SSs and assembly of half-probes

The SS is the sequence inserted between the PSS and the TSS to adjust both HPL and TPL.

(i) Using the following equations, calculate the length of the SS for each half-probe: length of 5' SS = predefined 5' HPL – (length of 5' PSS + length of 5' TSS); length of 3' SS = predefined 3' HPL – (length of 3' PSS + length of 3' TSS).

(ii) Paste the SSs of the appropriate length into the "Probe Set Assembly Table".

COMMENT: Although stuffers can be any sequence of appropriate length, we recommend using the appropriate fragments of the same universal SS in all probes. The universal SS used in all our MLPA sets is a 117-nt fragment of M13 sequence (AC# V00604) (Supplementary Table 1). This sequence was selected based on its GC content (49%), lack of substantial similarities to the human genome, and lack of any back-folding self-complementarities. The appropriate 5' and 3' SS fragments are generated from the 5' and 3' ends of the universal SS, respectively. Designing SSs in the way described above (and presented in Supplementary Table 1) substantially increases probe similarity in a way that extends well beyond the universal PSSs. This similarity significantly improves the uniformity of probe amplification and thus reduces amplicon-dependent signal variation. In all the designed MLPA sets, the relative signal intensity of most probes does not differ more than twofold (Fig. 2).

(iii) Combine the half-probe segments in the following order: 5' half-probe – 5' PSS, 5' SS, and 5' TSS; 3' half-probe – 3' TSS, 3'SS, and 3'PSS.

(iv) Again, BLAST all final probe sequences (combined 5' and 3' TSS) against the human genome to double check that no error was introduced during probe sequence assembly and handling. Use the BLAST parameters described above (step 2A vii). A correctly designed TSS should show (1) one perfectly matched sequence (the target) and (2) a lack of any other substantial similarities. Minor complementarities (e.g., <90% homology over 10 nt) to alternative genomic locations are acceptable.

3. Design of MS- probes

The MS- probe is a type of DS probe whose signal decreases in the presence of small mutations. Examples are shown in Fig. 4.

(i) Locate the position of the small mutation in the sequence of interest (Supplementary File 1; red-bold font).

(ii) Following the instructions described in step 2A, design a 5' and 3' TSS pair with a ligation position directly adjacent to the small mutation. The 3' end of the 5' TSS or the 5' end of the 3' TSS should overlap the mutation.

(iii) Add PSSs and SSs as described in step 2.

COMMENT: A single nucleotide mismatch at either the 5' or the 3' side of the ligation point will completely preclude ligation and subsequent amplification of the MS- probe. Note, however, that small-size mutations located outside of target sequences do not affect the probes signal and thus cannot be detected by MLPA.

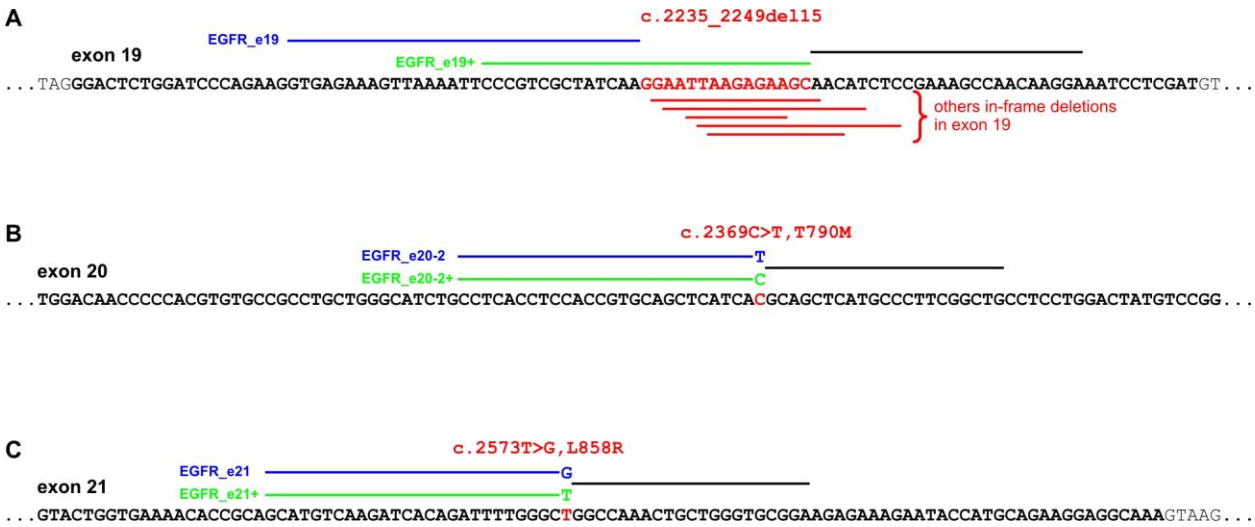


FIGURE 4. Pairs of target sequences for MS- and MS+ probes specific for (A) in-frame deletion c.2235_2249del15 in exon 19, (B) T790M in exon 20, and (C) L858R in exon 21. Black lines: 3' TSS shared by the MS- and MS+ probes; green and blue lines: 5' TSSs specific for the normal and mutant sequences, respectively.

4. Design of MS+ probes

The signal from MS+ probes appears only when a specific mutation is present. In the case of wild-type sequence, MS+ probes give no signal. Examples are shown in Fig. 4.

(i) Select a TSS for the MS- probe as described in step 3.
(ii) Replace either the 5' or the 3' TSS (depending on which one's end overlaps the mutation) of the MS- probe with the mutated TSS.
(iii) Add PSSs and SSs as described in step 2. One half-probe should be common for both MS- and MS+, and a second one should discriminate between the normal (MS-) and mutant (MS+) sequences (probes). Discriminating half-probes must be different in length (Fig. 1E).

5. Generation of half-probe oligonucleotides

(i) Order the oligonucleotide half-probes. There are many companies that provide oligonucleotide service suitable for generating MLPA probes. All half-probes used in our probe sets were synthesized by IDT (http://idtdna.com/) using the following parameters: synthesis scale, 100 nmol; purification IE HPLC; modification, 5' phosphorylation (3' half-probes only) (Supplementary Table 2).

CAUTION: To enable ligation of sister half-probes, all 3' half-probes must be phosphorylated at their 5' ends.

6. Preparation of the probe set

(i) Dilute all oligonucleotide half-probes with deionized water to a concentration of 100 µM (stock solutions).
(ii) Prepare a chart of a 96-well plate with individual positions designated for each half-probe oligonucleotide (Supplementary Table 2).

(iv) Aliquot 2 µl of each stock solution to the appropriate position in the 96-well plate. To each 2 µl of stock solution add 200 µl of deionized water and mix it well by carefully pipetting the mixture up and down (about 10 times). The concentration of oligonucleotides in the 96-well plate is 1 µM (working solutions).

(v) Mix 2 µl of each half-probe working solution in a 1.5-ml Eppendorf tube. Dilute the mixture with deionized water up to 400 µl (probe set mix).

7. MLPA reaction

(i) Use the prepared probe set mix as a standard "probemix" with SALSA MLPA reagents (MRC-Holland). Follow the standard MLPA protocol.

COMMENT: MLPA is a robust and easy-to-perform procedure. The MLPA protocol was described thoroughly in a seminal MLPA paper[16]. Additional information and troubleshooting can be found on the MRC-Holland webpage (www.mlpa.com). Therefore, detailed descriptions of the MLPA reactions and analysis are not part of this protocol.

8. CE of MLPA amplicons

The separation of MLPA amplicons can be performed on any standard multicolor capillary DNA analyzer (e.g., ABI-Prism 3100, 1700 [Applied Biosystems], CEQ-2000, 8000, 8800 [Beckman]). The general strategy for CE analysis and signal detection is similar with most commonly available apparatuses, but the detailed procedure differs from apparatus to apparatus and is described in detail in the appropriate manufacturers' manuals.

(i) Run the MLPA reaction under denaturing CE conditions following the detailed manufacturer's protocol. CE analysis of MLPA products is typically performed under the following conditions: sample dilution, 20–40× in deionized formamide; capillary length, ~42 cm; electroinjection voltage, 5 kV; electroinjection time, 5 sec; denaturing polymer, POP7 (Applied Bioscience); run temperature, 60°C; run voltage, 13 kV; electrophoresis time, ~20 min.

(ii) Using an appropriate program (e.g., GeneMapper [Applied Biosystems]), extract the probe signal intensity data (peak heights or peak areas).

COMMENT: We did not find any substantial difference using peak heights vs. peak areas as the probes' intensity representation; therefore, we routinely use only peak heights.

9. Analysis of MLPA results

(i) Divide the signals of all probes by the average signal of the control probes (signal normalization).

(ii) To calculate a copy number value for each probe, divide the normalized signal by the corresponding average normalized signal from a set of reference samples, and multiply by 2. We recommend the use of four reference samples.

COMMENT: The use of several reference samples in every experiment allows those reference results that show substantial deviation in relative probe signal intensity to be excluded, which reduces the effect of random signal variation occurring in individual (reference) samples.

(iii) The copy number values of all the analyzed probes can be visualized in the form of a bar graph, as shown in Figs. 1 and 5.

CAUTION: Most CNVs extend over long regions of the genome, and are often detected and validated by the simultaneous change in signal from multiple adjacent probes. Generally, for multiprobe CNVs, we recommend to assume that a signal-change threshold equals 2 standard deviations (SD) of a signal from unaffected probes[6,37,38]. In practice, multiprobe CNVs can be reliably detected by ≥20% increase (duplication) or decrease (deletion) of relative probe signals (assuming reasonably good-quality reactions with an unaffected probe signal SD of about 10%). This sensitivity allows not only for the detection of heterozygous mutations (50% signal change), but also for detection of heterozygous mosaic mutations affecting as little as 40% of cells from which the DNA has been extracted (20% signal change)[6,19,37]. However apparent CNV seen with a single probe may be artifactual or due to the presence of small mutations affecting the probe sequence. Therefore all single probe findings (including small-mutations detected by MS- probes) have to be validated by the use of alternative method.

TIMING: Designing a full set of MLPA probes (~25) takes 1–2 days (depending on the experimenter's skill and experience). Oligonucleotide dilution and probe mix preparation takes about 4 h.

ANTICIPATED RESULTS

The protocol described in this paper was successfully used to design several different MLPA assays, including a test for combined copy number and small-mutation analysis of the *EGFR* gene (EGFRmut+). *EGFR* is composed of 28 exons spanning almost 200 kb of chromosome 7p11.2. Except for the extremely large intron 1 (over 100 kb), *EGFR* represents a typical human multiexon gene (Fig. 5A).

The reference sequence of *EGFR* extracted from UCSC GB (Supplementary File 1) shows exons (blue font), repetitive sequences (lowercase underlined font), and SNPs (red font). Additionally, the red-bold font indicates positions of the most common oncogenic *EGFR* mutations (labeled manually). Using the described protocol, a set of 24 probes was designed. The positions of the *EGFR* probes are indicated in the *EGFR* reference sequence (Supplementary File 1) (yellow and green highlight for 5' and 3' TSSs, respectively) and in Fig. 5A. The probes included five control probes (located on different chromosomes), 15 probes specific for *EGFR*, and four probes located in two other proto-oncogenes (two in *MET* [chr 7] and two in *ERBB2* [chr 17]). The *EGFR* gene probe set contained seven DS, five MS-, and three MS+ (Fig. 5). The mutations covered by the MS- probes accounted for over 90% of all oncogenic mutations occurring in the TK domain of *EGFR* (Fig. 5A). For the two most common *EGFR* mutations (L858R in exon 20 [~40%] and the most frequent in-frame deletion in exon 19, c.2235_2249del15 [~20%]) and for T790M in exon 20, probes specifically recognizing the mutant sequence (MS+) were also designed (Fig. 5).

The *EGFR* amplification that frequently occurs in different types of cancer resulted in increased relative signals from all DS and MS- probes. Also, other oncogenic rearrangements of *EGFR* could be detected as changes in DS and MS- probe signal intensity. An example of such a rearrangement is *EGFR* variant III (a large in-frame deletion including exons 2–8). Regardless of the copy number status of *EGFR*, the occurrence of specific small mutations resulted in a decrease of the relative signal of the corresponding MS- probe. This decrease was proportional to the number of *EGFR* copies in which the small mutation occurs. Additionally, in the case of three mutations (L858R, an in-frame deletion in exon 19 [c.2235_2249del15], and T790M), a signal from the corresponding MS+ probes should also occur.

The results of the EGFRmut+ MLPA analysis are shown in Fig. 5. The electropherograms shown in Fig. 5B represent one reference and two cancer samples (sample 1 and sample 2). The overlay of the reference and cancer sample electropherograms clearly shows an increase in *EGFR* probe signal in both

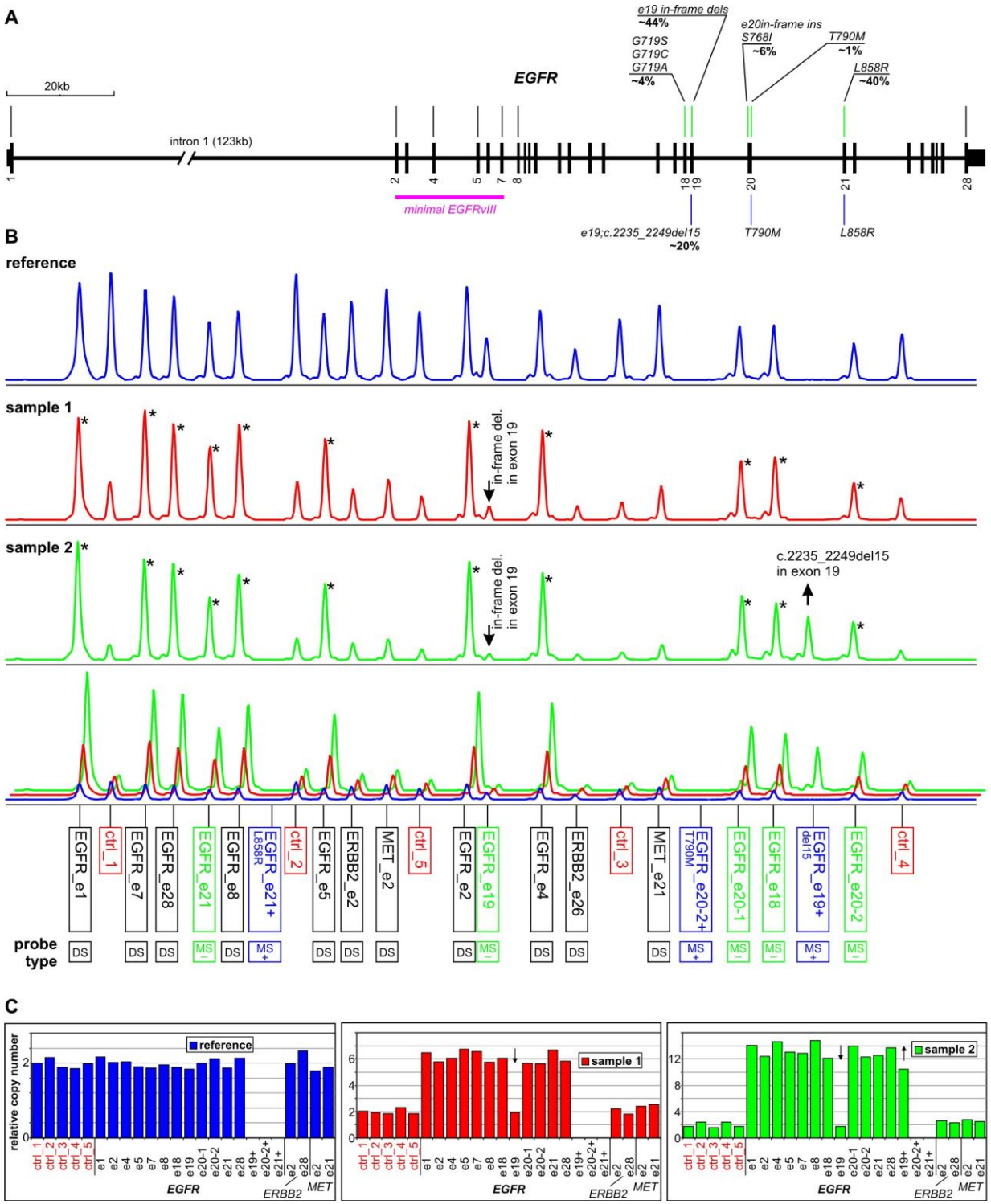


FIGURE 5. Representative results from the EGFRmut+ MLPA assay for combined copy number and small-mutation analysis of the *EGFR* gene. (A) Map of the *EGFR* gene with the positions of the EGFRmut+ probes indicated (vertical lines over and under the map). Black, green, and blue lines indicate the DS, MS-, and MS+ probes, respectively. The oncogenic small mutations (and their frequencies) covered by the MS- and MS+ probes are indicated over and under the corresponding probes, respectively. (B) Electropherograms represent (from the top) the reference sample, cancer sample 1, cancer sample 2, and an overlay of all three samples. Probe IDs and types are indicated under the electropherograms. Asterisks indicate amplified signals; arrowheads indicate reduced signal from MS- probe EGFR_e19 (specific for all in-frame deletions in exon 19) and increased signal of MS+ probe EGFR_e19+ (specific for deletion c.2235_2249del15). (C) Bar plots (corresponding to the electropherograms shown above [B]) represent the normalized copy number value (y-axis) of each probe (x-axis).

cancer samples (Fig. 5B and C), which corresponds to *EGFR* gene amplification up to six and 12 copies in samples 1 and 2, respectively. In both analyzed cancer samples, a lower signal from the EGFR_e19- probe is also clearly visible. The lower signal of this probe indicated the presence of one of the in-frame deletions in exon 19. Additionally, the signal of the EGFR_e19+ probe that appears in sample 2 and is clearly absent in the reference and sample 1 indicates that the in-frame deletion that occurred in sample 2 was c.2235_2249del15, which is the most common in-frame deletion in exon 19 and the second most common mutation in *EGFR*.

The protocol proposed here can be easily used to design an MLPA probe set for copy number analysis, or for combined copy number and small-mutation analysis of any region of interest in any genome. This strategy for parallel copy number and small-mutation analysis can be used to prescreen disease-related genes for large mutations and the most common recurrent small mutations.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Science and Higher Education, Grant No. N N302-278937 (PK and MM); Uniting Against Lung Cancer grant (DJK and K-KW); and Dana-Farber-Harvard Cancer Center Lung Cancer Specialized Program of Research Excellence (SPORE) grants P50 CA090578 (DJK, K-KW), U01 CA141576 (K-KW), R01 AG2400401 (K-KW), R01 CA122794 (K-KW), R01 CA140594 (K-KW), and 1RC2CA147940-01 (K-KW).

REFERENCES

1. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., Cho, E.K., Dallaire, S., Freeman, J.L., Gonzalez, J.R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J.R., Marshall, C.R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M.J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D.F., Estivill, X., Tyler-Smith, C., Carter, N.P., Aburatani, H., Lee, C., Jones, K.W., Scherer, S.W., and Hurles, M.E. (2006) Global variation in copy number in the human genome. *Nature* **444**, 444–454.
2. Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., Mc Henry, K.T., Pinchback, R.M., Ligon, A.H., Cho, Y.J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M.S., Weir, B.A., Tanaka, K.E., Chiang, D.Y., Bass, A.J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F.J., Sasaki, H., Tepper, J.E., Fletcher, J.A., Tabernero, J., Baselga, J., Tsao, M.S., Demichelis, F., Rubin, M.A., Janne, P.A., Daly, M.J., Nucera, C., Levine, R.L., Ebert, B.L., Gabriel, S., Rustgi, A.K., Antonescu, C.R., Ladanyi, M., Letai, A., Garraway, L.A., Loda, M., Beer, D.G., True, L.D., Okamoto, A., Pomeroy, S.L., Singer, S., Golub, T.R., Lander, E.S., Getz, G., Sellers, W.R., and Meyerson, M. (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905.
3. Aretz, S., Stienen, D., Uhlhaas, S., Loff, S., Back, W., Pagenstecher, C., McLeod, D.R., Graham, G.E., Mangold, E., Santer, R., Propping, P., and Friedl, W. (2005) High proportion of large genomic STK11 deletions in Peutz-Jeghers syndrome. *Hum. Mutat.* **26**, 513–519.
4. White, S.J. and den Dunnen, J.T. (2006) Copy number variation in the genome; the human DMD gene as an example. *Cytogenet. Genome Res.* **115**, 240–246.
5. Montagna, M., Dalla Palma, M., Menin, C., Agata, S., De Nicolo, A., Chieco-Bianchi, L., and D'Andrea, E. (2003) Genomic rearrangements account for more than one-third of the BRCA1 mutations in northern Italian breast/ovarian cancer families. *Hum. Mol. Genet.* **12**, 1055–1061.
6. Kozlowski, P., Roberts, P., Dabora, S., Franz, D., Bissler, J., Northrup, H., Au, K.S., Lazarus, R., Domanska-Pakiela, D., Kotulska, K., Jozwiak, S., and Kwiatkowski, D.J. (2007) Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. *Hum. Genet.* **121**, 389–400.
7. Stankiewicz, P. and Lupski, J.R. (2002) Genome architecture, rearrangements and genomic disorders. *Trends Genet.* **18**, 74–82.

8. Weir, B.A., Woo, M.S., Getz, G., Perner, S., Ding, L., Beroukhim, R., Lin, W.M., Province, M.A., Kraja, A., Johnson, L.A., Shah, K., Sato, M., Thomas, R.K., Barletta, J.A., Borecki, I.B., Broderick, S., Chang, A.C., Chiang, D.Y., Chirieac, L.R., Cho, J., Fujii, Y., Gazdar, A.F., Giordano, T., Greulich, H., Hanna, M., Johnson, B.E., Kris, M.G., Lash, A., Lin, L., Lindeman, N., Mardis, E.R., McPherson, J.D., Minna, J.D., Morgan, M.B., Nadel, M., Orringer, M.B., Osborne, J.R., Ozenberger, B., Ramos, A.H., Robinson, J., Roth, J.A., Rusch, V., Sasaki, H., Shepherd, F., Sougnez, C., Spitz, M.R., Tsao, M.S., Twomey, D., Verhaak, R.G., Weinstock, G.M., Wheeler, D.A., Winckler, W., Yoshizawa, A., Yu, S., Zakowski, M.F., Zhang, Q., Beer, D.G., Wistuba, II, Watson, M.A., Garraway, L.A., Ladanyi, M., Travis, W.D., Pao, W., Rubin, M.A., Gabriel, S.B., Gibbs, R.A., Varmus, H.E., Wilson, R.K., Lander, E.S., and Meyerson, M. (2007) Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**, 893–898.
9. Carter, N.P. (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat. Genet.* **39**, S16–21.
10. Chiang, D.Y., Getz, G., Jaffe, D.B., O'Kelly, M.J., Zhao, X., Carter, S.L., Russ, C., Nusbaum, C., Meyerson, M., and Lander, E.S. (2009) High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat. Methods* **6**, 99–103.
11. McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A., Elliott, A.L., Parkin, M., Hubbell, E., Webster, T., Mei, R., Veitch, J., Collins, P.J., Handsaker, R., Lincoln, S., Nizzari, M., Blume, J., Jones, K.W., Rava, R., Daly, M.J., Gabriel, S.B., and Altshuler, D. (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* **40**, 1166–1174.
12. Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., Macarthur, D.G., Macdonald, J.R., Onyiah, I., Pang, A.W., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W., and Hurles, M.E. (2010) Origins and functional impact of copy number variation in the human genome. *Nature* **464,** 704–712.
13. Conrad, D.F., Bird, C., Blackburne, B., Lindsay, S., Mamanova, L., Lee, C., Turner, D.J., and Hurles, M.E. (2010) Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. *Nat. Genet*. **42**, 385–391.
14. Kozlowski, P., Jasinska, A.J., and Kwiatkowski, D.J. (2008) New applications and developments in the use of multiplex ligation-dependent probe amplification. *Electrophoresis* **29**, 4627–4636.
15. Sellner, L.N. and Taylor, G.R. (2004) MLPA and MAPH: new techniques for detection of gene deletions. *Hum. Mutat.* **23**, 413–419.
16. Schouten, J.P., McElgunn, C.J., Waaijer, R., Zwijnenburg, D., Diepvens, F., and Pals, G. (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res.* **30**, e57.
17. Nygren, A.O., Ameziane, N., Duarte, H.M., Vijzelaar, R.N., Waisfisz, Q., Hess, C.J., Schouten, J.P., and Errami, A. (2005) Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res*. **33**, e128.
18. Kozlowski, P., Bissler, J., Pei, Y., and Kwiatkowski, D.J. (2008) Analysis of PKD1 for genomic deletion by multiplex ligation-dependent probe assay: absence of hot spots. *Genomics* **91**, 203–208.
19. Consugar, M.B., Wong, W.C., Lundquist, P.A., Rossetti, S., Kubly, V.J., Walker, D.L., Rangel, L.J., Aspinwall, R., Niaudet, W.P., Ozen, S., David, A., Velinov, M., Bergstralh, E.J., Bae, K.T., Chapman, A.B., Guay-Woodford, L.M., Grantham, J.J., Torres, V.E., Sampson, J.R., Dawson, B.D., and Harris, P.C. (2008) Characterization of large rearrangements in autosomal dominant polycystic kidney disease and the PKD1/TSC2 contiguous gene syndrome. *Kidney Int.* **74**, 1468–1479.
20. Eldering, E., Spek, C.A., Aberson, H.L., Grummels, A., Derks, I.A., de Vos, A.F., McElgunn, C.J., and Schouten, J.P. (2003) Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. *Nucleic Acids Res*. **31**, e153.
21. Kozlowski, P., Lin, M., Meikle, L., and Kwiatkowski, D.J. (2007) Robust method for distinguishing heterozygous from homozygous transgenic alleles by multiplex ligation-dependent probe assay. *Biotechniques* **42**, 584, 586, 588.
22. Ohnesorg, T., Eggers, S., Leonhard, W.N., Sinclair, A.H., and White, S.J. (2009) Rapid high-throughput analysis of DNaseI hypersensitive sites using a modified multiplex ligation-dependent probe amplification approach. *BMC Genomics* **10**, 412.
23. Liang, M.C., Ma, J., Chen, L., Kozlowski, P., Qin, W., Li, D., Goto, J., Shimamura, T., Hayes, D.N., Meyerson, M., Kwiatkowski, D.J., and Wong, K.K. (2010) TSC1 loss synergizes with KRAS activation in lung cancer development in the mouse and confers rapamycin sensitivity. *Oncogene* **29**, 1588–1597.
24. White, S.J., Vink, G.R., Kriek, M., Wuyts, W., Schouten, J., Bakker, B., Breuning, M.H., and den Dunnen, J.T. (2004) Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses. *Hum. Mutat.* **24**, 86–92.
25. Kantarci, S. and Donahoe, P.K. (2007) Congenital diaphragmatic hernia (CDH) etiology as revealed by pathway genetics. *Am. J. Med. Genet. C Semin. Med. Genet*. **145**, 217–226.

26. Sanchez-Mejias, A., Nunez-Torres, R., Fernandez, R.M., Antinolo, G., and Borrego, S. (2010) Novel MLPA procedure using self-designed probes allows comprehensive analysis for CNVs of the genes involved in Hirschsprung disease. *BMC Med. Genet*. **11**, 71.
27. Serizawa, R.R., Ralfkiaer, U., Dahl, C., Lam, G.W., Hansen, A.B., Steven, K., Horn, T., and Guldberg, P. (2010) Custom-designed MLPA using multiple short synthetic probes: application to methylation analysis of five promoter CpG islands in tumor and urine specimens from patients with bladder cancer. *J. Mol. Diagn*. **12**, 402–408.
28. Coutton, C., Monnier, N., Rendu, J., and Lunardi, J. (2010) Development of a multiplex ligation-dependent probe amplification (MLPA) assay for quantification of the OCRL1 gene. *Clin. Biochem*. **43**, 609–614.
29. Murray, S., Dahabreh, I.J., Linardou, H., Manoloukos, M., Bafaloukos, D., and Kosmidis, P. (2008) Somatic mutations of the tyrosine kinase domain of epidermal growth factor receptor and tyrosine kinase inhibitor response to TKIs in non-small cell lung cancer: an analytical database. *J. Thorac. Oncol.* **3**, 832–839.
30. Gazdar, A.F. (2009) Activating and resistance mutations of EGFR in non-small-cell lung cancer: role in clinical response to EGFR tyrosine kinase inhibitors. *Oncogene* **28(Suppl 1)**, S24–31.
31. Lynch, T.J., Bell, D.W., Sordella, R., Gurubhagavatula, S., Okimoto, R.A., Brannigan, B.W., Harris, P.L., Haserlat, S.M., Supko, J.G., Haluska, F.G., Louis, D.N., Christiani, D.C., Settleman, J., and Haber, D.A. (2004) Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139.
32. Paez, J.G., Janne, P.A., Lee, J.C., Tracy, S., Greulich, H., Gabriel, S., Herman, P., Kaye, F.J., Lindeman, N., Boggon, T.J., Naoki, K., Sasaki, H., Fujii, Y., Eck, M.J., Sellers, W.R., Johnson, B.E., and Meyerson, M. (2004) EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500.
33. Kobayashi, S., Boggon, T.J., Dayaram, T., Janne, P.A., Kocher, O., Meyerson, M., Johnson, B.E., Eck, M.J., Tenen, D.G., and Halmos, B. (2005) EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med*. **352**, 786–792.
34. Ercan, D., Zejnullahu, K., Yonesaka, K., Xiao, Y., Capelletti, M., Rogers, A., Lifshits, E., Brown, A., Lee, C., Christensen, J.G., Kwiatkowski, D.J., Engelman, J.A., and Janne, P.A. (2010) Amplification of EGFR T790M causes resistance to an irreversible EGFR inhibitor. *Oncogene* **29**, 2346–2356.
35. Itsara, A., Cooper, G.M., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R.M., Myers, R.M., Ridker, P.M., Chasman, D.I., Mefford, H., Ying, P., Nickerson, D.A., and Eichler, E.E. (2009) Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet*. **84**, 148–161.
36. Pedersen, K., Wiechec, E., Madsen, B.E., Overgaard, J., and Hansen, L.L. (2010) A simple way to evaluate self-designed probes for tumor specific multiplex ligation-dependent probe amplification (MLPA). *BMC Res. Notes* **3**, 179.
37. Wimmer, K., Yao, S., Claes, K., Kehrer-Sawatzki, H., Tinschert, S., De Raedt, T., Legius, E., Callens, T., Beiglbock, H., Maertens, O., and Messiaen, L. (2006) Spectrum of single- and multiexon NF1 copy number changes in a cohort of 1,100 unselected NF1 patients. *Genes Chromosomes Cancer* **45**, 265–276.
38. Zeng, F., Ren, Z.R., Huang, S.Z., Kalf, M., Mommersteeg, M., Smit, M., White, S., Jin, C.L., Xu, M., Zhou, D.W., Yan, J.B., Chen, M.J., van Beuningen, R., Huang, S.Z., den Dunnen, J., Zeng, Y.T., and Wu, Y. (2008) Array-MLPA: comprehensive detection of deletions and duplications and its application to DMD patients. *Hum. Mutat.* **29**, 190–197.

This article should be cited as follows:

Marcinkowska, M., Wong, K.-K., Kwiatkowski, D.J., and Kozlowski, P. (2010) Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: *EGFR* as an example. *TheScientificWorldJOURNAL* **10**, 2003–2018. DOI 10.1100/tsw.2010.195.

MATERIAŁY UZUPEŁNIAJĄCE DO PUBLIKACJI

Marcinkowska i wsp., *TheScientificWorldJOURNAL* 2010

Supplementary File 1

The reference sequence of EGFR extracted from UCSC GB (continued on the next pages)

```
>chr7:55054119-55242625
ctcctctgctcctcccgatccctcctccgccgcctggtccctcctcctcccgccctgcct
ccccGCGCCTCGGCCCGCGCGAGCTAGACGTCCGGGCAGCCCCCGGCGCAGCGCGGCCGC exon 1
AGCAGCCTCCGCCCCCCGCACGGTGTGAGCGCCCGACGCGGCCGAGGCGGCCGGGAGTCCC
GAGCTAgccccggcggccgccgccgcccAGACCGGACGACAGGCCACCTCGTCGGCGTCC
GCCCGAGTCCCCGCCTCGCCGCCAACGCCACAACCACCGCGCACGGCCCCCTGACTCCGT
CCAGTATTGATCGGGAGAGCCGGAGCGAGCTCTTCGGGGAGCAGCGAGCGGCGCCTCCGG EGFR_e1
GACGGCCGGGCCAGCGCTCCTGGCGCTGCTGGCTGCGCTCTGCCCGGCGAGTCGGGCTCT
GGAGGAAAAGAAAGGTAAGGGCGTGTCTcgccggctcccgcgccgcccccggatcgcgcc
ccggaccccgcagccccgcccaaccgcgcaccggcgcaccggctcggcgcccgcgcccccg
cccgTCCTTTCCTGTTTCCTTGAGATCAGCTGCGCCGCCGACCGGGACCGCGGGAGGAAC
GGGACGTTTCGTTCTTCGGCCGGGAGAGTCTGGGGCGGGCGGAGGAGGAGACGCGTGGGA
CACCGGGCTGCAGGCCAGGCGGGGAACGGCCGCCGGGACCTCCGGCGCCCCGAACCGCTC
CCAACTTTCTTCCCTCACTTTCCCCGCCCAGCTGCGCAGGATCGGCGTCAGTGGGCGAAA
GCCGGGTGCTGGTGGGCGCCTGGGGCCGGGGTCCCGCACGTGCGCCCCGCGCTGTCTTCC
CAGGGCGCGACGGGGTCCTGGCGCGCACCCGAGGGGCGGGCGCTGCCCACCCGCCGAGAC
TGCACTGTTTAGGGAAGCTGAGGAAGGAACCCAAAAATACAGCCTCCCCTCGGACCCCGC
GGGACAGGCGGCTTTCTGAGAGGACCTCCCCGCCTCCGCCTCCGCGCAGGTCTCAAACT
GAAGCCGGCGCCCGCCAGCCTGGCCCCGGCCCCTCTCCAGGTCCCCGCGATCCTCGTTCC
CCAGTGTGGAGTCGCAGCCTCGACCTGGGAGCTGGGAGAACTCGTCTACCACCACCTGCG
GCTCCCGGGGAGGGGTGGTGCTGGCGGCGGTTAGTTTCCTCGTTGGCAAAAGGCAGGTGG
GGTCCGACCCGCCCCTTGGGCGCAGACCCCGGCCGCTCGCCTCGCCCGGTGCGCCCTCGT
CTTGCCTATCCAAGAGTGCCCCCCACCTCCCGGGGACCCCAGCTCCCTCCTGGGCGCCCG
CGCCGAAAGCCCCAGGCTCTCCTTCGATGGCCGCCTCGCGGAGACGTCCGGGTCTGCTCC
ACCTGCAGCCCTTCGGTCGCGCCTGGGCTTCGCGGTGGAGCGGGACGCGGCTGTCCGGCC
ACTGCAGGGGGGGATCGCGGGACTCTTGAGCGGAAGCCCCGGAAGCAGAGCTCATCCTGG
CCAACACCATGGTGTTTCAAAATGGGGCTCACAGCAAACTTCTCCTCAAAACCCGGAGAC
TTTCTTTCTTGGATGTCTCTTTTTGCTGTTTGAAGAATTTGAGCCAACCAAAATATTAAA
CCTGTCTTacacacacacacacacacacacacacacacacacCGGATTGCTGTCCCTGGT
TCAAGTGTGCCAAGTGTGCAGACAGAACATGAGCGAGTCTGGCTTCGTGACTACCGACCA
TAAACCCACTTGACAGGGGAAACATGCCTTGGAAGGTTTAATTGCACAATTCCAACCTTG
AGCTGCGCGGGTTCCAAGAGCCAGGCCCGTACTTGCTGTTGATGTCATTGGCTTGGGGAG
TTGGGGTTTGGTGCCCAGCGCGGTCGTTGGGGGAGGGGCAAGGCATAGAAcagtggttcc
cagaccttgctgcacattggaattacctgggattaaaaaaaaaaaatcaaaacaaaaac
cagtgtctggctcccgccccagacattctgatttaattggcatggggcaagacctggac
ttgggatttttttaatgtcttcatgtgatctgttgggcagccagatttggggatcact
AGACGGAAGAAGGATTGTTAAAGTCTCCGGAGATGTTACTTGCCAATGCTAAGAGCTCTT
TGAGGACATCTGGAATTGTTACAATATTGCCAAATATAGGAAAGAGGGAAAAGGTAGAGT
GTGATTCCAATAATAAAGGATTCCGCTTTTCATTGAAGGAACTGGTGGAAAGGTTTCTTC
TCTGCTGAGCCTGCAGGCCCGTCCTGCCTGCCTGGGGTGCCCGGGAGACGCGGGCCTGCT
CCGGAGACTGCTGACTGCCGGTCCTGTTAGTCAGGTGTCAGCCCTGTCTCTGCCGAAGAG
ACTCTTCTCTTTATTTAAATTAAACCCTCAGAGCACCACCAAAGCATCACTTTTCTCCC
TCCATTGGTGTTCTCATTCTTTGATGTTACTTGTTTGAACACCACTATTAGTAGTTGGAG
ATTTGTTCCTGAGAAAAATATAAATACCACTTAATTTGCCTGTTTGTCCCGCATTCACTC
AAAACAGAATGCTCCTGAAGACAAGAGAGAGAGTAGGAGAACAGACGCTATTCCATTACA
GTAACATAAAAGACTGGATTTTCAGGGGCAAATTATTAAAATAGGAGATGAGCTCTTTTA
ACAGAAATTTGTTTAAGGCCTGTGTCTATCAAATTCAGTGGATTTTATTCAAGATGCACT
TTGTTTAGTGGGAGTTTTGTTTGGTTCTGGGACATGCTAACTTCTAGACTTGCTGCTCTT
AGAGGTAATGACTGCCAGACACCATTTCATGAGTCCtaatccccacattaagcataagag
gtgcacactctcctcctatggggaaactgaggtacgaagaactaaagtgactttcccac
agctggtgggaggcagacgggaaattcacaccaggggcttccaactccagatccctctct
caacttccaaactccactgCCTTGTCCGAGTTCTGGTTTCAGGAGATCCAAATCAGGTGT
GTGCAAATGTCTAATGTCAGAGCTGGCAAGGGGAAAGGGCCCAGGGAGCCGGCTCATGAC
GATGAGCCTGTCTGAAGCTTCAACGCGGGCTGTCCGGCAGTCTGCATTCCTGCCGAGTTC
CTCAGCCCTCTGTTGGGTCACCTTCCATAGAGGCAGCTTAGTCCTCAGTTCAGTGAGCAT
GGAGTGGAGACTGCTTGAGGGGTGCTGAGCAAAGCCCTGCCTCTTACAGGATGAAGGTGC
TCTCCAGAAGGGACACTGGAAAGTATTCCAAGGCGAGTCGAATTCCCAACTGAGGGAGCT
TTGTGGAAATAAGCCCGCCCAGCCCCACTTCTGGAGACGTTCCCATTCAGTAGGTCCGAG
CTGTCTTAAAGAGAAACCAAAGTGGGGATATTAATGGTATCCAAAGTGAGATCTACCCCA
CCCTCCCTCCTCAAAGGAGGTCAGATCAAGAAAGCCCAAGCCCGGCCTGGCAATTGGGAC
CTTTCTTCTCACTCCAGCCCAGGGTGAAGGTGGACAAGTCACTTTGACCCTTCAGGCTTC
TGAGCTGTTGTTTCTGAATTCAGTGAATATTTACTGAGTGCATAGAATATGCTAGATATT
CTGGGCTAAAGGTTGAAGGGGGGGTGAGTTTTAAGGGTTTCTGCTCTTGCTTCCAGATTG
CTTTCAAATCTGGAAAGGACACCAGTGGTTTGTGTGTTAGACCCACACTGCCGTAGCACA
GAATACAAGAAACTGGCTGAGAGCTCCAATAGGCTTTTAACAGTAATTTCTGGCTTCACG
TATTTAGTTTCATAACTCATGATTTTTCAAAAACTTCTGGTTTGAAGACACCGATTGCCG
AAAGTCCATTGTGCTGCATAATTACACTTGGTCCACGTGACAGCACTAACATGTTCTGAA
ATGTTTTTAGAAGTAGTCTCAGCAAAGATGAAGGATTCCTCCCTGTTTGAAAAGAAAATA
TTCTTTGTTTTTTCTTTGATCTAAGCTCTAAGACTAGCAGCTAGCATCTGAAACTTTTTT
GACGAGAGTGACAAACCAACTCTAATATTAAAGGCAATTGATGATTATGGGCACTGAAGG
GAAGGTAACCCCAGGCTGGTGCCCCGGAATAGGGATGGGTCACAATGTTGAGGACATTTC
GCCTGTTGCAGAACCCACCTGCAACACAGTGTGGCCCTTGCCATGTGACTtgtgtgtgtg
cctgtgtgtctgtgtgtgcgtgtTTTAATTTTGACTTCATAAGTACTCTAGTTATGAGCT
TATTTAACATTGGGTTTTACTAATAGGGGTATGTGTTGAGAAAATTTCAAAGTTTTAGAA
TATGGTCACCCACATGTTGCTTCCCTGTAAATATAATTTTAAAACCAGATTCTGggcc
gggcatggtggctcacctctataatcccaaaacgttgggaggccgaggcaggcgaatcat
gaagccaggagtttgagaccaggctgaccaacacggtgaaacccagtctctactaaaaat
acaaaaaaaaattagctgggcgtggtggcaggtgcctgtaatcccagctactcaggaggct
gaggcaggagaattgcttgaacccaggaggcagaggttgcagtgagccaagatcgcacca
ttgcactccagccccgcgcgacagtgtgagactccatctcaaaaaaaaaaaaaaaaaaaaCA
GATTCTGTTCCTCAGATCCATTCCATTTTTGTTTTCCTTTATCACTTATGGACATTTGAA
ATTATGGTAATAAACATTGTTAGTCTCAGTTAATTATTACTGGTTTATTCTTGAACCACT
AATCCATAGAGAATAGAGTGTAAATCTTAACTTGTTCCTGTAGGCCATCCCCATTAAACA
TCATAGTGTTTTCTCATTCGTTCTTTTTCGTTTTCCTCCTACAGGAATGAATTTTCTAAG
AAAATTCCAGCAGTTGGCTCTTTGGACGACATCTCTAGATTGTCCTCCATTGGGCCCATA
GGCACAAGCTGGCCAGTTTGAATTTGGGCAAGAATCCAGGCATTGGAACTTATTCAAATA
ACTAGTTTGCCTGTAATTTTCACTTTTTCAGAGTCATCTGATAAAGCTTTCTTGCTACAC
ATTTAGATAGATACACTCAATCCAGTTGTCTAGAAAGTTCCCTGAGCCAGCTGGGAGCAG
GAGGGGTAGTTGGGGCCAGGAATATTGGGGGTGTGTTTACTGAGCCCCTAGAAAGTAAGT
GCTAGATTTGACATTTCAATCCCTGAAGGCCCTGAAGTTCAGTATCAAATGACTGGTCCT
GTGGACTGAGCATCTGTGAATTGCATATGCTTAGAGTAAATTTTACTCCTACCAGTTTCA
GCAGCTTGCTTTAGCAAGCAGTATGGAAACACTAACATGGGGGAGTAGAATTTCTCTCTC
TGATCCAAGTTTTATCTCATTCTGGTGGGTTTTCAAGGAGAGACTCGGAGTCCAAGTGTC
CTTTCTGAATATATCTGGAACTTCTCATTAACAAAAGACTCAAGTTATAATTTAGGGGAC
AAGGCACCCAATGAGAATGCCTTGCAGGCAGCCCTAAGTACACCTGCAATTACACCATTA
CTAGCGCGGCAGCACACATGGCCCTGACTTAGTTTAAATAATTACGTAAGTCAACCATGA
TTGTTTGCCCTTTGCATAGAAGGGCAAGTATTGGTACCTGTTACAACTTAGGCTTTTTTT
TCTTTATGTTTGAGCCATGATGAGTGATTTACACTGTTGCATCCATATGTTGAGATGTAA
GAATAAATTAGACTTGGTAATTGCCCTTAAGTGTCTGGAAGTCAACTGGGGAAAGAGAGC
TAGAGATAATAAGTGTGAAACAATGTCACAGAATCAATGACGGAACTCTTCCCAGGACAA
AGGATGACTTTTGAGTTCAGTCTTTGCCTTTAATTCTACATGGGGAGGAGAGCACGTTTA
GCCACAAATGGAAGGGATTACTCATTTGAGCTATTTGGTTATATGATTATTTCCCCAGAG
AATAGGATGTGCAGGGCATTACACAAGCAGTGCCAATAGCAGCAAAGTTCTTGAGAGTGC
TAGTAATTCAAATGGCAGGAAGAGAAGGAATAAATGGTAAGGCTACCTACAGTTCACAGA
GAGCTCCATCCTcactgtggctttggattttgtcctgtgtgaaagagaagtgactgtgaa
ctgacatgctgtgtttggtgttttagaaagatggctgcagcagcggtttggggaatggac
tgcaggagtggcattggaaacaggaaggttcatgactattgccagagacagaggatgaag
caggagcaaggaagattcaggacaggggactccggggctgatcaggaggcagaactggtt
gataagtatatgtagcagcataagaaagaaagaatcccagattgacacccaggcttctca
cttggaagcctggatagatactgaatgcaatcacaaaggctgggaagtcaatgggactgc
agggaagggaagggaagggaggagaagaggaagggcaggagggtccaatatcaatattca
gcttttagatgtgttgagcttgaagtgctcagatggagaagtccaggaggcagtagaata
cggtggtccagagcacaggagagcaatgtggcttgagTTGTCATTTGCTCACATATTTCC
GTGTCAGTTACTTGTCTTAGATCACAGAACAAGTTCTCCTCTCACAGTTTCCTGGCTCCA
CCTGTCTCATGCTCACCGTCAGCATCGAAATTGAGCCACACCAGGGGTTCTGGATACCAG
CTTCTCTCTAGGTGAGGCTGCTATAGTCAGCAGCTGATTAGTTGCAGTTATCAGCAACTG
GTAATATAATATATTGTGCATATAAGTGTACCAGAAGTCATGTTTATATATTGCTGCAAA
TACTCGGAATGGGGATCTCTTGTTCCCTGCTTAAGACCACATCACATTACTTGGTTTTGT
ACGCTAGTGGCTGAACCAAAAAAAGTAGGAGATGATTTTTTTCTTTTTTCTTAAAGcag
tagcttttgaaccttgaccatgctttctaaccagctgaggggcttttgaaaaagagggtg
ccttactgtgccccagaccaggacaatcagtatttctggggaatggagcctggcacacac
acatttcttaaagctcccttggcaattcTGAGGAGTGGATTACATGTTGTATGTAGCTCG
TAACGAAAGAAATCTTGTCTTTGCTCTCAGACCCCCATTTCTACTCATCTCATGAGCTC
CTTCGAGATCCAGAAACAGTTGCATATTTCATTAGTAAATCAGTTCCAGAGTCACATTTT
ATTTCACAAGTTAGTCCATTAAAAGTTTCCTGCAGTGaggaaatagccagaaagaacact
ccacccctcctcctttttataactatagggtctggCTCgacagagcaggagcatcgccat
cttggacaagcccctcattctaaagttcaccttaataaaaaactgcctaaattcaaactg
catcagcctaatggctaaggtcagcatgaccataaaccacaaataacatctccaaccgga
aacattcgaaactcctcctcgaccagagacatgctagtcccgagataaccccccctccagc
agggaagatgccagtctcgggataacctctctctggccggaaagatgcctgccccaagat
aaacttgcctcctcccagagatattccaaccctgccataaaacttctccctcaaacagga
acattccaaaattctgataatctccctcaccctaaaaccaatatatactcctagtctgta
agagaaagcgctcttgaccaaaattcaccaggagtgcctcccaggttttaactaaagaaa
acctctctttaactgCCAAAAAAAAAAAGGGAAAAAAAAAAGCTTTCTGcagtggctttc
agcgggcccagcatggcagcagcacctgagaacctgttggagatgcacactcttggaccc
caccctggcctctgagtaagacactggaagggcaggccccggtctgtgcacacaagtcct
cagggagattctgactgatgcatgccagattttgagaactgctgATATACTCCAGGCACA
TCGCATGCTGGGATCTAGATACACCAAGGGAACAAAATAACTGCACTTGTCCTCTGAGGA
CCGACTTACCTTTGGAAGGGCTGAGAAAGAGAGACACATACAAGATCACTCCCTGTAAT
GCAATGTTTTATAACAGATGTGATTTGGGATTTCAGTGGGAGCCCAAAAGAGGGACTGAC
TAATTCAGCCTCTGTGACAAGGGGAGTTTCTCAGAAACAGAATGCTTAGCTGGGCCTCCA
GGCACAGGGACAGGAATGAGGAAATACTTGTAGGCCCTGTGCTCCTTCAGCAAAACCCTC
AGTTTCTTGTTATTTTTATAAATGCAAACATCTTATTAAAGTAGATGCTAAGGCATTAGA
ATTTCCTGCTTTATTTTTCTAAATGACCATGAGGAAACCTGGAATGTCAAAGATAAAGTG
CAACACATTCTGCATTTAAAAATTAAAATGATCCTTTTTAAAAGTAGCAACCAGATGTGA
AAAATTGGACTGGAGTCCAGGTTATAGTTGATAGCTTTAACTTTCTCCCCAACAGCAACA
GCACAATTTTCCCTAAAATGTGTTATGAATAAGTAAAATGACTACTTCACATCCTTTAAC
TCTTCCTACAGAAATCTAAGAGAGAAATGAAACAAAAGTTTGCACAGTTCTAGACACGAT
AAATACATGTGAAATCACACAACTCAGAAAATGTCCCTTAAATTAATTGAGCCATTGGTA
CTTGTGAATTAGAAGAGACATCTATGTTCTGATCCACTGTTGAAAGCTGTACAATGTTAC
CTATTTATTTGCAGACATCCTTTGGAAACAAATAGGTAGATTTGCAACAAATAAAGAGTG
GAGTACAGCTGCTGACATTACCTTGTATATTCATGCCTTTATGTAAAAAAAAAAAAAAAA
atatatatatatatatatatatatatatatatacacacacacacacatatGGAGGTAA
AGACCACTGCTTGCTTTGCAGTTGTTTTAAGAGCATTCATGAAGGATTTTATTTTATAAG
```

CAGAAATGTGATATCTGACGATTTTACCACTACATGCTTGCAGGCCAGTGCACAGCAGAT
GACGTCATGATTGTTTTAGCAGTCCTATCGTTTTACTTATGATGTCATTACAACCCTTTG
CTAAAATTTCTTTCCTTTACTCCAGGTTTTGGATAAAATTGATGCATTGCACATAGTCTC
TCTGATAAGACAAACTGGCATTTGTATGTGAAAAACTGTGCATGTTTTAGTGTCTCTGCT
GATACTCAAATTATCCATTATTTTAGTGCTGGAATAAAAACAAACCACTTAGTGAATTTG
TGCAGGTCCTTAAGGACAGGCAAAGGTGTCCTGAGATTTTCTGATCATTGTATACCAAAT
TTTAGAAACTTTTTCAAAAACATTTTTTTAATTTCAAAAACCTGGTTTTGTTTATTTACC
AGCAATCATTGAATACCTGAAAGCTTTCAGGAGATTTTATTACAATGGTTTCTATTCACT
TACAAAATTATCTCCTAGTTCATTCTCATACACTGTAAGCCATTGTAAATGCTTCAAATT
GTGCCGAACAAGATAAACTAGACAAACTATTTTAAGTTTGTTCTAGTGCTAACTTGCAAG
ATCTAATGGCTCCAACTAGattttaaaataaagtatattttaatatattattaGAAAGT
TAAGCAATTATCTGTTTATAGGTAACAAAAACCCTGGAACCCCAATGTCAGATGTCATCC
ACTTTTGATTAAGTCCAAACATATGACAGATAAACAAAAGATGGTTggctgggctcagtg
gctcatgcctgtaatctcagcactttcagaggccgaggcgggcggatcacaaggtcagga
gtttgagacttgcctgaccaacatggtgaaacccgcctctactaaaaatacaaaaaaaac
agctgggtgcggtggcacgtgcctgtagtcccagctactcaggaggctgaggcaggagaa
tcacttaaacctggaaggcaggggttgcagtgagctgagatcacaccactacactccagc
ctaggcgacagagcaagactcagtcaaaaaacaaaaaaaaaGTGGTCATTGGAGAATTAT
TGTGTCACCTGTTGTTTTTTAATGTACTAATTTTGAGAGGCTTTTAAATAGAGTGCACTA
TAGAACTTTTTCTTGGCTTCAATTTGCTACAATGTTAATAGAGAATCAGAAACCTTATCC
TTATAGATGTTTCTTGATTTTTTAATTTCTGGTGACATTTATGAGTGAGAATAGTGTAT
TGCCCTGTtttctttcttactccccttcttcttccttccttgctttctttcttcttcccc
ttccttctttctcttcctcgctccttcttttttACAAGCTGTTATGAATTAGCCTTCACA
GAGAAAGAAAAATTTTTATAAATAACTGGAAATGAAACTTTGCAAAGGACTGCAGATGAA
AAACTTTGTCAAATGACTGTAAAAATATACTATATAATTTTCAAAAGTTAGAAAGTACCA
AACACACTCAGTATTCATGGTTATACAAGTATGCATACACATGTATTGCTCCCTGAAAAG
TGGTGTTGTTAAGGGAGTTTTTCTTAGTACGCGGCTTAACATATTTTTTTCTGTAATTTG
TTGTTAGTTATAATGGGGAGAGAAAACAGGTTAGAGTCTCCCCTCTCAGTTTCACCTTCC
ATAAAACAGCTAAACTAGACGATCGTCAGACTCCTTCCAGCTGAAAACATCTGTAAAATT
AAAAACAAATCTAAATGTATGCAAGATATGTATTTAAACATGCTGGTAATAAGTGTGCTG
TCCCTATAATTTAGATGCTAAAACATTGATGTCATAATAATAACAACACCTCGCATTTGT
ACAGCACCTCATAGTTTACACAATGCCTTAACATTCTTCTCTCTCAGCCTCCTACAACCC
CACAGGATTGGGATAGCTTTCCAGATTGGGAGGTGAGGGACCCAGGCTCAGAGCGATTCT
GCTGTTGTCCGTAATCACCAGGCTGGTGATCAGTGGGCACTGGGTGCTCTCCTGCTACAC
AGCACTGTCTCTCAACATGCAGGTCAAGGTTACTTATTCCTCCTTCAAGACGTCATTGGG
TTTTTTAGCTATGGATGCCCCATCACTTTTAGTTCTATTTGTGAATCAAAGGCTAAATAA
AGTATTCCTCAAAATTTGTTATACTTCTGTTACTAATGCTTAATGTCCCTCACAATTTCT
GTATATTTCTGTGTATTTCTGCTCTGTTTTGGTTCCTTTCCCAGGTTTCTTTTTTGTTAT
GAAGTAGTTTTTAGACTCAAGTCTCTTCTGTATGTGTTATAACTGCCCATTCCATAAGAT
ACAGGGCAGTGAATTTGTGAGCCTTGAAAATATTTACTTTAGAAATGAGAAGTATGACTT
TTCAACGTTGTGTCATCAACTTCTGTAAATTTTCCAGACCTATAAATACTTGCAGAAAAA
AAATGAAAGGAGAAGGCAACTTGATTTAGCAGTTGGGTCAGTTAGCAATGCCTATGGCAA
GCTGTAGTAATTCCCTTACATAGATTTGTAAGACTCATTTCTATGATTTAAATGAAGGCA
TACACTTAACCTCTTTAGGGTGTGAAACAGCTTTTACAAAAAGAGACAAACTTAAGAAAC
AGTGTGGCCCTCCAAGAGTGTTCATTTTCCATATCATACCATTTGTAATAAGCTATTCTG
GCTGGGATTTACTTGCAAGCATTGGCTTTTAAGAAGAGATGGTTTCACACATCAAATTAT
TCACTTGGAGGCACTTTCTGGGTTGAAGGAATGGAATGGAGAGTGCGGCAGTGAGTAGAT
CTCTCAGTGACGGTGATGTGCCTCTCCCAGAAGAAATTTCAAAATGCAGTGTTCATTTTC
CTCCACAAGAAAGGAAGAAACTGTTTTGTTATTGTTTATTCCTAACATAGTGGAAACTTT
TCAGTACTCTGGCAGAAATTTCCCAAAAGCAATTTTCTATTTCATGATTATAAAGTAGCA
AAGGAAAAAGTCCTGCACTCCAGCTGAGCAATGGATCTCCAGTTGTTATCTAGGTGCTGC
AGGTTTAGAGAGGATTGCCAGGAGAACACATCGATTTTTCAGGCCTGTGATGACGTATCT
CTTGTTGAATAAGTAAACCCTTCCAGTAAACAGACAGTTAGTATATTGATTTCAGGGTGG
CTTTAGCCACTGAACCTGTAAGTCTTGCAAAGGTTACTTGGGCAAAAGCATCATTATTTT
ACCTTCAGTCAACAAAAATCTACCTGGCCAAGGCAGAACAGAAAGTTCAGCAATTTGATG
AAGTGGGACAACATGAAGAATCAGGTGAGTTGCCTACTTTTTCACTTCACTTTCCACCTT
TAGAGATTCTTGTTTAGATGCAGAGTAGTGACGTGCCTGGTGTCAGGGAGAGAGTTGAAT
GAGAAAAGTCCCAGAAGGGCAGAAGACTTGGGTGATTATCTGAGTCCATCTTTCCTTATC
ACATGACAGAGTTCTTGAAGTCTTGGCTAGGAATTCTAGGCTTTTAGATTCTTTGGGCAA
TGGCTACTAAATGTTCATAATGTTGCTCAGTTGCAAAAACAAGACATTCAAACTATAGCC
AGGGAGATAAGTAGTCACGAACTCAAGGCCTAAATTCTGCTGATGGAGCCGATGAGAATT
GGGTGCTAAGGCAAAGAGAGTTGCCAATATTATATTCTTCGGGGTTTTTTGTTTTTATTC
GCATTTTGGAAAAGGAAAATATTAGCATTCCTCTGACTTAATATTGAGAAGACATTGGGC
ACTCTTTTTCCTCCCACACTTGTCTTCTTTCACTAGGTGACAAGGGAAGAGGTAGCATGA
GGTGGTGGTCACAGGTGAGAGGGGCTGTTGTGAGCACAGGCATGTTGACTGCACATTGGT
CACCTAGTAGAAGTTTTGCAGGCTTGGTGACTTCTGAACACTGTTTTCAAGGTTGATTTT
TAGTTGAGAGAACCTCTAGGTACCACGTAATGTTATTAACAGTAGTACTGATCTCACAAT
CGCCCTATGTCCCATTCACAAGATGTTCTGCCAAGCCATAAAAGGCCCAGTTAAGTTTAA
GAGAAGTCTCAAAAGTAACAGATGATAACTAATTAATACCCAGTGATTTTGAAATGTAGA
CATCAAACATACCAATTCAGTGGTATCATCCTTAGAGGCAGACAGAGGATGATTAAATCA
TTCAGCCCATCTCTGTCTGAGGACGCAGCTTAGCACAGCATGGTGGAGGCTAAATGGGCC
TTAAGGGAAAAAATGATATCTGAAGATGCAATTTATTTCAAAAAGAGTTTGCTCCCGTGA
ATTTTCACTCTCTATGTAGAACGGCACCAGCACACACTTTTCCTGAGCCTTTGCATGTGT
GGCAGGCAGCGGCCTGGCATCCTGGGGAACTGAATGAGGACGCAGATGACCCGGACGTGT
TCACAGTTGACACATCTGACTCCCAGATCAGGGACAGCTAGCTTTGCTGGCTGGTTAAG
TTGATGATTCCATCTTTGCCTGGTTCTCTGACTGTCTCATGCTTTCTGTTATTACTATTT
TGCAGCAGATATTTCTGCTCATTTTTCAATCATATATGCATCCTGGATGGCATAGAGTTG

ATTCTCCTAACAAATCAGTGTCCCTTTGTATTTTTTTCTGGCCATAAGATAGAATATATA
TGTCATTTATTAAAAATGGAGAAAATGTTCAGGAGTTTCTTGACTCAGAGAGGGAAAAGG
GATACTCAGGGCACTTTTTCAGCCAGGAATTTACTACCTTTGCAGGGTAAAGGGGACTCA
CCACGCTGGAAGTCAAAATAAGCCACCAGTGCCAAGTGTTCAAAGCCCTTAGAATCACAA
TGCTCTTAAAGCAAAGTCTTCAACAATGCTTGAAAACTTCCACTGGTTCTCAGTATGTCC
AAAATTGTCATGTCTATGAATGATTTTCTCAATCTGAAAATTTTTATAGCAGGCTAAAGA
ATGAGATAGGTCAGTGTGATTCTAGAACTAATCATTAACATTCAATAGATGACTATTTTA
TTCTAGAAAAAGCAGCAACTTTCTATTTACTCTCTATTTTGAGGGTAAATTCTCTGTAAG
TAGAAAAAGCAAAATGTGGACATGGGACTAACATATGAATATACAAAGCAAATGTACCGA
AAAAATCTTAAGACCTGCCTTGTGGTGTTTTTTGTTTTGTTTTGTTTTCATTAAAGTGAC
TTGTTAGCCTCTTGCTCCCTGTGAAGCACAGGGAGGTGACGTGATGTGCACAGGGCAGAC
TCTGCCATATGCCCTGGCCTTGAACTCAGGGCCCCCTGGGGACTGCAGGGGATGCTGGCC
ATGCTGAGCAATGCctgtgggtgtcagtttcctcatctgcagaatgagggtaggcctggt
gcttatttcatagggtcgcagaggggattcagtgaCAGGGTGGTGTAGAGGCTGGAGCGT
GCCCCATGTGTGCACGACAGCCTTCCAACTAGGGGAGGCGGGCCTGGGCTCTCACCAGAG
AGCCTGTGTCTCCATGGCTACATGACTTTGCCCCAGACGTCCTTCCCGTGGTCTGGACC
CTGGGAAGTCGCCAAGAGCCAGACAGGAGAAAGGCTCCACTTGGCTCTCCTCTTTGGTGA
CCATCCCTTGCCTCCATGGCGGGACTCTCAGGTGACATCCCACCAACCCTCACTTTGCTT
CCCTGGTGGGTCTCACTTTCCCTCAAGAGTGTTGCTTTTTTGTTTCCTGCATAGTCCTGG
GCCAGTTTTGATAACCCTCTTCATTTCACTTCAGAAACCCTGATGATTTCTTCCTGTGCT
CTTTTTACCTTAGGACTTTTACTATGACGACTGTGACTGGCCCATTTCTTGTTTTTTTTTC
TCTTGCTCTGCTTTCTCCCCCATCATCACTAAAGCAGACATGGCAATGATGGCCATGCAC
ACTTTCCAAGGGTCCAGCTGTAGATCTTCATGGTTCCCCAGGTGCCTGGACCATCTTGTG
AGGAGGGAGGCAAACACACCCTGCCTGGAGCACTTGGCCCTTTCGGCAATGTTTTGGCTT
CCTCAAGTGAGAAAAGAATGGATTTGTATTCCCCCTCTGCATTATTGttttgttttgtt
tgtttgttttgttttgtattgagacagagtctcactttttccccaggctggagtgcagt
ggccccgacctcggctcactgcaacctccaccttccgggttcaagtgattctcctgtctca
gcccctgagtagctgggactacaggtgcccgccaccacacctgactaatttttgtatgt
tttgtagagacagggtttcaccatgttggccaggTGCCCATTATTATTTGATCTGGAATT
AACTGAGCTACTGCAGGAATTGCTTGATTCACTGATGACTGGTGTTGAGCCAGTACACAC
CCACACCCAAGGACTGTGACTGTCTTCTGAGGTCCATCCTCAGAAATTCCTGTCTCTTCA
CCTAGTGTGTAATAAGGCCTGCGCGTGTTATATGGAACTGTAAAAAATGCGCCAACCATC
TGTCCTTCCTCTTTATCTGATTACTTATCATTGTTCTCTAAGTTGCAAGTTAATAGACTG
ATCATAAATTAATGCATGCTGGAGACTTGCTGTTTCCTACTAGCAGCATATAAAAGTTAT
TTTTAAAGTTGTTTTAAATCTGTGAGTAAAAATAAATTGCTTTGCTGCAAGAAACACCAA
ACATGGAAAAGCTAACGGTTCAAAGTTAATAATTTATCTTATGGACATCACTAGTGGCAT
AGTTGCTTTAAACAGTGAGAGGATTTAATAGATATTTGATTTGCAAGTGGGATGAAGGGT
GGTCTAACCTTTGTCCTGTGTTTACCTTCCATGAGATCCTAGAGGTTGTACAGCACAGTA
GTGGCATGTGACACACTTGAGAGTGCCTGTTCTGTTTGGAAACCTGGAAACTATGAAGGG
AAGTGGCCTTCGAGCTTAACACATAAGACTTGGGAGGCAAAACCTTTTATTCTCTTTAAA
TATTCACTTTAGGATAAGCATTTTTTTAGGTGTTAGGAACAGGGAAAACTGTGTGGTTAG
GAAGGAAGAAAGAAGAAAGTTAACTGTTGTACATTCCCTAGGTAATGTTTTTAAGCATTG
TTATTCACTTTCAAAACACATTTTATTTATTTGGACTTAATATTTTGATCTTATTTTTTC
AATTTCTTTTAATTTAACAGACAGGATGAGTTTTTTTATAGTTGTATTACTTAGAAATTA
TACTAAAAATggccgagtgtggtggctcacacctgtaatcccagcactttgggaggccaa
ggcaggtggatcacttggatcacttgaggttgggagttcaagaccagcctggccaacata
gcaaaaccccgtcttcactaaaaaaaaaaaacaaaaaaaaaaactagccacgcatggtggc
aggtgtgcctgtaaccctatctactagggagactgagacataagaatcacttgaatccag
gaagcagaggttgcagtgagcagagattgcaccactgcacttcagcctgggtgacagagc
aagactctgtctcggaaaaaaaaaaaaaaaaggataaagaaaTCATACTAAAAACAAAAC
AGAATGCTGACCACCTTATAGAAATAGAAATAGTGGTTTGCTGTGATAGCAAATTTTCTT
GTTaactttttattttaaagaattgcacattcacaggaagttgcaaaaaatctactggg
aggtcctatccccttccccaacctcctccagtagtaacatcttagtagcaaagttttg
tatatttattttgatatcattatctaagtttgacatcattatctaatattaacctaagcc
AAAAGCCCACTATTTTAATTATCTAGTGATGCAGTGTTATAGAACTCATAGCCTTTCACA
GCATTATTTGGAAGTTAATTTTCTTAAGTGAAATGTTTTTGGTCTTTAAGGTTTGGAGGC
CATGGAGGCATGAGGAGAAATGGGATGAGGGAGAGAGAGCTAAGATAGATAAAGACAGAG
ATGGGGAGATCCACTGATTCGTTGAACAAACCAGATACTCCTTATAGTTTTTGGATTAA
CTTACATGAGCTAAGTTTATATTCTGTTCAGATCACAAGTGGTCAAGTTTGTGTGTGTGT
GGGGGGGGGGGGTGGGTGTGTGTGTGTACCACTCTACCCATCCTATATTTATTGTCCTG
TATTTGGTCTGTTCTGCCTTCTTTATTTTCAGGATAGGTGTCCTAAatgagggtctttgg
aaagctggtgaggccatgttgcccgtttcaggtgttccgtgctcaaatgtattcatttct
tgaaaaattcagggagtgcacacttttgtacatttccctatgtgtatatgataccattat
ataaatcttaaaaaTATATATGGTTCACCTGAATCCCCAGCCATTTGGTAGAGAAGATAG
AAAACCTACAGAGGAGGCTAAGATTTTATTAGAAAATTCAGCTTCTCGACGGAGGTATTG
GCTTTAAAGTCAAGGCAATGCATCTATTCTTTCTTTTGATATAACTAGCTAAAAGATCTC
TTAAATTCAAAGTGGCCCTCATCTTACTGTTACTGCAATTTACTCTTAATTACAAATTAT
ATAAAAATAGTTTTGAAATACTGTAGCGACAAAGTAACATACCTCTGCTCCATTACACA
GATAAAACCTCTAAGGAACACCTCCTCTCTTAACAGGCATTAACCAACTGCAGAAACTGC
AGAAGGACAGGGCTATTTGGGAATAACACAGCTCCCTTCCTTGTCTGTTCCCTCCCATTG
TCAGGCTTCTGTGGAGCCATATTCAGAGCAACATAGGGAGGGGGAAGAGAAAATCAACCC
CTTGGTGAAGGAAAGCTCCCAATTCACAGAGCAAACATGGGTACTCTTGTTTGTGGGAGC
TCCCAGGGCCTCCCAGCTCACCGAGCATTCTGAGCCCTGATCCTTACACTAATTGTATTA
TGCAACCATAAATGATGTCTGCTGTACCAGCGGGGACAGTTTATTTTAATAGATTGGTAT
AACTTGGCAGAATCTTATCTGCATGTTTCATCTTGGATTTTAGCTCAATTCAACTCAAT
AGGCATGTGTCAAATGTCTACTGCAGACTGAGCACTGAAAAGCTGCTGGGTACAGGGTTA
CATGGATAGAAAACGTAGCCTCTGACCCCTAAGGAGCCTGTAATCCAGATCCCCATTCTT

TCCATCCCATTCTCCCAAGCAAGAATTTACCTAATGTGGTTTGCGAGAATTTAAGAGCTG
GAAAGGTGGTCACGAGAAGCCGGAATGGGTTCGCTAAAATGTGTCTATATGATTAAGCAT
AACGTAGCTTTGCAGCACTCTTCACAGCTTCCTCAGAGCCTTCCGCACGCGGTGTCTCAT
TTGAATACTTGTGTGAGGATAGCCTCATACCCCTCAGTGAGCTCTTCATGGAGTGATGCA
GTAGACAGCAAGCCTCACACTTCTATGCTCACGGAAGACCAAATTTGCCttgaaaaatct
ttatagtctcttcacatttctaagttgacatcaaaaatcggttaccataaaatcctaata
gttgaagagatgtaatttcaattatttggtaaacctgaccttcattgtcaaagcaattag
tcaactcagatttactttctcccagataatagattctgacttctttttTCTGATTAAAA
AACTTAACACCTTCCTCAGGAGATCTATCTCAGTTCTGAATGCTGATTCTAACTAAGAAG
GATATTTGGCTACATGCTGGGAAGAGGGGTACTGAGGCACGCCGCGATTCCACTCCAGCA
TTTCCAGTTAGTCGGGTGCCTCTGCACTCCCGGTGTTCCGGCGCCCAGTTAGTTGTGTAC
TCTGGGCTGTCCCTATACTGGAGTCCTAAAACACTTACGACTGCAGATAGGGGGAGGTTT
TTCAAAACCTTGGTCTGAAAAGCCATagaagggagataggaaagcgggggggtggagcca
cagtacattcaggtggatccgtttttggaaatagtacaaactggaggtgaaaccctggaa
attgatctgtcgttcacatgcttcatgccgagtccttgtggacccacagagacacactcg
ccccagtttgaaggctgctAACTTGATTCTGAGGACACCAGTGAGGTGGTAGTGTGCAAA
TGATGTGTGAGGAAACTTTGGAGGAGTCTCACCCTGCCTGGAGCACGTGGCCCCTAAAAC
AGCGCAGCCTCCCAAAGACAGAAGATGTGGACTAGTGAGAAGCCAGGTATGGTGACTGCT
GCTGGATGAAGCTTGTCCCACCAGAGGCTCGCTTGTTTCATTGAGCACCTACTGTGTGCT
TGTGGGATGCAAACACACGTGTGGTCCCTGCCCTCAGGTTAATAGGCAGGGGTGGAACAG
TTATGAAACTGCTCTAAAGTCATTTTCTCAAACTGGGAGTGACAAATGTATCCACTTGGA
AAAGATTGAGAATTTTATAAGATTTTTAAATTTTTGTTTATTCACATTGAGGAGAATCTA
AATTCTTTTGAACTTATGTATAGATTTCACCATTTTATAGTAATAAATCAGTCCTCCtgt
gtgtgtgtgtgtatgtgtgtgtgtgtatgtAAACCTCACCTTGCAATATTATTATTTT
AAATAGCCACTTGCATCTTAAGGAAATTAAGAGGACAAAAGAAAAGCTGCTGTTTTGTAT
GTATCCACATATTTACCAGCTGCTTCCCTGCCGGCAGGTGCTCTGgttctgcactgcctg
ttgtcccttgcctgaaaatggttgcctccaatattttgctcagttttctgattgtttaca
gtggcagaggagggtagatctggtaccagttagtaattgccagaggtggaagtctGTGGA
TGAAATTTGTATAACATGGAACGTTAGTTCCACAGTTAATGCTACTCAATTGGAACCCAT
GGAAATTATTTTTGGTGAAAAGGGCCCATGCGTTATGAAATTTGAGATCCATCACTTTA
AGTGAATGTAGGCCCTGGATACAGTGGGAGCTCAGAAGAGCAAATCAGTTGGTCACCTTG
CTCAACGTATTTTACTAAGGGCATCAGTAAGGCTTTCTATGACCTGCTCCTTCAATGCTT
GGTTGACATTTGGGGAGCAAAGATAAACTAAGGATTCTAAGTTCTGTCCTGTGATGCTGT
AAGGGGAATCTCAAACCTCTAGGTGGAGGAGTGCAGAGATGACCAGGATGGTGGAAGCCT
GCAGGAGAGCTGAACACCTGAAGACACCCAGTGGGAAGACCAGGACCTTTAACGCCCATA
TCTGCTGCTCAAGACTGGCAGAGAGAAGAGGGTTTGTGATGAGAAAAGGTGGTGAAAGGC
ACAAGGAGGCACAGAGCATGTCAGGTCCCATATCCCAAAAGGAATGTGCTTGGGTGAGGG
AGAGCTCCTCCATGGCTGGAGGCATTCAGAGACCAGGCAGTCGCTTGTGGGTTTGTGATT
AGAGTGAGGTTCTTTTATAAAGGGAGTGAGAAGAGAAGGTCTGTGGATACTTGAGTGTAT
CGGTAATTAAGAAATAAATTGTGTACATCCCATTTCTTTCCACATTTTCCTGGGCTGTCA
CAGTGGCTGCAAAGAAAGCAGTCCGTGAACTGAACTGTGATCCCAGACAGGCAAGCACAC
CAGGAATCTCTTCTCAGCTGTTGATAATGAGGGAGCGCTGGGGAGAGAAATGGGGTCCTC
TTTGAGTTTCCTCTGTGCCGATACCTTTCTCTTTGTTAAAACAGCTAATTAAACACTGAA
GCAGTATAGCTCTCTTACTATACACTGGTAGTCATAGTTCTCTTACTGTTCTCTTCACTG
ACAGTTCTCTTACTATACACTGATGGTGACGCAGAAATTCAGAATTCCCCGCATGTGTCC
CGGTTTGAAAGCCACTGTGCTTTGCTGTGGATTAGGATCAGACAGTTGAGTCTTGTTCCA
ACAAGGAAAGTTGCTTATTGGAAAGTTTTGCTGCAGGGAGCCTTGAGTTCTGCATCAGGC
TTGGAAGTGGGCTCTGTGGAGGTCAGAAGGAGGATCCCCCACCCGCAGCCTCAAGAAAAA
TATGAAAAGTGGATTATGCCTCTGTAGCTATATTGCCTATAAACTTTCTGCAGAATGACA
GTATTCATATCCTACATTTTTTCAAAGCGATATTAATCCTGAGACCTGCAGCTAAAGTCA
AGTAGAATTTAGGGATAATTAATAGGAGGAAGGTGGGGTTGGAAGATCTGCATGATTATA
GTCCTCTGATATAACTGGAAAATTCTTTCCATTAGCAAGGAGCTTTGGTTAATATAAAAT
GGACAGATTAAACCTAGGCAATTTATTTACTCATTGCTGTATTTTATTTCAGAGCTGG
TTGAAAATATTACAAAGTAATATTTTAAAGTGCTTATCTAAACTCTTACTCTGCATTTTA
TCATTGGGTTATGAAATGACTGGGGAAAGACTTTTCTTGCTTTTATTTCTCAGTGTCTAC
TTATAAACATGTTTTTTGAACTACTGTTTTTGTGACAACATGCCTTTTTCCCAGAAAATC
TCAGGTTAACATTAAATAGGCACTGGATGTTTATCTGATCTTGTTTATAGAAACACAAGA
AAATTTTAACCTTGTATATACTTTACTCAATTAACTAGGTAAGAGGTCATTGAAACATTT
AGAATTCCACTCTACATTTCAATAATTATCAGGTGAAAGCTACTGCATCTACATCAGAAG
ATGTTTGTAATTTATTTAAGAATAAAATTAGCTATGCAAGAAATAGTATGTGGAGTCCTA
TGTGGAAATCACAGAAACCCTGACAACTTGATGATCTTTCCGCAAGCTAAAAATATCACT
CTGGATCACAGCAGTAGAGGACTCTGTAAATTTAATCTGTGTGTCTCCTGTAAATAAGTG
CATTAGCAGTACACAGGTGGTGTCAGAGTCAGTGATGATGGATAGAAATTCTACATAAAA
Tccaggctcagtggctcatgcctttaatcccagcactttgggagtctgaggcgggtggat
cacctgaggtcaggagttcgagaccagcctggccaacatggcaaaacctcgtctctacta
aaaatacaaaaattagctggatgatggcacatgcctgtaatcccagctattcgggaggcg
gaggcaggagaatctcttgaacctgggaggtagaggttgcagtgagccgagatcacgcca
ttgcactccagcctgggcaaaagagcgacactccatcgcaaaaaaaaaagaaGTAAGAAG
TTTTACATAAAAACGTGGAGTGAGCCCAAGGTGCCATTTATCCAGCCCATACACATCGTA
CCATGTACAGAGTGGACACCAGATAAATACATTGACTGCATGCCACAAACATATATATGT
AGGCACCGTTGCATTCAAATACACATCTGCAGCCCTAACACATCTTTATTTGCTAACGAG
CATCAATGTATTTAAAAACAAACATGTTTAAACTAGTGAATGATTAGATTATAATGATCT
TAATTCATAAGTTTTCTCATTGGCCTTTTGTATACTTCAATTGTAATACCTAGAAAAACA
GTTATGTCCAAAGGAGTGAATAGGCCTTATCTGAAACAGGTGAGCGTGACAAGTGTTTTC
TTACTTATTTTACTTTTCAGATAATTCATCCTTAAAGTACATTAGTTTAAAAGTACTGTT
TAAGGAAACAGTACTTGGATTAAAACTTGAATCATTGTTAAGGAAAACTATACCTTAACT
TCATGTAATCACAATTAAACCTCTTCATATAGAAGGATCTAAGAATTTTCTGCAGCATTC

ACCAGCACCAAAAAGCTCAGAGACATATATTTCTTTCTCTGTATATGTATTTTAAATTCA
AGTTAGTATAAATTGACAGGCAGGTCAGAGTAATATATGATCTTCTGAGTCCCCTTAGTA
ATTAAAAGAAATGATTATTTTTGCATGAAATATGATAAAGTGATTTTAAGTGCCTGATAA
AAAGTCTTAACCATGACAACCATTAAAGATTACATCAAAGAAAAATAAGTTTGACTTTCA
TTTACCTTGGAAACAGCTATTAACTGGTAACCTCAAGAAACACCATGAAGAGTCAGTTTG
CTCCACACATGTCTTGTAAAAGTCAAATAACTGGTGGTTATCCAGTAATGACAAGAGGTA
GAAGTTACATCCTTGCTGTCTGATTGAACCTTCCCAGAGCTGGCACAAGGCTGGGAAGAC
CATAGGTGCTAAATGAGGAACTACTTAAAGAAAGAAAATGGAATTTCACGGACAAGAAAA
TCCATGTCCATTTGGTTCTGTGACCCACATCCTTTGTATCCTATGCTTTTTTACACTTGG
TACATGGTTGCAAGATTGCCCCTGTTTTCTACTTATAGTTCCATGCAGCATGGATGTGGG
AAAAAGTCTCCTCGCAAAGGGGGTTAATGCAGGTCACTCTACGTATGTGCACGAGGTCG
TTATAAAGCTCGAAAATATGGGCTCACCAACCAGGTGATTTTTTTAATTATCCAACCAGA
AGACATAACATATAGGGGAATCAAAAGAAATCTCTGAGTAAAATAATGATAACAGGTCAA
ACTTTGCGGTCCCACGTGAGGCTGGAGATGCGTATTGTCTTGACTTTGCATCTACAAGTT
TAACAAATGATGCTTTCTCAGTTTACCTCTGGAAATGGAAATTAGCATTGCAAATGACTT
CATGAGGAGGTAGAAGCTATCTGTGAATTTCCTTTCGCTGTGTTTACGATAGACTCTCAC
GTCTAGATGTGTCATGTATTATGTTAAATTGGTATGTCTTGAAGTTATAAAGCACAGCCC
TCTATAAGTATATATATTCCACCTCTTTCAAATCGGATGGTACCTATCCTTCAAACTGCT
ATTTAATGACTGTCTGCTATGTTCAAGGCACTGCTCTCAATGTTAATACTTgatgagatc
gggcgcgttcaaggtggcatggccgtagacTCAATGTTAGTATCTGAAATATGGCCTACG
AGCTGAGTTGTGAATCAAGTTAATAGATTTTCGGAATGTTAAGGTCTAAACcagtagctc
ttaactgagacaatcctgtcctcatctcacctgggagacatctggcaatgtttggagaac
cttttggttgtcacactggggcatctagtgagtagaggtcagggatggtggtaaacaagt
tttttgtttgtttgtttttgtttttgagacagagtctcactttgtcacccaggctggagt
gcagtggtgtgatctcagctcactgcaacctctgcctcctaggttcaagcaattcttatg
cctcagcctcccaagtagtagctgggattacaggtgtgcaccactacactcagctaattt
ttgcatttttagtagagacggggttttgccatgttggctaggttggtctcgaactcctgg
cctcaagagaaccgcccccttcttggcctcccaatatgccgggattacaggtgtgagcca
ccgtgcccaggctaacattctttaatgcataggacagcccccaccatacagaggaatccc
cagcccagaatgttaatagttctaaggttgagaaacccAAGGTTAAGCCAAGTCAACTTA
TCTATCTTCTTTAAAATTGCATAAGAATGCAGTCCTGTTCTTCATTCCTCTTGCTTTGCA
GTTAATGATCCTTTGCCTGGACTTTCTAAGTGCCCAGAAGAGCAACAGCCAGCATGCAGG
ATGGCATTCCTGACCAGTTGCACTTGGCCTAGCATTCCAACCTCACCTGCCTCAGCTTGT
TCAACCTGAAAACCTACCAAGTGAAAGCAAGAGCCACGTGAAGACGCCTTAGTTATATGC
ACCCACCCAGACACTTGCTCAGAAAGGAATCAGTGGGGCCCTGGCCTTAGAAACTGGCTC
CTTCACTGCTGTAGAAACAACATAAATTTAACATAAAACACGTGCTTTTCTTTTTTCTTC
TTACTTTTTCCTGTCTTGGCAATGCAAGGATGCCATTAGGTAAAGAAATCCTTCACCACA
CTAATCCTGCAGAGCCAGAAGAGAAACCAGCTTGTTCTAACCCAGCTTTGTCATGGAGAG
AAGGCAGCTGCTCCAGTCTGAACTATTCTTTCTTTTGGTAGCAGCCTGCCCAAGGGTGAA
AGTGTGTTTAATAGTTTGAATTACACAAGTGAACAGTAAATGTATGCCTGTTTCTGCTTT
ATGGGACTTTGAAATAATGTTGTTTGTGCCAAGGTTTTAGATTACTATACCTAACAACCT
AGAAAAAGAAATGAAAAGGAAGCCTTCTGCCAGGCAGAGGTCACTACGGGCCTGGAGCTG
GGCACCTGACTCAGCAGCTGCCCAGATCCCCAGAGCTGAGAAGTCACCATGCATTTGTGG
TGCTTCGAGCGAGTTACCAGAGTCCTGGAACAGAGCAGCACACCTGCGGGGTGTCCCCTT
GGCATTTGGGCAGGGCAGGTGACCAAGGGTCTTGTTGGAACTGAAGTCCAGCTTGAAAAG
CAAATCTGGTTGTGAGCTAGAGTCCAGTAACACTTGTTTCCCGCCGCCCCCCGCATAACT
CGTGTGTCCTAAAATACAATAATTTCTTGAACTTCAGTCACTTATGCCTATAAGCGggca
tacaacaggggcacaataaatgtttgttaagtgaatgaatTCTTTCAGAACTAGATGGGA
TCTTAGTCCAACTCTCTTATTTAACGAGGTCCACAGAGGTTCTGCGATTGTCTAAGAAAG
AAGGCTGTGTTCATGGCCTTTGTTGTTTACGTGGCCCTGTGATTCTCTTGGCTCCGTGAA
AGTCCTGATGCAGACATTCCGGCCATCTAGAAAGGCATGCAGACAAGCCATCCAGCTGGC
ATGATCCTGAGTCCAGCtttctttaaaagagcttccaaaactgcttaagctttgactgca
caaaacctgcatcACCTCCAGTTGAGAAACTCAAGAGAATAAGTAAGTTATGGAGTTGGA
GACCCCAGCTTAACTACTAGTTTTAAAATAGTGAAATCAACATTTTCAAATCTTTGACTT
CACTAAGATTTAATAAAGTTTATTAATCATATATTATGAGTTATTGCTCTCTCTTTATGT
CTGTAATGCAGTTGCTCCTCTCTGTATAAATTAATAAGTTTTAGAGATCCAAAATGAGAA
TTTTAAAATAAATTACGTATATTTTAATCAAGTTTAATTTGACTATATCCAGCTAAACAA
TTGATTGAACTTCACTTGCTTTTCTATGACAGGTTTTTTGTTCTTAGTAAAAgaccccag
ttttctcacttgtgaacagaagggttagacttcatgacagctaaggttccttccgtctc
taacaAAAGTGGCCTGAAGAGAGGCTTCTAGACTATACTCACGGTGGGTTCTTGGGACCT
CAGAGTCAgctccatcacttaagtggctgtgtgattgagtggagacacctcaatctcttt
gtgcctcagtttcctcacctgtcgagtgtcaacatgatggcacctaaagctgttgagact
tcagaaaggtaatgtgtgaaaagtgaaaagtgcctggcatccaggaagtactcaataaat
accaactattttattGCTGCAGCTGTTCTTATAGATGTGATTTCTAGAACATTGCCTtct
aataggtagccatgggccacaattgttggctgttcggtgtttcacatatggttagtcca
aactaagatgtgttgtgagtctcaaatacacactggattgtgaagacttaggacaaggaa
aacaatgttaataaaatctcattgataacttttaaattaattacatgttgaaatgaaaat
atttgggacatattgagttaaataaaacaggagattaatttcttctgtttctttctactt
ttttattagtgtggctactcaaaaatgtgacattatgtatgcatctcgtattacatttc
tattggacagcAGCGCTCTAGACAGTACTATGGGTAGTATCTGTGGGGAGGTTCTCAGAA
ACATGTCGCATGCTCTTTTAGAACCTTAAAGTATTCCTAGTCTCCTCTACTTCCAGCCCT
TGGCTCTTGGGCCTCAGTCTTTTTACTTTTGCGGCTGTGTTTCTCTGAAGGCTTGGCATT
AGTAGATTGAAAAGAATAACCATCTAGGGAAATGTGAATTCAGTTTCTTTCTGACATTCT
GCTCTCTACAAGGGGATATTATGTACACATAAACCTACTTCCAAAATAATGAAGTGAGGC
CTAATTCCTTACTCTTCAGAGAGCCCACTGTGGAAGTGTCACTGACCTTGTGTATGGGCT
GCCCTTCATGGCTCTGGGAGTCATTATAAAGGGCAGCATTTGGCGTGGTGCGTCCTAAGC
CAGTGTTTCTCGGCTCTGTTCCTTAGACATGTGTTAGTGTTAATAGATGTTCTTGGAAAA

AAAAAAAAAAAACAGCATTCTGAGGTCAAACATGCTCAGAAAGCTTGGAATCTGCACTAC
GCTTCTCGtacacatttcatattaaagattttggaaagtcctgcaatacagagccctgtc
taatattgccacaacccacaattgctcaaaTGTAAATAGATTTGAGTTTATTCACATTCA
Gatcacctcttaaggccccacctcccaatgctgtcacaatggcaattagatttccacatg
agttttggaaggggacattcagaccacagcaGGGGAAAGCAGGGTACTTGCTGCTTTGCAA
GTGTGTCCACATCTAATTAATAGTACAGTTCTTACTCTTGGTGTGTCCGGTGATATTAAA
AATTAATGTGCCTTATTTAGATAAGTAACATAAAAATCACAAAATGTATGCCTTAGATTT
ATATGTATTTATAACTAGTCTATTTCCTGAAAACAGTTGAGACACCTTGTAAAAGTTACC
GGTACGATAGGGCCATTCCAACAAAGCTGTAAAGTGGTGATAACACAGTCATAAAGAAGA
GGAGATAGCTCTGGGAGAAAAGGTGGCCCAGAAACCAGCTCTGAGCCTCATGGCTGCAGG
CAAGGTCTGCAGGTTCCTGGTCCTGATTGCAGGCCATTTGCTGCCTTGAGTGGTGGTTAC
ACAAGGCCAGCCCTGGGGGTATCACCCAGAACACCTAGTACACGAATTTCAGTTTAGAGG
ACGAAGCATTACTGGAGTATTGTTATGCAGGAAAACTTTTTCCTAAAAATGCCCTGAAAA
GAGAGTAGCCTAATGCATTCAATCAAAATGTTTTTAAGTGGAAAACATATTGTGTGTACT
TGATCTGGCCTGCTGCTTTTAAAAGATTAAAACTGGgactggcatggtggctcacacct
gtaatcccagcactttgggaggcagaggcaggtggatcacctgaggtcaagagttggaga
ccagcctgaccaagatggtgaaaccccatgcctactaaaaatgcaaaaagttagccaggc
ttggtggcgcatgccggtaatcccagctagttgaggggctgaggcagggggaatcacttga
acctgggagccggaggttgcagtgagctgagatcgcatcattgtactccagcctgggcaa
caagagtgaaactccatctcgaaaacaaacaaacaaacaaaaaaaacaCTGGGGCCAAAGA
ACTCTGTGTGCTGTATCACCTAACCACATTTCATGACACGGCTAGAGAAGAATCATGCAA
ATAAAAATTTCCAACATGTTCGTAAACTGGGAAAGTATTTCACTGGGGAGTGAGCAGAAA
AGTAATACTATAACCTCTATATCTAGACAAATGTGAATTCAGTTTCACATATAAATATAT
AAGTGAAAaaatatataaatataaataatatgaaataatggttatctcaccactttctac
atcttttgtgaatattttatagtgctcaaatatattagtgcactagtatatgtacattac
attaaataactaatcatttattaggaggatgtgcttgtttttgctaataaagatgataa
taaaaaaaTCCTTAGACCCCCCCTCGGTTTGTTTTCAGTTAGGAATTAGGGATATTTATA
AGAATATCTTTAAATGACACATGCCTTGCTCTGGGACGAGGCATCTGCATGGGTGACACA
TATGTGTTGTGTGTACAGGCTCCCAGCATTTCCAGGGCCCTGCTCAGAATGTAGGCCTTA
CTGATTCTTACAGAGTTACAAGCGCTGGTGAGGTTGGCGAAGTTTAGGTAAACACAGCTG
GGAATGCCCCATGGCCTCTGGGTGACTTTGGACATCACTGAACTTTACCCTTAGAGATGC
ATACCTGCATCTTTTTTACCCTGATAGGGCCTTCCATGATGCTTTCAAAGTGTTTTTGTC
TGCTTTTCGGTTAATAGACTTTCACAGTAGCCAATTGaatatattggttaaatgcatctc
tttatacacagactggattcaaactgaggttgtgtctctccctggctgtgtgacgttggg
tatgatccaagtgtcagattactcaacttcaaaatgaggacagagcctttccttctagg
gctgccaggaacattgaatgagagagtgctggcagcttagtacagGTGTTCATTGCTCTT
GTATGGTACTGTCTGTGGCACGGCTAGATAAAATACAGTAGCCACTGATTCAAATTTCAA
CTGAGGAGTAAAATAAACTGAATAACTTAGAAAAGTTTTCTTCTTTTGAATGACTCTAAG
AATTTAAGGAGCATGTGAGTGTTGATGGCTCTAAAAGGGTAACAGAGCCCAACTAGCTCA
GTTCTCAGCATGAAAATAGTCATATGGCACAGACTCAGTGGAGTGGGTGCACTTCAATAA
CTGGAAGCACAGATGCCCTACAGCAGCATCAAAGATGGCACTCTAAACTACTTTCAATCC
TTTAAAATAAATGGAAACGCACATTTAGTATGCATATGACAACACGAAGGACTTCGATTT
TGCTGATGCAATACAGTTTTACAGGATTTTTTATACTCAAATTAGTAAAATTCTGTATTG
CATCCAAATTATAAATTATAATATCATCTAGATTGGACATAGGAATAACGACCACTGGTA
TCTGCCCAGAAAGCTCTACCGCCTGTTTATAAGCTCCTGCAGGAGACACAAAAAGAAGAG
AATTTGAATATAACTTGAAATGACCGTAATCTCCTGCCCCAACTCATTTCATTACCAAAC
CGCCTCTTTCTTCATTATTTCTCCTGAAGCACAAATCTATAGAGAACTCAGCTGCCAGTC
TCTCCCACTGCACTCAGCAGTGAAAGGGTTAGGCCTAGGCTTTTCAAACAGACCAGTGCT
TGTATCAGCCCTTAAACATCTCTGGAGAAGGAAATGGGATCCTTCTTTGGTAATTCATTT
TTGACAGTTGGGGATTAGGTGTTCTGTATCTGGGGGGCCTTGCTGTCTTCTCTCCTCCTC
CTCCCACTGCAGACCCTCTCCTCCCCTCCCTCTCCAGCTCTCTGATGACTGCTTCATGC
TCCTTCCACCTGAGGACTGCCAGCACAGCCTATTGCAGGAACAGCCAATGAGGGGCTGGC
TGTGCTCttttatttataaaattataaaCTCAAGCAAAATCTAGACTATGTGTCCCCAAG
ATCAGAGGAGCACAAATCCCTTGCTTACAGATTGCATGGGGGGCACATTCTTTAAAATTG
GTCCCTGATCTAGACTCTAGCCTGAGAATCATCTTTAAGTTCAGAATTTCCACTCATGAC
CTCACATCTGTGGGCTCCCACATTGTCTTCCAAAACACACATGGCATCTGGCATCACCTT
CACCCCCACCCTCAGAGCCTCATCTCCCTGCAGGTAGATAGTCAAGGCAACCTCTTCACT
CTTCTGCCAAGCCTCCTCTCCTCAGCTCTTCCCTTCCTCTCTCTTTTTGAAAATAttttt
aattgtggcaaaatatacacaacataaaatttaccatcttaatcatgtataaaagtggag
ttcagtggcattaaatacattcacgttgttctatagccataaacaccattcatctccaga
gctcctttcatcttgcaaagctgaaactctgtccccattaagcaatggctctgttttcct
ccgttcccccagcccctggccaccatcctcagttttctgtctctgtgagtttgattactc
taagcacctcttataagtggatcatacaatgtatctgtctttttgtgactggcttgtttc
actttccataatgtcttcaaggttcatccacgttgcagcatatggcagaacatctgtcca
tttccaggctgaatggtactcttttgtacgtgtggaccacatttcatttatccattcatc
cacgggagggcacttgggttgcttctgctttttagctattgtgaataacgctgctatgaa
catagctgtaTGCCTTTGTctttttaaagcccaaatctgatcaagtcactccccagcttaa
aaccttccactgctccccaGCAGTGGGATAAAGGCCAGTCTCCCCTGTAGGTCTCTCCCG
CCAGCCCTGCTCAGTCTTCTTGCTTGTCATCCTTGGCTAGGCCTTGCATTGCCATAGCCC
TCTGCCTCTGTTCACGCTCTCTCATCTTGGAGCATGAGCCTTCCATCATCTCTACCAGAT
GAACTCTCATTTCTTCTTTCAAAAAATAAAAAACCCAAAAAACCCAGAGATCCCAACTGT
CCTGGTGTCTGCATAGTCTGCAGCACACGCCCCCTCCATGGCCCTTCCTCCATAAGCAGA
ATCACTCCTCACTGTTCCTGCAGCACCTCCTGTGTGCCCACACAGCTGTCCTGCGGTGGG
CTGTGTGTGTGAGTGTGCCCCCTCTAGGACCTGAGCTCCTTCTGGAGGGTGGGCACAGCA
TCCATTCATTCTGGGAATCCTGGTCGGCACCATGCTAGAACTTCTGCAAGTGAGTGCCTT
TGGTGCTGGCCCATGGGAGAGCTGTTGGTAAGGCATACTTTTGCAGATTCCAGTTGCTGC
TGAGGTTGTTGCTCTTTGCACAAGTTTCTTCTAGTCACCAGTGAAGTGACATGTGTGGCA

GGCATGGCCCAGGGAGGCTTTTTCATAAAGAAGAGGTTGAATCTTTGGGgctgtggtttg
aatatgtccctcaagcttatgtgttggaaacttaatcccaaatgcaatagtgttaggagg
tggggcctaatcacaggtgattaggtcataaggctctgccctcatggatggcttaacatg
tttagtgaggcagtgggttagctattgtgagagtgggcttgttagaaaattgagtgcagc
ccccttcttgcttgctggctaccatgctctcttgcttttctgccttctgccgtggggtgac
acagcaagaagaccctccccagatgctggcaccatgccctgggactttccagccttcaga
accacgagccagacaaatttcttttctttataaattacccagtctgtggtattctgttat
agaaacacaaaatggactaagacaACTCTTCTTTCATCAAGTTAGGGTACCAACCTTTAAA
GACTGCCAGTCCAAGGTTAAAGGAAACTTTTCAAGAGCAGTCCAAacatgatctggccct
cagctactctccagggtcatgccaccctatcacccactggctcacacagacgctgaccac
tgcttagtttctcaaactgaagttttcctcctcagagcttttgcaaaacctttctttgc
ctggaaaactccccccacaaatctttagttgtaggttccttctcatcttgcagAAttatt
agtttgctcttcaaatagtctctccagctagactatcaactccaggagggcagagttctt
cttcgcttccttcacccatgtgcccactgagtccagaactgtATAGCAGTTTGATTGAAA
AAATCcacagggtggaggatgagaggaccctggatcccagcctcacagcctcttacttca
cctgtgtgattttggtcaagtcctttattcttcctgggctttagttttcccttatctaaa
atatgagAAAAGTTCCCCTCTCCTGGGTATTCTGGGAGACTCATGTAAAAGGCACTGAGC
CAGTGCAGCACATCTATGACCAGGAAGGGTCAGCTTCCTGCCTTGCATGAGACACACATT
CCCTTCTTCATGCACAGTTATTCATGAGTTAAATATGTATTGAGAAGTGGGTTCTCAGGA
GATGATGCATCCACAGCATTGTTTGTATGCCTCTGTCTTTGATGTCCCTGCCTGAGTCGC
CCACTTTAGAGCCCTTCTGTTCTTCAGAAACCAGACTTTTCTTTCAATAGTTTCAGTaat
caatcgatcaatcaatcaaccaatcaaCAGTGATAATAATCATGAGTGAGCCCCTGCCCG
TGCTGGCTGTGTCCTGCTGAAGGCACACTAAGTGCTGCCCTTCCCAGAAGCCTCAGGAAG
CTTGCGAAGCTCAGGTGCATGGATGCCTGGTGGAATGAGGAAGGGATGCAGCCAGGTAGA
GAAATGCCCTGCCATCACTTGCATCAGCATCTGTGAAGAGCTGGCCAGGCTTTTGCTCAC
AGTGGTTGACACAGTCAAGGAGCAAGGGCCCCGTAGGAGAGGGGAGTCAAGGGCTCCGGG
TGGGAATGGAGCTGGGGGCTGATGCTGGCTTCTGGAGCACTGTAATGTGACTGAGAAAGG
TGAAGGAGCCGTTCTGAAAAAGAAGAAGGCAGGAGCTCGCACAGCTCTTGACTCATCTTG
ACTTCTTTTTCCTGCTTCATCCAAGCAGGTCGACTCTCTCGTGATCTCAGAGACAGAGTG
AAGTCATGAGTGGGAGGGGAGCACAGAAAATAAGACCTTGATTCCCAGCATTGGGAGACT
CCCTGCTCCCCTGAGTCTCGGAAAATAGCACCCTTCAAATGTTTTAGGGATCCAGATTTG
ATGAAGAGATGTTATTTTGGCTTTTAGATTCTTAGGAGAGATTTGTCTTTCTCAGGTCAG
GAAGAAAATGCTGCCCGCTGCACATTCTTCGGGACAGACTCTTTTAATTATTACTAGTTT
AATGTATGTTTTGCTTAGTTAAGGAAAACCCCTGTGGTTTCTTGACGTGCTTCAGTATTC
TAACTCACAGCTGATTCAGTTCAGGGGGCTGGGGAGATGTCCTCGACCTTGGAAAGGAG
GGTGCATCTCTAGAAATAAGGCTAAGTATGCCACTGACACTGTCTGCATAAACGTGTGTG
ATCTCAGGTCCAAAGGATGGGGCCTGGTCTAAGCCAGGGACGTGGGAAATCATTTTCCTG
TGGCAACTTGTGAAGACCATTCTGTGACCTTGGTGTCTCTGGGCCTTCTCTTAGATTTTC
TAAGTTGGCTAGTCAGTGGAGCTGCCATCCCTCCTTTGCCCATGTTCTACTCCCAGAGTT
CCTCCAAGAAATTGCGGAGCAATGCCTGTTTCATGAGAGCTGAGTTTGCTGTGTCTTCCA
CTTAGAAACAACACTGTGGACCAGGAGGACACACAGCTCCCAGGGCCATCACCACACAAA
GTGAAGGCTGGTGAATCCGAGGCTTCTAGCCCTTGCCGGGCCAGGCCCGCAGCACTCCGC
TCCCCAACCCAGCCGCTGCTTTGTCGCAGGAACCTCAGCAGGGCAGGGTGTTTCCTAGGA
GGACATCCGATTCCCAGCCATTCCTTTCAGTGAATCACCTGAGCTCACATTCTTTTTTCT
TTTATTTTTGAAGCTCTTAGCCAATCTGCTTCGCGATGAACCAGTTTTGCTTGAAGCAGA
CAAACCCGATTGTCAGGAGACAGTGATGATTTCTTCAGTCTCTGAGGAAGAGTTTTCATT
TTCCCCAATTCGCAAAAAAAGTCAGGTCCCTCCCTCCCTCCCTCTCCGTAGAATATTTTC
CATGTGTGTTAACAATGGCTGAGCGTGGTAGATGCCAGGAATTTCTGTCAACCCTCAAAG
AGGAAAGCCCTGCCTAATGGTCTGCCCGTTCTTGTTCACTCCCTGCCCCAGGCTCCCCAC
CCGCCTTCTTTCTGGAAGGTATAAAGGCTCCTGCTTATACCTGGCACTGCACGCTTCGCT
CCCTCTGATCTCCTGACTGTCATGCCCCAGTGTCTCAGCCTATCATTCTACCTCTAACTCG
ACCTTGAGTGACCTTGAGCAAGTTTCTCAGGATTCCACCTCCAAGTCACTCTCCCTTTGG
GATATGCAGCACTAAGTTAAGCTTGCCTGGAAAACATCACTTGAAGCTGGAAAACCACTT
TTAACACAGCGGGAAAAGCTATTTGTTCAGACAGGAGTGGGGTGGGTCTGGGCAGAGCAC
TGCTCTAACTTGGCCATGCCGTGGCAGCAGCTCCTTTAATGCCACTTTTTCCTGGCGCGC
CCGCGGGGCCTGGAGCTCAGAAAGAGGGGAACGCTCCCTCGTCTCTCAACAGTTGCTCCA
GACAGGTCAGCAAACATGGAATTCAGAATGTTCATTAAACACTGGCTGTGTCTTTTGTGT
TCAAAAGCAAGACACTCTCTCTGAACCATGGCCCCACAGAGAGTGCAGAATGTGTGAAAC
CTGCCGGGAAGGTCTGGACCCCTTGCGGGGCAGTGGGCAGCACCGTGCCTCCGTTCACAC
CACTCACATGGCTGTGCCTCTGCTTCCTTCTGGCATGGCTGCTTCTTCCTCAGGTCTCAA
CCATCTCCCTCAGATGCTCTTTCCCATGTTTGTGGCTACAGGTCCCCGTGACCTGCAGAG
GCAGAGCACTCACCAGCAGCCCAGCCTCGTTGCGCACCCATGTTTGCATTTGCAGGCCCT
AGAACCACTCCAAGCTCCGTGTGGCGAGATGCACCCTCCTGCCCTTCACTGGGGAGCTGC
CCTCCTGTTCACAGCGGCACCTGAGTCACACATCTGGAGCCATCCTGGACTGCCTCATTT
CCCCGATGGGGGGTTTCCCTGACTTCATCCATCCTGTCTTTTGGGTCCCCATAATAACTG
ACATGGGTCGGCCCGTACCAGCCCCTGTGAGAAGGGCTTTAACTGCCTTCCCACCCCCTG
CTCATCTTAGAGTCTCTCTATAGTGCTGCTGAAAGAATCTCTAAATcagtggttctcaac
ctcagccgcacattgagaatcacctgggacccttaaaaaaatcttaactcttggtccaag
aattctattacaatcggtctgggatggggccctacaggtatttttttaaagctctccagt
tggtaatgcatagctagagttgagtatcgctgTTCTAACGTGCAGATCTGGTCATGTTAC
CAGCCTTTTAGGTGGTCTTCTTTGGCTTTCTCTATCTAAAGTTCAAAACCGAACATGTGC
GCATTCAGTGCACCCATTTTCAACTGTGCATTAACACATTCAGCCCACCAGCAAGATTTA
TGAACCATTTTCTGCTGTTGTATATAACATATCATATGCATAATGGCATAGGTTATTGTT
TTCTTCAAAATATATGAGATGTGAGTCCTTCTACGAACTGACTCACACTGATTGCCCAAC
TTCCTCTCTCGAGGTCTCATCCTCTTTCCCTGCAGCCGTCTCCCTCTTGcacgcacacac
acacacacacacacacacacacacacacacacacacCAGGGTCGATGCCATCTACCCT
GGACTTCATCTTGAACTCCTTCGAGTGTGAGTCATTACTCCTTTGTGCACCTCTGCTTTC

```
TCTTCTCAAGATGTTCACCTGCTTGAGGTCAGTTCCTTGAGCGTCTTCCACTTGCCATGT
TCACCACAGTGCTCAACATGCCTGAATGCATGGATGGCGacttctcagatcctcagtctc
ctcatctgggtaataaggcattgggttggcgggtccatctggtttcttccagctctgaga
gtgcatttgctctgtgattcATTCGTTCCACAACACTTCACCAATTAAAGAGAGGGTACA
AAAGGTGAACATCCTTGGCTCCCAGCAGATGCTCCTCAAAACCTGAAAAATCAGATAGGT
GAGGGAAGATTGAATGAAAGGCCTCTTATGATTCTGCAGCAATTTTGGTGGTTTAAGAAC
TCTATGGAAAAATCATCAGTATTTCTGGAATTGAAGTAAAATGGATAGTGAGCCTCTGTG
TATGTGAAGGCCCGCATCTGGAACATGAAAGAACCTGTCTGATGTGTTCTAGTCAGGAAA
GCAGGTAGCCAATACTATTTATAGAATTTACAGAAACTGAAGATTTTGTTTCTACTGATT
TTCAAAATAGTATTATGTCTGATTTTTTTCCTCAGAAATATACTTCCTGCTCTTCTCAAC
AAACTCATTTGAAAATATGATTAGAACATGATAGAATTTTACTCATTTGCCAACTGCGGT
TCCCATTTCACATATTGTTAGAATTCTGCATGGTGGCTTTGCCCTTTAACCACTAACTGA
TAAATGATGTAGTTAGCTTTTAAATGTGTGGAAAAATATAATTTCAGGTTCAACCATAGG
TCAGAAGTACACGTGTTTTGTTAGTCTATTTGTCTCTCAGTCATCTCATGGAAAATTCTC
AGCTTTTGGTATGGAAATAATTTTCTTGAAGGCAATATTTGTTGAGTGACTGACGGAATG
AAAAACGCCAGTTGCGTAAGTgtgaaaaagatctgggtgttttcattggatccaaattcc
acatgagccaacaacagcgtggtgtggaggctggagcacattaataagaacagtgtccta
aattcaggaggtaatgctctgcccatgccctgtgcagctcagacggtgtgtgcagtgcag
tatgtaacccagggcacatttcaggggcccacagggagctgcagcttgtaaggtggagtg
cagccaacagagcagagagtcagaatccccgcagagtggttgaaggcacaaggatgcgca
gcaaggaagacagacttataggtggtgcgactgccatcctctggtactgaaggtgctatc
atggagggagggaagtagattgaccctcctggctccagagtacggaactcagacaaacgg
tcagaagcttacagggaggccaattttggatcaactttaagaagaatttttaaaagcta
gagcaatcctaaaatggaatttgctctttataaagttgcgaatgcctcaccctggaattg
cttaagcaaagttgggacgggcagttgtgagtaatCTCCTTTCCAATCCATACCCGCAAT
CACCAGAAACGTGGACTTCCCTGACACTGAGCACCTCTTAATTAAGCATCTCATAAGTGA
ACAAAACCCAGCCCTTCAAAGAAGTCACTTTATTTATGTGTGGGTCTGCAGCTTGGATTT
CTTGATAATGTTAAATAAAACTCCATCTACTCTTCCACAAACACTTCAAGAAACCTAAGA
CTTTTGGCCAGAGTAACACCGAGGTTTGAGAGAAAGGATATGTGTGTGAGAGGTGTGGTT
TCATTAGAACATATTATTTGACTTCATGTTGAATCAACACTTTTGTGCAAAATGCAGTTT
TACCAGCCTCTTTCCTTGTTTTGGTCACATAATTTAACTTAACATTCTCGGTACTTGATT
TTCTAACATAAAATGGGATTGAGAGGGGAATTTTGAAGTTCCCATGGTCTGTCCTCTACA
TTCTGACAGCTCATTATCTCTGCGGTATTGTTCTCACATTTAAGTGAGGTTAGCGGAGGC
AGAGGCCTCTCAGGCCTGAAGATAGCCTCTGTTTTCAGGGAAATACTAGACTGTGAGATC
TGTGACACTGAAGCACTAAGTTCATCTCACAAAAGCAACGTGCTCTTTTTAAATGGTTGA
TCAAAGTTACTTTCAAAAGGAAGTGTTAGTTTTTGTTATTAGCCGAAACAAGAGCTGCTT
TAATGTAGTATATTTAAAATCATATCTCAATTAAGATGTTATTCaaatactatttgaccc
accaatctcattactggatatatacccaaaggaatagaaatcattctattataaaaacac
atggctgggcacagtggctcacgcctgtaatcccagcattttgggaggccgaggcgggtg
gatcacgaggtcaggagttcaagaccagcctggccaagatggtgaaacctcatctctact
aaaaatacaaaaattagccaggcgcggtggcaggcacctgtaatcccagctactcggaag
gctgaggcaggaaaattgcttgaacgcgggaggcggagtttgcagtgaacagagatgaag
ccactgcactttagcctaggtgacagagcgagactctgtctcaaaaaaaaaaaaagaacc
acttgcatatacactattcacaatagcaaagacgtggaatcaacctaaatgcccatcggt
gatagactgcataaagaaaatgtggtacatatataccacgaaatactatgcagccataaa
aaagaacaagatcatgtcctttgcggggacatggatggaactgcaggtcattatccttag
caaacgaataagaaaagaaaacaaaataccgcatgttatcacttataagtgggaggtaaa
tgatgagaacacaaggatacactggggcctacttgagggtagagggttgaagggagagaa
gcagaaaaaataactattggggtactaggcttagtaccagggtgacaaaataatctgtac
aacaaactactatgacacaagtttacctgtttaacatacctgcacatgtacccctgaact
taaaaaaatttttaaaaagaTGCTATGCAATAAAATTCTCAATTAAGAATTTAACTTGGT
AAATGTTCATTTAATGATCTAAAAATATGTGTCTGGATGGCTCTAGCAAAAAAATAAATA
ATAAGTTTCTCAGAGATGGTAAGGCTGAAATAAATGGGGAAAAATCTGAATTGTAATCCT
TTTTCTGTTGGACCTGGTGTTGGGGTTTCACACTTGTGGGTGAATGTGGGCCTCCTGTGA
GCACCAGCACAAAAGACTAAACTGAACAAAAGATTAAATGTCACCTCTAAAATTCTGTGC
AACAAGACTTCCAGCCACAGAATGTGCAACTCAGATTTCCAAGTAAAAACACACCAGGAA
GCAGATCTTAGATCTCTGTTATCTCCTTGGCACCAGCTGGTATTCATCCTCAATGCTAGC
TAGAGTTGAAATAAAGAGTGAAAGAACTTTCTCTTTTATTACttaataaacttcctttttt
tgagctgttttaggcttacagaaaaattgagtggcagtttcagggagttccagcacggcc
cctgtttctttctcatggtccctgcaggtttcccctattattaacgtctgtcattagcat
ggcacatttgttacaattaatgagccaatattgatacattattcactaaagcccacaggt
tgcgttaggggtcattcttggtggtgtacgttcttcaggtctggacaaatctataatgac
atgcattcaccattactatatcacgcagagtcgtctcctggccctacaagtcccctcctt
ccccacctgctcactcctccttcccaccctccccaaactgtggcaaccatttaactttg
actgAATGGATTTATTCTTATTCTGCCTTATTGTATGTACACCATATTTTAATAAGATAA
AATAATAGTCTATAGTAGACTTCTGTAAATACTCAATGAATAAATACTTGCATGAATGCA
GGAAAAATCAATCAGTCTTGCAGGATTTCTTATGCGTTACATCGTCCTTATAAGAAAGCA
GTCATTCTCACCGAGATGTGCTGAGCAGATACTGGACATGTTCTGACCCAGATAAGGGCT
GGGTGGAAGTAGGGCTGGAGACACAGAGACCCAGTGCCAACTTCCAGGACCTCGGAAGAA
CTGAAGGCAGAGAGGTCCTCTCAGTGTGGACTGGGCCTCTGCTGGCAGCCACCAGCGGGC
ACAGAGCTGATGTGTGTTATGCCACGTGGGGAAAACCTACAGACGATTCTGAGAAAGGCT
CACAGGGACACCCTCTGCCCCTAAAAGAACAATTTAACTCTAATTTATTTCTGTCACTCT
GCATTTTCTGACCTTTCCCAAGTGTACAGTTTTATATGCATTTAACTGCCAAATTGTCAT
GTGAGATTATATGGTTATATTTCATTAATATATTCTAGTTTGTTCAGCTGTTCTTACTGG
GTGAATTTGTGTGGTTTCCTGACATTTTTGTTTTTAGTAGTGCCTCAGTAGTTTTATACA
TAATTACGTTTCCCTTCTGGATTATTTCCTTAGTATCTAGTTCAAGAAGTGAAATCGCTG
GATTCTTGTGGTAAATTTTTGAATTTCACAGTATAATGCTGATTTTCTCAAAGTCTCACA
```

```
TTCTAAGAAAGTATAATGAGGCAAAACAAACAACAAACATCTTAAGTTGATTTTTTCCTA
GCATCTTTTCCTTCCATCTTTGCTTGTAGAATCTAGACTATTTCATGAACCCAAGATATA
ATCAGTATCCTTCTTCAGTATGGCCAAAGTGAGTTTCTCATTATTTTACCTCCCCTTCAG
GAAATGACTTTTCATCTTGTGTTTTGGGAGCCATAGATGGTTCTGGGCAGGAAACTGGCT
TTGGATAGACCCAGCATGTAGATGGCTATTTGGCCTTGCTCCCAGTATAACGATGCAGTT
CCCTGTGAAAGGGTATGAGTAGGTTTTGGGGCTCTGGATACCGTGTGGCCTGAAGAGACA
AGGGCTCAATGCCAACTCTGCCTGTTTCCAACTGTGTAACCATGTGAGCGTCAAAAATCA
TGGACGTGCTCTGGTTAACACTGAGTGGGAGCTCAACAAATTATTATTTTTAATTGTTAC
TTGGACATGGCCAAGTTGACTACACTTTATGTTCTGCTACCTGCCAGTCTGAAAGTGACG
CCACAGAAGGTGAACCGCATGTTGGGAGATGCTCCTCATCTGCTTAAATGAGGTGCAAAC
ACAGCCCATGCGCCTGCTCTTCATGACTGTATCTGTACCAGCAATATTTGTATTGGCAAA
TCACATGCCCCAGTGGGAACTACTTAAGGGGAATTCAATGGATTTCATTCCTTTTATGTA
ATTGGCCACTTAGTAATAGACGTGTAGGTCTCTTGTGTGGATAAGGATTCTGCCTTTTAT
GTAAGATATGTGTTGCAATTCAGCTTTCAGGTCCCAGCCCCGGGAAGGCTCCAGGCCTTC
ACAAACTGGCCCACCCACGAGAAGGAAAGCAATTGTCCAAATGTGGGTAGCTTTTCTTCC
CACTGTTGTCAGCTGCTTCCAATTAGCCCCCATATACATAATCCCAGTTTGTGTCTGTAT
CAGTACAATTCTCCCATGTCAATGTGAATTTTAAGCCACAGAGGGAAAGGGGACAGAGAA
TATGCTTTCATTCAGCTCTCCTCGTCTCACACCTCTTGCCCTGCATGCATTTCTTTGCTC
TGATTAAACGAGCATTTTATAAGCCACATTTGCTGTGTGAAAGGCAAAGTCTTCCCTCCC
ACGGATGACGGTCTCCAGGGAtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtga
gagagagagagagagagagagaCTGTAAACATATATCTCTGTGAAACTTCATTTTCCATA
TGTGAATTTTTGGAACCGAGACAAATGGAACTTAGCTAAAAGATGGGAAAGGTAGACTGA
CTCTGACTTAATCTACTTAACCTACCAGGCAATTTATAACTTGATGGCCTAATTTTTGCA
GCACCCAGAAGCAAGCCTGTTTCAGCACGGCAAAGGCTCAGCTGCTAAGTGGGCAGCATT
GTTGGAGGTGAGCAGCTTAGGCTGACTGTTCATCAAAGGACCAAGCGCTTGAGGTTCGCT
CATCGCTGGAGGCCAGAGTGGGGAGGGCCATTTAACTGCTCAAGGCCATGGAACTCTACT
GTCAGTTTCAGGGAAATTTGGGACCCTGGAGCACAAACCAAAACTCCAATTAACCAGGAG
AGGAACTCGATCCCCAGGAGATAAGTGAAGAGTAAGAAGTCTATCTTTAGAAACAAGAGA
TGTCCAAGGCTAGAAAGATGGGGAAGGAGGGTGGAACTGTTCTGGAAGTGGGTCTCAATC
TCAGCACCAGCAGCTCTCAAGACTTTCTAGAGAAGGAAACTTCATTTCTGAATTAAAATT
AGTCTTCAATGACATGGCAGGGATTTCGGCACACTCTCTTGCGTCATAGGCCACTGTGTT
GGAGGCAGGAGTGTTGGCTTTGGAGGCATAGAGATTAAAATTAGAGTAACACGTGAGCAC
TGAAAAGGTTAAACAGTAGAGACATGGAGGACTCCCGACCCCCATGTACCCCTTTCTTAA
CCCTTTAATTAAGATCACAGCCCTAGAAATAGCTTGCAAAATAATTAACTACTGATCATT
TATACCTTAGTGCTTCTGTGAGCATGTTTTCTCTTTCATTGCTGCTCATCTGCATGGAAA
AATGTGCATGGGTTTCTGAATATAACTCCATGGTGCTTGCTTCCATTATATTTGTGCCAT
TTGGATCATAACTGATAAGCAACCAAAGAGTCCCATATTACTGCACGTTCCCATCGCTAT
TTTATGTGAAGGTGGTCCTGGGGGCTGTTCTGAATTCTCAGTTTCCTTTTTTCCCCTCCC
CAGTTCTTTGAAAATATCAGAAACGGACTTGTGGCATCTTTGAAAAGCTACTTAAAATGT
GCTGCTGTGCTCTGAACTTGAAAATGTGCTTTTAATACAAAGTTTGTGCAGCCCTTGCTG
CTCATACGAGATGAATCTtaccatgtggtggatgcccgtctcatgccaggcactgtgctc
taagcccattggtttatttcagtgcttgaaattggctttcgagagaggcaccacggttcc
ctttttacaggagaggaaacaccagaggatcagagatggagagtctttctccacaaactc
acagaccccaaaggcaagctcagggttgtcagcttccaaagtctgcctgctccaggacct
catgttgcatctcCATTCTCTTCACTGAGGGTCAAATGGAAAGAACACATGGGGGTCAAG
TTTCAGAAAATAAGAGAAATGAAGAAATATGTGCCCGGAAGCAAGAACGACCGACCTCAT
TAAACTGGCTCCCTTCACCTCCTCTCACATCTTTTTCTGCCTTTTGGCCAAGTTTTCTCT
CCCCCGCATTTCCTCCTTGATCTCGTTTGAATCCTCTTCCCTGGTGAAGTCATTTAGGTT
CAGGCTCTTATTTTACTTTGGTCCATAATTTAGATCGAACCACATGTGCTGATGTGATTG
AAACGATGTGGAATTCTCTGGACAGAGATAGAATTATGGAGGGGTTAGTGTGTGTGTTTA
AGATTAAAAGACCAGGTGTATGGGAGGAAATATAATGAACAAAAAATAGTATTTTAAATG
AATACTAAACTTGCACTCATGGAAAAAGTTCTCTTCCCATGAGGTTCTCGCAAAGCATTT
TACCATCAGCACACGCAGTTTTTCTCAGTTTTCTGAGATGGGGCCATCTTGAATCCAACA
GACAACACACAGCATCAGCCAGACTAACACAAAGGACGTCATGGGCATGGACGTAAATAC
TGGTGTCAACACTAGGTCTGCACCTCGAGAGGAGTGGAGCAAAAGGATGGAGTGGCAGAT
GAAGGTATGCTGTTCAGAAAGGAGGCAGAAATGAAAGGAAGACCATCAGTGCGCTCCACA
GCTTGAGGACCGTCCTGGAGGGCAAATGCCAGCTGCTCACTTCTGAAAAGAAAAATTCCA
GTGAAATGAGTACAGTCATTCTTAGGATTACTCACTTGATACTGTGTATGTCTCTTCTTG
GCTTCTCATCTCCACACAAAACCCTCAGGTGGtaaaaatctaattaaaaaaattatataa
aGTCTTGTAGATTTATTAGCCTGAACATAATAGATTTTTTTTAAGCACGTTAAGTCTTCC
ATGGACTAAAAGAAAACTTGTAAACCTAAGAGAACCTCTATTTTTGATATACAAAATAAT
ACATTTCCTTAAACTATGATCTTGATACTAGAATTTTAATTAAAAAATACCTGCAGTTTA
TATGCAAAGTTATAGATTAATGCTTAAAAATAGGTTGTATGTAGTATCCACAGGTCATGT
TTGACTGTCAAATAGATGTAATTTTAATTCATAATAATTGTGTCGTGTTCTTCCCCACTA
GAAGCCAATTATGCAAGCTTCACCATTCACACATGGAAAATAATTTAATGGAGTACTCAT
TGCAATTTCACTTATCCAGAATTGGCTGTTGTTCTCAGAGCAGCTTGTGTTGCCTTGTTA
AGGAGAATATGTTAGTATCCAGACATCCAGAAAGGATCCTTTACTGTTTCAGAGTCCATT
TTCCCCACTTTTGAAATACACACACAAACACCCATTCATGCAAACCAAACAGAGATTGTA
AAGTGATTCCACTGACATTTATGCACTTCTTTTTTCTCTTTGGTTCTTCAAACTCTCAGT
CAGTGCGCATTTACTCTTAATTTAGATACGGTTTAAACCTAATTAGAAACCAGAAGCTCT
TGTATTTCCACAAAGGATTATGACAGCCCCAAGAAAAGATAGTGAAACCATTATATAACA
AGATAAAGGCTTCTTAACAATACAAGGATGGATTTTCTCATTGATCTTAGCCTTCTGAAT
TTTAGAAATTGCCATTTCAAAGTCTAAAACAAAGGAAAATCAGGGAATAAAAGAATGGTA
AGTAGACACAAACCTACTGGCTCCATCATTTCTGTTTTAGCAAATAACCTGCCACATATA
CCAATAGCCCAAGAGATGGGCATGTCCCTGCATTTCCTGGTCAAGGTGACAACACTGCGT
CCTCCTGGAAGAGGTCTGCCACTCACCATACCACAAACCAAATATAATAAAATCAGAAGG
CACACTATAGTGAATTTTTTAGAGGCATGTATTGAAAAGCATCTCAAAAAGCATTCTCGA
```

AGCTTCCAGAAGTCAACTCAAGTTATCTGAAAAGTGACACTTTTGATGATTGCTCGCTTA
ATACTGGGAGAGCCAGATGAAGATTCCTCCCCACTTCCTCAGATGTGCAACTCTGGAATT
TCTTAGTGTTACTGGAGATTCCTGCTGCATTCTGGGCCTTTAATGCATAAACACTGAGAT
GTTCTAAGGAAATTACTCCCTAGGGAGGAGAGGGGTGGACGAGGAGTAAGCTTTGCTGGT
GACTCATGCGCTGTGTGGAAACTCCCTGCACAAGTGAGCTGCGCAGGGTGAGTCTAAAGG
GTTAATGCACTTTCAAAAGCCTCTAATTTGTTATTCCAGAAGAGTAATTTACTCACTAGA
AGTATCTGGGTGGCTACTAACACATTTGTGTCTTTAAAAAGATCAGTTTTATTTTAAGAT
TAAAAATATAAAGCAAGAGCTGGAAAGTCACTAAAAACTGACAGCCAGTTTCCCATTTTC
AAGAGTATTTATTAAAAGGTTCTGGTTGCAGAAGGAATAAGAAATGGCTTGAGATCATGA
CACAGTGAATCATGTTGTAAACATGTTAGCTATGGCTGTGAATTCAACCAGCGATGAGTT
CAAGCGTCCCCAGAAGGTGTTGGGGGAATTAGGGACATGGCTGTGTTTCCCCAGAGAAAA
GTGGCCATTTTACTTTCCCTCTTCACTAACATGCTTTTGACATGCATGGCAGAGCTGAAG
GCAAGGGGAAGGGGACAACATAGTAAGTGACTAAGTGGCTTTTTTTTTTTTTTTTTGCC
AAGTGAAGCTGAGTCATATGGCCTCTGTCATTCCAAAACTATTCTCTACGGCTGCATTCC
TTTCGCTCTTGCCTTCCTTTAGAACCCTGGAGAAGGCCTCCTGAAGCCTGGCCCTATTAT
GTATCCTGACAAAGATAAACTTTTCCAAAAAGCTGCATGTTGTTTCTAGCACAGTTTTTC
CTCGCAGTGACTACGTGATGAAAGTACCATGCAGAGGAGGTGTCTGACTGAGGCGTTCGT
GGTGTGTGACAGAGTCCCCTGCACAGGACAGCCGCACTCCCCTCTTGCGTCCTTTCCTCC
CATGTTTGCAAAGCCTCTTTCCCTGTCAGCAGGGGGTGTTCTGGCAGTTGACATTTCTGA
AAACTACAGCCTACATTTTTAAAAAATCCAGTAAGTGAAAACTAAAAAATTAATACCGTG
GTCATAATAGTGTGGCATTTGATAACTAATGAGGCACTGTCGTGCCAGCTATTATTTTCA
GACATTTACAGTCCTTTTTTAAATACAAAGAAATATTTGGTGTGAAATGTTCCCCGGGAG
CTGGTGCAAGCAGAGGCGACAGGGCAAGGGAGCTTGGGTTGTAGCCTCGAATTCCTCCGG
CCAGGGCTACCGTCAGCCTGCGGCACACAAGTAAATCAAATATAAAACCAAAATTTCTGT
AAGCAAATCAGTTTCTAACTCACTGTAACGAATTATCTTTCGCACATCACAGAGGCATCT
CTTTTCACTGTCGAGTTTGGTTTGCTTGGTTACAAAAAGGGCAGTTCAAAAGCTTTGGTT
GCTATTGTGAAAGTCAGCTGAATTCCTTCCACCGTGCTGGGGTGGGGTGGGGTTCACGCA
GGTTCTCTTTTGTCACCAGGGGTGCTGTGGATTCACAAGTAAGCAAGAGGCTCCTCAGGT
CAAGCCTCTGGCTGCTCCCTGAGGTCAGCTGCCTAGCTTCTCCTCCTCTGAGATAGACGG
GAACAAAGTCTTTGATGTGTGCATTTCTCAAGCTTGACAATGATACAGCTACATAAAAAC
CCATGATTTCATATAGATATTCCAAAACGTAAAAGTAAACCATGCATCCACAGAGACATG
GAATTACAGAACTGGATGCTGAGCTGGTCACTTGGGAGGCAGGCGTCCTTGCCATTGGTT
TATGCCTCAGCCCCACCATGCAGTGGCTGGCCAGGTGACCTAGGCCAGTCCTGCATCCTC
GGCTCCTCACCTGCCTGGTGGGACAGTGACATCTCTCCTGCAGCACTGCTGTCAGGGTGA
GGGAGGTAGGGCGCAGTTTCAGAAAACCATTGGGCTGCACCTGCGTGAGCACAGCTGCAG
GAGCAAAAGTCAGAAAGGTCAGCAAAGGATTTCAGGAGCAAAGGTCAGAAGAAACCCTCA
AGGTGGTTGTGTCTGCAGGAAAGTGCTGTCGTCTCCTGCAATGCTTTCAAGACTATTCAG
AAGCACAGTGTGAAGGGAGAGCCGGAGCCCATGGGGAAATGACTCCAGAGTGTTCCACGT
GTTGGAAGGCATCTGTTGGAAAACGGACATTCAAGCAAATAGTTGCCTGCATAGACAACG
CAGAATGACTGGGAAAGCCCCAACAAGTTACCTACTGGTAAATGAGGTGAGAAGCTTAAA
GTGAGAACCCCATTGCTGCCTCTTTTTCACTTTAAAAACATTTAAGttttgaattatggt
aaaatacacgtaagatttactactgtaaccatttttaagtgtacggttcagtagtgttaa
gtatattcacattgctaaggaaccaatctGCTACTTTTGTTTATTAATTTTTTCCTGAGG
GGAAATAtttttaaattttaaaatatttaattgacaaataaaaattgtgtatattcaagg
tgtagaacatgatttcatatgcacgtacattgtatactcattaccacaatcaaagaaatt
aacacatccaaccccacccatagttgccattgtgtgtgCGCGGAtgtgcgtgtatgtgtgt
gtatgtgtgcacgtgtgcgcctgtgtgtgtCtgtgtgtctctgtgtatacgtgtgtgtac
atgtgtgtacgtgtgtgttcctgtgtatgtgtgtctgcgcacgtgtgtatGCATGTATAT
GGgtatgtgtgtacgtgtgtacgtgtgtgtgcatgtgtgtatatgtgtgtctgtgggcac
aggtgtgcctgtgtgtatgtgtatatgtgtatgtgtgtacatgtatgtacgcgtgtgcat
acgtgtgtgtgtgtgcacaggtgtgtatgtgtgtgcctgtgtgtgtgtgtgcatgtgtgG
TGGGGACACTAAAAATCTCTCATCACCTTTTTAGTCAAAAGAACAGTTGTTTTGGTTTGG
CTCTTCTGTTTTAAAATATCAGAACAATAATAATTTCCCACAGACAAAATCCTCAATCCT
CACCATCCTTCTATTTCCTATATTCATCATAAACTTCATGCTTGATGTTGAAATTGTTTT
CTGAAAATAGAGAATACAAAGAGGAGATTTTAAAATGTCAGTGGCAGCCCCACACTCCTT
TTTAATCTTATTTCCTGATATCTTGAGTTTACTTGGACGTAGAGTTTTCCTTGACTATGG
TTATTTCTGGTAGTAGCAGCTCCAGATTAGGCAATGGTTTTCTTCAGAGATAGCTTAGAG
TGAGCCCCAGAACAAGGTCAATGCGAAGATTGCTTGTGTCTGCGTGTCCAGGGCACAGTG
ATCCTCATCACTAGCCGGGGGGCTCCGTGAGGATCTGCTCCTGGTCGTTTCTGTTCTGTA
TCTTCTCTGCAGCCCTTACTGAAGCCGTTACCAACTGGCACAATTCAATTCCTACTGTAC
CCATCATGCACAGATGGCTGAAGTATTGAGAACGCTCCAGTGACCGGGAGGCAATAGTCT
GTCCACATCTAAGAACACACTTGGAATAACCTTAGAGAAGAGAGAGAGAGAGAGAATGCA
TGGTTAGTAGGTTATCAAACTCCTATGACTTTTCACAGGAAAAGCCCTCATCCACACCAA
CTTTAGGAATGTGTAGAAAGAAGGGTCAGGGACAGGGGTGAGTGGTGGGCAGAGCAGTTG
GAGGGCACAGGGAAAAGGCATCTGGTCATGTATTTGGAGTAGGAGGTCTTGCTTTACTAT
TGAATTGCAGGGACACTTTGGGAACAGTGTTCACTTCTTTTTGCAACCATTTCTTCAGAG
AAAAGTCATGATACTCAAGTCTTCTTACAAAGCAGTTTGAGGCTTTGAGTACCAGACTGA
TTACAGAGATGAGTATGAAGCATTATTGTAGTATTTTTAAGTGAAATTCACTAAATGCAA
ATAAACCTAGCAAATGCTCTATGGTTAATTTTTTTCTAAAATTCAGATAATTAAGACAAT
TCATTCTCCTGAAACTGCTGTTCATGTAAAAAGGAATTTTATCGAGGTGGCCCTTGAGTG
CCAAACAGCCTGTCCTCAGCTGCAAAATGAGTCGTTGATGATCCTCCAGCAAGGGATACT
TTTTAGCTCGTGTGGTGATTGCTGCACACGGGATATGTGCAGCAAGTATCTGCTGAGCTA
ATAATAAACAGCCTCAGACAGAAAGACAGTGGGCACAAGGTCATGCTTAAAAAGACCCCT
TGTTCTACTGCATCCCAGCTCCCCACCATGGGGCCTCACAGGCCCTGGTGACCAAGCACA
TCAGACCTGGTTCTTGCTCAGTCCTGGGAGCCACAGAACCCAGCACGTACTTTACCCCCA
AGACCAGACTCCAGCTTGGCTTTTGTCCTCCTCTCCAGGATTGGTGACCTCCTAGGTCGT
GAAGCTGTGATGAGCAAAGACACACTCCTCTCCATTCTCCCAACTTCAGGTCCCTTTGAC

AGTGTCAGCAGGCATTTAAATAGCAGACCACCCACAGCAGGGCTGGTAGATGCAGTGAAC
TCAGGAAGATGCCTGCATAGACTCTAGTGTTAAAGACAGAATCCTTACAAGGAACCCCCA
TAGTTACCTAACTGCTGTCTCCAGTGGTCATAGAAGTGTGATAACCCACTAATCATCATT
CTCTGTCTCTCTGTCTTTCTCATACACACTTACACACACATACACACAACCTTGttgctt
aattttcagagagtctactttcagaaaagccttcaggaatacatcatgtacaaaactgag
aaattACCTGAAGTATCTTTAAATTTAGTAAAAAGTTGCATTGTTTTTTGAACATCACAC
TTGAAAAGTACATGAATACAAACATACTTAGGAAAAAAAGCTTTAATTAATTTAAAAAGG
AGAACAATGCTATATGCTGTATCCCACCTTTCTCTGAATGTTACATTTTCTCCCCTATCC
CAGGCTGCATCTAAGAAAACTCAGAGGGAATATGCTATCTATCTTTTCCGAGCAATGAAA
GCTCTGGGTTTTTTCCTTGCTTTTCAGGGCACAATACTTCTCTTTCTTCCTGGTTAGACA
GGATAAGTTCTGAGTCCCCTGGTATCATCAGCTTACTTCTTCTCTGTTAAATATTCACAA
AAAATCACTAACTTTCATGCCTCAGCAAACCTCCACTGCCTAAAATATAGTGAGGTCATT
CATCTTCGGACAAATTGCCCCAACTACGGTGGGAAAAAGAACCAATGTGTTGGACTATTTA
TCTAATTTTTGTTTAGTTCGGGGATACAAATAAATGCATAGATACATACAAACATGCGTA
CATAATAGCAGCAGCAGCCTGTGAAACATTGACAAGACCTGGAGTTGGAAGAGGACTTTG
CCATCCTCCAGTCCAACAGTTGCCTGTCACAGATTAGACGACTGGGATGTGCGCAGGCGA
TTATTTGCAAACGGCCCTGAGTCCCCCAGTTTATGTCTTAATTCGCAGCCAGGGCTGATT
GTAGAAGCAAATTTGCAAACATGTGCAAGAAGAAATCACACATCCTAGAGCTTGGATTTC
CTCGTTTCTTGCTATTTCTATCCGTAGACAGAACCATTGCTGAGCTGTTAAATTTGTCTC
CTTCCCCTATACCAGTCTTGAAAAAGGAAAGGAAGTGGAGCAAAGAAAAAGAAATTAATA
AAGCCGGCAGATCCTAGGAGAATCTTATTTAATCCAAGCTTTGTAAAGTTTTGCTTTATT
CCATGGCAACATGGGTATACACATCCCACCGGCTGTTTCAGTGGCTCAGAGCAGGTAAGG
CCTGTGCCAAACGCCGCTAGCAGGAGGAACAACGTGGAGACAGCCCCAGAGGTGGAACGT
TGGCCCTTCTGTGGCTCCGGTGTCTCAGGACCTCCCTAAAGCCCAGCCCTGACACTGAGC
AAGTTTCCACCACTGTTAGGAAGAAGTAGAAAGGAATTTGGAGGGTTGGTGTTACTGTTC
AAGAGCTGGAAGGCTTCTGCCCCCATTCCCATTCCATTAATTGCGTGAGGTAGAGAACTC
ATAGAAGATAGGAACACATATGCTGATTTCCAAAATTGCCTTTGTATATTTTCACGTGAA
GACTTTAGGGGCAAAAGAAAAGAAGCAAGCATTTTGAATATGTGTTTCAATTTGCCTTCT
GTTATATAAAATTGTATTTTGCCTATTCTTTTTTCATTATTCGGAACCTTCAAGAAATAA
ATTAAGTTCTCTCAAAAATGTGTTTTTTGAAAAGAGGACTAAAACAGATGGCCTGGCTGT
GTTAAACACAGGGACCAGACCAGCACCCACCTCTCCACCTGCCCTGCCTTCACTGGCAGA
ATTGTGATCCATCATGTTCTCTGTTCAATGTCATCATCCCTTTCAGAGCATGGGTCTCTT
CCTTTCTAGGCAGTCTTACCAGGATGCATGGGTGTGCCTGCGTAGGCACACGCACAGCTC
CCAAGGACTCTAAAAAAAGATATTTTTCTGCTTATATACTAATAATATGTTAGAGATTTA
TGTTTCAAATTAGTACAGAATCACATGGTTCTCTCCAAATTATATTTGAGAGAGAAAGAA
TAGAACAAAATTTATTTTACAAAAATACTCAGTACATTTAGGGCATATACAAAGATGTTC
CAGAATGTAGCTTATCTCTTTAAAGACAATTAACACAGTTTCTGGGCAAGGCAAGGCAAA
ATATTCAGTAACTTAGCAACACCAACAGAAGACAGCCAATATTGCAGCACATTTTTCTCT
TGGATTGGGTCAGAGAGTACTGCAGAGAAAATGGAGTAGAGAGACCTGAAATACTTTCGC
ACACACTGTGGTCAGTGCAGCGTCCACTGTGTGCCACAGTAATACTAGAAACTCCCTGGT
TAGGCCTTGGAATCCAGCTCTCATTTCGTATGTGACCTGCAGGGAAGTAAGTTAAATGCA
CACGTTTTATCAAGTTCAAATGCAAACTTAATTTTAAATGTATGCAACATCAGTTTAAGC
GTTGTAGCTATTACTAGCAATTGTACCTATTACTAGTCTGTACTCTGCACAACTTTGGAG
TATACTGCCTACTCAAGGTGGATTTTAGAGCTCTATTTGTGGCATTATATCACGGACaaa
agcacgttcatcagagtcagaggaatgtggtgcaaatcccagctgtcccacttaccagct
gtgggacttgagtaagctcctgaagcagctgcacctgcattttctggtgggcaccatgga
gctgtcagcagtgctttcctcagagggctgcgggctggatgaggtttgctggtgcatgtg
aagtgTCAATCATTGCTCTCATGAGTGGTGATGCTGATGCCGTTCCCTTTTTTAGGGAAG
TGATTTTCCCTTACAAAGTTACCAACAGTTTCATGTTGGCCATTTTTCTATTAATTGTT
TCCACTAATAGGACCAACAGTGGTAGTCCCATCATTTTATTACTGCTTGTCGTAGCACAA
GCAGTTGCTTCATTGTGTTTAGATAAATATTGACGGCTGCTTTTAACAGTCTGCTGTTTT
GTCTCCTTTTGAGGTCCTTAAAGTAATCCTTAAAAAGATAGTGCAGATGGAAAGATGTCT
GGAGTCAGTGAACCTGCCTTCTTTCCTGTGTGCTTGTCAGTTTCTAAAATGCCATACACA
AAGGACTTTCATGATTTCTTTTTAGGTACATGATTACAGTTCAATTCACTTCACTGTCTG
GAAAATTTCCTTATAATCAGGATGAAATTTCTCATGTTAGCCTTTCACATTTCACTACTT
TTAGATAAGGAATTCTCAGGCTTTGCTATATCTGACTGCTCTTGGAGGCTGAGCTTTTGG
CTAACTACCTGACTACTTTGTCGTTTCTCTTCCCTTGGAATGAAGCAAATATCTAACTTC
TCACTCATTGTTTCTGCTATTTTACCATTTAGTCATCTGTGATTTTTCTAAATACTGAAA
GACTTCCCTCAATTCAAACTATGTGCCGGATCAAGGAAAGGGCAGTTGGATATTGCAGAC
AGCATAGTGCAATTGTGAAGAGTGTctgcttaccagccacgctgccttgcacaagttatc
aagcctctcaacccacttcctcaatctgtaaaataggtatgagtgtaggaccttcccagg
ggatttttttgtgactatagaatgaTTCTCAGAAGACTTTCAGGCAGTATGTGGGTGAGG
CACATGCTGGAAAGGCTTCTGCAGGTGCAGTGATCAATGCTTTTCTCAGTGTGTACATCC
CATAATACAGACACGTTACCAGAAACTCCCTAGCCAGGACTTTGATTGCAGCTCACATTT
TGTATATGGCCCATAGGGAAATGAAGTGTGTATTTTTTATAAAGTTCAAGTGTTAACTTA
ATTTGGAATTTACTATCAAATCTCAGTTGTTATGGGCATTTATAGCTATTAATACTTCGT
CCCATGTGTCCCATGAGGAAACCAAGGAACAGAAATTAAAGTTCTTTCTGGAGTCCCCTG
AATCTCGTTCCTGTTCTTTTGCACCCTGTTAATTACATAGAGACATTCACAGCTCTTCTG
ACCTTATCAGCGTTAAGGAAAACAGAAAACCAGCGTGCTATTTGTTCTGTCCCTTAGTCA
AGCCTTCTCAACATATATTTTTCTTCCAAGATTTTGCATGTGCACAGGGATGCCTATCCT
CTACAAGAAACACATTTTAGGCAAATTATAATTAAAATGCTGTTTACATCTCTTCACCTT
TAGAATTTAAAGAATGATCATTTCTTAGATTGCATCTCAGACACACCCTTCCCCTAGTCT
GGAGAGGGCGAGGCCCATGGGTACTGCAAACAGCCTGACGTTGTCAGGGGCGGTCTCAAC
GGCTCATTCACCACATCTGCCTCGCGAAGGCTAAGCCATGTGCTGTTACCCCTGCTGCGC
TCTGGCTCATTCTAAGGTACACGCTATTAACCTTGTGAGAAAACAAAGAGGCCAGCCCCA
CCCTTCCTGCTCACTCTGAGTCACGGTGAAAATGTTTCAGGATCTCGGGTTCGACCATGA
GTCCTGTCCAGGTCCAGGAGGAAATTCGGAAGGACCACATGTTCACTCTGAGATCCCACT

TTCATTTCCCTCCTGGTTGAGCAGCATTAATACTCTGGCTAGATTTAAATTCTGGCTTTC
TCCAGTTAGAACTGAAAGTTATGACAATGTAATCAAAATAGAATGTGGGTTTACAGCTGG
CCCCCTGGCCTGGTTTGTGAACATAAAACAGAAACAGAAAGTGTAAGTGGTGACATCATA
TTCTCTCATTCAATGTGAAAGGCCACCGAAGTCTTTCCAGAATTATTTTTGAGaataata
tgaatttttaaaaaatacctaattattttaaatatCGTCTTGCTTGCTCCCCAAATACCT
ACTGTTTTCAACTTGGATATACGACATGATTAAAGAATATCTAATATTTGGGAATGCATA
CTTTAACCTTATAAACTACCACTGTAAATAGACAGACTCATTAAAGTGAAAGGACATTTT
AAATCAATTAGTAAGCAAATCAATTAGGTGGCAAAGACAAGATTATTTTTCCTTATGGTA
GTTGAAGAATAATGCTTAACCTGTCATTCTAATTACCAAGCACGGTGTTCTCTTTGGAAG
ATCATTTCAACAAAACATTATTTTCATCCAGAATTTGAACCTTGAGATTGCATGGTATTT
TAGAAATCTATTTTAGAAATCTTTGGCAAAGGTTACTATTAAAACAATCACATTCATGGA
AAATCAGTATAAGAGCAACTAAAATAACTCACAATACCAGTAAAATCACTTTGTCATCTT
CTTAAGACTTTTAAAGAGCATTTGTAAGTAACTGAATAGAAGGCCAAAGGGTGTGTAGGT
AGCCCAGACCATCAGTGGGCAGCCAGGGCCAGGGCAGGGGCCACGGTTGCAGCCTGCATT
CTTCTAAAGGGCAGAGCAAATTAAAGTTGAAGCAGGAGCTAAAAAAAAAAAAAAAAATGT
TTCAAAGAATTCCACCAACCAGAGGATACTACCTAGGACAGTTTGGGCCTAACTTATCTG
TGAAGGCCTCCAGCTTCCTCCACACCGGTGGCCACTTTTCATTCACTCTGAACCCTTCTT
TGTATGGAGGTCATTTTATTAATTGAGCTGTGACCAACATGACAGAATTTCCTGTTTTAG
GGCTTTTATAATATAGATAGTTTATATCTAATTTCAGAATATATTCACTGGGGAATGGAC
TTAGCAACCACTACCACAACAATGCAACAATGTGTTTTGGAACAAATTTACCAATCTGAA
TTTCCCCCTAGATTAGGTCACAGGAACATTGCAGCTGATGTACAGCTATGTTCCTCCTGA
AACTTGGAGACACATCCTCTTGAGCTGGGTTATAATGGGCCACCCAAAGCTCGAGTTCCT
GTAATGGATACACTCAGGCAGCAGAACCTACCACCGTAGTGAGGACAGCACCCAGAGCCC
TCAGAGGCCATCACAAGTGCACCACAGCTGCCTTCTCTGGCACGCTCAGAGCTACACAGT
GTACTCTGGGATTGGAACTCTTTATTTTTTTTCAGTTGATTTGTAAATAAGATTGCACA
AAAATCCATGCACATCAACTCTCCAAATCAGAATTTGCTGAGCTAAAAAGAGCATTAAAT
TAGATGGGCTGGCTTTCAAGGGGTGGGGGTGCAATAGTGGAACTCTGCACAACAGTTCTT
TACAAAGAGACAAGCAAGCACATCGCGTGGAAATTTCCATTCAACTGGAAATGTCCAAGC
CTGTTTACCTCAATTAATTGTCCTTGTTCACTTGTCCAGCCTAGCAATTGTCCATTAGTA
ATTTGTTATAAATGAGACATTTGGTATTAAAGCATCTCTTTGGGATACTGGTATGGTTTA
TTATAACATTCTGTTAGTAGTGTTGTACAAGCTTGAGATGTATTAATACGAAATCCAAGC
TGCATGAGGGCTTTATTTTTCAAGCCTACACCTTGCTGAAATTCTGAATTAAAATATGAT
TCTCAGTACAAATGAATAAATCAACAGAAATGGTAACGCATGTCAAATATTCTTAAAACC
CAAGAAAGCCTTGTAACTTCCTTCAATCTAATGGGAAATGCAGGCAAATACAAGACTGAT
GTCCTTGAGTTTTATTATCAAGACTCAAGGGCACCAGTAAAATCTAGTTTCATTGGTTGG
AAAAAAAATCCTGATAAGCACTGTTAGGCATATTAACTTTAATGATTACAATTTTTAGGA
CACTCTGTGGCCTAGACTTAGAAACACAACTAATGTCCAGAAAAAGATTCCTCTTTTTAT
TCCATCATCTGATAGGCCTATTTTTACACATACACACCAACCAAAAGTAGCCAAGCAAAC
AAAACAACATACTCACACCCCTTCGCCTATTATCATCTAGGTGATTTTCAATGCTCATTG
CAATGAAACCTACTTATTGTGCATGGCACCCACCCCCACTGAGGAATACTGTAGTTTCTT
TCCCTTTGAACTTCATTAGTAGAGCACATGGTTCATTCACTCCTGAAGAGTTCTTCGTAT
GTCAGAATATATATACTACAACATAATTTCCATCAGAGCTCTGACCACCCCGCTTATCTAT
TTTCATAATGCCTGCCACTCCATCATTAGCTGTTGTCATGTAGGCTATCAATAAATATAT
GACAAATAAAACAGTTAGGGAATGAGGGAAATTGACTAGCAGCCAAAGACCTAAGCCATC
CTCTGCTTGGACATTAGAAAACTGAGTTCACTACAGTCATAAGATACACAAAGGCAGAAT
GTAAGCCATACAAAAATCCATGTCAATCCCAATATGTGAGTACAACTATTGAACACCATG
TACTAATGGATGAGTTGGTAAATCATTCAATGTCTTCATGAGGTCAATTACAGATTATTA
TTTAGACCCCAAAGATTCCAAAGATGGTATTTCGGTCAGATCTTCATCCTTTGTAAGCCT
AGCAGAAAATATGGCAGTTTTATTGACTACTATTCTTTGCTGGGTGTGGTATTTTTAAAC
TGAGACATCAGTGTGCCTAGCACAGGGCCTCAAGCACACAGAAAAATTCCTTGATAATAA
TTAAATAAAATTTCAGCAAAAAATATCATCTTAAGGCTGTGAAATTATCTTCCTGTGTGG
CTAAAATAGTGAATAAAATTCAGCGCAATATAAATCATAGTACAATTTCATCACTAAATT
TTCTGATCTTGATCTTGTCATTTTACATTGGAAGTAAAAATGTGTCCTCCTTTTTTTCTC
TGACAGTGAAAAGTGTGTGTGTGTTGTGTGCCCTTTTGCACACCCTGCCTCACACTTGCT
GGTCTAATTCCTTCCAGCATGATTATGATATAATTAAATGACAGAAATGTTTACTTCCAA
GTGGAACTAAGCCAGGGTAACTCAGGGTAGGGCAGCTGCTTGCACCGAAAGACCAAGACT
GCTAGAGAACTAGGAAACAGGCGGTGCAAGAACTCCAGGCTCTCATGGAAGAGCGGGAGG
CTTCTATGGGGCTGCAGAAACTCTTTGGTGCTTGGGGAAAAATGGGTTAAATGCTCTTA
AAAAAGAAACCTGGGAGAGGTAGTTTCCAGATGCAGGCCCGTCTTTTCTTTTAAACAGAG
GCAGCTCCGAAGAGCTGGACATTGAACCCTGAGCAGGAACTGGAGGCCGTCAGCGCAGCT
TTGTTTGGCGAGCGGAGCTTTGCAAGGGTGTAATGCTGCACCAGGGAGACGCTATCTGCA
GGGACCGGTGACGCCGTGGGTGTGGAGGGGGAGGCAGTGGCTGGCCCTCTTGGGGTAAGG
TACGCCCAGGAACAGTTTAGAATAACGTGCGCGAGTCAAAGGGAAGAAGAAGCTCCTGCA
GACCTTCTGGGCACTGTGCAGGGTTTGCTCCTGTCCACCGTGCCGTGTTCCTGTCCTGGG
GTATTTGGGTGTGTGGCGTGTGGGGAGGGGAGAAGGAGCAAGGCGGCAGGGAGGGGATGA
GGACCACCCTGTCCATGGGACAGGCCCTGGGCCCCGCACACACCCCAAGCCCCGCGTCCC
GCGTCCTCACTGTCCTGGGACACCCCCCACCCCACCCCACCGCCACAGCCCAGAGCGGTG
CCAGGAAGCCGCCTCGACGCAGCCGTATCTTGAGGCTCCAGCCCCATCCCCAGGGTACCA
CGCCACGTAGAGACACTATTTTTCACTTCGTGTTTGTCACTCCTAAAGCATGTGTGCTAG
CTGCACCAACCCTGGGATGCCTCGGTGCATAGGGTTTATGTGCGTCCTCCTCCTTCCCTC
TGAGCTGGTCCCCCGTGGGGAACTGCTGCCCAGACTGACCTGCGTCCTTCCGCACGTGCA
GGAAAATGTCCACGTGCACTTGTCAGGGTGGGGGCCACACGGGCACCACCACTGATCATC
TGTGGGATCGAGTTACTGCCCATGCAGATCCCACGTGCAGGGCCCAGTCGCTTTGGTGAG
AGAGTGGACGCTGTGGTGACTCCACGGTCTGTGGCTGTGCTCAGGAGGACAGAGAGGGGA
CATCCTGAGATGGTTTGGGCAGCCCGCGGATCCTGTGCATGTCCCCAGAGCGTCCACTTT
CTCCATGGAGCAGTGGAGTGGCGTTGCTGAGACAGAAAGTTCAGGTTCTCCACTCCCCAT
GCAGCCCCCACTCCCCTGTCTCCGGCCAGGCACGCGTCTGGGGTGGAGACTCCCGGTGCC

CGGGGCCCTCCAGACCTCTTTCCCCACCCCAGGGAGCAGGCGGGTACTTCTATTCCGTTT
GGCTTCAGAAGGGAAAAGAGAACGTAAGTTCAGGGAGTTCTCGTCCATTCCTCTCCCGTG
GGCCGGGCAGGCAGCAGGGACAGCCTTCAGGAGCCAGGAGGGGCTCGAGCTGCGAGGCCC
TGGAATGAGGCAGGCATGGGCTGAGGCTGGAGGGAAAGCCCCGCTAAGGCTGGGCGGGGG
CGGGAAAACTTACCACCAGGGGACTCGAGATGGGGAAGGAAAGGTCAGAAGAGGAGAGGC
CAGGCACGGGGTGTGGGCGGCCTGCAGAGCTGGAGCAGGTGCTCCGCCCAGAGCCAGGCA
TGCACACTCAGAGTAGGTGGCCTGTGCAGCGGGGAAGAGGGGCGGGTCGGCGTGCTGCTG
AAGATGCAGGAGCTGCGGCCTGCTCTGTGCGTGCTGAAGGTGTGGTGAGAAGCACTTACA
AAAAGAAATGGACTGTGTTAGGATTGCACATTTTACTTTGTTTCTCCCAAATACGTGTTC
TTTGAATTTTTTTCCTTCCAGGGCCAGGACTGGAGTGATGGTTGAGACAggcacgcactg
ggtcttgtctgcatttacattttgagattttgttcagcatggattttATGGCgttttttt
gtttgttttgtttgttcgtttTCAAAATACTGCACGGtttatcgtgaagacagggtccttt
gctgccgtcttaagttttgggcccaagaacgtgccccaccctaggcccgggcctgCTGGC
TTCATAGCTCTCATCATTCCCACGGAACCTTAAGACCTGAGGACAGAAAGGAAGGAAACA
AGCCCAGTAGTCCGTGAAAATCCAGGGTCCCGCCACTCCAGGTGTCTGCAGCAGAGCTGA
ACACACGTAGGCTCTTGCCAGGAGGGGCATTTGTATGTGCtgagcattccttatattctc
aatatgacgcctttgaaagatctgtggtttgcaaatatttactctcagtccataacttat
ctttccaacctcttaccaggctcttttgctgaataaaagttttaaattttgaagtctaat
atatttttaatttttttattttatggatcatacttttttgtgtcaggtttgagaagtctgc
accaaagtatgtcctgtggttttcccttaggtcatcttcaacaagtttcatagtattttg
tttagatgtaaatctgtggcccattttgagttagtttttgcacaagagttgaggtcaagg
ttctttttttgcctgtgatgttcagtggctctggcaccatttgttgaaaaCATGATAGCC
AATGTCAAGACTTAATAGTTATAATAATCAGGAGCTTTTGTTTCTTTTTGTTTTGTTTTT
AGTAACTGCCAGTCACTGCTtgtggtatacatacacaatggaatactattcagtcttaaa
aaaaaaaaaaagaaggaaatcctgtcatttgcatacctggaggacattatgttaagtgaaa
taagccaggcaccaaaagaaaaacattgcatgatctcactccttcatggaatctaaaaaa
ttgtattcagagaagcagagagtggaatggtggttaccaggggctgggaaggtgtgagct
tggggagatttggtgaaaggacatagaatctcagttagacaggaggaataagttaaagag
atctattgcacatcatggtaactgtagttagtgacaatgtattgtatacatgaaaattgc
taagagagtagatttaagtgttctcaccacaccaaaaaaaggtatgtgcagtaatacag
tcattaattagcttgatgtagccattccacaatggatacatatatcaaaacatcatgttg
tataccataaatatatactgtctctttatgtaaatttaaaaaTAAGATAAAATAAATGTT
ATTCACTTGTCGTGGATGTGGTGGGGACAGGTGTGGGATAGCCCTCCCTGTACAACTAGG
ACCCAGGGGTGATCTAGTGACACTAGCCATTTATCAGGACGTATGGGTGCCAGTCAGGAT
GATAAAGCTTCCTTTTGGCCACTATACTACTTAGAAATGCCCTGCAAAAGGTGCACATCA
AAGATTGAAAGCTCAATCCTGGATTTTAAGTGCTTCAAAAGTGCACTTAATTGCCACATT
TTTGTCAAACATTTTCCCAGGTAGTATTTTTCCTCATGTAAAACAACAGCAATTTAATTT
GAACAGAAAGCATTTTGAAACATACTTTTGGCAGGGTTCCTTGCAGATCAGAATGGAAAT
GATTAACAGGGCAATTATCAATCATGGACTTTTGGCGGCAGAAGGAACTGTATTGTTTGG
TACAGTCTGGGCCAGGGCCACACACCGTAACGGAGATACTCTATTCTGTGGACGGTTGGA
GGGGGCTGTGCTGAGCAGGGTAACTGCATCTTTTCCTAGACTGTTCACACTGCTGCCACG
AAGGAGTCTTGTTTAGACTGGACCTGGCTTTCTTCTTCGCAATGAGTGTTGCAGACTCCC
GACAAAGGCCAGGTGGTAAAGTGTGGTGTCTGTGAGCGAGAGCCTGAGATGCCTGAGCTG
ACCTGTCCTCAGCCACCTGCCATCGTGCAGAGGTGAGAGCAGCCCCTGAATTCTGCCCCT
CGGTCTCTCCATAGCTAAAGCAAAACCATCCTTCCGTGCTCCCAGGACAAGCAGGCTATT
ACCAAATCACCACTAACCCTGGGCGAGGAGGGGCCATCACTGCACAATTCATCAGTGTC
TGTGACAGGAAGAGATTGTTTTAGACTGGTTTTTTTTTTTTTATTTGCAAGCTTTTTTCT
CTCTCCAAAACGTGCTGTCAGTGTGTTCTAATTTACTCTGTAAGGAATTCTGGAGCTAAT
CATAGGCTCACAAAAAGCAGCACAGGAAAGTTTCCCAGATAACATCTATTTcagtggctt
tcaaacatttttgaccttaccaaagtaagaaatacattttaatatcatggcacacataca
gctgtatctaaactttcataatactgcctttacgatatcactctgatattgtctattctt
ttctgtttatttttcttttgttccttgttatgctggttgtgacccactccagtgatttc
acaatgcaggctgggtggtgtcccacagtttgaaaTCCCAATCTAGGGCCTTCCTCTCAC
TGTACAAAGTAGGTAACTGGGGACATTAGTGGATCAGTGATCAAACCAAAGTTATTTGAT
CTTACCAAGTGATATCAGGATGAGAAAGCTGTTAGAGTGTCAGATATGTGAAGGAACTTG
GGTCATTCCTGATACCTCAAAGAGAAAAAAGGTAGTCCTTGAACACCTCCTACTTGTAAA
GGATGCACAATCCTACATGCCCCTCCCTTTCCTTTCCTCCCCTCTGTACCCCACCCCTGC
CCACATTTTCTTCATAAGCAGCTTTGGTGTTTTGGCTTGTTTGTTTCCCTTGTCTCCTAC
CTGTGACTTTATAGCCTTTTGGAGACTCACAGCAATAGTTGTATTTAAACTCAGTGGGTG
GCATCCAAGGCTAAAAAGGAGATTGCCTAGACACAAAACCACCCAAGGGAGAAAGCAGGA
CAGCATCTTACTATGATTGTTTCTTGTTTCTTCCTGTCTCATAAGGATTATTACCCAGGG
TTTTCATTTTTTTCATTTCATGGTTCATTTTCGCTCCAGTGTAGACATACAATAGACCAC
TCGTCCCTGTGGCTCCGGGCAGCAGCCTCATCTGAGACCCTCCTGAGACATCTCGTGCAG
GGCAGCCGTAGTGTGTGGCttccccagggctgctctaacagatcaccatccttgccatgg
cttaagaagctgcagatttatttgcttacagctctggaagccagaagtccaaaatcaagg
tgtcagtagagtctctctctctgaaacctgctgaggatgatgcccctggcctctccccag
cctctggtgttcccagcagcccttggcattccttgccttgtagatgcaaaactccgatct
ccacctctatcctcacagtgagttctcctgcatgtctgtctctgtgccttcacattcctc
tctgtgtgtctgtgtttccatctccttatgaggacacccatcactgaatcagggcccact
ctataccagtaagacctcatttcaactccattacatcttcaaaaaccccattctcaaata
aggttacttcacaagtgctggaggttaggacttgaacataCCTTATTGAACAATCCAACT
GATGACACATAGTAATTTATGCACTCGTTCTTGGAGACGTTGACTTTATTTAGTAGCATT
AACCATGGCAATGTCACCAGCATCGCTGACAGCCTGAAGCATATGATCTCCAGAATGTAT
TTCAATCATCATGTTCACTTCCTTGGTATTCTTTAGACAATAACTCAGCCTTGAACTCCA
GTAAAGGGTTTCCCTGGGATTTTCTTCTTGACTCACTCCACTGTGGCCTCCCTCATCCAG
GACTGTAACAGACGCCTGACGTCAGTGGTCTAGACCTCTCTGCTGAATGTCATCTTTGGT
GAATGTCTTATGAGAAAACACATGGTTGGTCACTCTTAGAAGGGCATGAAAGCCTGTCTG

CAGTATAACCAAAACAGGCACATGGCGAGGCACACTGTGCGCATGTGTGTACAATTAATA
TCATGGTTTTAAATTATTTTCAGGCCAAGGGGAGATCTTTGCTGCATCTACTGAAGAAAG
CGAATCTTTTTCTTCCTGAAAAAAAATGGCTACTTATTAGTCGAATTTGTGTTTTAAAAA
TATGTGAACTAATATAATGCAGACATGCATTAATGTTTAAATATACTGGAAGTTTTTGGT
AAAATGAAACCCATTGTCTCTGTTGATTACTTTGATGAGTCAAGAAGTAACATCCTGGGA
ATGATTGGCCAGTTTAAATGAGTGCCTCAGGTTTTTGGAATACAAGAAATCAAGAGGAAG
GGATTAGAACATATAGGTTAGCAAGATTGGGATCCTAAAATACAGACCCAAATGAATGGA
ACAAAATCAGGGAATTTATTAATAACAGGGTCAAGGCCAAATCAGTAACAAATATCCTGA
GTGGAAGAAAGGTGGTTTAACAAATGCCCCTATGAAAGATAGAGATTGGCTTACCATGAT
GAGATGTAAGCCCAAGTTATGAGGTTGGCACacaaaaccacaaatgtcatagcttaaaac
aacacacacttcttatctctgtttctgtgggtcagggtctggttctcagggactcacaa
agtatgttttcatctggagctccaggtcctcttccaggctcataaggttcttggcagaa
ttcagtttcttgaggctgtaggactgaggtcctggctcctagaggccaccctctccataa
gcagttcttagcatggccgcctgcttctccaggcccagtgggaaagcatgtgcctcagg
agggctcagtccattcttcatggcttttacctggttaagtcaggcccactcaggataact
tcattttgtattaaatcaaaaccagctgatttgggatgttaattacatctgcacaacttc
aactttgccatataaccta accatgggactgatatttatcatGCATTTGGGTCAAGTTGC
ATTAAGAGATATAATAAAGCTGGACAAGCTTCTGTTGATTAGAAGAgttcagttacaagg
ctacacttgggaggaatgtttacaaactggaatggtcagaggatggggaagacacttgag
aaaagtcaagtgacggatgaaggcaaatgtggatatttatctgggagaaaactaagagga
gttataatagctgtcttcaaatatttaaagggcttttattaggaagaggaatttggcata
ttggatttgccttcagagAAGTGGAGTCCTGAGATGCTCTTAGCCATTCATTCCAGCCT
CCAGGGCTCACCTGCTGTCTTCTGTCCAGGTTCTCGGTAGCAGGGCAGTACAGCCCCATC
CGTGATCTTCCATAGTCAGGCATATTGTCACACTCAGTGAGCGGAGAGTCAACCGGGAGG
AAGGCACAGTTTCTCTGGAATGACCTACGGAATGGTACGCTCAAATGCAAATTCTCCTTC
CCTTCCCCAGTCCTTGTCCTTCAGATGGTAATTTAGGAGCTGAAGGTCAGGGCACCAGCA
GCCTTTGGAAGCCTACAGGACAACAGTCAGCCTGGCTAGAAAAAAAAACAATGTCACAGG
CATGTTGTGTTTAATCACATGAAGGATATTTGCATTGTTTTCCAACTGATGCCAGCAGAC
ACATTGTCAGTGGTATCATGCCTGGGGTATCAGAGTTGACATTGGGTTGCCCCTTCTCTG
AGGCATTCATGTAAATCCTTTTAAGTTTATAAAACCTCCATGTGGCTCCTGCATGCTTCA
TCATTTGCATGTGTCTCTTTTTCCAGGGGAGGCAGCATGGGGAGCAGGATGCTGGTGGGC
TCCAGGTGCAGAGAGCAGGGTGGGCGTCAGACCCCAGGTCCACTGTGCACGCCCTCTTGT
AGAGCCCGTTCCGTTGTCCATGAGATGAGGAGTGTTCTTATCTCTAAAGTATTATCATGA
AAACCTAACAATGTAGAAAGACTAAAGCACATGGGTGGTGCTTCATAAATAGTATTTCTC
CCACTTTCTGAAAACTCCTGCTGAAGTAACTGCACAAGAATCCTTGAACATTTAGAATTC
TGGTTTTAGCCATACCATAAAGTCAGTAGTGCGTGGTGGAATTCTGCTAACGAAAATTGC
GAAGGATCAAGGCAGAGTACAGAGCTGGTGTGTAGCGGGTACCTTCTGTCTGCTGGCACT
AGGTATTTTACACATTAAATCAGCTCGTTCTCACATCAGCTCTTTTAAAAATAAGGAAAT
GAGGAGCCACAGTGGCCCAACTGATGCAGTGGCAGAAGTAGAATTTGAGCTTGTGCAGAT
GTGCCTCCGTGTTTTGTCTCCTGAGCATGCTGCCCCAAGTTTGACAATACCAAGATTTGT
ACTGGAACATTCCCTCCCATCCCCACCCCCTAGAAGCCCCTCTTCCTCCCTTAGATTTGA
CACATAGTTTGAAACCACTATTAACTACCTTATGagagccactgtttgtgaagtgctgac
tatgtgccaggtcccgtgccgtgcaatttttgtgaattatctcgtgtctacagtgcctca
caatttctctgctcaatacctccatgttactgccgaggaaagggaagctcagagagagta
agtaatttgctcgagttaaagagctggccaggacagccaggggcttgcaccccggAGCCT
TCATCCACTACACTGTCAGCTGGTATCTCAACCAGCCATTACAGGCTGTAAAAAAATTAT
ATAAGATAGTCTATGGTAATGCAGAAAAGTGAGGTTATTTTGCTCCCTTTCCCTTTGAAG
AAAAAAGCCCTGGAAAGACATATCACTTGAGTATGGGAAAAAATGAAGCTGTGGCTTTTC
TGTGAGTCAATTCTTTCCTGGCAGCTTCTTGGAATAAGACCAAGTATAGCAGCAGAGTTT
TCTGTTTTAATTTGAGCTGCAGGGTGACTTTTTTTCTTCTATGCTTTCATCTCTCTGTGG
CTTCTTTTGCCTCGTTAATTTCATGCCCTGCCCAGGCGGGCTACTGTGCTGCCCAGTCAC
CCGGGTCTGGGGCGGCCACCGCTGGCCAGCAGGCAGGCCCTCCAGAGGCAGAGGTGGCCA
CGCTTAGGTCGCTCCCGCGTGTGGAGGCGGCACACTTGGGTGGCAGCACAGCTGTGATGTG
GCGGCAGCTGGCAGCCCCATGGGAAAGATGTGTGAAGTGTGGGGTTTGACGACCCATGGG
AGAACAGACTTTCTTCCTCTTCTTGTTTTCCCTTCAAAGCCGTGAGTCAACCTCAAATTC
TCTGTCTTTTTTCTCCACCCCCTCGTGCCTCTCTCCCTCACGCTCTGCATCTCTCATTGC
AAGCTTGCATTTTTTTGCACACAACACTATCTTAATATTTCTCTTTTCTGCAGGCAGGAA
ATGAGAAGTCATTTTTCAGGGTCATTCAGGAAGTCATCCAGAGTTATAATGGCCCATTAT
CTACTGGTCAGAGTTTACTTAGGCTTTCACTACTTCCACTGCCCACTTGAAACAGGGAAA
AATATTTTCCCCCCGCGCTGTGAGTGTGCTATTTAGAGCTGACCACAAGCGGGGGGAAGA
GAGGATGGCTCGGATGCTGCATTTCCACTGAGAACACAAGGCTGGCAAAGCTTGTCTGCT
GCCCAGCAAGCACTTCAGGCTCACACCATTTTAGGTTCACTTTAAGTAGTTTCTCAATTG
TTaaaaaaaaaacaaaaaaaaaaaaaCCTGTACTCTGAGGATATGCTTATAATCCCATAG
CTAACCCAGAATTTCTTAGAGAACTGatcaacatcagcagtggcacttactgaaaatgca
cattctcaggccctgcgtagggcctactgagttagaatattagagagcaggtctcagaaa
CATTCTATCCGGCAGTCTTATTCTATGCACCCGAAGGGATAAGAGCCATGCTTTCATGAA
ACATGGGTTGTGTGTAAAATGTTTAAAAGGTATGGCAAAATGTGTTTGATTGGCACCAAG
GATTTCTGGTTCCTCCTAGAATCATTAATCAAACTTTGAAGGAGAAATAAGAGAGTCGGC
ATTTTCTTGCACATTCTTTGTGATGTTGTGATGAGTTGGAAACTTCCCGATTGGGTTTAT
TAGAGCATGAACACCCAGGCACCCAGCTTCTAGCCAGCCCTGTCAGGCAGAGTCTCCTCG
AAGATGTGGAAAGGACTGACCAACAGCTGAGGCCTACAGGAACCTGAGCAGGCAAGGGGA
GAGGCACCCCGGAACCAGGAGCAATGGCCTTCCCACCCTCCCTCGTCCTCTCCTCTTCTC
CTTTTGGAGTTGCAGGCCACAGAAAGGAAGTGACATGAGTCACTTTGGGCCTTCTTAATT
CCTTCATCAAAGGCAGCACAGGTGTGTATGTGTGTTGGTGGCTAATTGAGGTAGGCCCAC
AGAGGAGATAACAGATGGACATACTATTTCCTTTCTTCCATTCTGATATAATTCAGGGTA
TAAacacacacacacacacacacacacacaTTCTCACTTCTTTGGCATCTACCACACCTGCC
CCAGTGCCCATTTCTCTCCCACCTGAATAAAAAGCCCCCACAAAGCCTGAGGTACATGGA

AAGGAGCAGTGGTCTGGCTCCCAGGAGTGTGAGAAGCAGCCATGTTTTCAGAGGCTGTAT
TCCACTTGGACTTGGCCCTACGCTGAAGGTAGGAGCGGATGGGGGAGGCCCCCTTCGCAC
AAAGAGCCCCATGAAAGAGTGCACAGTCCAGTCTATAAAACAGACGCAGAAAATGTGTGT
AGGACTTCTTCCTGAAAAAGAGCGTGGTGCGTCCAGTACCTCCATGTTCATGGAACTTCC
CAGTCTGCAGTTTACCCTTTTGTGCAACTCCCTTTTGGTAAAGCCCTGGTCACACTTCTG
GTTGTTCAGATTATACAGGGATAATTCCAGAGTGATTTTAAAGTCAACTGCCAGGCATCC
GCACTTGCAAATTAGATGCCTGGCACATGCTTGTGTTAAGGTAATAATTCATTACAATAC
AAATTACAGGGGAGTTCCTCTGGGCATGCGACCTTTCCCGTCATTTGGCTTTCCCTGTGA
TTATCAGGGGAGCTTCCATCGTGCTGCTAATGGGACCTTAACCATGTGTCAACCCATGGC
TGTAATGCTGACACTGTTTTCTTTCTGGAATGAAAGGCCTTCGCAATTGAAACCAAAATG
TTATCCAACTCAGTCCTGTCCCTTTGACGATGAAAACATCAAGTTCTGGAGACTGGCCAT
CCAGCCTCCCTGCCTCATCTCCCACGCCCTCCATCATTTTTTGTCTCTACTTACTTATTT
ATTTGGCTGTATTTTACGTACATCATGCAAAAATATTCCTCTTTGTAAAAAGTATAATGA
TTTCAGGAAATTAGAGGGTAAAAAGCAAGAACCATGCTTTCACTCCACTGTCAAGAGTTG
TGGAAGAATCCTTCCAGCATTTTTTCTGTGTATTTTACATACATACAAATATATGTACAA
ATAAAGGTCGATCATTTAGGTTTTGTTTATATTTTTGTATATGAGCTTATGTCATTCA
TACATATTGTTTTGCCTCTTGCTTTTTTTTAACTTAATTTTACTTTGCTTGAGAGCTTTT
TGAACTGAAGTACGTGTAAGTCAGCCTATGCATGTAATGGCTCCCTCATCTTCTGTGAGG
CTGTCACTAAAAAGGGGATTTAGCTTGTTCTGGGCTTTGCAGCCCGTACACTGGGCACTG
TTCATACGTACTTCTCTGTGCACGCAAAGGAGGGCTTGCTAGGGAGGCCTGGCAGAGGGT
GCCATTCAAATAGGATTTCAATGGAGGAATTTTTAAATTTTCAGTTATTTGAATAAGTT
TTAATATATATCCAGAACCCcaaatcatcaagtttgttttcttccacatctgtccttcca
tttctgaactattttaaggccagtcatgtctcatccaagaaatcccatcctttcacacaa
cactatctccgtttcatggttatgaatctctaaaagcatgatttttaaaacataatcaca
atgctgtcatcgaacttaaaaattagccataaatctcttatgttacccaacaaccagcct
actgacacatctccagttgtctcaaaaatgtgttttccattgtggtttgtctgaaacatg
atccaaaAGTCAGACCCACCTCTCACCTTTCCCTAACCTGCCGGAGCCCATGTTTCTTTC
CAGCCAGGCTTGGAGACCACCACACGGGATTTGCTTCTTGGGGCCTCCCTCTAACCAGCT
ATGCAGGATGCCCTCTTTCCTGTCAATACAAGCTGCTCAAAGGACTCATTCAGTTCAAAT
TCACCTATGTGAGCCTAGGTGATGCtactttatttatttatttatttatttatttatttat
ttatttatttattttgagatggagtctcactctgttgcccaggctggagttcagtggcat
aatctgggctcactgcaagctctgcctcccgggttcaagtgattctcctgcctcagcctc
ctcagtagctgagattacaggcacgtgccaccacgcccagctaatttttatagttttagt
agagacagggtttcaccatgttggtcaggttggtctcaaactcctgacctcgtgatccac
ccacctcggcttcccaaagtgctTCATGTTTTCAGGAGCTGTACGTGCATTTTTAGTTTT
GATGACCAGGTCCTTTTTCTGTTTTTTAAAGAACTTCAAATGATCTCCAGGGTACACAGC
GCTTGTGTGCTGATGAAAAAGCTGGCAGTACAAAGGCCACCAGCCAAGGTCACACAGCCA
AAAAGCCCCTGACCTCGGGCCCCTTCCCAGACCCTGGGTCTTTTGCTGCCACATGAATCT
TCTTCAAGGTCCTATGTGTAGATTTTCTTGACTTGGCCATATTATTTAGGATTCAGATAT
AATAACAAAATAGATGTTAAAGCATAACATGAAGGCATTTAAAAGGGTAGAAAGCACATG
ATTTACTAAAACCATAAATCTTATGACCTGAAAGTTTCACCTAATCTCTTAAAAAAATACC
GTACTAAACCCTGATTGAAAATCAGAGCTCAGACATACAGCCTGAGATGCCAAAAAATGG
CCAGGCTTGTCTGTTGAGAAAGCCATATGTAACTAACTGTTTGGAAATTCAAAATATATC
TTATCATTTAAAAACATCTTTCTTCTAAAGACAATCATCTTGGCTTCAGGAATGAGGCT
AGTAAAAAGTGAAATACTCCTACTTGTGGAAGAAATCCTCATTTTAACCATGAAGAACTG
AAAAATGCATTCTGATGTTGATGGACCCAACCTATATTTGGGTATTTTATGATGTACACA
ATATACTTTTGTATATGAGATTGTTATTAAATGTGACTTTGCTTTTTCAAGACATACAAT
GTTCCTCCGGGGGTCAGGCACTGTGTTTAGCACTTTGTCCTGACCTCATCTGACTTCTCA
GCTGTCCCTGAGAGGTACCAGTGTGCAAGATCGCTGAGTTGGCAAGTGATAGTGACAATA
TTTTCACCCCAATTTCTAATTTAAAGACCCCGATTTCTAGTTTTGTTTTGTATTGGATTT
GCACAATTTCACGTTCTGAAAGAGGATGCCCTCAACTTTGCAAAATGGGCCTTTTGAATG
AAAAGGATCAGTCATGTCAGGAAAAGCGCTACAATGATGAAATATGATAAATAAGTCAGT
CTTTCATCTGTAATTATCTACTATGGGGTAAAAAGTGATGAAAACTACCATCTTGAAAGG
TTCTGGTGATAGTGGTTCCTAATGCAGTGAAAGATGTGTAAGTCAAAGATTTGTAACCAG
CCAGGGAATGAGAGGCGAAGCCATAGCTGGTGGCGGGGGCCACATCTGGGTGTGGGGAGG
CCACAGTTGGGTTGGGGGTGGGGCCTGCAGTTATCCACACCCCTCCCACCTCCCTTCGAC
AGTACAGGCTTCCTGGTTACCTTCCAGAGAGTAAGGCCAGGGAGAGTTGAATAAGTTGAG
AAATGTCATGTCGAAGCTATTGGTGGAAAGAGTTCCATTAATTGACAATACAAGTCCCTA
CTACATTCTAAAATCTGGTCCTGACTAGTGGCAAGCCGGGCCCAGGAGTAGCACTTAAAC
AATGGCAGGCTTGTGTTGCTGGCAGGATACTTCAGCCTCAGAGGAGCTGTGTGCAGCTGG
GGAGACTCACACTCAGAGGATTTCAAAGCAGAGGGCATCTCGTAGAGCAACTTATCCAAA
CCCTGACCCACTGTAAacacacacacacacacacacacacacacacacacacacacac
CCTgagagagagaaagagagagagataactaaagagagagaACTAAAGTTTGGCAAAATA
ATACATGCTCTAATGAAGGTTTATTAATGATTAATCTACTCCTAGCatttcctagtccac
tctatctccttaaaaaaaaattctggttgcagcccactaacttgattgtacagctgctta
atggatagcaggctgtaatttTCAGAGAACTGTTTAATGCGGGCTACCTCTGTTCTTCCA
TGCTGCTTGTGGTTCCTGCTCTGCTCAGGACAGAATGGGGAGGAAAACAGGCTCTGCGGC
ACAATATTGGCAAGTGAAATTTTGTAAACCGGCCCTCCCTTCCTTTTGCATTTGGTCTGA
AAATTCAATTAGATGCTGAGTCCTACAATGTATTTGAGAAGCCCAGGAGTGCCCTAGAGG
ATGAGACTGGGTGGCTCCCTGTCAGGTTGAACATTTGCCTTAATTACTTTGGCAAGATTT
GCATCAGTGGTATTAGTCCCTGCCTCACTTGGAGGCCTGCACTTAAGTGGCCACATTCAG
GCTCCAATTTCCTGGTGATTTCATAGTGTAGGGCACTTGCAATCAAAACTAGGCTTAAAG
CCCAACCCTCTTACATTTTACCCACCCCCACAAATGCAGCAAATAAAATGACTCTGATTT
TCATTCCCTAGACCTCTTTTCTATATTTATTACATTATTGTTAAGACAGTTTTTGAAGAA
AGCTGTTTTATTTAACAAAATAGCTTTATGGAATCAACTTCATATATCTTCTCCGCCAGA
TCAAAACAAGCTCGTAGTATTAGATGTCACCGAGCACCATGACAGGCAGATGAACATCAT
CCCTGTGCCCGGCTAATGATAGCTCGGCCTGCCCCGGCGTCAGCCGCTCCTGGCAGGGCC

AGCGGGCGGTGTGGGACCGGCACCGTATCTCCAGCAATTCGCAGATAACAAATATGGTTC
TGATGATGTTACTAAAGATCTGTCCCTTTCAAGATTGGATTAGACATTAGGAATTTGGAG
GGCTTTTTATTGCTAGCATTTTTAAGAATAACCAATTAGAGTATTGATTCTAAAGTCTGA
AAGCCACATGGACAGAGTTCATGTAATTGGCTACTTTATGTGCCTCTTCCTAGATTGCCC
TGCATTTTCAAAACAAGAGCCTTTCTATTTTAATCAAAAGAATCCAGAATGAAATGAGGC
TTTGAAAACTCAGCCTATGTTTGTCTTGATTTCCTTAACTGACATCTAGAAGAAAATATG
AGCTCAGGGGTCCGCTGGGTTCCTTCCAGCGCCTAAGCCTGTAAGCTCTTCCTGCTGGAA
CCAAGCTTTAAATGCACTTGTCAGTCATGTCCCATGAGAATAGATACTGCCTTCCATGTT
TTTTTGTTCTGATTTCCGTGTTTGAAATGATGAAAATCATTTTTCTGTGCTTTTTAAAAA
TGGAATTGCTTTTGTGTTGGGAATTGTGCTGTTCATTTTTACTCTACCTCGTTTTGGAAT
CACTAATGTGGCCAATTTATAGCCAAAAATCAGTATCGTAGAGTGAGCAATGAATGGCAT
GGTGACTGTGTGAGCGAATTCATGCCCTCCCTCCCCACCGCTCGCCCCGCGTCTCAGTCC
TCAGTGATGGTAAACAGAATGAGGACCTTCTCCCGACCGTGATGCGCCTCAGCCCTACTT
CCCTTGTCCTTTCCTATCATAAAATCTTCTTTCATAGAAATGGTCATTTCTGTTCATATC
TGTGGACTGTAAATAACAAGGAAGTCATTTTTGAGGTGAAAACTGCACTTAGACTCATTC
CAATTTTGATGGAAACTTTTAGCTGGTGGATGGCATTTTGTTTTGTCTTAGTTTTGCAAG
GAGTTATCTTAATTTAGGGAGATGAAACTAGTCTGTGATCCGAGGTCTCACTTCCATACA
TTTCTCTCGGGCAGTGTGGCTGCCTGAATCATGCCTGGATGCCACAGGTGCTTAGCCAGC
TGGTCCTGTCGTAACTGTCACTGGTAGCTCAGGGAGTGCAGAGGTGCCAGCAGACACTAT
GAAATTGGCCTCGTAAAGCATCAGTTATGTTGTGATGGTGGCAAAGCTGCAGGCGAGATG
GGAAGTGCAGCCACTGAGAACTCACAGTAGAGCGTGTGTAACGTAAAAAGATGAAACCCA
TTGTACACAGCTGTGTACTGCCTCCTTGAAGTCAAATTTCCCCCATTACCAAGGAAAAGT
TTTTTCTGAAGGGGGCTGCTTGACAGGATGACATCTGGTGATATCATTTATTCCTTTGGA
AATCAATCTGTGGAAGTGAGTTTCCACTGACTGATGAGGAGAAAAATGAATTGGCTTCAC
CCAGCATCCAGCTTCTTATCCTGGGAGAGATAGCTCTTGGTCTGTCATCCACGCAGCTGC
CTGGTGCAAGAGCCAAGTTTGTGCAGCCTGCAGAGCACTCTTCCTGAGCTGTGGGCTGCC
AGGTCGGGGGGCAGGGGGGGCCTCACTGTGCAGCCTCCTGCCACCCACTGATCATCTGGG
GAGACTGGCCTATCCTGTCAGGAGACGCAGTTGCCCAGACGTTTTCAAGGGCCTAAGATG
TAGGCAGTTGATCCACAGATTTTTGGAGAGTCCTTGAGTTGGAGATTACAGGTGACCTCA
GAGGAGGGAGTGAGAACATCTGGGTCATGGGTTTCTACTAGGAGTCCACAGTGAAAACAA
GAAGAGGAATTTACGACAAGACAGTCCAGCAACTTCCTTTCTAACTTCTCCTTTCACATA
TGCTGGATACTCCAAGACTTTGCATTTACATGGACATCACAGATCCACTTTGAGAGAAGT
AGGGTAAAAAGAAATAAATACATAGTGCTTTAGGTGTATTTCTATACATCTTAATTGATA
TGGGATTACATTTTCACTTGTGTTTACTGTACAGACTCTAGACAGATCCTGCTCTTTTGC
AGGTAAAACAAATATTTCTTAAAACCTAGAAAGACCCAAAACAATTTAACAGAAACATTT
TGGACCATTTTGGACCTTGGCAGTTAGGCCCCAGTGCAGCAGCGGCAACCATAAACCTCT
CCATAGGTGCTGAACCCAGGTGATCCCTGGCACCGGCAGCCTTATGTCAGGGCTCTCTTA
TCGCTGGTTTTTATTTCTCCTAATAAAAGTGATTAAAAGATTCATCTTTTAAAGAAAGCA
AGGACACAGAGGTGGATTCTCCCTGACGCTAGCACAGCTCATGCCCAAGCCACTCCTGCA
GGGCTCTGGTCTAAGTGCAAAAGCTGGAAAAGCTGCAGGTCCCGCAAGACACAGAGCAAC
CCTGCAAGCCAGGTCACCTTCCCTCTTCTCTGCTGTCCGACTGGCCCTCCACCATGTGAC
ATTCAAAAGCTcaagttacttaacctctcaaaactcagcatcctttctgtacagtgggg
aagatactggactgttgtgaggattaagtgagGAGAGTGGCCCAATGAGGTTGACAGTTA
TTACTGTCATTGTCATTATTTGCCTTCTCACAGGCAGGCGTGCCACAGTCATTTTACTGA
AGCTGCTTCAGTGGGTCCTGAATTAGGCCCTGTCCTTTGGGAGAGACAGTCCTGGTTCAA
CACACAGCTCCCTGCCCAGGGCAGCTTGGGAGTGTGGGCCAGTTTCGCCTTTAGAACCAC
AATTCTCTGATATGTGCAATGAGAGAATTAATTATAGACTCAAAGGATTGcatgcagaca
cacacagatacaaacacatacacacaacacacagagttacacacagacatgctcacaata
cacagaaatacacacagacacacgcacacagcacacagagatacacacagacacacacac
acacacacacagacatacgcacagatgggcacacacagagacacactcacagagacacac
agatacacacaggcacacacacagagagacatacacacagcccacagggatacacacaga
cacacaGAGACATACCTACAacacacagagatacacacagtcacacacagagagacatac
atacaatacacagagatacacacagagacacagatacagacacagacagacatacacaca
gacacgggcacacacagagacacacagacacacacaggcacacacGTGCAGATAAGGTAA
TATTAGCTAGTTCAGGAGGAGAAAGAGATAAAGATAAAGTAATATTAGCTAGTTCAGGAG
GAGTGAAAGAAGCCTTGTTTTTCTCCACTTTTTATAGAAGAGAAAGTGAAGATTCGATTT
GAGGTGAGTTCAGCACAAAAGCGTATCCCAGGCCCTCTGGCTCCAACTGCAGCCCTTTCT
ACCTCATTCCCAGACCCCACCTAAGCCTTTTCTCTTCAAAATCTTCTCAGGCACACTGAT
ACACATACCTCAGATTTTTAATTCTCCGGTTGTGTTCACCAGGTGCTTGGTCATGATTAA
GAATTCCGTGATGTGTACCCCATGTGTTTAAATTTGCTGCTGAGTTAACTTTGTGGCGGC
CTGTGGACTAGACCTCTGCACATGCAATGCAGAACGGCAGGGCCAGATTTGAAATCCTGC
TATCTTTTCGGCTGCCTTGTAAAAATAACATCAGGCGATGGGGATACGATGCCAGAGGTC
ACCTGTGATAAGTTCTGTTTATGGCCATTTTACTTCTAGGAAGACAGGAAGTGTCAGGAT
CTCAGGGATCTAGGAAGCCAAAATGTTTTTCCACTCTGAAATAAAGTGACTGACCAGGAG
TTCCCGGCCACGCAGCCCTGTGGGAACTGCCGCACGGCCACTTTTATGAAGTGGACACGT
GTTGGTCCCACTGAAAAGAAACTCCCCACCCATGGCTCCCTCACGCTGCAGCAGAGGCCC
TGCCACAGCACCTGTCAGCCCCTGCCAGCTTGCAGGGGCGCAGGCGCAGAGCGGTTTGTG
CCCTTGCTGGAGCCAGGGAAGGGCACAGGGTCCCTCCTGGAGTCATGGGAGGTGCAGCCG
AGGTTCTATATTAAAATACAGAGGCTAGCACATGTGCTTGGGGAATGCAGCTACAGTAGT
GGAATGAAAGTGCTGTCCGTTCCTTACCCCCCCAGCTCCTCACCTGTCCTCCACACGCAT
ATCCCTGGCTCCCTTTCCCTAGTAAGGAGACTGAATTGAAATTGTGGCTTGCCCGAGGCT
GCATACCTGTGCTCTTTCTGAAGCCCAAGTCACTGGCTCTAGAATTCTAACCTGTGAGGA
AGCcactgaggatgtttgtcaaaatacatatttctgtgccttgccccagttccacggccc
aggaatctgcagttttcacaagcacccccaggtgattctggtgGTGTCTTTGCACTTCTT
CAAGGCAGTACTGCCTGGAACGCAGAATCCCAGCCTCCTCTATCCTCCTTGCCTAATGGC
CTGGATGCTCTCAGATCTACAGGGGAAGGGAAGGTCACACAGTCATCGCAATAGTAACCT
CAGCTGATAAATCCTCCCCCATAAAACTTATTCCCCAGTGTTTTTTAATAGGAAACAATa

aaactgtaaccagcccaaatatccatcaaagagaaaatggagaagtaaatcatcgcacat
tcaCCTGGACCAGATCTATTGTAAAGCCAATAATACTGAAgccccttcaaggccctggg
agtcctaacagtgcactggcagtgtctataatttatattatgaaatttgcataaggaaaa
catttttGTCTCATTTGTGCAATTTCTCCTTCTAAATATACGTGTCACTTTGTACCTGATT
TCTATAAGACCCAGGACCTACAAACCCTGTGTCTGCCCCTGCAGCCACCCAGGGAAGGAC
TGCACAGCAGCAAGACAGATTGCCATGGAGCATGTTGTGCCCAACTAGGGACAGCGCAGA
TAGATTCTGTAATTTGCCTAACAATGTCTATAGGATGATCCCATTTGTCAAAAAAAAAAA
AGAACTGGGCTTTATTGATGTCACCTAAATGCACCTAAACTTCTTTTTTGCCCCATGCTC
TTCTGTACTCTTGATCTTTCCCCAAATTTTTAAAAACATGACACTCATTCCCTTATTTTT
CCTACTTAGAAAAGTGTAGATGGTTTTATCATAGGAAGTTCAAAAAAATTAAAATATAAT
GAAAAATACTCAAATAGTGCCTCACAACAGTAACTACTGCTAACATAAATAAAATCCATA
TTTCCTCTCATACAGACCCCAGAGTTGCTTTGCCTGACAGTGTAGTTGATGGAGAAAATA
ATCTTTATCCTTAGCCTCCATCTGGTTGCAGACCATAAAGACAGGGAAAAAATGAGGGTG
TTGGTAGCTTCGTTAGAAACTGAAAGCTCACTGATTTTTTCAAAACCTAAATAGCCTGTG
TTTCTCCAAATAACTAATTTGCAGCCTTCGGCAGCCAGGACTGGCAGGGATGGGGCTAGG
GGGACTGGGGAGAACTGCTCTCTCCTGAGGGTGGTCTGACCCGACAGCACGCATGACCTt
cccacagtcaggaactgctcagagacgtgatggcaactccatagaatgaaatactcttca
gccagtaaaatgtatttttggataaatatttgctttaaaaaactttactatatgttgtta
aatgaaaaaaaaaccttaaggcatcagaaattatgtgcagtaaaatctcacttttgtaaa
taaatatacctgtttactacgtatgcataaaaagaatcctgagaaatataAGTACTGTAT
GCATATTGTTGTTAAGTATTTTTTCTGTTTGCTTATCTATAATTCTAATTTTGCTTCAAA
GAACAAGTTACTCCGGCAATATAAAAATAAAATAACTAATTTGTCTTGTCATCAAACAGA
TAGTAAGAACAGGCAAACCTGGCCCTCCACACTGCCAGCCTTTTGTGATTCAAGGCTTCA
GTTTCCTCCACTTGTTAAAAAGATTCAACAAAGTAGTTGAAATAGTATGTGAACCAGTAA
ACCCTAAAAGGTGTCCAGTGTTGTCTGTGAGCTAATTAAGTGATTTGATTCTGACTCCCC
GAGTCTTCTGATTTCGAAGCAGTGGGGAGTCAGACAGGAGCCTCAGGTGGCCTCTCCTGA
GAGGCCCTGGAAAGTGATGAGAACCTGGCCTCTGGCAGCTCTTCATAAACGTCCATGTTT
TCCCTCTACTCTCTCACTCTTTTCCCAGGGCCTCAAACAGAAGATGAAAATCAATTTCTA
AAACAGCCCTCTGTGTGCTCTCTCGTATCTCTCCTTTTCACACATCGTGGTGGTGGCTTT
CTCTGTGTTCCTCTGTTGATTCAGTCTCTGGAATTAACGGATCAGGATTCCATGCCCAGA
ATGCTACAAAGACTGTGCTTGAGTTCTCCCACATCTCACTCAATTACACAGAAGTTTCAG
ATTATGTAACAGATGCTGTGCTGGGTTAGGCAGAGCCATCTGACTTGTTTTGCTTTATTT
TAGACCATGAGATGGGTGAGTTTTTCTTTTTAATGCCACATTCTTTTAAGAATTAAAAAC
CTCCACTTGGCTGTCAGCATTGGAAATCAGAGTGATGGTGCAAGCCCTGATGAGGACAAT
GTCCTTGTCTATGAAAAGGTGAAATCATTGCTTGAAATCGCTAAGCAGGACATGCAGTCC
CAGATGGAGGGGGGAATTCGGGAGCTGGTTGGAAAAGAGTATTTGGCACTTTGCAGCCTT
GAGAGGTGCAGAAGAGACACCGAGGGGTTCACCACCAGAGCCACCATTGTCAGAGAGGCG
TCCAGCTGTGTCCACCTGGGACTCTGCCTTCAGGGCTTCTTGCCTGGCTGGGAGCTGCAC
AGGCAGACTCCTGGGACGGTGTGCCGACAGCTCTGGGCACCCCCTTCTAGGATCTGATTC
CTGAGGAATCACAATGTGGATTTCACAATCACTTCCAGTGTCTTTTGCCAACCTCTGTGA
ACAGATGTGCAATTAAAAAAAAAAAAAGAAAGGGGCCCAATTCTCAACACTGTAAGTGGA
AACTTTTTAAATGGAAAAGGATAGGCTAATGAATTGAATTTGAAATCTGAGACAGAACCGA
TGCATCAAATGTGCTGGTGTTTACAGATAATACAAGGGGGGCTGCATCTTATGGTTTCAA
TCCTTTTTTAAATTTTTGTTCTGAGAGACCCAGCCAGCAGACTGCCGCCAGTCTTGTCAG
AGATGTCAGTGGTGGCCACTCTGAATGGAAAGCAGCATCTCTCAGCATCTCTGAGGCACT
GCTCCTCAGCGGAGACTGTGGTGGCTTTGCCTTTCAGCACGCATCCTTTCTACGATGCCT
GACAGTGCCCAGGGAATGGGCAGAGCTGGGAGCTCTGAAGCCCTTTCACCTAAACCACCC
TGGGTCACCTGACCTAGTTTTCCTCCCAATTTTAATTATGTCAGGCACTTCACAAAGGCC
TCCTTGGGGACACCATGAGCTCACTGTCATCAGATTGCTCCAATCACAGCTGTGGCTTGC
ACACAACCGCCATCTCTGCCCCAGCAGATGCTGTGTGTAAACAGTTGTATTAATTACATC
TCAAAAACATGGTTCTTGCCAGATCCTCAGGATTTGGGTGCAGCCTCTGAGGTGGGTGGG
AGGCCCTCGAGGGAGAAATGTCTGCAGGAAATTCTTCCCCTACGAGAGGTCTGTTTTCTA
AGTTATCTAAGAGCTACTGCAGCTGTTTACTGCAGAGTGACCCTGCTCAAAGCTGTGGTC
ACCCAAGGCTTTGAAAGGGGACCTCCACTTCCGCCCTGGGTGGAGCACCGTGCTGGAGAC
CCACGCCTGCCAAGGCCTCATTGTCATCTCCACACGCCGTCCTTGGGGTGGGCCACTCCT
GGGACACGCAGACAGGAAGCCGGCCACCTGAGCCACTCGGAGGCTCTATCCAGAGTCAGC
TGCCAAGCCTCACGTCACACATCACTGTTAGTCTTGGAGGGCTGGCGGGGCCCTGAAGTC
AAttgaacacttggatgacagggaacttgccactgccagaggcaatatgctccatttttt
tgacagttccaacaatttttcttttaaactgTCATAAAAAATTGCTGCTGTGAATACCAGT
GTCGGCGTCCCTGCCTCACCTTTACCTGGTGCTTTTCCACCACACAAAACTGTTTCTCCT
CGTGCTGGCCTTGGGCTTGCAGACAGCTGATTCTTCTCCTCCCGCGGCTGAGCAGCCTCC
TCCGAGCAACCCTCTGACAACTCTGCTCCTTCTGACAACCTCTGCAAGGGCTGCCAGATG
TGAACAAGGGGCCCGGGCAGAAGGTATCCAGGAAGACTGGAAACTCGAGGAAGCCTGCCC
TGTCCTGTCCACCAGACTTTACGCTTGCGTCACTGGGCTTTGGGACCTAAGTCCTCGTCA
TTTGTTCCTTTTGCAGTTCCTACTGTTCTCAGCACTTCCTTCCAGCTTACTGAGGTACAC
TCAGATGTGATATGCCATCGGTACAGACACAGTTCTGCTCCAGCATTTCCCCGTGTTCTT
TCTGTCGCTCTATTTACTGAATTACCGTGAGGATGTGGAGCGAGGCTGAGTTCTGTATTT
TAACACCATTTAATTCTCACCTACTGAGAAATCCATCCTCTTATCACTGTGCTTTTTTT
AACCTGTCACGAATCCATGAAATCCTATCAGCCAGCCTGCATACTTCCTTTTAAGGTGCA
GTTGAATCAGGAGAAACTTGCCGCACATGCTGCGTCCGGGCACAGCATTGGCTGAGGCTG
CTGCCCTGACCTGTCCGCTTTGTAGTACTGCCCAGCTATGAAACAGGTTAGCCACACATG
ACCTGCATTTAGGAGTAACAAGTCTGTCTGTACATGCACATACAGCAACTTTTTTAAACT
GTCTATATTTTTTCCTGAGATAGGTATTTATAATATCTCCATCTTCTTTCCCATTTTGAA
ACTTAGAACAAGTTTGCCTGTCAACAGTTCTCCACAGCATACTGTGTATTCTAGGATTTT
CTAAGGTTGAGCAACGGAGGTTCAGCAATTTTGACTTAATTCTTCCCATCCCTTTTCCA
CGCAGCCCAGAAGCCTTGGATCACGTGGTGAGGGGAAGAGGTTGTGCTATGTCGGGAAAC
TCTGTATCGAAGCTCGGCTCAGATCATGACATTCTCTTGACTAAAACCCTCAGTTTCCAT

CAAACTTGTCACTCTGGCATTAAAGCCTGTCACTGTGTGGCTCTGAAAACCTCTCTGAAC
GTGTTCCCTGCCTCTGCCCTGCAGGTCCCTGTGCTCCACAGAAGCCCACTTATGTGACCC
ACCCCCACTCATCACCACCTTCCCTCACCCAGAGCCTCAGCTCCCCACTCCCACCTGTAA
GACCCCTACTGGAAAGATTCCCACCTGCCCCTCAAGATTAATCTCCAAGGACATTTCCAA
ATTCCTCTCCCCATCTCTCAGCCAGATGGCTTTGCTCCCTCCAGGAACCCCAGCCACCTT
CGACCTCCAGCAGGGCACTCCACTCCACATTCTCCTGGTCTGTCTGGCTCATCTTACCTG
AGCCATGCTCTCCAGGTGAAGGACTATGTCTAACTCAACTCTGCTTTAAAAGCAGCTAAC
ACATTGCTCTTTGCATATTGTTCACTCACTAAGTTGAACTGGACTTGGACATGCACACTG
AACTGCAGCGTCTGCTGCTTCTTGGTGGCCCAGCTCGTCAAAAGAATAAGATTTCAGCAA
AACAATGTAACAATTTTTTTTACCAAAAGTAATGTTAACAATATATGGTTTTCCCCTGAT
GTTTGCGTCAAAATGCTTTTTGGAAAAAACATTTTTCAACTCTTTAGGGTCAGAATTAAG
CAATGAAATTTATATACCACATGTATAATGTGTATGTTTATCTAAGTATCTGTTCattta
tatatcttaaatagaaattttaaaaattttttaaaaCTCCTGATAAACATTCTCAGGAG
GCACACTATGTAACTGTTGGTTGATATACCTAGCTAGATGGTGAAATCAGATTTTGTTTA
AAGCATGGAGGAGAGGGAAAAATTAAATCTTGCAGATTCTGCAGTCCTTAACATCTTTGA
AAGAGGAACATTTCAGACAATGTAATAAGAAGGCCACGTGCTTTGACTTCTGTAGATTTT
AAAAATACTTCTGTATAGTTTCTTCTTCCTTTGAAGAAGTTTGGGGAGTTTGGGAAGATG
GAGAAAGATATAAGAATAGACTCCCCATATGGGTCATGAATTATCTTTTTGCATCAGAAC
TCTTAGTGCAGTTTCAGTATTTTCTTCCTCAGGAGGGTGAGCTGCTTCCGAATGTCCTCC
CCTTCTTTGAGGCATCCTCTGTTGGTGAACTTTGAGAGCATCCATTTATGAAGTTGATGA
CCTTTCCCAGTCTCTGCAAGCCCTTCAGTGTGTGTCCTCTCTGAGCAAATCTGAATTGTG
TGCTTAATACATGGAAAGGGATTTGGGAGGGTTGCTTTTTAAACTGATTTCTTAATTAAT
ATTATGGTTTAGTTAACTAGACAGTCTCATTGCAGAAGTGCATAACCATAATATGTCTTC
AAATATATCTCCCTTCCTAACACCCTGTAATATACTTTTGTAAAGATACCCTTACAGAAT
GTGATCCACCATTTATGAACCTGCAGCATTGCATTCAGAGACTAAGTGAAAAGCTGGCAG
ATTTTCATTTAAAGCACAAGCTAAGGAAGAAAGCTGGTCTAGAAGGAGCTACAGAAGGGT
AATGCTTAGGGAGGGAATGATGTGCCTGTGGGTGGTGGTAGTTAAATCTAACCAAAGAAT
GATGTCGTGGGTGTTTGGATATTGGATGGTCCACATTGGGCCACATTCTTTCAAACATAA
GAGTCTGTAGAAATATGACCTGTAAAAGACTCTTAAATATTCTGGAAACTGTTTCTTCCT
TGTCACATCCTTATATATACTTGAACCTATGCCTACCAGACATGACATGTGACTATTCAT
ACAGATTTCATCATCTCTGGTTTAAGAATAAAGGATGCTGCATAGAAGGCTCACATCTTT
TAATTCACAAGACTGAAACTGTTCTGAAATGACATTGTTTCTAAAAATTCATTACTTGCA
TTATATTCATTTTTATTTTTCCATGCCAGAAGGGTAGAAGTTCCTGTGCTCATATTAAGA
AACAGCAATGTCAATCGAGGCCCAACTCAAATCCAATTTATAGGAGTTATAAAGGGCGTG
TGCCTGTTTTGTCTAGAAGCAGTGTTGGGCAGCACTGAGTAGGATAGACCACCTGTTGCT
ACCGATAAAGGAGCAGCTTCTCGAATGCTCCTGTCTGGTAGGCACTATCCCGAGTGCTTT
GGCCCCTCATCCACAATCTGTGTGGCAAAAGGCATTGCAGGCAATTCAGTGAGGAGACCG
AGGCATGGAGAGCAAGTGCCATGGAATTCCCTAAGGCCGTGCAGGGAGCAGGTTGCCAAG
CTGGGTTGAAACCGTCCTCCGTAGGCTCCCAACTCCGCCGTCGCTGCTACTGTGCTGGAT
GATGCCTGGTAGATGCAGATGTGGAGCCCCAtggattctgagacaggccgggtttcagtc
ctgccctagctgcctattggctggatgaccttggcaagttgactttcgtgagcctcattt
gtctcatctctcaattaagaaaacctagagcctatctgtgggggttatctgaaggattcc
agggatgcatatggcactgtctaccgcatgcggtaactgtttcacaaatgatGAGGAGCG
ATTTATGTTCTTAGTGGAAATATGTCGGCGTGTGAAGTCCCAAAGCTCTGCCCTGCCTGG
CTTGATCCAGTGCCTAGGCACTGCCCCTCTTCCCCTCTCTCCCAACCCACTGTAAGAGGC
TAGGCTGCCTCAGTAACTCTGAGGGGCATTGACTCTTTTCATCCAAAAATTCATGTTACT
GCCCCACATTTTTTCTGTTGTTTTACAACGCAGTAGGAAGTGGGCAGACTGTCAGGAAAA
GTGATTTATAGTCATGTATTGCTTGTGCTTTGGCTTCATTTGATCCAATGCAGATCAGCT
GCACTCAGAAAACTACTCAAGTGAAAGAGAAAAAGTAACTGAAGGGGGAAATCTGGATGA
GTAAGAATTCCAGGGATAGGAATATTAATAGCAAGCTTTTTGCCTGATATAGTCACTTTA
TGCTGCAGGGGTGCCCCTTTATAAAGTGCTTGTACAATGGATGTTTGCTTTTGATTTTGG
ATTTGGAGTCTAATGAATGTTCTAAATTATTATTAGAGGAGCTTGCGGTTGTTACATGTC
TGCCTTTATTGCTTATTTTTAGCCATCTCCCCTGATGTCAAATGCTCAGGCAAGAATGAT
ACATTCATTTATAATGTGGCTCCTTCAGAAATATACCACATACCTTTTGGTGTGGTTTGT
GGCTGAGAAGAGTGGGGAATGCACAAGTGGAAAACTGCAGAAAGATTATGCCTTCATCAC
TTCAAGTATTTGAGATGAAACTAGATCATTTGCTGTTGCTTTTTATTCTCATTCTAAGTG
CTTTTCAAAGTCAGCGCTAAGATTTTAAAATGGTTTTCTGTTGTTGGCAGAGAGGGAATT
ACTCTATTACTTTCTGATAAAACAGAGTCTTTCATGATCAAAGAGAACCAGGCTCTAGTA
GTTCCAGTATCCTAACGTGGACACTAATTGTTTCCCTCCTTTCTTCATGAAAACAGCTT
CTGCACAAATGATAGCCTTGTGAACTAGCCATGGGCACAACTGGAGAAGCATTTAGGGAG
CTTTAGTGCAAATTGAGACCACCTACACATCTGACTCTACAGGGTTTGACAACATCCAGG
GTGAATCACAAAACATCAGTCTAATCAGGGCTTATATAGAAAGAGTGAAAGAACTCTGAT
TTCATCCTAAAGATTATTTATATTAACCATTGTTCCAAATGCATTAACTATTTTAATTTA
GTTGTTTTGATTGTTAAAAAAAACACATCTGTTTGGTAGATAAGACATAATTTAAGACAA
ATGTTCTATTTGATAAGCTTTTAGAAACAACTTATTTTTATTCTTTCCTGTGAGATAACT
CAGATGTGGAGAATGTGACAAAATTTTAAGCATAACATGAGAAGGGCTGACACACATAGA
TTTCTGTGTGCTTACTTGAAAACAACAAAATTTAAGAATTTGGTATAGGAGTTGTATCAG
GTAGTGCAGAGTCCCCAGGAGACCTAGAGACCCAGGTCTGGGAGCCTAGCGGCAAGGGCT
GAATGTGGGATGACATCAGCAGAAACTCACAGCCACTGCTATTCCAAAAACCCAGCAGCA
GCTCAGTGCAGGGCAGTGCTGATAGTACAGTGCCTGCAATCCTGGAGTGGATTTGGATGT
GTCAGGTACGCACACGCTCACTGCTCCCCCAGCAGTACGTTGAACAGTGTGCGTCCAGGT
GTCTGTAGGGCCCCTCGCCCTAACTCACAAAACCATTCTGGGTCAGAAGCCACCAATATT
GTCATCATCCTCCCTTTTCTGAGAACCCTAGTAAGTCCCTCCAGTGGGGCAAGCCCACCT
TTTCCCTTCATTCTGTGGCAATATGCCTTCATTTCCTAATCAGTTTTGCCCTGCTCATTC
AATGCAAAATGGATCTGCTTTCCTTGGGCACCAATATGTCCAGGGATTGTTTATCAATCT
TCAGTTCTGTTTCCTTTACATATCCCTCCAAAAATCAGGCCTGCACTGCCTGTGCACTCC
ACAATCCACAGGCCTGAAGGAAATGTTATCTTTGATGTAGAGACTTAAAGTAAAACTCTT

CAAATTAATTATTTCATGCAAAAGGCTAGTCCTGACTCTAATTCTAAGACATGTCTCCTA
AACTCTGGAAGTCTGATGTATCCTATTATCAACATTTATCCTTAATGTGATGGTTTATCA
TTTATCCTCAAAGCTGCATTGTAAAATGTACACTGTAAAGTGTACATTTTAAAGTCGGTT
TTAAAAAATCATATTTAGAGATCCTGGTAAAAATCTATCAAGTCAAGACATTACCTTATT
ACCCATGGAATTGTCTTCAACTCTTACAGTTCAAATATTCCTGAATTGGCTTTCACAATA
AACATCCTAAATATGTAAGTAGAAACATATATATTGCCAACTTTGTGCCTTCCCAAGCAA
AATTAAAATACAGGAAAAGTCAGTTTGTTTTGCCCATAAATAAATATATGTGTGTGTGTA
TGTGTGTGTATACACATACACACTCAGAAAAGATAGAAGCAGCAGCATATTTTGGCAGCA
TCTGGTTTATTGGAACTCAAACGTTCTGATTGTGCATACAGACTAGTTAATGTGGTAACA
ATTATGTATTTCTTCCCTGCTCCTTGCCTTCTTTCCCTCCCCAGTTTTTTTCTTCCTGAT
AGTAGGTGTGTACTTTTTTCCTATTTCCATTGGCAAGCCACATGACAAGCAAAACGATCA
CTCGAAGAATATTGTTCCCTCAATCAAGAAAAATGCCCATTGGGTTTTGTTATTTGATGT
TATTTGATGACAGAGACCTATTGTTTTTCCATTTTTCTTTTTTTGTTTTCCGTGGCACCT
ATGGAATTAAGCAATATAAAAAATCTATTATTTCAGATGTTCACGTCTAATGAATTTCAT
GTGAAATACTGGCAGTATAACCCCAAATAGAGGAAATTTGTGAAGAGTGGATGCTGCAGG
GCATGAGACATCTGCACAGAGTTCATCTCTTCCAGCATCTTGCATGTCCCAAGCACTGCC
CTGCCAGGCAGAGAATGCTGCAGATCACGGCAGTGAATTCCAGTTGTTCAGAGCACATTT
GACTTCCAAATTCTCAAGGCCACAGATTTGAGGACAGAACAATATTTGCATTTGAAATTG
GAAGATTATTTTTTGCACAAGTGCCTATATGCTATATAGAGTTTGCCCACTCTGCATTAT
CTTCCCCCTGTTCCCCCGTTATCTGGCACAAGCTATTCAAAAGACACGCCTACTTGTAAA
ATAAATGGTTTGCAAACTAAGGAAAATACTTAAATCTCATGTAAATGGTACTATACTATG
TATAAAAATGTGAAGAAACACAGAACAGCTCATGAACACCTCCACTGCTGTATAAAAGAA
CCATCTTTTTTCTGGCTCCTATTGGATGCCTTAGAAAAATCTGTATTTCCTCTTTAGTTA
TTGTGTTTGAAAGATGAAGTTGAGACAAAAGTTCTATTCTTTTTAAGTTGGCAGAACTTC
TGAAAGGTGATTTTTAGCTGCAGTGTGACTCATTCCAAATGCAGAAATCTCTGACCCTGA
GTTAGTCTATTTGTCATGCAAGAGCCTAGAAAAGCCCTGAGTGATAAGAAATGGCCATAG
GCCATTCCCACAGAATTTTCAACAAAAATAGAATCATGCTTATGTTCTAGTCATGACTTA
GAACTTATAACTCATGTTCGGAACTGTCCATGTTCACGCACAGGGGCCGTATCACTCCGC
CAGAGCTGCCCTGGGTGCCGGTGTGCAGAGGGGTCCGAGAGTGACTGTCTCTTCCTCTGT
TGTCGAATGTGTGGGTTATCTCCATAAATGGCTGCCATGAGCATCCTGTTCACACATTT
TTAGGTACTTGAGTGAGTGTCTGTGGAATAATTTTGGGAAGTGAAATCTGTGGTCAGAGG
TTTGTGAGTTTTACATGCTACATTTTCAGAAGTTGAGAAATAGCAGTAGGCTGAAGGCAA
GTCGCCATGCCTGGAATTCATGAACACTAGTTGAAAGAACTGGCGTGAGTTAGTCATGAC
AGGAGAGATGGGGAAGGGAGTTGCAGGTAGGAGGGCCATCTTCAAATTCTCAAAGTATAG
TCACTCCAAACCAAAATTCGATTTAATCTGTAGGACTCCATTCTCAAAGCACAGTCACTC
CAAACCgaaattcgatttaatctgtaggactccaggtggcagaataagaggcaatggatg
ggtggaagcgaaacagggccaaagtttgacttcatgtgcaacttcctaaggagtgatttg
aactccacaaacatgaactaagcACCTCAACACAGGCTGGGCaagttgctgttcttttgg
agcttacatcttagtggggaaagagaaatgcctatgtaaacatataaatCAGCAGGATAC
ATTGTGAGGACGGTCATTGCTCAGTGAGACTGCAATAGAGTGATACGCTGGAGGGGGCTG
CAAGGGAGAAGGTGGGAGGGACAGCATTTAGCAGAATGAGCAGCACAGTCCCATAGGAAG
AAGAATTTATTGCCTCCTTAGGCAAATAAATTCCCAAACCTTGAACATCAGAAAGGAAAT
AGATTAATGTGCACAGAGGATTAAATTATGTGATCTGCAAAGTCATTTAAAATCTATTTC
CACATAAAACATATTAATGCAACCTAAACAAAAGGGGTCTGGATACCCTCATCTTCTTCC
CAAGCATCAAGTCTTTCTATAGTTAAACTGAGATGCTTTTATTCTTGGAAAATTTTAAGG
ACTATCTACAGCAATGGAAGAATCGGGTGTTGGGATGTGTTCCCAGGTAATAATGACTGC
AGGCTGATTTGGCCCTTGAGGTGTGGCCTCATGGCCCTCTCCAAAAAAAATCAAGGACCT
GCTACAAAGCACAAAGCCGACTGCAATGCTTGCTGCTTACTGGTTAGGGCAGCTCCTCTT
TGCCAGCGACCAAGCAGAAAGCAAGACAAGACAGGTTCTGAAGCAGTAATTCAAAGCCTT
CCTCGCTTTCCCATGTGAGTCATTGCTAGTCAGAATATTACCTTTGCAGAGAGGCTTAAT
TCCAAATTTGCTCTTAAAGGGATATCCTCTCCTGGTTTAGGTATAAACTTTTGACTCACA
GGACAAATTCTATCATTCCTTTGGGCCTAGGATTGCATTTATTTCCATGACAAAAGGGCC
TGTCTGGTGTTTCAGCAAATGAAAACAAAAATATAAAGCCCATCTCCTTTTGAATGAGCT
CTAAAACAGTTCTCCACTGGACTTCAGAACAAGAGGGGAGCTCTGGGCTGCTGGCTGGTTG
TGCATTTGCTGTGGGTTCCCTCCGGCAGGCGACCTCTCCGCGCTGAGAAGGTTATCCGGA
TAACCAAGTAAGAAAGTACATGAGGAGGCACAGAAAGAAAAATGTGAGAGATAACAGCAT
AAACACACAGTGTATGTTGTTATGAGGCATCACATGATGAGATACTGCTGGGGAGGGAAG
AAGTGAGGAGATTCCTAGGAATCTTATGAGAATTTCCAGAGACAACAAGTTTTGAGCTTT
TTTTTAATTTAGAAAATTTACCTTATTTTTAAAAGAATATGTAACATATCCCATGCTATA
AAATTCTAGACATAGTAGATTTAAAACAGCATAATGGAAAATATAAATATCTATTTTCTT
TTCCTATTTATGTATTCTGTGCCAGTAGGAATGTAGCCAAAAAGAGAGAAAAGGGGTCTC
TGCAGACATGGATGTCTCTGTGACTTGATCACTGCTAACCCAAGAAGATAATAAAGCAGA
AGCATGTATCCAGGTTGCTGCAGCCAAGCCTGCCCGGTCTGCGGGGCGTCCTCACACATG
GGGCAGCTCTCCCACCCCACACACTGGGAAAGGCGGACAGAGGCTGGGCAAAGCCCCCAA
TTTTCGTTGGCACTGACCCCGATGATTTATAGGCCTTTGTTTCCCATGTTAAATGTCTTA
CGATCATTAAATTATTTATAGCTCAATTAGCATGTGTCCAAAACCAGGAAGTTCATAGGA
GACTGTGTGACTGGGAATTAAGGAGCAAAGCAACTTTCCAGTCTGTGATTTACTGGGTTT
CCATTCTGTTTCCTGTTCGGATCCGGAAGTAGAATTTCAAATATTGCTTTTCATGCTTTA
TTTGGGACCGATTTTAGCCCCGCTCTCCTTTCTCTTGCCATTCGCTGGCCATTAGCCACC
AGCCTCTGCACAATGACCAGCTGGCCCCTGGCAGATCTTGGGCCCAGGTGTGAAGTCGCT
GGAGAAGCATTTCAGGGCCAAGATGGGAGTGATTTCATTTTCCATTGACACTATGCAGAA
ATGAAGGGGATTCAAGTGCCTTCAGAAAAGCTTCCTTCCAGCGAATGGAGTTTTGGGGGT
TTTCCAGACTTGCAACTGCTTTTATTCTTGGAAGCATCATTGTTGCTTTTCCCCCCTTC
CATTTATATCCCAGGAACTGATTCAGAAACCATAGAAATTGGATTTGGAATCGCTGAATG
CTAGCAGACAGCTGACTGCACTCTTCCCAAGAAACCCTGCCAGCTGGGTTCGGGTATCGC
GCGGTGTGTGCTCTCTCTGCCTGGCCCGCTGAGTCCTCTAACTCTAATGGATTCCTTCTT
ACACCAAAGTGCACTAGAACTAAAGTGTTTTGCTTCATTCTTTAGACATTTTGTGGTTTA

GGGCTCAATCAGCCAGGGTATGATTTGCAATCCACAGTAACCGGTTTCAGAGCAGCTGCC
CAGCGAGGCAGGTTTCATCTCGCTTGCTAGACGTTTTGTTTTTTTTTTTTTCTAAACCTC
ACACCTTTTATTTATTAGACTTGGATTCCAGTTTCCTGAGCCTGTTTGTGCCACTGATTA
GACAGGCTTGAAGCAGAACCCACCAGGCTTCCTGAATAAAATGCAGCAGTGATTGTATTA
GGGGGTTTTAAATTGCTCAAAATACTGTCTAAAAAACACTAAAAATCATGTTACTTTCTA
GATTGAATAAAATCCTATAGAAATGAATTCCTGGACTTGATATGTAGCAAGCTGGCATTG
GCTCGGGAGTGAGTGGGCTCAGTTAAGTGAGCTAAGATGAGATGGTGCACAGGCGAGCAC
CCACCTGAGGAGTGTTTGGATGTTATGATAGCCAGCTCCTCTGTAAAGACCTGTCCTTCT
ATGTCAGCAGCCCAGCAGATAAATGACGTGTAAATACCACATTTAGGAGGGCTTATGATG
ATGCCAATTAATGGAGACCTTTTTGAAACAGGAAGGAGGTGAAACATATTCCTTTGCTTC
TACATCACTGTGTGCCAGGCACTGTTTACAGCATCTCGTTTAACCAGCAGTCACCACCTG
ACGGATGGCTGATGTGGGGTGGGGTCCCAGGGTGGGATTGCGTGATGGGCTTGGGGTCTC
TGGCTGATGGGTGCCAGAGCTGGGACTGGAACTCCTGGCGTGACTGAGGCAGACACCTGG
GCTACCCAGCCTCACCCACGACGCCCTCACTAAGTGACCCACAGGACTCACCGGAAGCAG
GGCAGCAAGGTCCCCCTACAGAGGTCCCCACTGCAAACCGATACCCAGCTTAGACAGCAG
TTCTGCAGTCGGCGTCTCACCCCTTCGGGTCTCATTGTGACTCACTTTGATAGCCACACG
ATTTAAGGGTGGTTCAGTAGTGATTTGATGAGTGCTGTGGCTCAGGGTCATTCCCCTGCC
CAAGCATTTCAAATTCCAGAAGTTCATGCCCTGCATGGTGGGTGAAAAGTCTCAGGCCAA
CCATGAGCACACAGCAGCCAGGCGACTGAGGCAGCTGCCCGGGGTGGCACGTTGCTCAAA
CCCATCATTTGGAGTCAAAACAAACAGATGATTAGCTGGGGTGGTCACTTTCAATCAAGA
GTTTTCACATCGCCTAGACATGGCCTCAGAATCAGGCCTGGTGTGGCCAGGGGCTGATCT
CACAGTAGACAGGAAGTGTGGCCCGAGGGCCATGGCTGCCCCCTCAGAAGGCCCTGTGGA
GTGGCTGGCCGAGCCTCAGCAGCCTCCTGTGAAGCGAGGAAGGGTCTTCCTGCCGGCCTC
TGGAGATCAGTATGGGAATGCACAAGTAGGAAACGCTGGATGGGAATCCCTCTGCCCTGT
GATACCAAGGCAGTGAGttgtagactatggaattgctgtcggagggctctgtaaccggc
caaggtcacacaggtagccattggtagagcagggactggaatcccagacccccaacttcc
aggactgtgacctttctttatcccatacagccTTACAGTCAAGTGCCAGTGCAACACCT
GATTCCCAGGTTCCAGCCTTTGTCTTTTATAATGGGAATCAACCTTATCTTGACGATCCA
GAGATAGTCATCAAGGAAGATTAAATTATCCCCTTAGACTCagagtgaccatatcatttt
ccctccacacaaggacacttttgagaatgaaaaggagGAGATGTCTGTACCAGACGCTGG
ATGACAGGCACCGACAGGCTGTCTGCCAGGGGAGCAGCGATTCCTGTATGTTGTAGAAAG
TTTTTCAAAAGTCACCTTGGAAAGAGGTTTTGTTCCTTAACCTTCTGTTAAATAGGAAGC
TCCGTGAATGAAAACAACTCCCTTCCCTAAACATTCTAGTAATGACCCAACACTGCCAAG
CCTGCCAGCTCTGCCTCATGGTCGTGTTGACTGTGTGAGACTATGTGAGTGCCTGCTACA
CAGTACGCTTTCAGTAAACATGGTATTGCCTCGATAATCCCACAAAAATGTCCTATTCAA
ATCACCTGGCACCCAGGAAATTTCCTTCTTTTTTTTCCCAGGTGAAATATACAGTTGAAA
ACACCTGACAGCAATTCCCCTCTCCCATGTGTTTGCAGGATGGTGGTTTTGGTTCCTCCA
TCTTTGATGTGTACAAGTGTGATGTTTTCCCCCCACAGACAAGTAAACCACATTCTCTTC
ACATTCCCAATGTTTTGTCAATGTACCTCCTTCAATAGAGGATCGATAAGGAAAAAAATC
ATTGACAATCTCAATTAGATTCACTATTTCATCCAAAAGCATAGCTTAGAACTCTAGTTT
TTGTTCAACACTCTTGCCCTATGAGTGCACAGAACTTTAATTCTGATACAAACATCCCTG
AATGTTTAGCTTTGACAGAGATTCCAAGGTGATTTGATAAGAAGCAGGGCTGTGTTTGGG
CTCTGGGAGTTTTTGATATGGTTTCAAGCCCCATCCAAAACCCACAGACCTCTAGAAAGT
AGGTGCCTGCCTTCCTGCAGCAGCCCTGGAGCCTGCTGGGGGCTTTGAGCAGCTGCTGCC
AAGCCAGGCCTCACCCGACACTCTGATGGGCACGGCCATGGTGGCAGGGGCTTGGACGCT
GCCAGGTGACTCTAACTTGTGGCCAGGGTGGGAAGCACTGCTCCACAGAGGTGCCAAAAC
CAGGTTCCTTCCTGTGTTCTCACATTTCACAGCCTCAATGTAAAAAGTAAGACATGGGCA
CTCTGGAATATTACAAAAATATAGAAAAGCATGTTATAGTAAATAAAAGGCTCACAGAAT
TTTGTCATTTAGGAACAATGATTATTAATATATTAGTGTGTTTTTGCTCATTAACAGT
ATATCCTGAGATATTTCCTATACCATTTAATATTTTAAAAGATGTTTACactggccacag
tagctcatacctataatcccaacactttagagggcaaggcaggaggatcacttgaggctt
aaaaattagccaggtgtagtggcacatgcctgtagtcccagctactcaggaagctgaggc
tggaggatcacttgagcccaggagttcaaggctgcagtgagctataattgcaccattgca
ctccagcctaggtgacacagtgagaccctgtttctaaaataaataataaataaaTTAAAA
CATTTAAAAATACATGATGTTTAATTATTAGAGGACTCAATTTTATATCTATGTATACAA
TAATTTTAAGTTTCTTAATATTGGACTTTTAGTACCttttaaaaatactatttttaaa
aaaatCTGTATTTCTAACTTTTTATAACAAGGAACCTTTGGCTTTGAGATGACTGGGGAA
TCCATTCTTTCCTATAGTATCCATGTCCAATGGACTTAAAGTATTAATCAATGTGTTTAT
GTTTTGTTATTTTTCTGGCATTACAAAAAATTCTAAATATATTGTTACCGCCTGTATAAA
TATCAGCTTTTGAGAGAAGGACATTGTGTAGAAATAATGAAACACTGCAACTTGTATTTG
TATTAttctttttttttttttttttttgagatggagtctcgccctgtcacccaggctgg
agtgcaatggtgcgatctctgctcactgcaagctccgcctcccaggttcacaccattctc
ctgcctcagcctcctgagtagctgggactacaggtgcccgccaccgcgccgggctaattt
tttgtatttttagtagagacggggtttcaccatggtctcgatctcctgacctcatgatct
gcccgcctcagcctcccaatgcactgggattacaggcatTATATTATTCTTTAAATTCAC
ATGAGAATTTAGTATGGCTTCAAAAAATACCATAAGTTAAAATATCACCAAGACTCTGTT
CAGACAAAAGTATCAGAAAAGTGAGCCAGGCACTCACATAGTTTATAGTTTATAAAAGTG
AGACAGGCATGATCTCTTAACCTCACTATAGTCCTGTGAATAAGGTTTATTTACATTTCA
TTTTACCTGCCAGGATTATTGTAAAAACGCCAAGCACATTGCCTACACAAACTAAATATT
CAGTCAATGGCTGCTATTTTCATGAGTTCGTTTTAACATATATTTATTGTCCTCTACTGG
ATTTAAGAAGTTATATTTATTATCATCTAAGATTTTAGCTATTCCTTCTCTTAAAAATAG
ATTTTATAATCAATGGCAGTAAGGGAGAGTAACTCGCAGTTCTCTGAATCTCAAGGGGTT
CCTGGAAGCCTTCCTGAAGGTATAGTGAAATTTCAGCTTCACATTCCCATCCATGAGCTC
CCTGCAAATATCCCGGTCTGCTCTCAGGACCCAGTGACTTACCTATGCAGAGGCTGTAGA
TAGCACCTGGAGCTTCCTGTGTGCCCTCCTCAAACTCAGCCAATGCCGTCATACAGTAGC
AGGCAGGTGTCTTTGCTGGGTAGTTGGACTGGATGTCCCTGGGATTGCAGAACTGGAATG
GGGAGTGACATCAGGAAACTATAATCATCAGGACAACATGGTTTGCCATAACTTTAAGTT

TTAAGCGACCGCAGATTATGCGGAGAGAGATGCATGCCCACAGCCATGCTTCCCATGTAA
CTGGAGAGGGGTCTGAAGTTTGAAACAAGTGTTCCTAGGCACGGGTTACAGTGTTTGTTA
TCATCATACTTGATTTAGAATGGGGCACAACATGTGGATTCATGGTAACTGTTACAACCT
TACTCATTTTAATACCTGAAAACATGCTTTCCCCATGCTGGGAATCGAAGATTCTCCTA
GGAAAAGAAAGGCTTGACAACATCGATTCAAAAAGGGCATGCATTTTCCTCATTTAAATA
ACTCTAATGTGCAAGTAGATCCCCTGACCTCAAGCTCAGAAGAGTCCAGGCCTTCACACC
TTCTCTGCTTCTGCTCTGGGGCCAGCTATTGAGATTCCTGTGCCCACGCAATGCGCACAT
CCCACCCCTGGCCGCTGTCCACAAGAAATCCAGTTGCACCAAGCACCCCACTTTTTGCAC
CTCTCATTTATGTACTCCTAAGAGCCTCACCACAACTCCCTTCTAAAAACATGAGTTCCT
GACTGGGAATTCGATGCTGCCCAGGCAGCTTTGCTCAGAGGGAGCAGCCTTCTAGAAATG
TTTCAAGTAAACTttcaagtataactaaattcaaaaaaaacacatacacacacacacaca
cacacaagtcaaaggtgtgtaatttggccaatatcacaaaccaattagccctttgtaagt
ggcacccagatcaggacagctgaccataccagcaccctagaagcaccccgtgctgcctcc
tgggacagggctaccaccatcctaaggccagcacgatgggccagctttgcctgctgttga
attttgcttacatagaatcctccagtaggtactcctttgggtcaggttctttcactcaac
attatgtgttgatatttttccatgctgtgctgcaaaattgtatttcttgcattccataac
tgggcagttccatcataggagaataccacactgcgttcgtccattctaccgccaatggac
atatgggttctttctcttttcttgcagttacaagtttatgaatattgtcccacgtgtccc
tggtgaacttttgtttgcatttctgttgggtacctcagagtggcgttgctgggtcagagg
gtactggtcgctttagtagctttgaaagatattgccaaaacattttccagcgcagttata
gcaaattatacaccaccagcagtAGAAAACATCTCCTAATTGCTCACAGTAAACCCCCAA
AGATTGCCACATACATCTTCCATATCAATTACTTAACTATTCAGCAAATTTGAAGGGAAA
TATATTTAATCTTTTTATTCAAATAGTTTATAAAGTGGAATAGAGATGTGGGTAAAAGTT
GTCTTGCCACCTTTTTAGATCGGTAAAAGTTTGTTGAATGCAGGCAAGAAAAGATGAGAA
ATAATGGTACCCAATGAAAGACATAGCAGTCTACAAGGAGGGGGCATTTCCCGGGGTGGGG
GGGACCCACACTCTGTAACTCCCACATTCAATTAGCATGTTATAGGTAAGCTGCAGAAAA
CGAGGCAGCTTGTCAAAGAGGAACGGCTCTTGGCCATGGTTGCTGCCCTAGGAGGATATT
TGATACTAGCAGAGCTGGGGCAACCCTGGAGGAAACCACCTGGAATGATGGGAGAACTCC
TCCAGGGAACATGGCCCTTTAATAGATCTCTGTTATAAAAAATAATCCCAAAGCAGCCAC
CAGGGCATACTGCTGCGATCAAGTCCTAGGCGGTATTCCCTTCTGCGCCATAGACCCTGT
GCAGAGTGCCCTCAACGAAGGAGCAAGGAAGACCAAGTCTCCCGAGGGTTTGCATATGTG
TATGTGATTCTGCAGTCATGGTGAATGACACAGTCAGGGCTGCGGAAAAGCATTGGTAAA
GTGTATATTTGAGGCTTCAGAAGTTTGAAAAGGCTAGATTTCCTAGGCCAAAACACTGAA
AATTTGCAATTAGAACTTCAGTGCTGATGCTGGGAAGACTGGAGTTAGTTTGAGACATGC
ACCTGTGCAGAACTGGGCCCCCAGAAAAGGAGAAGGAAGGGAATCCAGACCAGAGTAGGG
CCTGACACCACTCAGACTCGGCGTGTCTATAAATTAGAATTGCGTTACAATTACACTTTG
ACATTTTAGTGGTTTTTAAAGTGCCCAGCACAAGTTAATTTTTCATTAATGAATCCTTTA
TTCATAAAATGCTTAGATGGAGATTACCCTTTTGAGCATTTTGCCAGTGCTTCTGAAATT
AATGGGGACCTCCTGTTGGAGGACACAGTCTGTTGCAATAGGTGACCACTGCTCTGAATC
TATGTCACCTCtccaggaccacgggcacaaccatcacctgaggcatgttggagatgcaga
tggtcaggccctcctagaatctcagaatctgcattttagcaaagtcctgggtaattccta
tgtccattggagtttgagaagcaCTGGTAATCTCAAATACTTTAAAAGATTACTAGAGTA
AGATAGGCTCAGTAGGTACCTGAAGGCACCATCCCAAAGACCAGAGTGGTAGAAGCAGGT
GGACCAGCCTCTGAACACATTTCTCCCCCACTCCCCGGCTGTGTGGAAGGTTGCCACCTT
TGGGGTAGtcattcaacaaacacgtgtcaactgtccactatgtgtcaggccaccactggg
cactggctgtggctagctggatagacaccatttctgccctccagaaatgtcatgtccact
ggCACATGACAAGTCACTAAGTCATTCAGAGCCATGGGTGACAGCTCCAGGGGCCGACAA
AGGAGCTGTGATCTCACAGATCCACAGAGAAGTGTCCCAGGGCGGGCGGGAACCAGGACT
GCACAGGGAGGGGTGAAGTGACACATAAGAAGTCAGCCCATCAGCCTGAAATGCTCCCCC
AAATCTTCCCATTCAGTGTTTTCTCAGTAGCAAACTCGTGGGAAAATTGGTTATTTTACT
TAAAAAACTCATACTAGAAAGCTAGTTTAACtttaaaaataaattttaaaaacattttta
ttaaCAAATCCTACCTTTCCTCCAAAGTCAAGGAGAAAAGAATAGAAGTGAACAATGGAC
CAAGTAAGCCTAAAACTCTGCTCTTTCCCCTGCTCATTTTACAGTTCAAGTGCCATTCAA
TTTATCCTGGCAAGAAGAGGAAGGCATCATCAAGACCTTAATTTTCTAATACATCTGATC
TGAGAAGAATGTGAAAGCTATAAAATTAATTTTTGATCAATAACTACAGGCCTTTTGAGA
GAGTGCCCTCCTAATGAATTGAGTACCTATTTCTCCATACACAGTGTCTATCATGACCTA
CAAACCCTTTTCCCATGAGGTGTAACAGAGAGAGATTACAGCCTTGGAACTGGATGTCAG
ACTCTCCTGGTTTAAGACAATAAGCCATGACATAGAGCCTGAAACCAACACAATCTTCCG
AGTGGTTCCAGAAACATATAGGGGATAATGTTGGCTCTGATGCTGTACATCCCCAACAAC
CATCAACTATTTGGAAACTAGAATTTCAGCATAATTGGAGTTGGTGTTACCCTAGCAAAT
GCTGTGGGAAGAGAGTCTCACTGTGTATCTTCTCCTGTTTAAAGCCTGAATTTGTTCAGA
ATGTAATATCTCTGTTTAGCCACTCTACTGAAACTGATCTAGGAAATGTTCAAAAAAAGG
TATCCCAAGGATCCCTTTGTAGCTACATCTGTGGGATTCCCCTCGCTCTGGCGTGGCCTG
GCCCCTCTGCATTTGACAATACGGTCCTATGCTTTTGTCTTCCTGGGCTGCGTGAACCCA
CCCTGCCCTGGTTCACCTCTCCTCTTGACCCATCCTTATCAGTGTCTTGAAAGGTCCTTC
TATTGGAGGACACATTCTGTTGCAGCAGGTGACCACTGCCCCAAATCTGTTTCACCTCCC
CAgggccatgggcacaaccatccctggagtgtgttagagatgcagttggccaggtcctcc
aaaatctcagaatctgcatttttgcaaagtcctgggtaactcctatgtccatgagagttt
gagaagtaCTGGTCTCATGAGTTCCTGACATACAAATAGTGCTGAGGCCAGTATGCTGAC
TGGGTAGCCAGATACAAGTGAAAACCTTCCTGTTTTTTGCAAACCTGGATGGACCCGAGG
CCGCTGACGTGGGCCAGGACAAGCTACTCTTTTTCAGTGTTTCTGTTGCATCGCTGTGTC
TCTCTGTGATCAGGTGCTGCCCTCCCTGGCAGGAGGACTGCAGACAGGATGACCAAGAGC
ACTCTACACAGCCTGCTCTCCAGTGTTGGGGGACGCCACCCACCCTCGTGGTTCCTGTTC
ATCTGCCTACACGTGGAGGGCCCAAGAGGGCTAATATGTGACTATCTCCACTTCCTGGTA
CCCTGTGTGAATAACTTCACTTACTAAAGGGATGTTGAGCAACTTTATTAATAATGAAGA
AAGCACTTTGGTTTGACAAATAATCACTCCATTTTTTCATTTGAAAGTTAACTCTTGTTA
GTAGAGAAAGCAATGTATTACAACCACAAGGACGTTTACATGGAAATGAACCATCTGCAA

AGCATCCCCCATTTTCCTTTTAAATCAGCCAATGGGTGGTGGTGGGAGAAATATTCACCA
GAGTATTTAacatctatccccctctagactgtcagctccatccgggcggagactgttg
gtatctccacagcacacacagggcctggcacacatccggggctcagtgagcacttgctga
atgGTGAACAGATTAGCTCTCCTGGGAACGTTGTTGACACATCTCATAACACTGGTTTGG
AGTGGAGGGCATTCATCGGGCTGCATATTCCTATTTTTAATTGTATTCTCCACTGGTTAC
AGCACCTACAGTTATAAAGACATTGTTAACATTGCTTATAGGAAGACATTTGATGGAAAT
GAGTCCAAAGGCATTACGGTTAGAAACTGGCCAGGTGTCATTTTTGAGAGATTAGATAAC
TGTTTTCCGGTAGAGTGAATTGCCTGTTTGTTGCAAGTTGGGACTTTGCTGGGCTGGTTT
ACAGGGCCAAGGGGAAAGAGATAAGTGGATCTTCTAGTGAGAGGTCATCTGTTTTGAAAG
CCTGGAAGATTCCATGAACTAAATCCAAGTCTTACAACACAGGGAAGTGTGTCATACTGT
GCAGGGATGAAGTCTCCAATTTAGCATGAAAACAAGAGCTCCTCACACTGTCCTCTTCAG
AAAGCCCATACAATCCAAACTTCTGAATGCTTAGCTGCTTACAACCATACATAGATTGAG
GGATAAAACTCTGATATGGAAGAGAAGGTAAACATTTTTTGGCAGACATTCCCAGGAAAA
GGCGGCTCTCTTCTCTCATTGCTGCTGCTCTTTCAGAATCCATTTCAACAGAGGAGGAGT
CAATGGGAGCCCCGTGCCTCTGGCAGATATCATATGGCGTTTCAGTGGCATTGTGTGTTA
CCCTTCTTAGGTAACAGCTCAGCCATTAGAAGAATGTCCTACACACCTTCTCATTTTCTG
TGATGAGAGGAATGTGAGGTACTGCCCTTCGAGAGCTGTCATTTGTCCTAGTAGCCAGCA
GCGTGACTGTGCTGTCTTCTGCTCTGTCTCCCTGTCAGCCTTCTGCCCAGCCACCACCAC
TATAGTTTTGTTCTCTCCATTGGAACTCCTGGTTCAGAGAATTACCATAAAAAACAGACC
CCTAGACATACAACACTCTATCACATAATGGTGACTTTGTCTTCTATTTTGGATTACTGA
GCTTTCTTGGGTAACTTCCACTAAATCGAAGTTAATATTAGAAGAACTTCCTCTTACTAG
AATCGAAAAGCATTTAAGTGATGCAGTCAAGTTTGTACCATAAGTAATTCAGTCATTTAA
CAAATATATATGGCCTCTGTGCGACAGTGACCTTGACTGGGAATGAAGCTGTCCCATGTG
GGGCCTGTTCTTCAAAGGCAGTTCCCTGCTGCCCAGTTCAGTCCAGTGGATCTGGGCATC
TCTCTTTAATCCGCATTAGGGGCTCTTTACTGATTCTTCACTATCCAAAAAGACTTGGAG
GGGAGACCTGAGCCCACTTCTGGAAGGAAATGATAACAATTTATTTAGATAATCTTTGTG
CAACAAGTCAATTCACTGAAGAGATCTGCTCTCTAGGAGCCTCTGTGACCCCACCATAAC
TGGGAAGGCTCTACCTCTCCAGTCTTCGGGCCACATTTCTCTCTGGCCTGCTGTCTTCCC
AGCACTCTCAGCCTTGCTCATGGAGCACTCTAGTCCTCCGTCGACCTTGGCCTTTGGTAA
CGTGATTTTTCACCTGGCAGCTCCCATCTGGTCTCACTCCCTCTTTTTGTCCAGTCTGCA
TGACACAGCCTCACATCGTTAGTGTTCCCTCACTCCCCTCTTACTGCCCAACCTGCAAAG
TCCATGCCTGGGCCAGTGCAGCATGTGTCCTCAATGGGCTGCTGGTGGCAGTGGGGGGAA
CCGCACAGCCACGCTGTGTGCTGCTGAAGAAATGCACAGCCTCCTACCCTCGCCCTCAAG
AGGCAGCCATGGCTGCGCATTTCTGCCCTTCTGAGCTCCGCTCACTTTTGGCAGCAGCCG
TTCCAACCTGCATGGGATCTTCACTCTCTCACAGATGTGCTGACTCCTCCTgctgcctcc
cctctctgtgccttctcactctctgttccctttgccctttctccccttttctcctctgcc
tacctccaagccatccatcacaggacagctcaagcatcagatcctctgggacactttcct
tagttgttcagtctgatgaggtgtccctcatcctctCTTAGCTGAAAATCAGCAGCTGCC
TCAACTTCTTTTCCAGCATGTCTCATGAGTATTGCCACAACAGCATCTGTCACAATGTGG
GGTAGTGGCTGACTTGCTTTTCTGCCATTCAACTGAGTTCCCTCAGTGCTGGGGCCAGCG
TGCAGTGTCTTGTATTCAGTATATAGCTGATTAATTGATGAATTGATTAATTAATGGTTC
ACACTAGCACAGTGCAACCTTCAATGCAAAGATCTCATCAAAATAATTCACATGGTGGGA
TATTTTAGAAGGATGACCAGGCTAGTTTGTAGTAAGAAAAAATCAACAAGACTAGGTCAG
GAATTCTTTTTTTGTCTACAGGCTTGCTATAGAAGATATTGAAAATCATCTACCTAATTA
CCTTTATTTTATCAGGTTGTGTATTAAATATCACGTCTGGGGGAAGAAAATGTGATATGT
GATTACAGACCTTTCCTGGTACAACATAGTACGTTTCAGATTAACTCAAGGTATTGTGGT
GATATTGCGGTCAAAGCCAGGTGATTAAAGAGTCATTCTTTGAAACAAATATCTGTGCAA
TCAATTAAGAAATTAATTTGCAAATTTTATTTGCTTAGAGTAATTGATATATCATTCCTT
TTACAAACAAATATAAAGAAAACTTAACTAAAAATACTGCATATCTCTTTCAGATTATAT
ATCCCAGAAAGGATATATTTTTCTCCTTTCTGGTCTTCCTTTTTGGTGTAGCATCTGTAG
GAAATGCATTTCTTCATAGCTAAGTGTACCTCCTTGTGAAATATCTTCAGAGTCTACTGG
TGCACATAAGCAATTGCTGGCAGCAGCTTGAGGGTCTCCATCTCACATTTATCATATGCC
TTATTGCATGAGGCTTTGCAAGAGGAGGTCTAGAGCTACAATATCTCATGGATATGAATG
TCAATTCAAATCCCAGTGGCAGTTTATGAGGGGGAAAGCCTAGAAGAGAAGAAACCTAGA
GGAATCAAGCAGGAGGGGAGAGTAATAAAAGACTAGAGCAGCAGGTTTTTCTTAACTCAA
ACTAGAATTAAATCTCtgtgtgtgtgtgcatgtgaatgtgcccgtatgtgcatgcatgca
cgtgtgtaaatggatgtgtgtgtgtgtgcatgtgtgtgcaagtaagtgtgtatacgtgtg
tgggcatgtattgtgtacatgtatgtgtttttatgcatctgtttgcaagtatgtgtgta
tgcacataaaagtgtgaatgtacatgtgtgcttggtgtatgtgtgtgtattaatgtatgc
gtgtAGTTCTAGAGTCTAGTTAGAGAAAGTGCATAAAGAAATAGGGAAATTAACAAGAAA
GCTATAGCTTAAATTATAGGAAAAACTTTTCTCCCTATCAGTCATGGTTTTAAAATGTTC
AGACTTGATATGTTTCCCAGTGCTATTGTCAGAAAATGTCCCTATGACATTCCATACTAC
TTCAATCAAATCTAAAACCTTTGTTCCAACATGTTTTATTGATATGAGTATATTTCAAAT
TTCTACCAGGTTTTTGGAGAGGTATTTTGGCCATAAAATTGACTAAATTATTCAAAATAA
AAAATGAATAAGCCTGGGCCAAGGCTTGGAGACTTGCTTAACTCAGTTCTTAAATTTTCA
GATTTTCAAAATTACAAATTTAAGCTCTAAAATCATGGTGCTGTGTATGATATTCTTTGA
TTGCAACTTATGGTTGAAAAACTATAGAGGGCTTTATGCTAAGAGTTGTGGATCTTAGGA
TTTTCATGAAATCTGCATTATCATCATCTGCAAGTTTAGATGGGGCATAACTGATCCAAA
GGATGGATCCCTCGGGGGCAATTCAACTGGCTGATTCCAGCCAAGATGACAACAGTCAGG
ATCCGTTCCCTTCTGATCATCCATTGGGTGCCCTGATTTCCTCTACAGCCCTAGCTGAAA
GACCAGACACTATCTCAGGCTGGCTGCCCCACATGCCTTGCTCCACACCAAATTCACAGT
CTATAAACCTGAGCCTCCAGTGCTCCTACTACCATACTCACTCGAACATTCCCGATTCTG
ACCTGGAGATGTCAACAGCTACTTGATGCCACTCTCTTCTATCTTTCTGTAGCTAAGCCA
TCCCCAAGTTTGTCGATTCACCCTCTTTAACCCCTGTCGGGGTGTCCATTGTGCCCCTTC
ACCCTGCCATCTCCCTGGTGCACtgttttgcaaagttcagcatacatgagcgtcacctgg
gaacctaataaaagtgcagatgttgattcagcaaatctgggatgccctcggctgcattt
ccagcaggctcctggggatgtccccgctgctgtgctgcagatgacactctcagtggTGGG

ACTCCAGGCTCTGCTGTCGCCTCCTAGGGGTTTCTCCACACTCCCTGGAGGCCTAATGGG
CCCTTCTCCACATGGCAGTAAGATCTGTTTTTGTGTTTGTGTTTCAAGTTGGGAGAAGGA
GATTATTTAATACTAAAATGTGCAACATGGGATTGAGAAAACTAATTATTAGTCATAAGT
TGAGTATGCAACATTGAAACCACATGCTTTAAAAAATTATAAGAAAAAATCATAGTATTT
GAAAGTTACAAGCTATTATGGCTAACTCCATTTATCTCAGTTAGAGAAGAAGAGTCACCT
GTCACCAGGGCACTGCCAGAAGCCAGGCTCATTTCCAACAGCACTGGGTGCTCCAGCTTT
GGGGTGCCAGCTCCTCCCATAAAGCAAACACATACCTAGGGATGATATTTCTTTGCAAGG
GCTCTGCCCTACAGCTTGTACATCTCAAGAAGTTATGTaattaaactgtctgttttgaga
aaattgtagattcacacatactagctgtaagaaatgatgcggataaatccagcgtaccag
ctttccccacggagacgtcttgcagcgtcacagccaggatgaggcattgacccaggcgaa
gtccagagcacctgtgcgctacagggccccttgcactgtgctgtcacagacacgcccact
tccAGATGCCATCTAGGACCCCCTCCAAAAAGcagaggcattcttaaaaacacacatctg
cacatgttcctcttcatttgaatctgtcagtggcttctcagtgcctttcaaatgaaatct
aaagtccttacaagccttgcagcaggaacctctccatcccacttcccctcacactctcag
cttcatctctgctaggctctgttcagccaggcagcctttcacagtccctctcctcctgcc
ctgccaggaaggtcccctgccccccaactcttccccacatgtggcggggccccgcttgtcc
ttagaagcccagctgaactgcttcctgaaggaacccctccagaacctctcagaccaggtc
aggtttctgcactcttagatcatccccatggcataatcacagttgtgatgttgtgatgat
tcagtgaatgtctgtctccccactggatggtaagcttcctgagggcaggaacagcattgg
ttccagtcaatgctatgtcccAGGACTGTTCGTTTTTGCACATACTAATCCTAAAAGGAC
GATGACAACAGCAACcacttacatgacctagatgctcttctgggtgttgtgcaaatatta
acaatttaatccttgcaacaatccacgagggaggcattcttctactcccacttaacagac
aaggacagtgaagctagtaaagagaagtcatttgcccaaggggaccccactactgttggc
agagctgggtgcaaacgcaggcttgtgaagccaggacccatgcattcaaagaccatgcca
GGTGCCCCACTGCACACCTCATCCCCACATACCAGTGAGGGGGAGAGAAATGCTCCTGC
ACTGCCTCTGATTAACTGCTTTCCTAGAAGTCACACATATAAAAGGGATTTAATTCTAGT
GGGATTGAATCTCAATAGTTTCCTTATTAGGTTGATTTCTGTTAATAGTTTAAGTACTGG
ATATACATGAATTAGAAAATCTAGATTATTAGCAAATGCAAACTATAAAGTATTTTATAA
ATGTTATCTTGTTTGTCAGGGGATGAGTGAGATATTCATTATACAAAAAGTAGTGTGGAT
TTTGAGGTAGAAGGTTTACTAAGGATCATACCGTAGTATGAAATAGCCACAAACATTCAG
TGAAACCAAACACCCCCGCTTAACCTCAAACTAACACTAAATAATAAGGAATAGACTTGG
GGGCAGTGCAAGTGTATTTCTAATGGTGAAAACCATTCCCCAGTGAAAACTAATGTACCA
TCTAGTTAATAAGAGCTCCTCTGACCCACGCACATCAATACTTACATCCCAATGGTGATG
TGACATTTTGGGTTTTGTATTTCTTTTGCAAATTGAGCTAGCATTTTTGATGAGTGGCAG
GGCTCTGCTACCCAACCTTTGGACAGTTTCCAAGCATAAAATCACAATTCCAGATAATTC
TGTCACAAAGATCTGGGTCTCATTAGGAAGGAGAGGAAGCTGGGAGATGATCCAGTCCAA
CCTCCCCCAAACCAAACATCACGGCCTTCTCAGTTGTTTCACCAACCATCTAAATGTTTT
AGTAATTCTAAAAATTGATGCGCTTTTTCCACGAAAGGAAGTGTTACCACATTTTCCAAG
TGGGAGGCATCTATATCCTTACTCCTTCATCCTCTCCTTCCCACCCCCTCACCCCCCACC
ACCCACACAACATCTGCAATTCTTAAACTAAAGCACAAATTGTTACAAAAGTTAATTGCA
CTTTCAAAGGAATGCTTGTATAGAAACTTTCTCGGCTTCAAGGAAAAATAATACGCTTTG
AATGGCTGTTCAACAGCATAGAAATTAGCTGAGTAGAAGGCACTCATATAGCCATTAGGA
CCAATCCTTTCTGCCGCCAACACCCCCCTTATAAAGACTTGACAGTGGGCCAGAATAAAC
AACTTCAGGATGAATTCAGTTGAGACACAAAGTACACACTTCCAGTTTTTCCCTTCTCTG
GTTACTGGCCTCAATAACCAGGCAGTCAACTTAAAAAGAAAAACAAAAGCTTGCTTCAGA
TTACAGATTGCAGACTTCTTATAATATGTCCATTTCACCAGGCCCCGCTCTCAGCCCCGG
GAAAGGCCACTGGAAACCACCTCACATGGTAGGGCCTTGCGGGAGCCAGTAATAACCTTA
TCTCCGTCAACATGTTCTGTCAGATTGAATGGGGCAGCCAGAGAAGCCAGAGTTGGCACA
GGAACCAAAACAAAGGCTTCCCATCCTCCTGGAGTGAGCGGTTGAGCCTGGATTGGTGCT
TAGACCTATAATGGGTGCAAGCAGCGTTCATTCATAGTGGCTTTCTAGACCCAGGGACTT
GGCCCCAGCCCTGCTGCTCCACTCCTCTTCTTGCTTCATTACCACGAGTCTCCTAGACCA
CCGAACGATGCCTGCATTTGAAAGACACTTCTGCTGATCAAAGCAGCTGATGTGTCCCTT
TGCGGTTCATTTCTAATTGTCCCCAAGGAGGAGAAATTCAAATAGTTTATTACTGAGAGT
TAAAGAAATCCACTGAAATATTCTTTGGTCTAAAATTACTGTCATGGCGGAGCAGCTTCA
CCTTAGTCATTGCCCTTAAATATGAAAGCTATTTAAGAAAGTTTGCCCTTAAATATGAAA
GCTATTTTAAAAAGTTTAATGAAAGAAGAGAATCACAAAACATTTTCAAAAAGCAAAAGA
AAACCTAAGAGAAAAGTTGAAAGTAGGAATTTTTTAAAGAATATACGACGTGTGTTCTGT
GACTCACCCCTGCAAGTTATTTGTGTGTATTCCCTTGCATAGTAATTAATAATGAAGCAA
AGCATGGCAATGATATCTTTTCTTGTCTAGTATTCTAGAAGACTCCATGTTTTTGGAAAA
TATCACTCTAGTTAGATCTCAAATATATTCAATCAGAAAATGGGTTTTCTACAAGATTCT
ATATCTGTAGTCAATAGCAAATATAATTCTATTAAGCTAGTAGGATGTGATAGGAAACTA
AAACCTAGGGGAGACCAAAGCAAGGAAAAATACTTCCTCATCCAAACTTGAGAGCAATTT
ACCGTCAGGCCTACTATTAATAGATGGAATACAGATTCCATTTTCATTACTCAACTGCCA
TATTCATTATTACACTGTACAGAAAAGGGAATCACATCTGTTGAAAACTTATATATGATG
TTCATGCATGCATTCCAGTAATTCAACAATTTTTATTTATCTTTTTATTGCTTGCTAATT
TTTCAAAATAATAAGCTAAAGAAAACAAAATGTTTGTGCTGTTCTCAGATGACATGTTAT
CTCTTTAAAGGACAAAATGTGCTGTGAAATAATAGAATGCTTTCAGCACTCAAGTGTGAG
TGAGTGCTCATACATGAGAGAAAGCCGTGGGGACTACAGAAGCCAAGAAGCAGATCTAGC
TGGGGAGGCCTTTGCAGAGGATGTAGTTGTGTGGAGAGGCCACACACGTGGAATTCCCAG
GAGGGCTGTGGAGGCGGGGAATCTGCAGGAAAGCACTGGGGTGGAGAAACGTgatgagaaa
caattattgtcttaaaatatctgcagggctgtaaggtagagaagcaatacgttgcatctg
tgttaagtcaaacaaaattatcaagggactggtttcagcttaacataaggaacaattatg
tgatagggttgtcaataacaagagtagactgcttcttcacacactcctagtcactcagaa
tggtccaggaggagtggacaaccatttggtagagTATGGGAAGGCAGGGGCCCTGGGTGG
GAGTGGTGAGGGTAGGGAGTGAGTATCCCAATCTAGAAGTAAATTGTGCCCAGCACGGAG
CTGCAACACTGCCCTGCACACAAACACACACAAATAACAATCCCCAGCCCCTGCATTTCC
CTCTCCGGTTTCAGGACCTTGTATCTTACTTCAATTCCTTTATTTAGCTGATGATGAAAT

AGGAAGAGCTTAGCACTAAGAAAATCCTTTTGGAGTTTGGCCTTGGGGGAAAATGAATCA
CTCCAACCAGGTCTGTCTTCTAGAAAGTATAGGATGAAAGGGCTCCTCATCACATACTTC
CTGACCTCCTGCTAGGCCTTTCCCTAAAAcaggggctggcaaagcacaacctgtgggtca
cgcctagcctgccacctgtttttgcaaataaagttttattggagcatgactatatgtatt
tgcttacagtctgtggctgcgttcacactatcccagcagagttgaataattgggacaggg
accatatgatgggtgaagctgaaaacattactctctggctgtattcagaggaggtttac
tgagccctTCTCTGAGACATGGCAAGCGCTGCTTCAGGCTCATGCTTCACTAGATTCAGG
CCTGGGGCAGTAAAGAGCCAGCTCAGGATAGCACTCCCGACTCACTCATTTTTTCAGGCA
GGGGAGCCATCTAATGTCAAGTGCCTACGTGCAGGAACTGGTCTGTTAATTAGCAGCTCT
CCTCATGGAAGGGATAATATATTCTAGAAACAGGAGTGCGGCCCTATTGCAAGAATGTCC
TGAGCCAAAATTAAGATTCTTCTATGGCAGAAACTTGGCTGGGGCTTCTCCTGAGTTAAC
TTGGTAGTTGTTAGTGATTTTTGAGTCAGTTTTTCCTTGTCAACGACCCCAGGAATGAGT
TTGGGATTACAGGGTAGCCAGGGAAAGGGAAAGCTTCACGCCCGCCCCCGGGACAAGGTC
TGTCTTCACACTGCTACATCCCTTCACCCACTTTAAAATGAAACTTAAAAGGAGGATTTC
AGTTGAGTAGGAAGTGAGAAGAGGGCTCATTTTAAAACAAGCGTTAAATGAAAACCCACA
CACACTCAGAGCACACAAATCCAACCACGCTTACAAAACCATCACAGAGGTCAGGCGAG
GCCCTTTTCTAAATGAAAAAGAACAGGGGTGGAGACTGTTCTGAGAGCATGCTGGGTTCC
CTGAAGGGAATTCTCAGCTGTATGTGCCCCGCACAGGATCCCTGCTAGACACAAGGCCAG
CTGCCTTCCTTTCAAGCCGCAGACGCATCCCTGTGTCCAGGCGGGCTGGTCAGCTGCGGT
CAGCACCAGCTTCCCCGCTCCATGGTGAGGTCATCACAACATGTGAGCAGGAGGGCAGGC
CGGCAACCTCTGAGTGCTTAGAGAAAGGGACGGGATTCCTCCTGTGCAACCCCTCTAGTC
TCACTCAGACTCAAGTCTGACTAAGGGGCCAGGTGCTTTGACCAGGGACTCTCCCCTCTC
ACTTCCCTCCCAGGAGTCACAGGTACATGAGTCCTTGTTTTACAAATGAAGAAAACAGAC
CCAACATGATTAAGATGTTGCCTTCATAGGGGTGGCACCAGGATTCCAAACCATGGACTC
CACTGAGCCCAGTGCCCACTGACATGTGCCAGTAACAGTGCAGCTGCCTGTGGTTCTGTC
GACTAAACTGCCGGCAGAGGCTGGCTTTCCACCTTCTTTTTTTTTTTTCACTCTTCAAA
CACTTTATGACATGAACATAAACTACTGGCTGCATCGTTCTGCTGACAACATGACATGTT
TCTATAACTTGAAAAAAGCAAGCAGTGGACTGCTCATTGGTAAAATTGAGTCAGTAATCT
TTTAGGAAGGTTATTTTTCTTCCTTTTACTGCTTCTCATCTGTTCCCCGCAGTAAAGAGG
ACAAGATGACGACGACTCAGGGAACACCTCCAGCCTGAAGCAGCACCATGCGAGCTTAGA
CCTTAGGGTCGGCTTAGAAACCACAGGCGGGGCGGCTTGGGCCCCTCGGACACTCCCTCT
CGAAGCTGCTTCTCCCCAAGCTACCCCAAAGGCACTGAGCGCCCTCTGCCCCCCAGCAAT
TCAATTCACTGGCTGTCCTGCTCCTGTCAGTACTGAGAGTTGCATGTTTGACCCTCGGGG
GAAAAGTCCAGAGGCCCTGGGGTGTCCAGCATGCTCTGAGGTCCCTGCTGCTGACCCCTT
GCGCTGTCAGCATTCAGAGACATTCACACAGCACAGCCTCCCAGGCTAACAGCTGTCATG
GAACAGTGGAGCAGCTAGACGTGGCCATTCTGTGGCCCAGTGCTGCAGAGGTCAAAGGGA
CAAGCGCAGGGAGCATCTTTGCTTTCAGAAAAAAAAAAAAAAAAAAAGAAGCACACTGGTG
CACTGACCTGCTCCTGGTGTCTTTGTGATTGCTCTTTTCTTTCGATTTTTGGTTGTCTTT
TTTTTTTTGAAAGAGGGGCTTTTATGCTTTTTTCCTAATGTTCATGGGTAAACCAATGTA
AATGTGTGTATGTTTATAGAGATGGCTTTAAATCGCAATTCTGCAGTAGAGATTGATTTT
TTAAAAAACATGGGTAAAAATTGAAGAAAAATTTTAAAAGAACATTTAAACCATCTTGGG
CTAGGGGTGGATATGCACCACCCCACGGAAGCCAAACAAAATCTCTCTGCAGATAAACAT
TTGCAAAAAGAATTTCCAATCCCAATTTTTGAGTCAGAGATCTTTTATTTCCTTGCAAAT
TACATATCTGTTTCAGGATTTTTGACTATAAGAAGAATGAATGAAGATGTGTTTCTTAACA
GATAACTATGAACAAACCAGGAAGGATAATAACTTGTATCCCCCAATTCGAATCCAGAGG
ATGGGAAGGCATAAAAAAAAGAAATGGAAGAAACTTTATTTTTAGTGGTAAATGGTGGGA
CTATGTATTTTACGTATGGTGAAGTCACCAAGCCCAACACTTGGCACTTGTAGGCAAGGT
AGTCTTCTAATCTGAATGTGAAGTATTATGTTTTCATTTGCTTGGTAATGAGGAATATTG
GTGCTTTCGTCCCAGTTCTCGAGCTGACTGACTTCTCTTTCTGACGTGTGTTCCTTTAGC
ACACCTCTACACTGCATGGCTCTGAGATGTCCTGTGACTGTTTCATGTGTAAAGTTGCCT
CCCCAAAGGACTCACATATTCCTTCAGGGCAGTGAGTACTTCTGATTCATCCTTAGCAGC
TACCTTCGCGCTACTTTACTAGATATGTTGTAGTTGAATTAATGAACAAAAGAACAAGCA
ACTTTGGTGCCTGGTGTGCATCTCAGAGCAGGGTGGAGTGAGCCTGGCCAAAGGGTCATC
ATGCAACCTCTGTGGCTGACTCCATCTGGCCACGGAGCTTCTCAGCCATGCTTGGTATTC
ACATGACTTCTAGGGCGACAGCTCAACCAGCAAATAAACAGCTTCATATGGGAAATATTA
CTCAGCCTTTGTCATCAAGGAGTGAGTCACGGGCCTGAACTGAATAGAAGATAGAGGAGA
AAAGGTGTGTGGACTGGGTGAGACAGCGCCCAGCGAGGTGAACTCCCGGCAGCCCTGCCT
GTCTTTACCTGCACATCACCTTGCTAGGGTGCCTTCGGTTGTGAGGGCCTGTCTAGGAAG
AGAAGAGTTGCACCCTGGCAGGCAGCACTGAGCTGTCTCATGCAAAGCTGAGGAAGAAAG
AGTGAGCTGCCCAGTGAGCCTGCTGGGGTGGTGGAGGCTGGGCTGGGCTGTGCAGTCTGC
AGCCCCCAGCAGCCCTTGGCACCTTTCTACTGCCTGGTGCTCACCAGCTCTCCAGTAACA
AAGAGGGACGTGAAGTCAGAGGGGAAGGGAGGTAGCACAGGGCAGTCTTGACTTTGAACA
AAGAGCTGGCTTCCTGAAGTCAGCTGGCCGGGTTTTGAAGCCGATTTTCCAGCAGTGATC
TTTGATGCCAACCCCATTTAGGAATTCTGTATCTCCCCCTACCTTCTACCAGATGTCTCT
GAGCTCACCTTTGGTGATAATCATGCAATCTCCGTCATCCCCACGTCCACACTGCCCCAT
TCTGTCCCACCCCGGGTTCTGTGGTGCTGTCGGCTCCCCAGCGAGCCAGGAAGGGAGAGG
CCAGCTCTGCTGGGGCTCCTGCCGCCCTGGCTCTGCACTGCCCTTCTCTGGCAGGTCTGA
GGCGCCACTGGAGGAGCCACACGGCCCTGAAGCAGCAAGGCAGATGCCCTGGACACAGTG
GAGGCACAGAGTGCAAGCACCGGCCTGGCCCACAGACTTTTGGAGGGGAAGTGGTATTAT
TCAGTTCAAAAGTATGCCTGTGTGTAAAGAGAGAGCCCCTGAACATGAGTAAGCAAAAGT
CTCAGCGCAGAGATTAGACAAGTAGAATGCTGGCCCGAGAGGAGGCGTTTACTCACCCTC
TGTCTAGGAAGGAAAGCCAGGCCCAGCACGCTCACTGCTATCTATCCTCTCACACAGAGG
GATTTTGAATCGAAGCCAGCATCCTGTCCTTTCTCCAATGTCCCCTGCTCAGGAGTCAGG
ACTCAGCAAGGCCCACCCCAGCCACACACAGATACAGTTCCAGGACTCAGAACTCAGCGA
GGCCCACCCCAGCCACATGCAGGTCCAGTTCCAGGATTCAGGACACAGTGAGGCCCACCC
GAGCCACATCCAGGTCCAGTTCCAGGACTCAGGATTCAGTGAGGCCCACCCCAGCCACAC
ACAGGTCCAGTTCCAGGACTCAGGACTCAGCGAGGCCCACCCCAGCCACATGCAGGTCCA

GTTCCAGGATTCAGGACACAGTGAGGCCCACCCCAGCCATATCCAGGTTCAGTTCCAGGT
AAATCATCTGCCTTCCTCCGTCCAAAAGCCTTGTTTCCTGTGTGTCCTTGTGTTTAAAAT
GGAAACGTTATGAGAAACTGCCTGCCAGGGCAAAGGGTGCTGCCCGGCACACAGTAGGGA
CTCAAAATGAAACTATTGTATTGAATACATAACAGATCAACGGGTATTGCTTTCTGAAAT
CTTTTTTAGCCCAATTTTGTTTCTTATAGTCCAATAACAGGTCAAATTCATTTCTGATTT
ACTAGCCATTCAGTTGCCCATAAAAAATGGAAAGTGATTTAAGATTATTAGTTTAAAAAC
CAATGAAGGTAAAACAGTTATCATTGAAGGCACATAGGCAGAAATAGATTGCAATAGTTG
CTGCCATGTGAAGCCTCAGTGTCATGCTCCATATTTAGAGAGATCTATGATTTCTGAGGC
CCTTTCATGTCCATGATCTCAGTACTGCTCACAACTGCCCTGTGAAATTCGCCGAGCTGG
CCCCATGTCAATCAGAGTACACTGAGCACTGAGACCCAGCATGTTGAGATAACTGGCTAG
AGATCATCCCATAATGGTACCATCACAATCTTCACACTGTAGAAGTTTGATGATGTCACT
GGAAGCATATTCCACAGTCCCTTGTGAACTGGCCTTCCTGTGATCAGAAGCATCAGTGAA
CTCCCAAGAGGGTGGGAACTCCCAAGAGGTATTCTCACTCTACTTAGTGTATATTTTACA
AATCACAAGCTTGGCTTTGGATTCTTTTAATGGCTAGAAGGAGAATCATGGGGTTGGAAG
TCCACCAGTTTGGGTATTCTGTTCCCTAACTCAAAATAAAGAGATGTTATTTTCAAGTCT
TCTGCTTGTTAACTTAATTAGAGATACATGAGTTTGCAGCTGTGCTGGGCATGCCGCAGC
TTGGCATGTTTAGTCCAGAAGGCATATTATAATGTACATGGAAGATTGTCAGAAATTCAA
AAGGACTTTTTGAGTATCACATGTGTATTTTCAAGTTCCAATATAGATTCACATTCAGTT
TGACAGGTATCTTTGGATGCCTATCAGTTAAGAACTATTTATTAGTTGTGGAATAAAATA
GGGTAAAATAAGGAACAACTGAGGAAAAAACATAAAATTTGCTTTGTGAATAAAAGTTGT
CTTCAAAATTATGACTTTTTCCATCCCACAAAAGTTTTGATTAAACCCACAATGAAAATT
TAAATAAGTGTATTTACTTTGGTTTAACCACTTATTTCATTATGACTCACAACTATAGGT
TTTCTAGTTTCCATTATTACAAACTATTGTGTGGTTTAAATCAATTTCATAGACTAGTCT
AGTTCTATAGTCACAATTTATAAAATTTTTTTATGTGGTAAATTGAGTGTCTTCATAGAT
GTACATGATTATTTCTCAATTTTTAAGGAATGTATTTTTAAGATAGCCTTCTTTAGCCT
TCTTTAACACTGATTTTTGTAAATTTTTTACAGATTTTTTTAAATTTTTGGTAATTTTTT
AGCATaaagtaatacatggtcactatggaaaacataaaaacacaaaaactatgaagagta
aataagaaaaacacccagaaatttaccattcagaaaaggtcattgttaacaacacggtgt
atcttcctcctgtcatgcttccgtgcatttgagcacatttgagatgtgtaTACATGTTCA
CTTTGAGATTTTAGTATAGCAAAAGAAATGACCGGTCCTGATTCAATGAAACCTCTGGCA
AACTCGCTATATTTTCCTTACATATTTTTAAGTTCATCCTATAAATGAACTATCCATTCA
TCTTATTTGAGATTTTCTTAAATCTTTCAGCAAGAAAGCGGGAAAAAAATCCTCCTCTGG
CCTTTAAAGCCTAATTAAATATATGACTAAGCTAGAAATATTTTATAATGACCAACCAGA
AAGTGGCAAGGACTGTCACTCTTCCCATACAGCCCACCTCCTCCTCTATCTCCCTCAGGC
ACACGGAAACGAGAAAGGCAGAGAAACCCAGGACAAGTCATCCAAGACTTTGGTCACATG
GCCATCCATTGCTTTCACAACAAAAATATAAATCCAACATGTGTGTGTGCATTTCATACC
AGTAGGTCCAATAAGCTATCTATATATACACATATGTGTACACACACACACACATCCT
TAcagacactccccagcttactacagtttgacttaagatttttgactttacgatggtgt
gaaagcaatgcacattcaatggaaaccatacttctaatgttgaatttttatctttttctt
gggttagttgatgtctgatatgttacttcttgcgatgccaggcaatggctgggagccag
agctcccagtcagccatgcaatcaagaggctaaacagctgatactatacagtggactgtg
tcaccagcattttggggatattgtgttttgtgtttttgaatcctatcatgtctacaaaat
gccattttcgactgctattttcaatttagggtgggtttatcaggacataacccctatggaa
agttgaggaccatctgTATATCTGGTAGGGAAAGATGGATAACAAATTCATAGGCAAATA
ATAATTTCATGATTATTATTAAGTTATTCCTACTTAATAATAAGTAGTGATCACTGCCAG
GGAGCAGAGAATGCAGGATAATGTGACAGATGTAATGGTGGGTACTTAAGCTAATGTAGT
TGCAGAACAGGCTTTTCTAGAGGGTAGGCCTTTAAGCGTACCTCGAAGATGCAAAGGAAG
CAAAGATGCGAAGATCTGGGCTGGGGATGGAAGCAGAGACAACTTGGAGGCCAAGGGGAG
AGACTGACAACAGCCCAGCTCATACCTCAGCAGCCTTTAATGCATAGCTAAGAAAACAAC
AAATTAAAACAATTATAGTTTACTTAGACGATTCTAAGTGTCTAAGTGGATTTGGGCAAA
TCTGGAGAAACTTGTTCTAATACTGTGTCTTAATAAGTAATATAGATTTGCCCAGGCTTG
TGGGCAGAGTGGTATACACCCCATAATAGCAGAGGAAGGCCACAGGGCCTACCCTACAAA
ACCAGAGGCATTTAAAAACTTAAAGGAGGCAGATTGCTTTTATTTTCAGTTAAAATAAAG
TGAGGAGTTTCTCAAGAAAAATAATAACGAGACCACCGGCCCGCCCTAGATGTCCAACAA
GAATGCACAGATAACTTCGTATATCCACTTTCCTGAACCTGCCCCTGACAGCCAAGTGGA
Gcacaacaacagagatgaacctcaaaactactgtgctgtgacataaggcttgctcaagag
gacagtgtggtgtgagtccatctatgttctaaagcaagcaaagctattctgtagtgaaaa
tggatcaggacagcagttgcctctggtgtatgggggcagggatcgactgggaggggcatg
agggatgacagttaggtttcgatcatgacaggaattcagattactccagcatgtgcatt
tgttaaagctcatcaaatgctacacttaagattaatcctctcacagtttgtggatgttac
cttaaAAACAACAATGATGACTGCAAACTAATATTGAACTCTGGTTAGTGATATACCAAT
GTGAAGTATAGTGATATCTCTACTTTACTTTAAAATGCATCCAAAGGCAGACTAGAGGAC
CATATCTGACAGACAGAAAAATAGATATGTGATAAGGTGAATGTAGTAAAATGCTAACAT
AAGGATGTTTGCGGTACAATTCTTTCAGCTTTTCTATACATTTATAAATCATAATAAAAT
TTTAGGACAAAAAGTTAGTGCTTTGAAGTCCTAAGTCATAGGGCCTGCTGCTCTTGATGC
AGTAGAATTTGTCTTCAGATTTGCAAAGGGTAAGGCAAACCACTAGCATTTTGTATGGAA
CTTGATGCAAATACTTTTAATTGTCTGGTTTTCAAATGTATAGACTTAAAGTAATATCAA
CTCTTTCTTTGAATCAACTACTGAAATACCTAGTCTTAAATAAATATTTTTATGTAATCC
TTAAAGTACTATGTATTCATTTTTCTTTCTTCTTTCTTTTCTGGTTTGATAAATATTCTA
TAAAGTAACTGTGTTTAATGGCCAACATTTGAGTAAGTCCATATGCAGATCCAAACATCT
CAGTTTAGACAATAACTTAAGACAATATAGAGTGGCTGACATCCCCTAACGTGGGTCCAG
ATGCATGTTATGTTATGTTTCTGTTGCATTCTCAATAGTTAACTTTAATAAAAGAAAGTC
AAAAGCTTATAtatttttcaatcttcaaaacatttctgggaggttgtcttagttaattt
tatgttgctatacccatatcacagactgggtaatttataaagaaaataaatgtatttggc
tcatggttctggtggctgggaagtccaagagcatggcattggcatctgcttggcagctgg
tgaggccttcatgctgtgtcaatctatggtggaaggtcaagagagcatgcatgtgaggt
ggtggggaagagaaaaagcgggtttaactcatccttttatcagggactcactcccgtgat

agctaacccattcttacatgaatggcattaatccattccttagggcacagctctcatgac
ctaattataatacctcttaaagtttccacctctcaacactgttgcattggtgattaagtt
tccaataaacgcactttggaaaacacattcaaacccacagcagagatcaacgttattgtca
ccattttcatatttgaggaaagcatggcacagagagcttggagaagtacttcaaggtcac
ccaatgaggaagtggctaaacaaaaaccttatcttaaattaattaaaaacctcttgctct
ttgcaGTTTTGTCTTAAATCTACCTAATTTGTGACTGTAATTTTTAAGTAATTTACTCAT
ATAAGTGGTCTCACATTAAATTTTCTCATTGCTTTATATTTCTAACATGAGATATTTGGT
ATAAGGATGGAACCAAGATCATACCTTGTTTTAATTAGAAAACCTAGACCAAGTCATTGT
GATCCTCATCCTAGATTTCAGTTAAATGCTGCTGTCTCCTTTTGGGTATGTGACAGGGGA
AAGCCTCAGAAGAAACAACCTTATGTGTTTTCTTTTGATACTTTAGTAATTAACCCAGGA
TAGTATTCAAGATTGACATGCCTTATATTGAATCAAATAGCATATCAACTGCCTTCTTAT
TCTCAAGTATAGACATGTTGGGTAATTGGGCATTTAAGTTTCTTTGCAATTTTTTCCATT
ATTAACAAAATTAATGAGCAACATTCTGCATAAGGTCTGTTTCCTCAGAATACGTTTCCC
AAAGTGGAATCATCATGACGTAGAATTTAAGCATACTTACTTGTTTAAACAAATTGTCCA
GTTGCTTCCCAAAATGTTTTGTGAATTAAGATTTACATCAAGAATATGTAATGTTGTTAC
TGTCTCCCAAATACAGGAtctttttctgaatataaaagttatacatgctaattgtagaca
atGAAGGGTCATTATCCTCATAGATAATGAAGTGCTTCTAATACTTGTGCTTTTATTCAT
TTATTCAAAAAGTGCTAAATAAGCCCTGAAGGGGCTTTTGGGGGTCATTTGGGGCTTAT
TTAGCACTTTTTGAATAAATAAATAAAAGCACAAGTACAGTTTTTTTAAAATACTGTTTT
CTATAATAGATTAATCTTAAATGGCATGTTTTCCTTTATTTTACTGACAAAAGTTACTTA
CTCTGTGATTGAATAATAAAAATTCTTTGGTTCAGCTGAGAGAAACTTGCAAGCTGACGT
CCTTGATTATTTAAAATGAAAGCAGCTGCCTGTTTTCATCTCTCTGCATCCTGAGGAAAC
TCTTCTGCAACGTGTTCCAGCCCTAGGTTCTAGCTGACCCTGTTCATCTGTTTGGCACGA
GGGGCCCAACTAACACTTGCGGCTACCTGGACGACAGCCAATCTAGTTGGAATGAGAGTT
AGAGGCCATAGTCTGTCAGCTGGGAAAGCAGCTTTTATTCCAAGGTGTGCCAACCGAAAG
GCCACATGTTATTGTCACAACCTGGTACCTACATCAGTGCTGACATCTTTAAGAACCTTA
GAATTGGGAAATCAGTTTAGCCCTATCTGCATGTGTAGCCGACAACCACACAATTGTTCC
AACTTGAGGTTGCATTCAGAGCAACCTCATTTCCCCCATACTCCTGAGGAAAAGCAGACC
AGAGACGCTGGGTCAATCCAGAGTTATGGTTGGAAAAATGATGGAATAATTCTGCCCCTG
GTGATAGGAGAGAGGGACTCCATCTTGTCAACTGTCATGGTTCCCATGTGAAAGCTATCA
TTATCACTGAAATTGAATGAGAACACAGAAGGGAAGAACAGGGAAATCCCCACAGAGTTA
AAGAGGATGTGAAGATTGCTTCATGTTTAATGTTTGTGTAAGTGCTTTGGGTTGGTTATG
TGCTGTCTGAACATGTGCTCATTTCCATGGCTCATTGAGAGGGCAGACAGTCCAATGATA
CTCTTTAGAATCATTCCCATGGGGAAGGAACAAAGAAGCCTGTAAAATAGAAATGCACAT
GTAAAAAGCATTGAAGAAAGTGCCAGTGTATTGATTTTGGCCATGGTTTGTGCTCTACCA
CCTGGTTACTGTGATTGCAGAAGTGCCTTTGCAGATGAGGAAGAACCTGGCCAAGGCTCA
ATCCAACATCCAAAGCCAGAGGCCATATTTCTTCACTCTTAAGATAATTTGGGTTCAAAT
TATAGTCCCTTTACACACTCTCTGCCTCAAAAGGCCCAAGACTCTCTTTTGTTATGCTTG
CCTAAACATGCCTTTCAAAGAACTAGTTCTGTAAATACAACTTATTATAAACCTCTCCT
TTGCTTTTAAAAATGGATCACCACGTCCATTTCTATGGTCCAACTTTGTCCCTTAATTTA
AAATTTTTTCTTGGATTAAGTTTGATGCCTTGAAACATTAGGAACTCAAGCATACAAGAT
TGTATGCTGGTGGTGAGGGAAGTAACTGTGCCTCCGCCTGTGCTGGGTGGATCAACATGG
AGTGTGGACGAGCATAGGGATGTGTGGGTTTCTCACTAGCTGAGAGTGTTTTTAAATGTT
GTATTTTGATGTTTGTTATTTTCTGAATATTCTACAGTTAGACCTTTGATTTATTCTTTG
ATGCATTCATTTGAATAATATTTTTAATCTCCAGCCAGTTAGGTTTTTAATTTACACTTT
TGTCCCTGATTTTAGGTGTAGTGTTGTGTACACTACTGCCCAGTGTATGTTATGTTTGTA
AACATTCATTGCACGCACAACAATGTGACTCACAATATTTTTGAGAAGTAAAAAGTTCAT
TATATAGTTATTAACTCAACCCTACAGTTATATTCGTGAAATACCTTGTGAAATTTATTT
TTTGCCTACTGGAGCTCTTACAGGTTAATCCTGTCTTCAAGATTTTCATAGAATTTTCAT
CTACCACCCACCCCTTTAAATTTCAACATTTTTTTATTTTGGCATTTTAATGCAATTCAA
TGCATTATAGGGACAAGCTATCTCTTATTATGAATTGCACCTTATATAAACTTAAAGATC
TTTTATCACAAATTTCTTTGCTGTGTCCTTTAGTGAGAATTTGTATTATCAGTCACTAAA
GCTCACTAAGTTAGTAAGCTTTGCGCCCAGATGACCTGGGCAGGAATGGGTGAGTCTCTG
TGTGGAGAGAGTGAAGAAACTGCTACCCTTAATACCTGGACCTTGAGGGATTGTTTTATT
TTAGTTTTTCTGCATTTCTCAGTATTTCATGTGATATCTGTCTTTTTCTTCCAGTTTGCC exon 2
AAGGCACGAGTAACAAGCTCACGCAGTTGGGCACTTTTGAAGATCATTTTCTCAGCCTCG EGFR_e2
AGAGGATGTTCAATAACTGTGAGGTGGTCCTTGGGAATTTGGAAATTACCTATGTGCAGA
GGAATTATGATCTTTCCTTCTTAAAGGTTGGTGACTTTGATTTTCCTACACAAATAAAAT
TGGAGAAAATCTAAGTGGAGAAAGGCCTGGGCAGAATTCCACTTGAAGTGTGTTTATTTT
TGCTATGGCAATGACAAGTCTTACAGAGCTACAAACGAGAGTTTTATGAGAAAGCCATTT
TACCAGCTAATGTCAAGTAATAACTAGAAAAGGATATCAAATAGAAACAGGCTAATCTGG
AGTTCCATGTCATCATAGACACTGACGTTTATCCCTGACCATTACCTCAGTCATGATGTG
CTGCCATACTCGCTCTTAAAAACTTTTTTaaaagccctgctttgcaccatttgcctatt
cccttagtgtaaatactcctactatagctgatttcaaggtaccaaGTTTCACTCAGCTGG
TCACAGAATTCTTATTTCACGATAGGCGCTAATGACCCCATAGGAGCCAGCTCTGAAGGC
TTCAGAGTTTCACTGAATTTTGGATGGGGTTTACTTAGCCTTCTTCTGTTTTTCTTTTAC
CTTTCCTTTTTAAATAAGAAATAATGCAAGACAGATACAAAGTAATTCTTTTTAATTTCC
ATTTTCACTGGAGAGTGTTGAACCCCGTGAGGCATGAGAGCACAGTGTTCCAGAACAATG
CTTACTGCTCATTATCACAGGGGTCAAAGGCTAACGTGCAGGGATTGTTGCAGATCGTGG
ACATGCTGCCTCCTGTGTCCATGACTGCAATCGTCTACCTATTTTACAGTTGTTGAGCAC
TCGTGTGCATTAGGGTTCAACTGGGCGTCCTAGGGCTCCCTGGACCCATTTTAGACCTTG
AGTTCTTGAGTTCCTCAAAAGAGAAATCACGCATTTATGTTTTCTCTTCTTAGACCATCC exon 3
AGGAGGTGGCTGGTTATGTCCTCATTGCCCTCAACACAGTGGAGCGAATTCCTTTGGAAA
ACCTGCAGATCATCAGAGGAAATATGTACTACGAAAATTCCTATGCCTTAGCAGTCTTAT
CTAACTATGATGCAAATAAAACCGGACTGAAGGAGCTGCCCATGAGAAATTTACAGGGTG
AGAGGCTGGGATGCCAAGGCTGGGGGTTCATAAATGCAGACAGCAGTTCCGATGGCTCCC
AGCGAGCTTGTCACTCAATTCCACCTCGGAGAAGGCTTTTATTTTTACCCAGTACACGTG

CACTGAGTGCCGGCTGTGTGTAAGATACTGCAGGGGAAGTTACTGAGAAGATGGCAGATA
CTGGAATGGGAAGATTTAAGCGGGGTACCAGTGTTTACATGGACATGAAAAAATACTGAG
AGATAGTAAGAAATCGTAAAGATTCTGAGTAAAAGAGAGTATGACCAAACAAGCTGAGCA
GGAATCGTGAATCTATGTGTGTAGGCAGTGAATAAACTGCCAGTCTTATTACCTGGACCT
CAAGGATAAAAGACATACAGTAAAAATCAACCCACATTGAGGACAGTTTCGAGAGTCGCG
CTGCTACACAGAAAGCCCTGTGTAAGTTAAGGATAGAGAATGAGGTGTTCTAGAACTTTG
AATTTTTGTGAGCAGGACTCGTGAGGTTCCTGTGAGAGGAAACAATGAAGGATGATAGAA
AAGAAGGGAAATTGATTTTAAAAAACTGGAGATAGCAGTGATTGTGCCTCACTGTGCAGT
GGGTTTGGGGCCAGGAATGTTAAATTGGTAACTTCATTTAACGCCCACAACCTTTCTTCA
AAGTAGGCACTGTACAGATGCCCCTTGACTTATGATGGCATCCTATCTGGCTGGACCCCG
CCGAGGGTGAAGGCGTCATTAGGTCGGATTTCAGGGCTAattgaatgtatattgccttca
caccatggcaaagtcgaaaatctgtgttaaatcatgctaagccggggactggctgTGCTC
TGCCATCGTACAAATAAATAAATGGAAGTCAAGTAACTCCCTTGAGGGCCCCAGCTAGTG
AATGGAGAGGCCAGCTATGGCCACCACTCTCTGCCCCAGGGCGCTCAACGCCCCTCCTGT
GCCATGCAGTTCTGACAGGGAGGCAGTGCTGGTAGGAAAGGGGTGTGATGAAAGGGGTGC
CCAGCAGAGGGAGTCATATCCGGAGTGACAGGAGCCCAACAGGGGTGCAGCGCTGGAACC
CAAGCCAGCACCTCTGGTCATGGCTCCTCAGTTCACCGCCTATAAAATTGTGTGGTTCCC
CCACACCCCTTGCTGCTCAGAGCAGCCGCGCACATGCTTGTGCTGTGCGTGCCTCCTGTG
AGATGGCCTGGTACACCGGTTCCTACAGTGCGCCTCACACGCTGTCTCGGAGGGAGGCAG
CCTGTGCGGGTgcctggacctccgagccagaccctctgggttcctgcctggccccgtccc
tcagcagccagatggctcgggagcacattctccaatccctccgtgtctctgtttcgtcat
cttcaaaaatgtggatggcatagctgctaaaaaatggtgacatacttcctaggtggtgca
gaaaattaagtgaCTGTAGGAACAGGCCTCAGCAGCTCCTTCCACTTCCTTGGTATGATT
GTTTTTTAAACCAAGGCTGGGATTGTATAGATGCAGATTAGTTAATGTGATACCATTAAT
AGCTAACCTAGTGCCTGCTGCAGGGTGAGCCTCCCCTAAGCCACCGGGAAGCGGCTCCTG
CAGCCTCCCTCACGTGTGCTGGCCCTCCTCTGGCAGTCATTGCCTGTGGTGTGCTGAAGG
CCCAGCTCTGACTGTGCCTCTGTGCTCTCCTCGCCCCGCCCCCTGCTCTCTCTCAGGTCT
TTGGTCTGTTGTCCGAGCTGCCACAGCAGCCTGGACATCCCTGTTGGTGTTTCCAGCCCT
GTCCTCTCCTGAGTTCCATCCACCTGTGCATGGCTTTTTCATGAGTGTTTTCACGGATGG
TTCTGCTGTCATCTCCAACCTGATAAACAAAGCACCACGATTCAGCCCTTATGACCCCAA
GCTTCCTTCCTCAGTTCCTTGCTTCTGTGCATCCACTGAAGAAGCCTGTTCCACTGTTTC
CCTGCACTGGGTCTCCTGTCTGCAGGAAGCCTTCAGCCCTCACTTCCACACTCCTCTAAG
ATGTGTGCCTGTGCCCTTCTGGGGAAGCTCATTTTCCTAGCAGCCTCCAGGATCTTCAGG
GGTGAATCCCTCCTTTCCCACGTTGGTACTCTGTACACACAACATGCCCATTCCCTGCCT
GGGGAGCTGGGCATTGCTTCATGAATCAGAGGTCAATTTTTTCTCTATTAAAGTCACAGA
TGCTCATTGCACCATTGTGAGAATGAATGAAGATAGTGCTTATAAATCAGCCAGCAAGGT
ACCCAGCCTCACTGTGTCAGGGTCTCCCTGGGCATGAGGTGGTTAGAGtgtgtgacatgt
ctgtccccaagcctgtcagctcccagatcgaagccagtggatctcattcatcctcgcagc
gcccacagcacttgcacagggttttgtacacataagtcattctgtcaatgttcatgttta
atgTCATCAGTGGAACACTCCCACTTTGTAAAGACTTGAATGTGTTCATCCCTGACTTTT
CCACATCTTGTTAGTTCTTCTTTGGAAACAGCTGTACAGTTTCACCATCCTGTGCATCCC
TGGAGTCTACCTGTCTCTGTCATACATTCAGATTCTTCTTGTTTCGTTCACTCTCATAT
CCTTTTCTCTAATGAAAAGCTCCGCCTGGGCATGCAAGGTGGAGCCCTGGATGCCAGCCC
CTCACCTGGCATCCAGGGCTGTAGCACTCAGGAACTGCCTCCCTGCCCTGCCTACCCCCT
ACATCATGCGACCATTCCAGTCCAGCCAATCAGCCCCTTGGGACCCAGCTTACCACATGC
ATATCATTTATGCTGTGACCACTGACTAAACCATTCTCTTCCTTCCTCCCCATATTTCTA
AATTTCTAATCATTGCTCAAAGCCCAATTCAGAGAAAACCCTAGCTCCTCCATGGCACCA
TCATTAACAATTTTATCTGGCCGCCCCCCGGGAAGTTCACTGGGCTAATTGCGGGACTCT
TGTTCGCACCATGGCATCTCTTTAGCAGAACATAAATGCGAAGAGCACATGCATCCTTCA
TGGGAATTTAAAGGAGCTGGAAAGAGTGCTCACCGCAGTTCCATTCTCCCGCAGAAATCC exon 4
TGCATGGCGCCGTGCGGTTCAGCAACAACCCTGCCCTGTGCAACGTGGAGAGCATCCAGT
GGCGGGACATAGTCAGCAGTGACTTTCTCAGCAACATGCCGATGGACTTCCAGAACCACC EGFR_e4
TGGGCAGCTGTAAGTGTCGCATACACACTATCTCTGCCTCCAGCTCCTATGGGGGACAGC
TCTACAGCACTGGGGCAGGGGAGAGAAGCCATGTTTAGTAAGTCACATTAATCAGAAACA
AAAAGTAGTAAGCAAAATATCTGACCACTAGAAAAGCATGTATTTACCACGGACATAGAG
ATCGTTTTTTTGTGGCGGGTGGCAGCCCAGCTGGTTGGCAGTGCAGGCCACCGGAGGCAG
ATCCCCTGCAGGGACAGCAGAGCACTTGTGTCCTGAGAAGAGCTGCTGTTCATGGGGCTG
GCAGCACCAGGGCCTCTCCTAGCCTGCCCTGCTGACACTGGCCAGACTCCTACATGCTTC
TGAGTCTCCAGAGGCTACCCGGCCCTCCTGAAGCACCAGGGCTGAATCCACCCCCAGCTG
AGGGCATGAACACTGCCACATGGAGTCACACACACAGCTGGGCACTGCCATGGAGAGGAA
GTCTGTCCATGTTTCCTTGAATACTGGTGGCCTGGTCCCTGTCCCATTCCCCAGTGAGGC
AGCCTGTGGGGAAGCCTGGCAGGGAACCAGGCGCAGGTCAGCGTGGCGCCCTGACTCAGG
CCAGCACTGATGGGGGACTCTGAGACGCAAGCTCACACTCACCCAGCTCCCCTGGGCTGC
GCCCGTTCCTGATCGCTTGGACTTTCTGTTCTTTAGAGTAAGAAGTGATCACCATTTCCT
GCTTCTTTGTTTCTCCACAACTGTGCAGTGGATGCCTGTTTGTTTTCTGCCCTCAGAACA
AAAAAAAAAAAAAAATAGAGCTGACGTGAATCTTCAAAATCATCAACTACAGGGCTTTGGA
TTTTTGTGTATTTGTTTTATTTTCATTTTATGGATGGattgtgatgaaatgcccgtaata
caagattttccatcttaaccattgtaagttacaatgtcagtggcattatacatccacatg
ggtgtgtggccatcaccaccgtccacacacagaactctttatcttgcaaagctgaaact
ctaccattagacagtaactctctgctctcccttccttcccagcctctggccctggcagg
caacagtccacttgatgtctctatgaatttgactgctctggggctctcatacaggtggaa
tcatgtagtatctgtccttttgtgtctggcttatttcacctagcaaaatgtcccgaaggt
ttatccatgctgtagcacgtgttaagaatgtccttcctcttcatggctgaataatattcc
attgtatgttgacactacattttgtttgtccattcacctatctacagacactggggttgc
ttccatcttttgactgtttgaataatgctgctgtgaacatgggtatTGAGGCTCTTTGTT
TTATAGACATATTATTCCACCAGATACCCATCCTGACACCTACTATGTTTGCAAGAAACT
GAAAGCTTTATTTTACATTGCAAAATTTCATATTATGAGATCAAGGTTAGCATTTCCTCA

GCTGTCTGGTGGACAATGGGGAGGTTAAACTGTGCACATTTTATTTTTTTTTAATGAACC
TGGAACGGTTATGGGGCCAGTGTTTGCCATGGATCAGGTCAGGCAGCCCACAATGGCAGG
TCTCCATGTTCTGTACAACAACTGTGGGAAAGACCCACAGAGAAAGTGCTGGAAAGGGGA
ATGATGGGTAGGTTCATGCAGTAAAAAGATTCAAATACTACAGGGCATTGAACTATAGGC
CAATATAGCATTGCTTTAAGAATAAACAAAAAATAAGACAGTAAGAATAAGCCTAGCAAA
ATCAAAAGTCTATAAAGAACTGACATTTCAAGCCAATAAGAGAATAATCCTTATTCAAT
AAATTGTCTGGAATGACTTAACTATTAGGGGTGAAAATATCAAAGTGAGAGAACTATAAA
GGGTTTTTAAAAAGGAATTAGGTATGTTGGGTTAGTCGCATTGGAGAGTGCAAATTCACC
ATCGACCTGATACCTGAAATTTCCTCCTTACCATCTAGAGGCAAGTTGGGAATGCTGCCA
GGCTCCTGTGGTAAAGGAAGCTCCTCTCTTGACTGGTGCTTTATGGCTACACGTTCCTGC
TCAGAATGGATCTCATTTAGTCTTCACCAAAAAAAAAATCTCATGAGATGATTTAAGTG
TTTTATGGACAAGATGTCTAAAACTCAGAAAAATTTCACAGTGTGCCTAGCTTTTATGTT
TATGTTGAAGTTGGGCATTAGAAGTTAGAATGAATGGGTTTACTTCAGAGAAAATTAAAT
CCATCACCCACTCCTTGTACTATGAATTCCAAATACatattaaatacatataataaaaata
tttaatatatatGTAAGTGCCAGAAGGAAACATAAATATGAATATTTTGTAATATCAAGT
TGAAGAAAAGCCAAAATCTGACATCATAAAAGAAAACTTTCAAGTAAAATATGTTAATGG
CTACCAGGAAAATATTGTGCAATGTCTGATTGCCATGAAGAGGGTTAATATCCTTGCTAT
ATCACTCTGTGAAGTCATCTTTAAAAGACTAAGAAAAAGATGAATCTCTTAATAAAAACC
TGGCCCAGAACATGAGCAGCCTCTCTCTCTCACTCTCACTGTCTCTCTTTCTGTcacaca
cacacacgcacacatacacacacacacacaAAATATGGCCAAGAAATAAAGTAAAATGTTA
TTTCTAATGTAATAAGTAGGTCAAAATAGAAAAAGAAAGCATCACACCTTCCTTTGCAAA
GTATTTGGGTTCCTTTTGCTTTTAAACACCTGGGTCAGCTGgggtgtcgagaaacagaaa
ttctcacgttctgcttgtgggcatatatgttaataaaaccaagcttggcaatatgcctgc
aatatgtatctaaagcttcaaagtatgtatagctttgaccaatcaatatcacatttcgga
ataagagaaaaagaaataatgaaagtgaaaatcataagagatgtagaaacatattcttat
acaagaattccttgcagccttatttataataaattttgtgaacaaattatatatctaaaa
ataagagattggttgaaaaaattatgcagcagccatgctattgataatcatgttagatag
aagcatatttaaaggcatggaaaaatTGCCATGTTTTATATGGGTTTTAAGGTTATAAC
ACAATGTATAGTGGGATTCCAATTCCTGTATATACATAGACTTATATGTCTATATTGATT
AACTCTGGATGAGTCTCATGTCTTCTTTTTGCTTTCTTCTATTATCCATATTTTATACGA
TGTGCCTGCATTTCTTTTTTGTAACAGATGGTCAATACTAGAATCATAAACAGATCTTGT
TTGTTTATTGGCAAATGTTTCCCGTTAGAAAAAGATGCATTTTTCttttaaatattttta
ttttataCAATGATTACAAGCTTATAATAGAAATTTGAAAATTATATGTGAGTACAGGGT
AAAAAGTTGAAAGAATGGGATTGCACGCTACAGATCTAGCTGCTTTTAGCACGCCTGCGT
AGGACCTTGCTTTCTCTAGACCTCTGTTGCAGTCTCTCTGCCTACCTCCTCACAACGTCC
ATCCCCCGCGGTCACTGTCGTGATGCCAGCCTCCCCGGCCTTCATGTCTCTAAGGAGCAC
CAGCGCGGCAATTAGCGCCCTTTGCCTTGGTGGTATTCTGGCTTCACAGTCACATGGGAG
ATCAATCGTCAGCTTTTCTGTTTGAAATCTAAATTCTTCCTGACTGCAGGGGACCTCGGG
ACCCATGAACACCTCTAGTTTACTATGTCTTCACAGTAAAAGATATCTGCATGACTGGAC
TCTTTAACAAATTTGGTGGTTAACCTACTCTTTCTATATAGATATAGCACTTCGACCTTC
AGACTTCTCAATACTGATAAAAAGAAAACACGACAGATGACAGGAAAACCTTTGCAGCTA
TAATTTGTAATCGGCCAATTATAAAAACTGCAAAAATTGACCAGATAGCTAAGGTTTTAC
ACAGTCATGAAAGTGATCTGCACTGTTAACATTTCACCCTCTGTGCACCATTCTGTGCTT
CTCTCTGGTTTGGAGTCTAGAAGGTTTTATTTACAGGCTATGACTTAACAATCCCAGAAC
GGCTGACACATGCAGTCACTCAAGACTGGACACAGCAAGGAAGTAGTGGGTCCATGCCAA
AGGCTCAGCCAGACGAGACACTCTAGCTGTGGCAGGAGATGCCAGGGAATGCTCCAAGCC
TAAGCAGATTGTAAACAAGGAACCTCAAATTCATGAAAAATTCTTGCTTATGTGGCCCAT
GTCAGtaattactctctgcctcagtttccgcagctgacatgtaaataaaagcagttcatg
gttcatCTTCTTTTCTTATCGGGGTCTCAAGTGATTCTACAAACCAGCCAGCCAAACAAT
CAGAGAATAAGTTGAAAAGATTGTCTTCATTTATTGAATGTGCTTAACTCAGGCCCGGGA
AAGGGCGTCATCAGTTTCTCATCATTTCACTGAGATATGCATCTATTACTTTTACATTTC exon 5
AGGCCAAAAGTGTGATCCAAGCTGTCCCAATGGGAGCTGCTGGGGTGCAGGAGAGGAGAA EGFR_e5
CTGCCAGAAACGTAAGTCAGTGAACAGCCTCAGACCCATGTGTGACCGCCCCTCTCTTCC
TTCACTTGCTTAGGTGATTGGATTTGTTTTCCCTCTGAAGACTCCAAAGAGTTACTTTAT
TACAGGGTCAGATGTGAACCAGTAGGTGAAGGACAGTCTTGCAAATCTCACCGCATGCAG
TTAATCCAGGGTGGGCTATTTTGGGAGCTTCAGCCTATCACAAATAAGTGAACATCAGCA
GGggctgggcgcggtggctcaccccctataatcccagcacttgggaggcggaggcggtcg
gatcacgaggtcaggagatcgagccattctggttaacacagtgaaacctcgtctctacta
aaaatacaaaaaattagccgggcgtggtggcgggcgcctgtagtcccagctactcgggag
gctgaggcaggagaatggcatgaacctgggaggcggagcttgcagtgagccgagattgtg
ccactgcattccagcctgggcgacagagcgagactccgtctcaaaacaacaacaacaaca
acaacaacaataaGTGAACATCAGCAAGTACCCCAGCCCTGTCCTCTGAACACAGCACAC
TTTCCCAGGAATGGAAGACTTGCTCCTGTTGACAGCAGTCACCAGACTTCTTGTTTCCTC
TCCCTCCCTGGCTTTCTTTGGTACCCACCTACACAGAAGCCTGAGCACGGGTTCTCATGG
GGACTTTTCCATGTGGACCCTGCTTTACGATGGAGAGGGCCATTCTCCTAGGTATGGTTG
TCTGGCTCAGCCTCTCAGTGGCCAAGGAACCTGGGGACATGAGCTCAAAAACGGACACTA
TGTCCTTAAGCTGAATTGTGGGGGGGCTGTTAGGCCCTTCTAAACACTACTTCCCAGCAG
GTATTTTTGTTCTTTGTATGTGCTTTCTGCATTGCCCAAGATGCATCTAATTATTTAGCA
GGTCTCAAAGTCTAGACTTGATCTCATGAGTTCTCTTAAGTGATTAAAAATAAATCAGGA
GAAAAAAGAGGCAATCAGAAAAGGGCATGGTTTGACTTAGTTTGAATGTGGTTTCGTTGG
AAGCAAATGTGTCTTCACTTTTTCATGAAAAAGTCTGCAAGTGCTCTGCGACATCCCTGG
GAAATGATCCTACCCTCACTCTTCAGCTCACAGGGAACCTTTGCTCTTTTTCAGTGACCA exon 6
AAATCATCTGTGCCCAGCAGTGCTCCGGGCGCTGCCGTGGCAAGTCCCCCAGTGACTGCT
GCCACAACCAGTGTGCTGCAGGCTGCACAGGCCCCCGGGAGAGCGACTGCCTGGTAAGAT
GCCCCTCCAGCAGCCTCCCTGGAGCAGGCTGGGGCTGCACCCGCCCCACCCACACCAGGA
CAGAAGACTTCCTGTGGGGGAGCTGTCAATTAGCATTTGTCATAACAGACAGGATATTGC
CCTCTGCctggtgacaaagtatctttagtatcctgcctccaccactcactgagaccttgg

gaaaatgatgggactaccatgcctccatttccttacctgacaatgatgcataacaaagtc
tctcccagttgaatgcttaaatgatgagatgcctgtgatgtccgtcattaggacctgggc
acagaacaagcactaaataCTACATGCAAGTATTTGTCATGAATGTGCCTTGTTGCCAGC
AGCACACTCTCTTTATTGTTTGACTTCGGCTATACCTCTAGAGACTTGACACTGTGAGGT
CCCTAAGAGACCCATGGAGAGCCACACAGGTCTTGCTGGCTGGGGCTGGGTTAGGGCCTC
CTGACACGGATCCCTCGGCTCCTCCACCACTGCTCAGGCACCTCCTGAGCTGCACCCTGC
CCTCAAGGGGTCCTGAAGTACTCACTGTCGCCCCATTGCTCCAGAAAGTGCCAGCAGAAG
CCTTGCTGCCCCAGCGGGCTCTGAGCAGCACTGGAGGGTACAGGTCAGAAGCGTCTTGGA
AGTCCTGGAGACGCCAAGGCTGGTGGATGTGACTCCTGGAGTGGGAGCTGGTGTGACGAA
GCCCTTCCTAAGACTAAATCCAGAGCACTCTGTGGTTTCAGAGAAGATTCCTAAATTCCA
GAGTTTGGACCCAGACCCAGGAATTGTGACTTGGTTGGCCTGAGCTGTTTCTAATGTGAG
CCCCAGGGAGAAGACTGTGCGTGGGGTTGGTCCTAGGAAAAGCCCTCGCTGTATTGGGTC
TGGCTCCTTTACACGGCATTGTTCTAGCAAGGCTTTCTGCCATTCAGCAATACATTATAA
AATATACCCTCAATTGTACTTTATAAGGGAAGCCCAATGTCCTTTATAAGGGAAATTAAA
CATAATTTCATTCCATAGTCACCGCTATAATGTGTGAACTCCATCATCTATACGTTAGTA
AACAGACGTATTTTTATCATAATCCATAAATTATGATAGGTGGGACAGTGCACCTAAGAA
AAAAATGGACTTTTTAGAGAAGGGTCTTTCTGACTCTGCAGAGGGCGCCAGCTGGGTTTT
CCCACACTAGTGGAACACTAGGCTGCAAAGACAGTAACTTGGGCTTTCTGACGGGAGTCA
ACACCGTGCTGCGCTTCCTCCGTGTGTGGCGCTGAGTGTACTTACCTCACTTGCCCAGCG
TGTCCTCTCTCCTCCATAGGTCTGCCGCAAATTCCGAGACGAAGCCACGTGCAAGGACAC exon 7
CTGCCCCCCACTCATGCTCTACAACCCCACCACGTACCAGATGGATGTGAACCCCGAGGG EGFR_e7
CAAATACAGCTTTGGTGCCACCTGCGTGAAGAAGTGTCCCCGTGAGTCCTCCTCTGTGGG
CCCTCTAACTGGTCAGGCATCCTTGTCCCGCTCTGTCTCCTGCTGAGCCCTGGAGTATCC
CATCTTGGAGAGTCTTTGGGTGGATGTGTTTGCCTTGCTTGGAGGAGGCGACCCTGTGCC
CGTCCAGGCACACAGGCGAGGGGAGGGGCTGGCTTGCTACCGAGGAGCGGGCAGGTGGTG
GCCATCTCCACCCATGGGGGCTGCTCAGTGCACAGGGCAGATCTGGGTGGCCAGGCCACC
TCACAGGAGAAACACCTGCTGCTCAGCCCTCACCACTCATCCAGCAGCCACAGCCGTGGG
TATTCAGTTGTCTGCTGGGCACAAAGCCGTGGGCATGCCACTGTTTAGTGCTTGTGCCAA
GCAGGTATTTAATACACCGAAATCAGAGAGTCTATCAGAAGACCTGCCTTCTTGAGTGGT
TAAAATTCTAGTGAAAGTTATGCCTCTTAGGAGTATTGCAGAGGttttgttttgttttt
atttgttttgttttaatggtttgggtttgagttttgcttgtttgtacttacatttgtAC
TGGTGGCTCCAGGGTTTAGGGAAATTGTGACATAAAATAATTCCTGACAGAGAAAGCAAA
ACTTTGTCTAATGAAAGAGTTTTAGAAGCCACTCTTGATCTCTAGAAGGGGAGATTAACT
GAGAAAAAAAATTGAAAGAACAATTATGAGGGGGAGATTTTACCCTGCCAGATTTGTGTA
CATGAAAAATTTTACATTCCGTATGGAAAAAAAAACACAAAATAATAAGCCATTATAAG
GTAAATGACAAACAAAGCTAAAGAAAAATGTGCCACAGTGATGACACAGATATATCTTTG
AGATAGGGCTTAACAGAGCTTTAAAATCCATAGGAAAACACTTCGAGCCTGAGATACCAA
GAGCAGATGGTTCACAGAAGAATCATCAATGTCCTATAAATATTTTTGAGGATCTTCTTG
GGGAACTTAAAACAGGAACAGGCCAGGCACAGTGGCTCATtggctcatgcctttaatccc
agcactttgggagactgaaggggctggattgtctgaggtcaggagtttgggaccagcctg
gccaacagggtgaaacctcgtctctactaaaaatacaaaaattagccgggcgtggtggcg
cacgcctgtaatcacagccgctcaggaggctgaggcaggagaattgctttaacccaggag
gcggaggttgcagtgagctgagatcacaccactgcactccagcctgggtgacagagcaag
actccatctcagacaaacaaaaaaGGAAGACATAGAGCTCCTAAAAATAACGCAGAAGTC
TGCTATTAATACAAATGAATTACTTTAAAGGTGAGAGCAGGTGGAGGAGAGGGCTGAGGT
GCCTGCTGGGACGCAAAACAGCTGGCCCCTCAAGGGACCCAGTGTTTCCTGCCATGATGA
AACACCTGTATTGTCCACATTGCGGCCTAGAATGTTATTAAACTCTTGAACGGGATTCCT
TCTCTATTTGCAACCTTTCATTCTTTGTCCTTAAAGTAAATAAAGCCAAAGGAGGATGGA
GCCTTTCCATCACCCCTCAAGAGGACCTGGACCGCCTGTGTGAGGCCCGAGCACCTGGTG
CCACCGTCATCACCTTCCTTTCATGCTCTCTTCCCCAGGTAATTATGTGGTGACAGATCA EGFR_e8
CGGCTCGTGCGTCCGAGCCTGTGGGGCCGACAGCTATGAGATGGAGGAAGACGGCGTCCG exon 8
CAAGTGTAAGAAGTGCGAAGGGCCTTGCCGCAAAGGTAGGAAGCCCGCCGGTGTGCGGAC
GAGGCTTGTTCTCGGCTGCTGAGGCTGGGCTCTCATGCCACCTCCAAAGGAACACATCTT
CCTCTTCTCATTAAAAAACAACTATACATATCGTTTCTTTAAAACAGAAGATAAAGCTGT
AAAGCTAGGTTAGGCAATGGGAAGGCACTGAAGGTTGTGACGGGGTGGGGGCTCTGATG
AGAACAGTCACAGAGCCAGCCCCGCTCAGCAGCTGCCAGGTGCCCAGCCCTGGGGAGAAT
CCAGGGAAGGCAGAGCTGGAAGCAGTGCAGCTCCAAGCGGCCCATGGGAAATAATGAGGA
GAACGCAAGGTCAGTGTGAGGTGACAGGGATGGCATCTCCTACACCGCCGTAGCCCCAAA
GTGTACTATAGGTCCTGGTGTCCCCCCTTCCCGCCTGCACTCTCCCCAGCCCCTTCAGTG
TTTGTTGAGTGAATGAAGGATGATGTGGCAGTGGCGGTTCCGGTGACCGGAATTCCTTCC
TGCTTCCCTCTGCCTGTGGATCCCTAGCTATTCTTAATCCAACAAATGTGAACGGAATAC
ACGTCTCTCTTATCTCTGCAGTGTGTAACGGAATAGGTATTGGTGAATTTAAAGACTCAC exon 9
TCTCCATAAATGCTACGAATATTAAACACTTCAAAAACTGCACCTCCATCAGTGGCGATC
TCCACATCCTGCCGGTGGCATTTAGGGGGTGAGTCACAGGTTCAGTTGCTTGTATAAAGA
AAAACAAAATCTGCCTTTTTAACTGGTAGAGATTGGTGATCAATAATCACCCTGTTGTTT
GTTTCAGTGACTCCTTCACACATACTCCTCCTCTGGATCCACAGGAACTGGATATTCTGA exon 10
AAACCGTAAAGGAAATCACAGGTTTGAGCTGAATTATCACATGAATATAAATGGGAAATC
AGTGTTTTAGAGAGAGAACTTTTCGACATATTTCCTGTTCCCTTGGAATAAAAACATTTC
TTCTGAAATTTTACCGTTAATGGCTGATGTTTTGATATTTTTCAAAAGTGCAGTTTCTCC
TGCAGGCAAAAGGGGACACGTTAAGTCCAGGCTTGGGTCATTCACTGCGGTGTAAACACG
CTTTCTCCCTCCCGCCCGGCCCCAGCCAGCTGCCTTGGTGGCCCATAACCCCTGAGGGTA
GAGGGAGGGGACAGGGGTAGGTGACAGGCAGCCTGGGCCTCAGGCTTTTGAAACTGGACG
CCAGAGCCTTGTGGGGCCACGGGCAAGCCTCGGGTCTATGACTGCCGCctgagctccgct
tccttcctctctaaaatgggaagattagaccaaaataacaagactgttttaaggttggaa
tcaaataaggaaaatttgtaaagctccttgtatgtgataccagatccacaatTGGCAGAT
AATCGCAGCAGGAGCCTCTTCGGGGTAATCAGATACGCGGCGCAGCAGGGGTCTCAGGGC
CACAGCCAGGGGGGCGGCGGGAGACATGCGGAATCGCAGCGGAAGGCGGGAGGCAGCTGT

GAACTGTGGCTCGGCCTGCGTCCGCCCTGCGCATGTACACTCAGAGAAGATGATAATGAA
AAAGAAAGCAAATCCAATTTTCCCACTTACTGTTCATATAATACAGAGTCCCTGAGAGTC
TAGAGTAATGTCTCATACAAAAAAGAAACTCCTACGTGGTGTGTGTCTGAAGTCTTTCAT
CTGCCTTACAGGGTTTTTGCTGATTCAGGCTTGGCCTGAAAACAGGACGGACCTCCATGC **exon 11**
CTTTGAGAACCTAGAAATCATACGCGGCAGGACCAAGCAACAGTAAGTTGACCACAGCCA
AAGCCTGGTAGATTACATTTGCCTTTTTAGTTGGAAAttaggcttaacaggagagttgct
aagatagggcacagagctcctgcatctctcgccggcattcccaaatgctatctcacatga
gcaggcacagggagcaagactgcacgaccactggcacaggctgtccgctaaaccacagac
ttctcagcgctcgccagtgcttctgcttctgtgtccactccagatcccacattgcactta
gttgtcaaatcttttcagtccatttctaacctatattagctcctgtgtctttccttgtct
ttcacggccttgacacttacaaaacgtgtgggtcaggtactttgcacactgtctaaccat
gtctgttcagctggtgttttctcaggatgcaattgaggttatgCACATCTTATCACAGGG
ACCAGAGAGACTTTTTAGCACCACTCTTCAAGAATTTCCACTTTTTCAGCTTTGACAGTG
GAATAGACATGCAGGTGCTCACACACAAGCATCTTTAATATGGTAATGGTAATCATCAGT
TTAGTGGTGTGGAGGAGGAGATGGGAATCTCTTAGTGAAACCCGCCTTGGAAGCAGCCTC
GTTATGAGAACTGCTGCCCCTACTTGACTCTTAAAGCACTAGATAATACTGTGCAACATT
AAAGAGAATAAGAGTGCGTGAAATATGCATTGCCTCCCATAAACTCCCTTGGCTCTGAAT
CTCTGATACTAAATATGTGGCTACCGTTGCTTCCCAGAAAGGCCTTTTTGCTCTGAATTC
TCTGGAATGCTTTCTTTGACCAAGATTCTTATAAAAATAAGAGATTTAGAGCAATTTTCT
TGGATGGCTGGTATGAGCCAGTTGGCTTAGTTGTAGGGATTTAAACAAGATAAGGGTTAC
TTACTTTTCACATTTAATGAGAAGTCTGGTGATTCCAGCTCCTACTGAGACAGGGTGGCC
ACACGTTCCAGGGTGTGACTCACTGAGGCCCCAGACCTGCCCTGCAAGGAAAACCTGGCT
CTGCCCTGGTGTCCTGGCCTCCCTGGGCATATGTGGGGGAGAATTCCTAATGGTATTGGT
TACAGGCTCCTATGCGAGACCACTCATCTGTGTAGGAGAAAGGAAAAAGATGGGGGAAAG
AAGAGCAGCAGGGAGAGGAGAAGCCTCTGGATGATACTCTAACCCCCTGCCATCCAACAC
CTGAACATCAGTCTCTTCATCCAGTGCTCTCAGCTGGCCCAGCCCCCAGCCTGGGGTCAG
ATGAGAGCTTCCTGCAAATGCAGATCTCTTTCCTGTGGCTCCTTCTCAATTACAGACAGC
TCCTCCACAAGGTGCACTCTGGCCTTGTGCTCCCTCCCCAAACCAGCCCAGCCCTCCCAG
CCTGCATCATCGTGGTCCTGTAGGGGCTAGAGGTTCTCACACCCATCGTGGTCTGGCAGA
GGCTGGTGGTTCTCACACCCATCGTGGTCCGGCAGGGGCTTAGTGGTTCTTATACCCATC
GTGGTTCAGGAGGGGCTAGTGGTTCTCACACCCATCGTGGTCTGGCTGGGGCTAGTGGTT
CTCATGTCCACCGCGTGCTTTCCTGCTCCTCCAGGTGGCTGAGGACATCCCCCCTTCGGT
CTGAATGACTTCCATCCAGTCATCTGATATACACATTGGACCACCCAATAGCATCCTAGT
GTCATGTTGGATGGTGAAGAAAATGCCACAGTTACTGCTTTCAGGGCCTCACAACCTTGG
GCATAGCTTTTTGGAGGAAGGCCCCACTTCCCAGGCATCCCTCCCAGACCTGGTCAGAGG
CCCCTGCTCTTTGCTTCCATGTTGCCCACACTCACTGTGCTCTTCACACCGGCTCAAAAT
GATCTGCTTACGGGGTTGTGTCACCACCAGATCAAGCGTCCTGGAGAGGAGGAAACATAT
TTAACCTGCACAGAATTTGGGACAGAGAACCTCTAGTGTTTGTTCAATAAATATATGAAT
GGATAGAGGGACAGGTTGGGTGGTGGATAGATGGATGAACCCACACCTTTGAAGTGTATT
TGGCTGTTTGAGAGGTTAGAATATGTTCTCAATTTCCAGGCAAAATGAAAATGGAGAAAA
TATAATGACATTAAGGCATTTTATTCATCCTCCCCATCTGCCACTGGGTTAAAGATACTA
AATAAACAAGGAACTATCTTTTGCCTGGAGGAACTTTAAAAACACCTGCAGTTTTCAAAA
GGTGCAGTGTGTGCCTCCCACAGCATGACCTACCATCATTGGAAAGCAGTTTGTAGTCAA
TCAAAGGTGGTCTGGAGAAACAAAGTTTTCAGGGATACATTGTTTTTATAATTTTTCACC
ACATGATTTTTCTTCTCTCCAATGTAGTGGTCAGTTTTCTCTTGCAGTCGTCAGCCTGAA **exon 12**
CATAACATCCTTGGGATTACGCTCCCTCAAGGAGATAAGTGATGGAGATGTGATAATTTC
AGGAAACAAAAATTTGTGCTATGCAAATACAATAAACTGGAAAAAACTGTTTGGGACCTC
CGGTCAGAAAACCAAAATTATAAGCAACAGAGGTGAAAACAGCTGCAGTAAGTCACCGCT
TTCTGTTTAGTTTATGGAGTTGGTTCTAATGGGTCCTTTATTTGTATTTAGAATATTGAA
GGGCTATTCCCATTTAAATTACTTTTTTCAGTTCCTTAAGAAGCAAATTAAAATCTTAAG
ATTCCTAACTGTGAAATTACCATGTGAATTCCATTAAAACTTTTTCCAGATCATTACCAT
TCAATGGGATGAATTTACCCTGAGGTTTAGGCTACCAATTATTTGTAATGTAAGTAACTA
AATTTAGTATTAGTTATATTACCTTTTAGTTGTAGGTCACTCTCTGCTCATTTCAGCCTG
TAAAGACTACAGCTACACACATACACACACAGAGGGAATGGAATGAGCACTTTACATCAAC
ACTTCCTGTTCTGGCTCTAGAGCCTCAGCTTTTGAAGCTGGTGAGAGCCTGGCCTGTGCT
GGGCCTTGGCcacgggcagcgtcagctttgagtcaagtgctggtctggcctccctagctt
tgagcctctgtcaattccctttaatctgtttaggctttggcttcctcatccatagaatgga
gatatgaatgattcctacgccgtagtgctttgagagaattcagtgaaattcctgtgtgta
aaacccttccatggtgcctagcacacagcacacagccaatGGCCCAATGGCTCCTATCAG
CTGTGGGATTTGTCATCAGAACACCACCAGCTCTGCTCCAGGCTGCCCTGGGTACCATCA
AAACACACCCTGTGCCCAGCAGCACCTGCTCCTCTGCACACCTGGTTCCTTCAGCAGGGG
CAGTGGCCGTGGGAGCACAGAAAACATGGAGTCCCATCTGGTTTAATTGATGCCATTGCC
AAAGGGGAGGACTCACGGCACCCCCTCTCGGGTGCCAGGGTGCCTGGCTCCCACCAGGAG
GAAGACCTGTCCTCCACTGTCAGGCACATTTCAGTCTTCCCAGCAGCCAGCACAACTACT
TTGTCCTTCCAGTCACGGTCGGCCTCTGGGAAGCCCAGTCTGTGTCCTCCTCCTTCAGGG
GTAGCCAGCATGTCTGTGTCACCCAAGGTCATGGAGCACAGGGCCCCTCCCGGGAAGGTG
CCGTCTCCTCCGGCCCCTCGGGTCCCTGCTCTGTCACTGACTGCTGTGACCCACTCTGTC
TCCGCAGAGGCCACAGGCCAGGTCTGCCATGCCTTGTGCTCCCCCGAGGGCTGCTGGGGC **exon 13**
CCGGAGCCCAGGGACTGCGTCTCTTGCCGGAATGTCAGCCGAGGCAGGGAATGCGTGGAC
AAGTGCAACCTTCTGGAGGGGTAGGAGGTTATTTCTTTAATCCCCTTGCGTTGATCAAAA
ATAAGGCTCCAGGTTGTTGTTATAGCTTTACAGGCATTCTGTTTGATTTTCTCTTCCTTT
TATTCTTTGCCCTTGGCTTTTGGAGGTTTTCTGTGGGGAGACGGGAAGTTGTT
TGATTGCGTTATTTTTGGCAAATTTAAGCACAATAGGAAATAAGCAAGTATTATTGCCTA
ATATAATCCAATAATTTATAGAATCTCTTTTCCTGGAAGTATCTTAAATTTTTCTAAGCT
ACAAAAAGTTCCTAAGACAAATGAGACAGTCATCAATGGTTCATCTAGCCAACACCGTGG
CCATTTGGGCTTTTCTTTGTAGTGCCCGATTCCTGGTGTGTGAAAATAAATTAACACAAA
TTATATTGCCAAGTTAATATCTGTTTTATGTGCCCCCAGCATGTGTTGAACATCAAACAG

TACCAGGGACTTTAAATATACCCACGGACAAAGAAATAATTCATAATGATGTTTGTTGAA
TTTAGTTGCAATCAATAAAAAGTGCAGTTTGTGAATGCTCTGAGGTTCTTGATATTGATG
TAAGGCTTTGAACGACAAATGAGGACAAAACATAAATAGGAAAGTAAAACTGAAGGATAG
AGGCCAAGGCCATGTTTTAGAAGATTTAAAGAAAAAGGGAAATTTGGTGAGCACCATAGG
AATTACAGATGGCTGTAGGAATTCTTCCTGTTTTACTCTCTGGGCATGGACCACAGCTTG
GATCCAGAAATATTTAGGAGCAGGATAAGAGGACCAAGTTCAATTCTATAGGAATCCTTT
AGCTGATAGGCTCAGAACAAATCACATAATTGATAGTGCTGCTTCAACTTCAAGTAAGGA
ATATTGATGCAATCCTTACAGCTACAAATGGACAGTGGTCTCATGTTTTCAGTTTTCAAG
TGTTTCTTAAGAGGCAAGGTGATGAAAACGCCCACGTGGGGAGCCCCATGTCCTTCCATT
AGTGTAGAGAAACCTGGTGTCCAGCAGCACCTGCTCCCTCTGCAAGCCCAGCCCCCTTCA
GCAAGGGCAGTGACCCAGAGAAGAAGCACAGAAGACACAACCCTGTATCACATTTGTTT
AATGGTGCCATTGACCAAAGGGGAGGATGAAAGGCACACACttttttgttgtttttttgag
acagagtctcacgccatcacccaggctggagtgcagtgatgtgatctcaactcactgcaa
cctctgccccctgagttcaggtgattctcctgcctcagcctcccaactagctggaattac
aggtgtgcaccaccatgtccagctaatttttttgtagttttagtagagacggggtttcacc
acgttggccaggctggtctcaaactcctgacctcaagtgatctgcccgcctcggcctccc
aaagtgttgggattataggcataagccactgcacctagccAAGGCACACACTTTGGAGAA
TAAACACTCCTTGTTCGCTGCTGGAGGGTAGAACTATGCTTGACTACTAGGCAGAGTCCA
GTCTTACTGACAAACAGCCGTACATCTGTTCTGTCTTTTCAATCAAACATCAGCTTCTTG
CTTAACATTGATGTGTACATCTTGAGGGATGTCAAAATATTGTAAGCTAAGTTTTTCATA
CCTGTGTTCCACACTCACCATTTTTAGTAATAACCATTGAGCGAGTTCATTCTCCCTCCT
TCCTTTTTCTATCACTTAATCTAAAATTATCATTTTTCCAGCTTAATTTTGATAACCATG
AATCTGGTATTAGAGGCAGGGAACACCTCCTCAGGACTATCTTTTCTTTTATCATTTGGC
TTGCTTACCCAATATGCAAAAACTATGCTGTAGAAAAAGCAGAAAAGATATCTTGATTAT
GAATGAAGCTCCTGTGTTTACTCAGAGAGAAGATGACCCAGGATTCAGTTAACAAAATCA
GCTGATTATATTACTATATAGTCCTGGAGTCCCAACTCCTTGACCATTACCTCAAGTTAT
TTGGAATTTTGAAGAGGTGATTTGTGTTCCTGCAATAATGTCTCAGGGGTGGGCTGACGG
GTTTCCTCTTCCTCCTCTCAGTGAGCCAAGGGAGTTTGTGGAGAACTCTGAGTGCATACA **exon 14**
GTGCCACCCAGAGTGCCTGCCTCAGGCCATGAACATCACCTGCACAGGACGGGTAAGAGC
CCCTTGCTGCTATCCACGTCCATTTCATGGGAAGGGCCTTCACAGAAGCCGAACAGTGAT
GATGGCCCAGGGCATCCTGTGTGGGCAGGACGGCCATCAGAGCCACTTCCCAGAGGAGAC
GGCAGGCGCTGACAGCGCTGTCCGGGCAGGGTGTCGGTGACATTAGCACACACATTAGCC
TGCGATGAACATTCACTCTTTCTGCTGACACCCCCAACCTTATCTAAGCTTATCAAATCC
TCACATTTAACGGAGGCTGTTTTCACCTGGTTTCCCCCATCCCTGACCTAGTCAGCATTG
CTTTATCGCTTTCATCAAACATCCTCAAATTCTTAACATTAGCTTGTAATTAATTGAAGA
ATTTTTAAAGAAATTGCTAGCAAAACTTTTTAAACTGCACAACTTTGTATCTATATGTTC
AATAACATATAGATACAATATTCTTTACAATAATCTTTTAAAGAATATGAGTGAGAATTC
GGGCCCCTCTCACACCAAATGTCCTGATGTTGTTAATTCTCAATGTTATTATATAGGGAG
CTCTGTTTTCTTGTGAGCTTCAACAGCCAGTTCTAAATCTACTAACTGAAAACATTTTTT
AGACATTCTCTAAATTGGGCAGAAGATGACAGGACTGTGTTTTGAGGGATAGGCTGCCAG
CGTGGCTGCTTACAAAGTAAAGACTTGGTTTATAGGTTTGCATGGTGTTGGGTTAAATTT
CTGTCATTAAAATAATTGGCGATATTGACATAGTCATCTAATTATGCTGGCTCTGGGCAC
ACACAGCCCTTGAGTGGACAAAACCAACATGAGAGAACTTAGCCAAGGGGAAAGCCTTTC
CCTGCTGGTTTATTTCTGCTACTTCTGAAGTGTGGGGCACACAACCTGAGCAGTGCTTT
TATTTGAGTCCCAATGCTTTTATTTGAGTTTTGCAAGGTTATTCCAAGTTTTACAAATAG
AAGGTAGCGTATGACTCAGTCCTTGATATGCCAACCACTGCACAGAGACTTGCCACCTTC
CTGTCACTGGAGAAACACTCATGTGGGTTTTCTTAAATTTGCCTCCCTCTGAGCTTCCCT
TTAACTTCAACTATAATATGCAAGAAAGACTATCTGACCATAAATACACATTTGGGCCAA
TCAAGATGGTTTTGCCAAGGAAAGATGCCCACAATGGTTAAGCAGAATGCAATAATGTAG
AGAATATCATTTCTTTCATGCTGGTGTATATCATATGCATTCAAAAACAGGGAGAACTTC
TAAGCAACTAACAGTGACCATATCAAGCAGGTGCAATCACAGAATAACTGGTTTTCTCCT
TTAAGAATTTTTCTATCATTTGGCTTTCCCCACTCACACACACTAAATATTTTAAGTAAA
AAGTTACTTCCATTTTGAAAGAGAAAAGAAAGAGACATGCATGAACATTTTTCTCCACCT
TGGTGCAGGGACCAGACAACTGTATCCAGTGTGCCCACTACATTGACGGCCCCCACTGCG **exon 15**
TCAAGACCTGCCCGGCAGGAGTCATGGGAGAAAACAACACCCTGGTCTGGAAGTACGCAG
ACGCCGGCCATGTGTGCCACCTGTGCCATCCAAACTGCACCTACGGGTGAGTGGAAAGTG
AAGGAGAACAGAACATTTCCTCTCTTGCAAATTCAGAGATCAAAAATGTCTCCCAAGTTT
TCCGGCAACAAATTGCCGAGGTTTGTATTTGAGTCAGTTACTTAAGGTGTTTTGGTCCCC
ACAGCCATGCCAGTAGCAACTTGCTTGTGAGCAGGCCTCAGTGCAGTGGGAATGACTCTG
CCATGCACCGTGTCCCCGGCCGGGCCTGTGTTGTGCAATGCTGCACATCACAACAGGAGG
GTAGGGGGACAAAAGAGCACAGGTCCTGGCAGCTGCCACAGTCTCCAGGGGCTTTTGCGT
TTCTCTCCAGATTTCTAAGGTTAACATGGGGATTAGCTGTTTTGCAATGAATAAAAGGTA
ACATTGCCTGGAATGTTGCTTAAAGACACTTTTTTAAAGCTAGTTGATTGTTAAGCTGTT
GCTACTTAAATTAAAACTACTTTGggccagacgcagtggctcacgcctgtaattccagca
ctttgggattccaaggcaggcagatcacttgaggtcaggagcttgagaccaggctggcca
acatggtgaaaccccacctctactaaaaatacacctgtagtcccagctactcaggaggct
gaggcaggagaattgcttgaacccgggaggcagaggttgcagtgagccaagatctcgcca
ctgcactccagcctgagcaccaagagcgaaactctgtcgcaaaaaacaaaaacaaaaaaa
aaaGCTACTTTGACTGGAATTAGCAGAAGCACTCTGATTGTGTGTATCTTATTTACTGGA
ATAATAAAGCTGTCAATCAAACTGGATCCCACTCAACAATCAGAAAGAGAAGTTGAGCTG
TCATATAGTAGTTCACACTTACTTCTGTTTCTCAAAATCCTCAGCTTTGTTTGGAACTGT
TACTCattcttctctgaatccatctgtatgagttgtgtgcccttgggcaagggtcttac
cttctctgtgcctcacttttcttttctgtaaattgggataataatgctgcatagctcacag
gatttttatgaccatgagttaagatatgtcatatacttaaaatggtgcctggaaaatggt
gaatactgagtcaatgatagcatcattGATGGTGGGATGGTGATGAGGAGGTGGGAGTCA
CAAtggtggtgttgatggtggtgatggtggtgaggaggtgggagtcacagtggtggtggt
gttgatggtggtgaggaggtgggagtcacaatggtggtggtgatggtgttgatggtggtg

```
aggaggtgggagtcacaatggtggtagtgatgatggtgttgatggtggtgAGGAGGTGAG
AGTCACAATGTtggtggtgttggtggtggtggtgtgaggaggtggagtcacaatggtg
gcagtgttggtggtgaggaggtgggagtcacaatggtggtagtgatgatggtgttgatgg
tggtgaggaggtgagagtcacaatgttggtggtgttgatggtggtgatggtgatgAGGAG
GTGGGAGTCACAATGGtggtgatgagggtggtgatgatgatgaggaggtgggagtcacaa
tggtgtcagtgttgatggtccgatggtgatgaggaggtgggagtcacaatgttggtggtg
ttgatggtggtgatgatgatgaggaggtgggagtcacaatggtgtcagtgttgatggtgg
cgatggtgatgaggaggtgggagtcacaatggtggtggtgatgacggtgttgacagtggt
gacgaggcgggagtcacaatggtgtcggtggtgatggtggtgAGGAGGTGGGAGTCACAA
tggtggtggtggtgatggtggtgatggtggtgaggaggtgggagtcacaatggtggtggt
gttgatggtggtgatggtggtgaggaggtgggagtcacaatggtggtggtgttgatggtg
gtgatggtggtgaggaggtgggagtcacagtggtggtggtgatgagggtggtgatggtgA
TGAGGAGGTGGGAGTCACAACGTtggtggtgatgatggtgttactggtggtgacgaggtg
ggagtcacaatggtggtggtgatggtggtgaggaggtgggagtcacagtggtggtgg
tgttgatggtggtgatggtggtgaggaggtgggagtcacagtggtggtggtgttgatggt
ggtgatggtggtgAGGAGGTGAGAGTCACAAtggtagtggcgatgatggtgttggtggtg
aggaggtggaagtcacggtggtggcgatgatggtggtgAGGACGTGGGAGTAACAACAGT
GGCAGTGACGGTGATTGAGACATGATGATGATTTGTCAACTTTCTAGGAAAACAATCATA
TAATCTCCAACAGTGATATCTTAATATCTTTTCCAAAAGTATCAGATCAtattataaggg
ccaagtttccagaataatatcagacataatgacagtggacatcagagcttggcatctaaa
ggtaatgggaatagctctaatgtctcagcgtgaaaaacaacatttgctattagtctgaga
tactaattatctagttaaggaagtactcacctatacctagtttttaactgtttttaaaa
tctggaattgattttgaattttaacaaatatttCCCTGGGAACAATGTAAGATTCTTCAT
ATTTTCGCCTTTGGGTATACCAACATGCCAGCTCTGTTGGCCACTTTGTGAGCTCGATGA
AGCATGGTATAAAAGATGCTTTGCTAGTGTTTCACGTAATCTATTTCTATAAGCAATTTT
GGAGCTAAGCCTCTGAAACAGAATTATATTATCTGTATAGAATAAATGTTTTATCTTCCC
CCTTTTCTTTCTTCTGGAATAGATGTGCATCAGTATCTCTGCATCAATATCTCTATATCA
GTATCTCTGTGTCAGTGAGCATATGTTGCTGGGCTTAGGGGAGGTCCAGAAAGTGATTGG
GTTTTGGCATTTTCAATACACTTACTTTGTATAAGAAATAGTTTGCCAAATATAGAAAGA
GGGGATTTAGTCAAGATTTAAATTAAAAATGTTAGTGGTCATTTTTCTAATGTCTTTCTA
TTTTTTCCCAGGTCCTAATAAATCTTCACTGTCTGACTTTAGTCTCCCACTAAAACTGCA exon alt.
TTTCCTTTCTACAATTTCAATTTCTCCCTTTGCTTCAAATAAAGTCCTGACACTATTCAT
TTGACATATGGAATTTTATAAATATTTTCTTTAGTATGTGTGATTACATTCCTGATTCTG
AGCCTTTTTAGATGAGTATATAGTTTGATATAATCTTGTTATTGCCACCTGTGTCTTCTC
CCAAAGCCATTAATTATATAGGAATTACACGATAGAAATGGGTTAATTTTAAAATACG
GCCAAGTGTTGATGAGAGGGAAAATTTTTTTAATTTCTTTCACTGAGTATTTATGACGTG
CACAACATTCCTGAATATATTGTCTCTCTCATTTCTCAGATGGGATGTATTGCCTTCTCC
ATTTCTATTGTTAAAGAAACACTTACAGGGGTTTCTTTAACAACTTGTGAACAGCAGCAT
CAGAGCCCAGACTACAGCATAAGCAGCTGCTGATTCCAAAAGCCCTACCTTCCAACCGGG
CAGGTGCAGCCACCCAGACGAGGGGGAGGAACCCTGGAGGAATAGCTAttctttttttt
ttttgtcgagacggagtcttgttctgtcaccctggctggagtgcagtgccgtgatcttgg
ctcactgcaacctccacctcccaggttcaagcaattctcctgcttcagcctcccgagtag
ctgggattacagacacctgccaccacgcctggctaatttttgtatttttagtacagacag
ggtttcaccatgttggccaggcttgtcttgatctcctgacaagtgatccacacaccttgg
cctcccaaagtgctgagattacaggcgtgagccactgcgccccagcAGGAATATCTATTTT
TAAATGGAACTGTGTTTTCATAGTACACGGTGAGGAGAAAGTTGCTTTGAAATCTTTATC
CTAATAAACCAAATAATATGAAAATTTGCCTATTTTAATTATATGTAACAAAGTTTAGTT
ACTGCTATAATTGCAAATATGTATAAATTCCTTACCAAAAAAAAAAGAATCAAGTGGGAG
CCAGAGAATAATTTTTCTGACAGAATTAAATAACATGCTATAGCTGCTTGAGTTCATACT
CAATAGTCATTTCTGCAGAGTTACCGAGGGCCTCATCAGCGTCAGCAGGAGCCCCTCGCC
TTCTGACGCTCTCACATCCTTCTCTCCTGCAGCCCCGTCCTGCCACTGTCCTTGTCCAGC
TTCTCTTCAAGGGTCAACTGGTCTACCTTTCCCTACAAGTCTGTCACAGCTTCTTGTTAG
CAATCCCTATGGTTGCCCAAAAGCATTTTCAGAGCCTGCATAAGACTGCATCTTGTAGAA
AATTTGCAGTTTCAATCTGCCCTCCCTCTGCCGGGTGTTCCCATTGTATTGCATTCAGCA
GGCAGGGAGAGACTGCTATTAGGTCTGTTCCTGAGTGACTGCTTTCTGTCTCAGACTGTT
TGGTGTCTGTAGGAGGTAGTGGGGTGGGCAGTAACGAGGTCTCCTGTATATTCCACCCCT
ACGAAGCCTGTGTGTTTGGTTTATGAACTAAGCTCAAAAGCACCACAGGGGTAAGACTGC
AGTACATGACACCATGGAAAAGAGGGAGCACCCAGACCCCCAAATTAAGAAGAGCAGTGT
AGAGAACAGAGACCTGGAGAGCAGAGATAGAAACTGTTAGGATCAGATTATAGTGTTACA
CCAGGGCTCCCCAGGCCTCTCACATATTGAAATGTACTTGTCCATCTTTCTCCAGGCCAG exon alt.
GAAATGAGAGTCTCAAAGCCATGTTATTCTGCCTTTTTAAACTATCATCCTGTAATCAAA
GTAATGATGGCAGCGTGTCCCACCAGAGCGGGAGCCCAGCTGCTCAGGAGTCATGCTTAG
GATGGATCCCTTCTCTTCTGCCGTCAGAGTTTCAGCTGGGTTGGGGTGGATGCAGCCACC
TCCATGCCTGGCCTTCTGCATCTGTGATCATCACGGCCTCCTCCTGCCACTGAGCCTCAT
GCCTTCACGTGTCTGTTCCCCCCGCTTTTCCTTTCTGCCACCCCTGCACGTGGGCCGCCA
GGTTCCCAAGAGTATCCTACCCATTTCCTTCCTTCCACTCCCTTTGCCAGTGCCTCTCAC
CCCAACTAGTAGCTAACCATCACCCCCAGGACTGACCTCTTCCTCCTCGCTGCCAGATGA
TTGTTCAAAGCACAGAATTTGTCAGAAACCTGCAGGGACTCCATGCTGCCAGCCTTCTCC
GTAATTAGCATGGCCCCAGTCCATGCTTCTAGCCTTGGTTCCTTCTGCCCCTCTGTTTGA
AATTCTAGAGCCAGCTGTGGGACAATTATCTGTGTCAAAAGCCAGATGTGAAAACATCTC
AATAACAAACtggctgctttgttcaatgctagaacaacgcctgtcacagagtagaaactc
aaaaatatttgctgagtgaatgaaCAAATGAATAAATGCATAATAAATAATTAACCACCA
ATCCAACATCCAGACACATAGTGATTTTAATTATTTAAGAGTAGTTTAGCATATATTGCT
TTATGATTTAATTAAAAATCTCCAAAATATATGCCAAAGAAGTAGAATGAGAAAAATGTA
TATTTCTCTTTCACTTCCTACAGATGCACTGGGCCAGGTCTTGAAGGCTGTCCAACGAAT exon 16
GGGTAAGTGTTCACAGCTCTGTGTCACATGGACCTCGTCAAGAATGACCACACTGCTGTG
GGTGAAGATGCTTTCCTGCATTTCTGACTGTCCTCTGTCCTGATCAAGTTTCTATGGCTC
```

```
TGGGCCAGCCTACCCTCAGCCAGGGTTTCTGCAGAGACTGCCCAGCTGGTTCCACGTGGC
TCCACGTGCCAACTTTGTCCTCAGTGGAGGGAAAGTTGGACACACAGTGCTGGGGCTGCT
CCCTGCTCCGCCGTTGCTCGATGCATGGCCTGCCTCTGAATTCCTTGGTTCCACTGGTTT
TGCTGGGTCCTTCTGTGCCTCTAGCTCCTCTTTTTTCTGTCCACTTACCCCATTGGTCC
CATCACAAGCCTGTGTGTGAGTGGCCTTTCTGTTCGATGACAACCTCCAGCATAGGGGAG
TGTTTCTCCTTGCTTTCTTTCCCAGACACACTGCCCAGCAAAGGCAAAAGGGCTTCCTTC
AACATCAGCTCTGGCCAGTTTGCCAGAGCAAAGCCCTGAGAAAAGCAAGGTTGAAAAGTC
TTATTCAAACTCACCAGGAAAGAGTGGTGTTACTCTCGATGGCGTCTAGCCAGGAATCAT
GGAATTATACACCGAGCACCTGTTTGCCATTTTGGATGTTTCCAAACATGAACCAAACTT
CCAGGCCCCTCTGCCATCTCTGGTAACATTTACAAAGTCCCTTCCTCACCACTGCCCTTC
CTTCATTTTGGCATGCTCCTCCGCCCCCGAGTTGACAGCCATAGCTCTCTCCTGCCAC
CAGTGTCACATGATCGAGGAAGAAGGCAACTTCAAAAAGACTGGGTCCCCTTCCACTCCC
ATCTCTTCAGTGAGCTGCTAGGACACCCAGCAGAACTTCCCCACTCCACACTGCAATCTC
AGGGATCTTAGTCACGGGGCTTTCCACCATGTCTCCACCTGGAAACCAGTCATGGCCATT
CCTTCTTACATCTGCTCTTTTCCATCTTTTTCTTCTCCTCCTGTTCACCCGCCCTTACTC
TTGTGGCGCCCTATGGATATGCGCTCCATAGCAAATGATTCTTTATATCTTACGGTATTC
TAGTGAGCTGGCACATGTGGCTTCTGGTTTCCTCTCTCTGGAACTAGACATGACCTCTGT
GGGAGGGAGGATTAAATGCACCCTACAGTCTGAGGCTGCATGATGACATCACTCATCACA
ATGATGCTTTCTATGTCTGAATCCTATTCCTTTATAACCCCTTTCAAGCTCGTTCAGAGA
GTATTTCACACAATCCATGTGCTCATCTTAAAAGCCAAGGACCCAGAGGAGTCTCAGCAT
TGCCAAAAAGTCCCTTCACCCAGCCTGGCCAGAGGCAGTGCCTGGTCCATGTGTATGGAC
TATGGCACTTCAATTGCATGGAAATACTCTTGGAATGAACAAAATACCAATCCATGAAAA
AGCATTATTGAAGTCTAAGTTATTTTTTGAATCATATTTTGTTAATCAACAAATTGAAAA
ATACTCATTATATGGAGAGGTCCAGATAAAGCCTCAATTTTAAAAAATGAGGAAAAGTGT
GCCTGGTAGGGGACTGGGGAGAGCTTGAGAAAGTTGGAAACGTTGCCTTAGAAGCCTGTT
TTTTCTCCTTTTAGAAGCTACATAGTGTCTCACTTTCCAAGATCATTCTACAAGATGTCA
GTGCACTGAAACATGCAGGGGCGTGTTGAGTGCCAAGGCCATGGAATCTGTCAGCAACCT
CACCCTTCCTTGTTCCTCCACCTCATTCCAGGCCTAAGATCCCGTCCATCGCCACTGGGA exon 17
TGGTGGGGGCCCTCCTCTTGCTGCTGGTGGTGGCCCTGGGGATCGGCCTCTTCATGCGAA
GGCGCCACATCGTTCGGAAGCGCACGCTGCGGAGGCTGCTGCAGGAGAGGGAGGTGAGTG
CCAGTCCTGGGTGGGCTCAGGAGCCCTCGCACCCCGACAGGAACAAGGGCCAGCCCCGAG
AACGGGCCATTAGCAGTTGTGTATGTTAGATACATAATTGTATTATGATGCAGAAAGAAT
CTCTGAATGTGCAGTTATACCCAGTTGGTGACATGTTGGTACATCCATCCGAGGAAATGG
CAATGTTTCTAGGCTGCACCCTTCAATGTCCACAAAGCTGTGTGGCATCTGCTTAGGACC
CGGTGCCTGTGTGTGCATAGGAGGGAGGCCAGGAAGCCTGGCTGTTGATCCCATGCTGGC
ACTGTGGCGAAGGCGAGAGATTCCTGCTTTGGAAAACACCATTGTCCACACAGTGGCTTT
GTCCATGATGGACTTCGCCACAGCCCAGTCCTGTGCTGGAAGCCATGTTCTCTGGAAAGA
GCAACCCAGCGGCTCATAAGCATAAGCGCGTGTGATGTGCCCCAACCAAACGACCGCCAT
GCACAACTTCCCTACCGGAGTTTTCAATCCAGTTAATAGGCGTGGAAACAGACATAGAAA
TTGTGTTTGTTGAAAGGTAGCTGTTCAGTTAAAGAACACCTGTATCAGAGCCTGTGTTTC
TACCAACTTCTGTCAAGCTCTGTAGAGAAGGCGTACATTTGTCCTTCCAAATGAGCTGGC
AAGTGCCGTGTCCTGGCACCCAAGCCCATGCCGTGGCTGCTGGTCCCCCTGCTGGGCCAT
GTCTGGCACTGCTTTCCAGCATGGTGAGGGCTGAGGTGACCCTTGTCTCTGTGTTCTTGT
CCCCCCCAGCTTGTGGAGCCTCTTACACCCAGTGGAGAAGCTCCCAACCAAGCTCTCTTG exon 18
AGGATCTTGAAGGAAACTGAATTCAAAAAGATCAAAGTGCTGGGGCTCCGGTGCGTTCGGC EGFR_e18
ACGGTGTATAAGGTAAGGTCCCTGGCACAGGCCTCTGGGCTGGGCCGCAGGGCCTCTCAT c.2155G>A,G719S;c.2155G>T, G719C;
GGTCTGGTGGGGAGCCCAGAGTCCTTGCAAGCTGTATATTTCCATCATCTACTTTACTCT c.2156G>C,G719A
TTGTTTCACTGAGTGTTTGGGAAACTCCAGTGTTTTTCCCAAGTTATTGAGAGGAAATCT
TTTATAACCACAGTAATCAGTGGTCCTGTGAGACCAATTCACAGACCAAAGGCATTTTTA
TGAAAGGGGCCATTGACCTTGCCATGGGGTGCAGCACAGGGCGGGAGGAGGGCCGCCTCT
CACCGCACGGCATCAGAATGCAGCCCAGCTGAAATGGGCTCATCTTCGTTTGCTTCTTCT
AGATCCTCTTTGCATGAAATCTGATTTCAGTTAGGCCTAGACGCAGCATCATTAAATTCT
GGATGAAATGATCCACACGGACTTTATAACAGGCTTTACAAGCTTGAGATTCTTTTATCT
AAATAATCAGTGTGATTCGTGGAGCCCAACAGCTGCAGGGCTGCGGGGGCGTCACAGCCC
CCAGCAATATCAGCCTTAGGTGCGGCTCCACAGCCCCAGTGTCCCTCACCTTCGGGGTGC
ATCGCTGGTAACATCCACCCAGATCACTGGGCAGCATGTGGCACCATCTCACAATTGCCA
GTTAACGTCTTCCTTCTCTCTCTGTCATAGGGACTCTGGATCCCAGAAGGTGAGAAAGTT exon 19
AAAATTCCCGTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAACAAGGAA EGFR_e19+ c.2235_2249del15
TTCCTCGATGTGAGTTTCTGCTTTGCTGTGTGGGGGTCCATGGCTCTGAACCTCAGGCCC EGFR_e19
ACCTTTTCTCATGTCTGGCAGCTGCTCTGCTCTAGACCCTGCTCATCTCCACATCCTAAA
TGTTCACTTTCTATGTCTTTCCCTTTCTAGCTCTAGTGGGTATAACTCCCTCCCCTTAGA
GACAGCACTGGCCTCTCCCATGCTGGTATCCACCCCAAAAGGCTGGAAACAGgcaattac
tggcatctacccagcactagtttcttgacacgcatgatgagtgagtgctcttggtgagcc
tggagcatgggtattgtttttggtatttttttggatgaagaaatggaggcataaagaaatt
ggctgacccttatatggctgggataggggtttaagccccttgttatttctgactctgaaac
ttgcattcaattcactccaccaagttatctcATCTTTGAAATGGCTTTTTTTAAAGGTGC
CTAGAATATGATGGCGTGCAGTCTATAAACTGTTGCCCACCTTCTGTACTTTCTCTCAGA
ATAATTCACATTCTTCTCCAGTGTCTGTTGATTGTTACTTTGTTGGAATAAGTTCTTGGAA
AATTCCACAAGATTATTGTTATCTTCTTACTACCAATTCTATTGAACTTTCTCCACCTTC
TCTGGGCCTTCCCCAGCCAGTGGTGGGAAGATGCTGGCTGGAGTCTGACAGAGCCTCTTC
TACACTGGCCTGGGCTTGCTGTGAGTTGGTGGAAACCTTTGCTCTTGTCCCAACACAGAG
CAAGTGAAAGAGGAGGTCAAGGGGCTCAGGCAGCGGACTAGGGAAGCAGAATCGAGGAAA
AGGAAAAATGGCTGACTTATTACCTCAAAACTCTAGAGAATTTAGTTGATCTTACAGCCA
AGAAGGACAAAAGCCAGAGAGTAATATCCTCCGCCTCATGTCTAACCCACAGAATACATA
GCAAGTAAAGAGAACATGGGCCTTTATAAAAATGTCTTAAGATACAATTTTTTAATTGGA
GGAAATCTACAGTTTAATTTTCTCTGGGCAGCTTTTCTTCCTTTTATTATAGTAGGGGAA
ATCCCATGTTGATATACTTCTAAATGAAAGATGATGAATTGATATAATACAATAAAAAAT
```

CTGTAAAATTGATGATATACTTATCAAGAAAAATTAGCTTTCATTTTAACGGTTTACAAA
TTGAGTCAAGTCCTAGTAACAAAATGTTAAGTCTATTAACATAACCACAAGAAATACAGG
AAGACGGGCAATCTGTGAAGCCTTTCACTTACAATCTCTGGCCCCTCACCTGTGCTGTGT
AGGAAAATCTTTGTGCACAATTTGCTTCCTTAATTCATTTTTTATTCATTCAACACATTC
TAATAAATTATACAAAATCATGTTGAAATGTGAATTTCAGTGGTATTTATAAATGCAGTG
TGAGGAGGGTTTGGATGTATTCTAAGACAATAGTTGTGCTTTGGGAAGGAAGCAGTGTTC
ACTGAAAAGTGCCCCCAGGACCTTTTAATTGGAGGAAATATGCTTCTGTGGAGTTGGAAA
TGGGGTAGAAGATAGATAAGGTCAAGGCTTAAAAGTTAAGTGCACCCAACATCTGAAGCG
TCCATGggcctggcatggtggctttcgcctgtaatcccagcactttgggaggctgaggca
ggaggatcccttgagcttaggagtttgagaccagcctgggcaacatactgagacccagtc
tctacaaaaaataaaaaattagctgggtgtggtgtctcatgcctgtagtcccagccactc
aggagatgggaagatggcttgagtccaggagatctaggctgcagtgagctaaaatctcac
cactgcactccagcctgggtgacaaagcaagaccctgctcaaaaaaaTAGTTAGATATAA
ATATTAatatagatacctatatatatctgaatatagatatctatatatactctgtatata
gttatttagatatataaatatatatgatatatatttagagagatatatatttagagagat
atatatttagagatttatatatatttatatatatttagagatatatatctctaaatata
tatctctctctaaatatatatatatctctctctaaatatatatatatccctaaatatatt
aaataaataAAAGAAATAAAAGAAAGCTCAGTTTGGCCTCCTGCTTGTCCTGTCTCCTCA
TCCCCTCTTCCCCCTCCATCATTTTATTTCCTTGCCCCATGTTTCTTCACTGCGGCCATG
TCCCCCCTCCTCTCCAATGATGGATGTCATGTCTGCTGCAGTCAGAGGGCGACAAGCCTG
GAGTGTTCCCTGAAGCCTGTGGTTTGTGGTTTGTCCTGCAGCTCAGGCTGCCCAGGCCTC
ACCAGCAATCCTGGCGGGCAGGGCACCACACTGGGATGGAGAGGGGGAAGCTGGAGGAGG
CACTTTCTGGTAAAGAAAGCAAAAGCCAGCAGTGCCCAGGCCAATTTCAACAGGGAGTTA
AATAGCACCTTAATCCTGTGGCAGGACAGCTCATGGGGCCATGTGTGCTCTTAGAAAGAC
TCACATGCACGCATGCACGGCAGCAATGACTCCATACTCACGTTCCCCTGCAGACACCAG
GCCCCCACAGCCGGCACACACACTGCAGCCCCAGTTCCATGTTGCTAGCAGTGGCTTAGT
GAATGAGTAAAGTTCTTAAAATGCAGGGGACACCTGCCCTTCATTCATAAGGCTGGACGT
ACACCTCTCCTTAAGGAGTTCAAGAGCTAGTGGAATCCCAATTCATACGGTAGAGCCATT
CACAGATGAGAGAGACAAGCCAGAAGGAAGGAACCAAAAGTCATGTCAGCAGTTAGGACA
AAATAACAGGCTTTCAAGGTCACAAAGCCTCAGGGACACTCCTGCGGTGGGACTGGGCTA
GGAGCCATGGGGGCTCCAACTGTGCGCTCTGCCTGCCAGCCTGTGGGTGCTGGGGCTCCA
CGAAGATTGTTGTGGAATACCAAGCATGCTTGCTGTAGGTCACGGTGCACGTTTACTACT
TCCAAGACAAACAGCCGAGAACAAAGCTCGCTTTAGCTTCTGCGTACACCGAACGGGACA
CACGACTGAACAGCGTTCCCATTGTGCCTGCTGGGTGGGGAGGAAGTGATGGCCCAGTGG
GTCTATCAGATGTTAGTAGGATGGGGCCTGGCGGGGCTCCAGGCTCTGTGTGGCCGACAC
CCACGCCCCCCGCTCTGCTCCCCATTCCCAGCCCCAGGTCAGCCCTGCGAGGCCCTGCAG
CAGATGGGCTGCTCAAACTGCTCTGGTTTGCAGATTTTCTTCCCTCTCAAATGAATACA
ATATGTTTTCAAGTCTCAACCAGATCTTGAGAAAATAGGAAGAGCCAGAGGGTTTCTTTG
GTGTTATGGTTGTACAGCTTCCCAGACTCCGGGGGAGAGATGTGATTTGTGCTTTCTGGC
AATCCCATGGCGTATTAAATTTTCATAGGCTTTCCAGTTTAAATTTAGGGTAGGCAATGG
AAGGGAACGCAAAACAGATTTCTAGGTGTACTGTGTGTGTGTCTCCCACGTCTAAAGTCT
GTTAACTGGAGCACCCAACAGGCCCCACAGGCTGCCTTCACACAGAGGACCTGGGGCGCC
TCCGACCCATTGgggtgagcagtgggccatggagggagccagggtcaggagacctggttg
tgggcctgacctgaccctgctcagggtggcctcaggtgggccgttcacctcgtcagcctc
agcttaccctctgactacagtgacctcagacaaaatacgcttcctggccctgtccagttc
tgactttttataaACAAGCACTTATCCAAGTTAAAGGGATATTTTCAATATCTACTGAGT
CCACAGATATTAAATATCTCCTCTCTTCTTTAAAATTGTGGCATTATCTTTAGAATATAA
AAGGAAAATAACACACACTCTCCTTGAAAATAGAGAGCCTAAACACTCTGCAGGAAATAT
TTAAAGCTATAGTTTTTGTTTGTTTGTCTTGAATGCAAGTGGCCTGGACTTTGACTTGCT
TTGAGTCTTTGACCTTCATGACTTCAGTACAGTTCAACCCTGACAGTTTTGAAGTAGGTA
TGTGCCTAGATCTGCCCTAGTCCCTGCTGGAATGTTGAAGAAGCAAAGGTCCAGGCCCTC
AGAGCACTTGCCACGTACTTGCCAACAGATACGGGGCGGAGACTTGAGTCAACGTAAGAG
CAAGTGTGTGCCGGGTGATCCGACACTGCAGAGCGCCAGCTAGACCCTAAGCGTGTGCTA
GGGGCTGACCAAGCCGTTCTTTCCTCAAAAACTTGGTGGGGAGGGTATTTTTAAAATCAC
ACAAATATTTAAGTACAGATTATGATGACTGCCTCAAAGcagtggctcttcagcttcatc
aagcttcagagtccagagggtttgttcatatggaaggctaggcctgtctcctgcatttca
ccctcttggcctgggggcgggacccaagaatgtgtggctctaaaaggttcccaggcaatg
ctgaggctgctTTCTGAAGGAAAAACTGCAAGATACCAGGAGAGTTTCATTTAGATTGAA
GAGTCGAGGAAGGCTCCTCTGAGAAAGAGTCTGCTAAGGAAGGAGGAGGTGGGTTCTGGG
GACAGAGGTTCTCCCGTGGGTAAGGGTGGAGGGAAGCTCTCCTGGGGAGAAGGTGGGCAG
GAGGACCAGAGGCTGGAGGGAGGAGGGCAGTCAGCCTCGGGGCTTCCCAGGAACAGGGAC
GGCCAGGGCAGGGTTTAGGGCAAGGAAAGCGTGTGAGCATATTTGTATTTTAGTAAATAT
TTAcagtttgccctccatgtctgcagtttcatatccatggattcaatcaaccacaatgaa
aaacgttggggaaaaaaattgcatcggtactgaacatatacggactttttttcttgtcat
tattcctaaacaatacagcataacaattattcacatagcatttgcactgtattaggtac
tataggtaatcaggagatgctgtagatgggaggatgtctgtaggttacacacaaatgctg
tgccactttatatcaggggcttgagcatcctcacattttgatatttaagggaggtcctgg
aaccaattccccagatactgagggtccactgTCTGTGTCCCCTCGCCCCACCTTGCCTTT
GTCTCCTGTCTCCTATCTCCACCCTGCCTCCCGCCAGCCTGTTGCTCCTGACCTGCCCGG
GCACCCTGGAGCAGCACCCTATCTCAGAGCCTGGCTCAGTGTGTTCACTTCTGCAGAGAA
ACTAACTTGCCCAAGTCCACACTCAAAACATAGGCATTGCTGAGATGTGAAAAGCAGCTG
TGGATGCTTTCTGCTACAGTCTGTGTGTTCTTTTCCATATCTGAATAAAAGGTCACCACC
ATTTGTATTTTAAAGAGAAAGAGAATTTATGGGTGGAAATTGGGGATTCCCTCATTCTCA
GTCAGACAGAAAAGAGGGCCCCATTGTGTGCCTGATTGCAAATAAATTTAGCTTCCTCAG
CCCAAGAATAGCAGAAGGGTTAAAATAAAGTCTGTATTTATGGCTCTGTCAAAGGAAGGC
CCCTGCCTTGGCAGCCAGCCGGAATTAGCAGGGCAGCAGATGCCTGACTCAGTGCAGCAT
GGATTTCCCATAGGGAGCCTGGGGGCACAGCACAGAGAGACCACTTCTCTTTAGAAATGG

GTCCCGGGCAGCCAGGCAGCCTTTAGTCACTGTAGATTGAATGCTCTGTCCATTTCAAAA
CCTGGGACTGGTCTATTGAAAGAGCTTATCCAGCTACTCTTTGCAGAGGTGCTGTGGGCA
GGGTCCCCAGCCCAAATGCCCACCCATTTCCCAGAGCACAGTCAGGGCCAAGCCTGGCCT
GTGGGGAAGGGAGGCCTTTCTCCCTGCTGGCTCGGTGCTCCCCGGATGCCTTCTCCATCG
CTTGTCCTCTGCAGCACCCACAGCCAGCGTTCCTGATGTGCAGGGTCAGTCATTACCCAG
GGTGTTCCGGACCCCACACAGATTCCTACAGGCCCTCATGATATTTTAAAACACAGCATC
CTCAACCTTGAGGCGGAGGTCTTCATAACAAAGATACTATCAGTTCCCAAACTCAGAGAT
CAGGTGACTCCGACTCCTCCTTTATCCAATGTGCTCCTCATGGCCACTGTTGCCTGGGCC
TCTCTGTCATGGGGAATCCCCAGATGCACCCAGGAGGGGCCCTCTCCCACTGCATCTGTC
ACTTCACAGCCCTGCGTAAACGTCCCTGTGCTAGGTCTTTTGCAGGCACAGCTTTTCCTC
CATGAGTACGTATTTTGAAACTCAAGATCGCATTCATGCGTCTTCACCTGGAAGGGGTCC
ATGTGCCCCTCCTTCTGGCCACCATGCGAAGCCACACTGACGTGCCTCTCCCTCCCTCCA exon 20
GGAAGCCTACGTGATGGCCAGCGTGGACAACCCCCACGTGTGCCGCCTGCTGGGCATCTG EGFR_e20 c.2303G>T,S768I
CCTCACCTCCACCGTGCAGCTCATCACGCAGCTCATGCCCTTGGGCTGCCTCCTGGACTA EGFR_e20-2 c.2369C>T,T790M
TGTCCGGGAACACAAAGACAATATTGGCTCCCAGTACCTGCTCAACTGGTGTGTGCAGAT
CGCAAAGGTAATCAGGGAAGGGAGATACGGGGAGGGGAGATAAGGAGCCAGGATCCTCAC
ATGCGGTCTGCGCTCCTGGGATAGCAAGAGTTTGCCATGGGGATATGTGTGTGCGTGCAT
GCAGCACACACACATTCCTTTATTTTGGATTCAATCAAGTTGATCTTCTTGTGCACAAAT
CAGTGCCTGTCCCATCTGCATGTGGAAACTCTCATCAATCAGCTACCTTTGAAGAATTTT
CTCTTTATTGAGTGCTCAGTGTGGTCTGATGTCTCTGTTCTTATTTCTCTGGAATTCTTT
GTGAATACTGTGGTGATTTGTAGTGGAGAAGGAATATTGCTTCCCCCATTCAGGACTTGA
TAACAAGGTAAGCAAGCCAGGCCAAGGCCAGGAGGACCCAGGTGATAGTGGTGGAGTGGA
GCAGGTGCCTTGCAGGAGGCCCAGTGAGGAGGTGCAAGGAGCTGACAGAGGGCGCAGCTG
CTGCTGCTATGTGGCTGGGGCCTTGGCTAAGTGTCCCCCTTTCCACAGGCTCGCTCCAGA
GCCAGGGCGGGGCTGAGAGAGCAGAGTGGTCAGGTAGCCCTGCCTGGGTGCTGGAGACAG
GCACAGAACAACAAGCCAGGTATTTCACAGCTGGTGCGGACCCAGAAAGACTTCTGCTTT
TGCCCCAAACCCCTCCCATCTCCATCCCAGTCTTGCATCAGTTATTTGCACTCAACTTGC
TAAGTCCTATTTTTTTCTAACAATGGGTATACATTTCATCCCATTGACTTTAAAGGATTT
GCAGGCAGGCCCTGTCTCTGAGAATACGCCGTTGCCCGTCATCTCTCTCCGACAGCAGGG
CAGGGGGTCCAGAGATGTGCCAGGGACCAGAGGGAGGGAGCAGACACCCACCCGGCCTGG
GCAGGTCCTCCTCATTGCTTGCATCCGCCTGGTTAGCAGTGGCAGTCAGTCCTGCCGAGT
CATTCGTGAGGCGCTCACCCAACTCCAGGCAGATGTAAAAGGTGACCTACAAGAAGACAA
ACAAAAACATCTGGAGCGCTCTTATGCCAGCATCTGCCCTTGACACCACCAGGCAGGCTG
TTGCTGGGAGCCGTGGTGCTTGGGTAAGCTCCTTCCCATGGCAGAGCTCCTGGGACGCAT
TGTAGAAGCAGGGACCACCTCCCAGGATAACCAGATAGCAGCACACCCTGCACAGCCCCT
TTTACTCCAGCATCATCGGGCATTGATATCTCAGCTGCAGCCACAGGCGGCCCCCAGCAC
CCCAGGAAGTGGGGAGCGCTCATGCTTCTCTGAGCACAAAAATCACTGAATATTTTTGCC
ATTCTCATGGTCATAACCCGGGCCACAGAGTAGAACACTCCTATCACTGTTGTTAGACAG
TGGTCCTGGGAGAGGGTCTTGTGTGCCTCGGATGCCAGGGCCTCTTTTTATTGGGAGGTG
CTTGTTATTTCTGTGTGTGGCtgcatttgtttcccaagactgccacaacaaatcatcacc
aacttggtagctcaacatagcacagctttattccctcctggctctggaggccaggtgtct
aaaaggccatgctcccacaatggttctgaggaggatccttcctgcctctctggcttctgg
tggctccagcatccctgggctgtggctgcacctccccatgtcaacctccgtcttcacaag
gccttttcctgtgtctctgcaaccacaggcccctctcttctcttaataaagataccag
tcatTGAGtttgaaaattgctaagagagtctgttgtaaatcttcttagcacaaaaaaaaa
tgacagatatgtgaagtggtagatatattaattagtttgatttgatcactccgctatgtg
tataaatgtcaaaacaaacattgcactccataaatatatataTTAAAaaagatcccagtc
attgcatttaggacccaccctaaatccaggatgatttcatttcaagacttttaactagat
ttgcaaaaccccatttccaaataaggtcacattctgcagttttgggTAGACGTGAAATGT
GGAGACACTGTGCAACCCACTGTCTTGGGGAGGGGGTGGTCAGCCTGGGGCAGATGTTGC
TGGGTGTGGAGCTACATCCACTCATGCCCTGACCTGGAACCCAGACCTGCTTCCCCAGCT
CTCCTCCTGGTTATCTGAAGCAGGGAATGGAGAGCACTGCCCTCCTTGCCCAGGCAGTCT
CTATCACCTGGTTTTAGTTTCTTCTTAGCACATATTGCCCCAGAATATCTGGTTGGTTTA
TGGCTTACTTGAGTTTGTGCCTACCTGTCCCAACCGGGAGGTGAGCCCTGGCTATTCCCC
AAACCCGGCCCTGCATGTGGGAGCTGCCCTTCCTCCGTTCATCAGAGGGGGCCAACAGTC
CACAGCTGTTCTTAATCATCTCCCAGTAACCCCCAGCTCCACAAAGGTGACTCCTTACAT
GGTGGAGAGGTGGTCGGGCCATCCGTGTGAAATGTGTATGTGACCGTTTTCCTTAAGGGG
CACGTAGTCTTGGCAGGTTTCGCTCAATATAGGATgagctcaggactccagtggactgtg
gattcagatctggattctggcgcattcgccgtgtgaacgggggcacgttgctggcctgtc
tgcgcctcgtctcccgactgtggagtgTGTTCTGCCCCTTGTCTTTCTGGGAGGTAGGGA
GGGCAGTGAGCCCCTTCGCATCGCCCACCACAGGCCCAGCACATGGCTGATCCCCACTGA
GTGTTCTTTTCCTCCTTTGATCCCCTTTGGCTGACCTAGGTTGGAGCAGCCACTAAAATA
TACCCAGAAACATCTTCCTAATCTACATCTGTGCCAACCCTCATTCCCTGGCGCAGCATG
ACCATCACATGCCCGCCATTGTTCCTGATCTCTGCTGCTCATGACCTGCTCTCCAGCGCT
CCTTCTCATGCTCACATTCCAGTTGGCCTGACCTAGATAAGTGGAGGTTTATTTGACCCC
AAAAATTAGCCTTCTACAAACGAATATAATAGTGTCCATTACAGAGAATAAACTTAGTGC
GTGTCCCATTTAAGCAGAAGTTACTGAAAGCCTGAGTTTAAGTTTCCAGGGCCTGAAAGT
TTTCCATGACAGTTTTCTGCATAATATTACCTACAATTTCAATCTGTTTATTTAAAGCCAT
TCTTGTGTTTGTTGTACTTTGATTAGCTTTATTTTGATTTGAAGTCCTTTTACATTACGG
GCAGTTAACGCTTTGTCTCTGTTAGATTTGCTTTTTAGTTCACAAGAGAAACCTCATTCC
TCTGTATTTGAATAGTTGCAATGATGGAACAGCTGTCCCTGGAGGGAAATGAAAACAGTG
ATTCCCCAAATTGTGACAATAGAAATTTGCTCTTGGGTTACTTACAATGTATCTGAGTAT
TAAAAAATTTTCTTTTTAAACGTTTGAAGTAAAACTACCCAGAAACACTTAGTGGCTGAC
CAGAAACTAAACTCCTGGCATCCTCAAAATGGGATTTATTGGCTTATAAATGTCCTGTGT
TGACTCACAAAGGCACAAACTATCTAGGTAAGTTTCTTCTAAATGTTGATGGGAGAGCT
GGCCACTGTTATGCAAGTTTCATTGTCCTGACTAAACTGCCAAAGAGATTACATAAAATT
ATATCAACTAGACAAAAGGAAAAAGGAAAAAAAACAGAGGTGTCTTGGGAGGAATCCATA

TGAGACCAGTAGACCATGAGAGAGACATCCCTTGCCATCTACAAGGAAAATGGATTTTGT
TCTCCATATGCAAAACCATCTCAGGAGCTTGCGGAGACACCACTTGCTTACTAGCCAGAA
AGAGCAGGTGCCTCCTAAATTCCCCACACAGGAGCTCACAGTGGCTTTCATGCACTGGGA
TTAAGTTAGACTTAAGAAAGCCTGTCTACTCTTCCTGGGATTTACAAGCCAGCTAGTAAA
TCCCAGAATAAATCACACGGCACAGTCATCCAAAGATCCCGTCATCCGTGCCGTTTGGAA
AGCCCTGCTCCTGTGCCACCCTCTCCCCGTGGAGCCTCCCATGCCCAGGACTGCAGAGTC
CTGCCATTCAGACTGCAACTCATCTCACATTCTTCCAAACTATTTGGACAACAGAGCTTT
CTCATCACCTAATGCAGATTACAGTCTCACAGAATTGAGTGTTCAGGCAGACACTGATGT
GGTTCTGTAGTACAGCAAACAATATCAGTTTACAGTCCTGAGGCCAGGCCTGGTGAACAA
CGCACGGTAGCGGTGGGGCAGGGTTCTCAGAATGAAACTGGCTTACACATGGCACTCTCT
GACCACAACTGTATAAGCACCAAACTACACTTAGTTCCATCTATGAGGTAAAATTTAATG
CAGATGAACATCAAAGAAAACGTCAAAGGCTCCTTTTTACAAGTACGTGGGCTACTTAAT
TTGGTCCAAGTCCATTTTAAAAAGCCCTAGGTGCTTTCACGGCTCTGCTACTGACAAGAA
GCCCCAGTGCCTGTGAGCTGCTAATGGGAGGGAGAGGAAGATGAGCTGAGTGGGCCGGGC
TATCCCGTCCACACCGGGAGACAGGGAAGGAGACTCCAAGCTGGTGGTGCCAGCACATTC
CAGGCCACTCAGGCCTATTCCTAGGTGCCAGGTCACGAAAACCACGCTGACAGATCGTGC
TGTGTGCGTGTCATAGCACACAAGCAGGACTGTGAGAGAGTGAAAGTGACACTGGGTGGA
GCACTGAGGAAGGGCCACAGTGTGTTGGTGGAGATAGGCTGTCATGGAGAAGAGACCCTG
GCTTGCTCTACATTGCTTCCAATGCAACTGCAAGGCAGGTCCCAGAGGGCTCCGGCCTTC
GTCATCCAGGTTGCTCCCTCCCCTCATGGCTTTCCCATCCTCAGATGAGGACTCGGCAG
AGCCTACCCCTGCTGACTAACTGTGGCCCCAGGGTGGTGACTCAGCCCTGCACCTCCTGA
TCCCGTCTGCACTGGGCCAGAGAGGATGACTTACCCAGCACGTTCACATCACACAGCTTT
GTGGATTCCTAGGTCCAAGGACCAGAGATTTCAGTTATGTGAGTTATTTTTTTATTTGT
TCTTGCGTATTCCACAAAGGGTCGCAGCTAAACTTAACCTAATGATCACTTTAGTATATC
ACTAAAAAGACAAAGCTCACAGTGCTGTTGAAGCACATTCATCATCTTTAGACATTTTGA
CTAGTTATTTCTTAAGCATTTACCTGCTAGTGTTAAGCATCACATGAAATACATATAGAA
GTAAGACAAAATTTCTTATCTCCCCAAGTTTGCCAACAAATACAGAGCAGGAAGGGAAGC
AGGTCAGAGCAGGAGGCGCAGCTATAGTGAGGCCACCATGCAAGGCACAGGGAGGGTGAG
CTCCAAGTTTGAATGGAATGGGTCTGTCAGCCAAGCCCCCTggctctgggaagatagcag
tgaacaagccagatggcccctcaccctccagagccgtgagtcctgcagaccaaacagCGT
GACAGGTCCTTTCCCTGTCCAGGAGGCCTCTGTGGGTGAGAGTTGGCTGCGGACAGGGCG
TGAAGGCACTTGAGGGTGGGGAAGTGACTCTGACTGGGAGATGCTGAGGACAGGGAGGAA
ACCACCAGATAAGGGACACTGGGGAGGAGGGGTGGACCCCTCAGGGCCAAGCACATGGAG
CCTCATCACAAAGGCAAGATGGTGGCCAAATTCAAGGTCGCTGCAAAAGGAATGGAGAAG
AGAGAATAGATTTGGCATTTGGAGGAAATGGTGACAATCATGAGCACCTACCCGGGACTC
TCCATGGGTGCTATCTCTACATAAACTCATTCCACCCTCTGATTAATccattctacatat
ggggaaacaaaggcatgcggtgtttacgtcacttgccaagatctcaggatttgatccagg
tggcctggttccatggtgcagcctctcagccTGCATGGATGCCCCAGCTCAGAGCATGAC
TCTCAGGACAGGGGTCCCAGCAGCCCTCCCTCCCTGAGCAGCAGGGTGCCCGTGCTGCAC
CACTTCTGTCTAGGAATAGGACATTCTGACACTTTCCTGCCTCTTCCGAGGTCTAGCACT
TACTCTATGCCTGCCTGGGAAGGTGGCAAGCTGGCCTGAGGAACAGACTCTTCCATTTTT
TAGGGAGCTCAAGGCCACAGATGCTCTGAGATCTGGAGTCCAGAGACAGGAGCGGAGGCT
TCTCCTGGTGACCACTCTGCTTAAAAACTTCATCAGATCCGTAGTTTCAGAGCCCCCCTG
AACCCCATCCCTTACCTCTACCAGTTGCAGGTGGGTCTCTGGGGTGGGGCTGCCCTCCCC
ACCAGCACCCCAAGGGCTAAAAGGTTGAGGGGAGAACACCATCATTTGTACAGGGGGATC
CTGGAAGATGAGGCCTGAGAAAGCCCTGCGGGGCCCCTCACCTTCTCCCTAGCTGTGGCC
AAGAGTGTCTGGCCTTGCCTGCCTCAGGACCAGCCCAAAGTGGAGGTGAGAGGTGAGCCC
CAGCCCCCAGGGGAAGGGTGATGGTGGTCTTGGTCTCAGCATGGTTCTGGTAGAGGTGGG
TTATTTTGAAGATGATGAACCTTAAGCCTCTTTCTGATCTTGCTTTAAATAAATACTTCT
GAACAACAGCAACAACAGAATAGTGTTGATAGGAAAGCCCTCCACTCCACCAGAACCACG
CGGCCTTCTCGTCCTCCCCTCCTCCACTTCCTTCCTAAGTCACTGCTCCATGAGCTCTTC
CACAGGAGATTTACAAAATAGAACACAAACAATCCAGTTCCCGCCTCTCACTCTGAACTC
CTCCCAAGACTCGTGGGGTGCGGCAGCCCCTGGGAACACCCAGCCCTTCAAGGTCAAACA
CAGCCCCCGCCCCTCACTCTGGGGTACCCTGCCAGAATAAGCCCCGACAGCCATGTGGAG
CAGAGCCTTCTTTTTTGTAAGTGGAAGTTCCAGGCTGGCTTTTCAAATCCCCTTTTAACC
TCAGTGCTGTATTTCAAAATTCATTCCAGTTTCCTGTAGtaattaacaaaataaatat
tttaatttcaattaaaGTGAGGGTCTCGGAGAAGAAGCAGGAACTGAGTTTCCTGAGAGG
CCCCGCTGAGGCTTTGTTGATATTTCTTCCTGCGACCTCTGCTCGGACCCTGGGAGCTCA
CAGGCCGTATCGCAGCTCTTATCTTTGGGGACCAGTTAAAGCATAACTGCGCCAGGCACA
GAGTTGTCCTTTCAAATGTGCCGGCAGTGGGACGGAGACCCATGCGTCAAGTCTCCTCTA
AGTTCACATGGGATTCTCTCCTTGTCCCAAAGCTGTCTCTGACTTAAAACCCTCCAACTG
ATTACCTGAATTCCAGAATATGTCCTGTGCTCTCTGCCCTTTCCCACGCCTTTGGTGAAG
ACCGGTGTTCTGAGGAAACAGACACTGTGTAGAAATGGCTCAGGTCCTTTAAAGCCCTGG
TGTGAGGAGTGGGGAAGGGCTGGGCCAGAGGTCAGCTGGATTTGTTAGATTGACAGAGTG
ACGCGGACTTCCCCAGAGGCACGGGACCAAGGTGCATGCTCACGCTGTCTCATGCTCTCA
CACATAAtgtgtgtgtgtgtgtgtgtgtatatatatatacacatatacatatatatatat
acacacatatgcatatatataAAACCCCAAGCAGCCTCTGGCTTAGCAGGTGCATTTCCC
AGCAGGGCAATTAAAGCCATGGTCCCAGTAGTGGTCTTGGGGTCTCAGGGTATTTGGTCT
GTGCAGCCACATGCTTCAGTCTCTGGACCCCAGGTCATCTAACGAGGTGGTCGTGTGGGG
ACTGGGATAGAAAAGGTGTCTGCACGGACGTGTGTGAAAGGGCTGGCACATCGCCAGTGC
TCAGCACTGTCAGCTGCTATCACCAGtcattcaatcattcattcattcagttgttcattc
TTCAACAGGCCGTTTTAAAAATGTGCCCAGTATACCAAAATCTCCGCTAAGCATTTAAAG
AGGCAGAATGAAAGTTAGCAGTGGTGGTGAAACGAAGCTGGGAATGTGCTCTGAGGGCCT
CCTTGTGGGCTTAATGAATATGTAGAAACCACGCATTTTAAATAGAGAGGGAGAAAGGGA
GAGGTTCCTGGTCCTCTGCATGGGGACTTGTGTGTGGCTCTTTACTGTAGGGCCTGTGCCA
CTCCTGCTCAACAGCTACCACAGAGGACGCCTTCAACAAATGTGAAGAACGAACAAAAGG
TACAAATGTGAAGAACGAACAGGGTAGAAAGAAAGGAGAAAGCAAGGGTGAGGGTGAGAA

ATCAAGGGACAGAGAAGAGAGAAGAGGAGATAGCCTGGGAGTTCACACAGCCAAGAAGGT
AGACACTGAGTGAACCAGCAAGAGGCTGAGCCTAACTCTCCCTTTCGAATGGGCAGGAG
TTCATGATATTTAATAAACAGAGGCCTTGCTCTGTAAGAGACAGGGTACCAGGCAGAGAG
CAAGTCAGCATCGCAGGAGTCAAACGAGGCAGACAGCGGGGGCAGGGAGCTTGCCTCTGA
AGGAGACCCAGGCTGCCAGAGTAGCAGGGAGTCTGGGCCAGTCCTCTTTTGGGAAGCGCT
TCCTCGGCTTCTGCCCCCCCTCTCCTCTCCCTTTCCACCCACCATCCTGACATAATACTT
CCTAATCTGGAAGTGTTGTCCAGAGAAGAACCTGCTCATTTCCTCTTAAGTAGGCAGGGA
AGCACTAACGTCCAGCAGCATCGGAAACCCGTAGGAGCGCTCTCGGCAGTGCAGGGTGAG
GGGACAGTCCATGTAGTCATGAGACGTGGGTGTCAGGCAAGCGTCTCTTTTCCAAAAGAG
AAAAACATTAAAGGCCTCACAAACGGCGCCCAAAGACTAATTCTGCATAGCATCTTTGCG
AGACCCTAGGTTCTTATGATGACTGGTTTTGCCTGAGAAAGAAAAAATTTAATTTTGCT
CTGACATGCCAATTCAACAAATCATTTTCACATAATATTCATGCAAAAAAAAAACAATTT
GCCAGAAAACTTGGGAATCCATCCACATCTACAGCTTTTCCCTGCAGTCACACTACAGTG
GGATCCCTCCATACAGGAGCGGCAGAGTGGAGCAGGCTAGAGATGCCTGTTTGTTTCTGT
TTGCTGCACCGCAGCAAGCATTTCTGTCGTGCCCACTCTGTACTAGAAAGTACATGAACA
TCAGCCATAAAGGGAACTAGAAAGGTGGCCCACCCTCTTGGTGGAGAGAAGAGAGTGT
GGTAGAAACAATAATAAGAAGTCTGCAGAACTTGACCCCTCCCAGCCTCTCCCACCTGCC
AGCCTGGCCCTTGCAGAGAGATGCAGGCTGCCATTCTTAGGCCAAAGCCTGGGACAGTTG
GGCTCAGCAAGGTAGGCATCCGTCAAGCAAGGAGGAGCAGGGGTCAGCAGTGACCCCAGC
AGCCAGCAGGGAGAAAGGTGCATGTGACAAGGACACCAGAGGCCGTGGGTCAGGATCAGC
CAGGGTCAGGGTAGCATTTCTAGGAATTCACTCTGTTGGGCGCTGTGCTGGCTGCTTCTC
ACATATTATTCCTTTCTTACTCTCAGAGCAGAGATTTCAATTGCAGCGAGATTGTGGAGG
CAGCCAGGGAGGTGGGGAGGGTGGTGTCTTCTAAAAGCATTTTCAGTATCCATGTGGTTT
CAGTAATAATAATAATAATAAACCAGTGAAAAGTAAAACAGGACAAAAATCTTCATAGGC
AGTGAACCATATCAGAGAGTCCAAGAAAGCACAATGAGAGTGTGGCTTAAAAACCCTGAA
CGACATTCCTTTGCACCAGCTTGGTGAGGAGGGCATGGTCCCCGCCACCCCCCACCCCCA
CTTTGCAGATAAACCACATGCAGGAAGGTCAGCCTGGCAAGTCCAGTAAGTTCAAGCCCA
GGTCTCAACTGGGCAGCAGAGCTCCTGCTCTTCTTTGTCCTCATATACGAGCACCTCTGG
ACTTAAAACTTGAGGAACTGGATGGAGAAAAGTTAATGGTCAGCAGCGGGTTACATCTTC
TTTCATGCGCCTTTCCATTCTTTGGATCAGTAGTCACTAACGTTCGCCAGCCATAAGTCC
TCGACGTGGAGAGGCTCAGAGCCTGGCATGAACATGACCCTGAATTCGGATGCAGAGCTT
CTTCCCATGATGATCTGTCCCTCACAGCAGGGTCTTCTCTGTTTCAGGGCATGAACTACT **exon 21**
TGGAGGACCGTCGCTTGGTGCACCGCGACCTGGCAGCCAGGAACGTACTGGTGAAAACAC
CGCAGCATGTCAAGATCACAGATTTTGGGCTGGGCAAAGTGCTGGGTGGCGGAAGAGAAAG **EGFR_e21 c.2573T>G,L858R**
AATACCATGCAGAAGGAGGCAAAGTAAGGAGGTGGCTTTAGGTCAGCCAGCATTTTCCTG
ACACCAGGGACCAGGCTGCCTTCCCACTAGCTGTATTGTTTAACACATGCAGGGGAGGAT
GCTCTCCAGACATTCTGGGTGAGCTCGCAGCAGCTGCTGCTGGCAGCTGGGTCCAGCCAG
GGTCTCCTGGTAGTGTGAGCCAGAGCTGCTTTGGGAACAGTACTTGCTGGGACAGTGAAT
GAGGATGTTATCCCCAGGTGATCATTAGCAAATGTTAGGTTTCAGTCTCTCCCTGCAGGA
TATATAAGTCCCCTTCAATAGCGCAATTGGGAAAGGTCACAGCTGCCTTGGTGGTCCACT
GCTGTCAAGGACACCTAAGGAACAGGAAAGGCCCCATGCGGACCCGAGCTCCCAGGGCTG
TCTGTGGCTCGTGGCTGGGACAGGCAGCAATGGAGTCCTTCTCTCCCTTCACTGGCTCGG
TTTCTCTTAGGGACCCTCACAGCACTAAGGGGTGCGCGTCCCCTGTCAGGCCCTCGAATG
CCCTCCCACAGCCAGGCCCCTCTGAGGTTTCACTCTGGCCTGCTTGGCTCCTAGCAGCCA
CCAACCCATGATGCTGGGCCCTGAAAACACACGCAGACCTGGATGAGTGAGGCCACTGGG
CACAACCAGGGCTCCCAGCTCACCAGAGCAGCCTGGGACACAGAGGGTGCTCAGAAACCT
ACCAGAGCAGCCCTGAACTCCGTCAGACTGAAATCCCCTGTTGCCGGGAGGAGGCGCCGG
GCCTGGGGGACGGGTCCTGGGGTGATCTGGCTCGTCTGTGTGTGTCACTCGTAATTAGGT
CCAGAGTGAGTTAACTTTTTCCAACAGAGGGAAACTAATAGTTGTCTCACTGCCTCATCT
CTCACCATCCCAAGGTGCCTATCAAGTGGATGGCATTGGAATCAATTTTACACAGAATCT **exon 22**
ATACCCACCAGAGTGATGTCTGGAGCTACGGTGAGTCATAATCCTGATGCTAATGAGTTT
GTACTGAGGCCAAGCTGGCTTTTATTGTTAGTTAATTTACATTATATCCTCTGACATGCA
AGTATTTTCTTTCGAGATAATGACTAATGATAATGTAATCATTGCTGTCTATCTATTGTA
CTGAGAAAACACGGCAGAGGAAATCGAGTCCAGCTGCCGTCCAAAAGTCACTGGAGATTG
CAATGAGCTCGTCTGGCAGGGTGGGGGGTATGGGAGGGAAAGAGCTTAGGAAACGGCTCT
CCCTGCAAAGTCCAACCAAACTTTAACGTTAACCAAACCATTAATGTTGCCATGAATTTG
AAGTGAACCAGAGGGAGGTGGCAGAAGAAGCTTAATGGGGAATAGTTCCGGTAGAGAAAT
GAGGCTTAAGATGAACTACCCTGGCCCTTATGTGTCAGAGAGAACGGCTTGACAAACACA
CACTGAGGATGTCTGCAGGGATAAAAGAAGAAAGGGAGATGACCCTTGCTTCTCGCTCTC
GGGAGGACCATCTGGTCCGGCCCTGGGGATTCTCTGTTTCCTCTTCTGAATCCCAGTGTT
GCCCAGCACTGGCCTGTACCCATCCTCACGAGGGCCGCTCTCCTCACCCGGCCCTAGGTC
CCTGCCCTGTCCTGAGCCTACAGGGGCCTCCCATGTTGAGAAAGTGTTGCTGACACATTG
TCTCTGACCGCTGTGCCAGGCATTTTCTGCTGAATTACCGCACTTGGTCCTTGAATTTCA
CCCAGCAACTTACTGAAAGGCTGGAACCCATGAACCTACCCCTTCACTGAGGAAAATAAG
TTACCCCAGCCATCTACAGCGACAGGAGCAAGGGAGGAGTCGCCTCACCTCTCTAGAAAT
GTGTATTTGAGGAGAACACTATTGAAATGAATTTCCAAGAATAATCTAGTCAGTATTACA
AAAGCAAAATTATTTGGGATATCGTCCTTTTTTACTTAGTATTTTTTCTTTTTCCTATAG
CATTATTAACTTTCTGATTTTCCAAATACATACACATTTTTAAATTTCCTGAGTCTTTAT
CTCTTCTGTTAAAATGTAAGATTTATGATACAAAGGCAGAGATTTGTGTCCATGAATAAG
TGAAGTTTGGTGTGCACCTGTGAGCTGAGCCACCTCAATTAATGGAACAGATAAGGAAAT
AAAGGTCTGCTGATGCATTGTTATTTACAGCCATTTTCAGAATGTATCTCCTCTCCACGA
GGGAACTGCAGGGTCCTGCCCCAAGCCATTTATTTTGTCCTCAAGCAGCCCGCCCCTCCC
ACTCCAGGCACAGCCCGgtctcctgctggtctcccctcttcccacttgctccccctcatc
tatgctccagacagaggccacatatatttttttaactttttttttttttttttgagacag
agtcttgccctgtcacccaggctggagtgcagtggtgcagtctcggctcactgcaacctc
cacctcccgggttcaagtgattctcctgcctcagcctcctgagtagctgggattacaggc
gcacaccaccatgcccagctaatttttgtatctctagttgagacagggtttcactatgt

tggccaggctggtctcgaactcctgacctcatgatctgcccgcctcggcctcccaaagtg
catattttttaacttttatcagacttttcattctctgctcaacatctttctttggtcctcc
aggtatgttcagataaaacctgagcacctggccatgactgatgggttgctgggccatctg
gccctggcaactctcccgtccaccaggtccccctcccgtcacgctccaggcataGCCTGT
GTGTGCCAGCGCAATGCCCACACTCCATGCACAAGTGGAAGCCCTCTCAAAGTCAGTGGC
TTAGTGCCTTGATGTGGTCACACCCATTCTCAGGAAGTCCGTTCCCACTGAAAACATTGT
GTGTTTTCAACATCATTGAGGCTGCCACGGCAGATTATAATCACTGGCCTAGGCAGCCCA
CTGGAACTACCAGACCATGAGCCTGAATTTTTTGTTTAAAAATCATATCCTGTTTTCTCT
ACTCTCTAGTCTCTAGTCAAGGTGAATTATTCAATTTAATAAATTAGGGGCCTAGTGTGT
TGTACCAAGGAGCTAAAAAGAGAGAACTCGCAACACCTTCCAGCCCATTCTCCACCTAAC
ACTGGCTATACTGGCTCTCCTCTCTCTCGCTGTTTGTTCCAAAATCTAATAACCTGTCTT
CCCACTAGAATTCATCATACATGTTTAAAAACCTAGTTAAATAGTAGTTAAACTGACTGC
ATAGATCTGGAAATGAGACAGTCTTTCTTTTACAAATCCATATAGACTATGAGTTGGGGG
CAGGGGATGACACAAGAATCTATTTTCTTGCCCCCAAACCATTGCTTTCCTTCCAATGTT
AAGCTTGTATTCTGTGTATTAATTCAGGTGGTTCCGTTTGGGAATGGCCTCTGTTACCCA
GAGATGGGAGGGCCATCAGAACTCGGGGTTGTCTGAAAAAACACTGGTTCTAAAATTATC
ACTGCTTTCACTTGTTTTTAACCATCATAGTTGTTTGATTTTGAAGGAAAAACATGAGGG
TTTTTATTCTATGCTTGTTATATCTATATTGTGGTTTCGTATTTTTTAGATTTTAGTACC
TGACAttttttaacttttatttaggttcaggggtacatgtgcaggtttgttatatagg
taaatttgtgtcatgggggtttgttacacagattattttatcacccagggattaagccta
gtacccattagttatttttcctgatcctctcccctcctcccatcctccaccgtcctataga
ccccagtgtgtgttgttccccctctaagtgtccatgtgttctcatcatttagctcccactt
ataagtaagaacatgcggtatttgattttctgttcctgcattagtttgctagggatgatg
gcctctagctccatccatgttcttgcaaagtacatgatctcattctctttttgtggctgcc
tagtgttccatggtgtatatgtaccacattttctttatccagtctgtcattgatgggcat
ttaggttgattccatgtctttgctattgtaaatagtgctgcagtgaaaatacgcatgcat
atgtctttatggtagaatgatttatattcctttgagtaatgggattgccgggtcaaatgg
tagttctgttttagctatctgagaaattgccacactctttccacaataattgaactaa
tttacattcccaccaacagtgtaaaagcattcctttttctccacaacctcaccagcatgt
gttgggatttttttttttttactttttcaataatagccatctgactggtatgagatggt
atctcagtgtggttttgatttttatttctttaatgatcagtgatgttaagctctttttca
tatacttgttggctgcatgtatgtcttcttctaaaaagtgtctgctcatgtccttttgccc
actttttaatgggattgtttaatttttcttgtgaatttacttaagttccttatagatgc
tggttattagacccttctcagattttgtagcttgcaaaaatgttcacccattctgtgggtt
gtcttcactctgatgatagtttctttgctgtgcagaagatcttcagtttagttagatcc
catttgtcaatttttgcttttgttgcaattgcttgatgtgttttcatcatgaaatcttag
cccattcctatatccagaatggtattacctaggttgtcttccagggtttttatagtttgg
ggttttacatttaagtctttaatccatgttgagtttatttttgtgtatggtgtaaggaag
gagtccagtttcaatcttcttcatggctagctagtcatcatttattgagtagggagtcct
ttattcattgctttttttttttgtcaactttgtcaacgatcacatggttgtaggtgtgc
agccttatttctgggctctctattctgtttcattggtctgtatgtctgtttctgtactag
taccatgctgttttggttactgtatccctgtagtttaaagtcaggtagcatcatgcttcc
agctttgttctttttgcttaggattgccttggcaattcaggctctttttggttccatgt
gaatttttaaattgtattttctagttctgtgaagaatctcattggtagtgtgataggagt
aacattgaatctataaaatactttgggcagtatagtcattttaatgatattgattctttc
tatccatgagcatggaatgttttccatttgtttgtgtcatctctgatttctttaagcag
tgttttgtggttcttattgtagagatctttcactttcctggtttactgtatttctaggta
ttttattctttttgtggcaattgtgaattgaattgcattcctgatttggttctcagcttg
actgttgtggcatattggaatgctaattattttttgtacattgattttgtacaactgagt
cttcactgaagttgtttatcagcttaaggggttttgggtcaagactatggggttttctag
atatagggatcatgtcatctgcaaacagagatagctgttttcctctctcttcctgtttggatg
tccattatttctttctctcacctgatttatctggccaggacttccaatactatgttaaat
aggagtgttgagagagggaatccttgtcttgtgtcaattttcaaggggaatgttttcaac
ttttgcccattcaatatgatgttggctgtgggtttgccatagatggctaatatgttgagg
tttgttctttaaatacctagtttattgagaattttaaacatgttgaattttattgagagc
cttttctgcatctattgagatgatcatgtggcttttgtccttagttctgtttgtgtggtg
aatcacatttattgatttgcatatgttgaaccaatcttgcatcccagggatgaagccgac
ttgattgtggtggcttaagcttttgatgtgctgctggattcgatttgccagtattttgt
tgaggatttttatgtctatgttcatcagagatattggcctgaagttttctttttttgttg
tatctctgccaagctttggtatcaggatgacattggcctcatagaatgagttaaggaaga
gtccctccttctcaatttttttggaatagtttcagtaggaatggtaccagctttttttgt
acatcttgtagaatttggctatgaatccatctagtcttaggctttgttttggttggtagg
ctatttattactgattcaattttggagctcattattggtctgttcagggattcagtttct
tcctgAGGTTTTTATTTTTATCAAATGGAACTTAAGCTTTTTCATTTCCAATTTTTTTAT
GATCTAAAAATGTGCAGTTTACAGCCCTGTTCAGAATCTGCATCTTCCTCATTCTGCAGA
TACAGGTCCCTCAGAGCAGGTGACTGAGTGTGTATCCTGTCTGGAGCATAATACTTATGC
TAGTAGAGTTACTGTTGTCTTTATTGTTAATTACCAAAGTTtaccacttatcagtcactt
actacttgctgggcattgcactaagcatttcagttgtattatcttgtgggtccttacag
caatcctgtgaaacagatactgctattaccccactttatagagaggtagactgaggcttc
cagcattgaagcaaattgcccaagactacagaaatgtaggtTTCTAAACATCAAGAAACA
GTAACCAGTAATGATGACTAAAGCAAGGGATTGTGATTGTTCATTCATGATCCCACTGCC
TTCTTTTCTTGCTTCATCCTCTCAGGGGTGACTGTTTGGGAGTTGATGACCTTTGGATCC **exon 23**
AAGCCATATGACGGAATCCCTGCCAGCGAGATCTCCTCCATCCTGGAGAAAGGAGAACGC
CTCCCTCAGCCACCCATATGTACCATCGATGTCTACATGATCATGGTCAAGTGTGAGTGA
CTGGTGGGTCTGTCCACACTGCCTAGCTGAGCCTTGGTGGCTGCTCTTAGCCAAACAGCT
GAGGCCTTTGCATCCCTGGAGAAATGTCATCACATTACTTAAGGCAGGCACACAAATCCA
GAAACATCTGTAAATACCCCTTCAAGCATTCTTTTAAAGACACTTCTTGACTCATTGGGC

AGTATGACCTGACATTTGCCCATGTTTGCAAGCAAATAAATAAAACTAAAGTCTTCCGCA
AGCCATTACACCAAAATATTCTATTCGCTGAGTTACTCAATGAAATACCGAGTTGCCCTA
TATTTTGAAGCCTGTTACCAGAGAGACTGAATGTTTTTAAATGCATGGCAGTGAGTAACA
ACATAAGGCTAATAGAGTCAACATTTCTGCTTTGACTTAAACCTTTTAAACCAGTGGATT
TATGTGAAGTCTCTGCAGTGTGGCATTTAAACATTTCAATCTAAATAAGAGTGTGTAATT
TGATTGATGCTATTATTCTACCAGATTCACGAGTGCAGTGGGCTCTGGAGGTAGCATTAC
ATGCATGGGATGAGCATTTGCAAAAGAAAGTTGTATAGGGAATATGACAGAGCCAAGTTA
ATGTAAATATTAATGCCTTTCTGAACTCTAGGCCACagagttgatctttttaacttcct
tggtttgggctaaggaagctgtgatccagagaagccacgtgatttgtctaaggtcacata
gcAGTCTGGCCTAAAATAGCTTGATATGCTGTGGATGGAAAATAAATGTGATCCCTCAAG
AGGCATGAGGATTTCCAGGCAGTAGCCATACCTCCAAATTGTTTAATCTGGATTTAGATT
GTTGGGTAGTCACATGCAGCAGCACAGTTAACAGTGTGTCCTCCTGTGGAAGTTGCCAGC
ACAGCCAGCCCTCTCACTTGCATGCATGCCCACCAGCCTTCTCACTTGCATGCATGCCCA
CTGGGTATGTGCTGTACTGGAGACGCCGGGGGTAGGGGCCCAGTCCCAACCCCAAATTCT
TTAAAGCCTATTTTTCTAAGTTGCATCTGGTTTCCTACCTGAAGGAATGCTAAGGGTGGA
TGTTGAGTGAGGACCTTGGTGCAGGGCACCCTGCAGTCAGGATAGTTCATGGAGAGCAAT
TGTACAGACCCACACTGCTCCATCCCCTCAGGCGTAACACAGGATGCTGACCCCAGGAAG
AGTGGGCGTAGAAAAACTAGAGGGCATTATTGTTATTCTGATTCAAATGTACAGTGCTGG
CATGGTCTTTAAACAGTAACCAGTACTAGCTGGCCAAGACAGAAAAGTCTACCACAAAGA
CTTGGTTCTTTCATCACTTATTTGACTGGAAGTGTCGCATCACCAATGCCTTCTTTAAGC
AATGCCATCTTTATCATTTCTTCCAGTGTTCTAATTGCACTGTTTTTTCTCATTCCTTCC
CCAGGCTGGATGATAGACGCAGATAGTCGCCCAAAGTTCCGTGAGTTGATCATCGAATTC **exon 24**
TCCAAAATGGCCCGAGACCCCCAGCGCTACCTTGTCATTCAGGTACAAATTGCAGTCTGT
GCTTCCATTGGGAAGAGTCCCTCTAATGAGCATCTCATGTCACTGTGTTCTGTCACATGC
CAGCCTGGCCTCCCTGTGTCCCAGATCGCATTATTAAACCCTCCAGCGCATTAGAGCAAG
CCTCAGTAAGGCGCAGGCCACATCGTGAACTAAGCAGCATCCGTGAGTGGGGCCCACCCA
ACTCCATCTCCCCCTCCCCGTCTGAACTCTCCTCTGGTGCTCGTCCTCACTGTCCGGCTA
GCCAAAGCCTCAGCTGGGTCTAAGAGAGAAGCATGGTCTATTgggctttggtgtcaggca
gacgtggcttcacacccctgactctccacttcttcgcatcacccaggcagccgatccacc
tatctccttccataacacaggaataccaaaaccaagctcacaggattgtctcaaagattc
aataaaatatgttgcaaaatacgctccctaacacctcacagcaaggtgcacactcgatga
atgctgcagctTCTTCCCTTTCTGTTTCCTCAGAAGCTATTTGAATCTCATGTAGGGGCT
TTCAAGCATCAAAGGATGGTTCATGTTTTATTTTAAGGCACCCACATCATGTCATGAGGG
GAGGCAGCTATAATTTAGAGAACCAAGGGGGATTTCATTATAACAAAATTGGCAAACACA
CAGGCACCTGCTGGCAATAGACCCCTGCTCCTATAGCCAAGAAGTGGAATAGCATCTCTA
CGGGCCATTCTAATAGCCTCAAAATCTCTGCACCAGGGGGATGAAAGAATGCATTTGCCA **exon 25**
AGTCCTACAGACTCCAACTTCTACCGTGCCCTGATGGATGAAGAAGACATGGACGACGTG
GTGGATGCCGACGAGTACCTCATCCCACAGCAGGGCTTCTTCAGCAGCCCCTCCACGTCA
CGGACTCCCCTCCTGAGCTCTCTGGTATGAAATCTCTGTCTCTCTCTCTCTCTCAAGCTG
TGTCTACTCATTTGAACAAATTGAATTTTAGGGAAAATAACCATCTAGTGAAACTCACAT
GGATATGAAGTCAATTTTAACCAAATGGTAAAATCAAAATCAAAATAAATTAAGTGTATT
AATTATTTTGTTGCAATTGCAACAACTTGATTGTAAGCCTTTTAGGTCCACTATGGAATGT
AATTAAATCAAAACTAAACCTAGTTGCTCTAAAACTAACGATTAAGACAAAAATTAAACA
CCTTCACAATATACCCTCCATGAGGCACACCACCTGCATTCAGGAAAAGTGGATGAGATG
TGGTACAAGCATTCCATGGGCAACTTCTCTGTTTCTTTTTCAGAGTGCAACCAGCAACAA **exon 26**
TTCCACCGTGGCTTGCATTGATAGAAATGGGGTATGTATGAACACCTTATAAGCCAGAAT
TTACAGCTCTCCACTATGGCTCTATTTTACATGGAAAATGCCTTAACCTAAATAATTTTA
ACCCAGATAATCTTGAGTTTTCTTCCTGTGTGGGTTTTTCCCTGCACGGCTGTCACGCCT
CAcagtgccgttcaaagcgtgactcctggaccagtagtagcatcgcctggccttgttaga
aacgccatttttcaggccactgccccagtttgaccaaatcaggacctctgggggtggcac
ccagtagtctatgtttgagccactttccaggtgatgctgatgtctgttgaagtgtgagGC
CGTGGTCTAGAccgcactgtgccatgcagaaaccactagccacatgtggctacttcaact
taaatgttaatggagttaaaatgaaataaaatataaaattcagtttctcacacatgtgaag
tgtccagtagccacacgtggctagtggtgaccgtattgaagagcaccgctcatagcacac
ctccctcactgcggaaagttctgctgtacagcacCCAGCACAGCCCTGCTGCCCACCCTG
CAGCCTGTGGCCCAGTAGCACCAGCACCCACCAGGGtgcagactctcaggcctgcccaac
ctactaatcagaaccagcatctcaaggagatctcgggtgatttttgcaaacactgaagtt
ggggcagcCCTGACCGGAGTAACCTTCCCTCATTTCCTCCTGCAGCTGCAAAGCTGTCCC **exon 27**
ATCAAGGAAGACAGCTTCTTGCAGCGATACAGCTCAGACCCCACAGGCGCCTTGACTGAG
GACAGCATAGACGACACCTTCCTCCCAGTGCCTGGTGAGTGGCTTGTCTGGAAACAGTCC
TGCTCCTCAACCTCCTCGACCCACTCAGCAGCAGCCAGTCTCCAGTGTCCAAGCCAGGTG
CTCCCTCCAGCATCTCCAGAGGGGGAAACAGTGGCAGATTTGCAGACACAGTGAAGGGCG
TAAGGAGCAGATAAACACATGACCGAGCCTGCACAAGCTCTTTGTTGTCTGGTTGTTT
GCTGTACCTCTGTTGTAAGAATGAATCTGCAAAATTTCTAGCTTATGAAGCAAATCACGG
ACATACACATCtgtgtgtgtgagtgttcatgatgtgtgtacatctgtgtatgtgtgtgtg
tgtatgtgtgtgtttgtgACAGATTTGATCCCTGTTCTCTCTGCTGGCTCTATCTTGACC
TGTGAAACGtatatttaactaattaaatattagttaatattaataaattttaaGCTTTAT
CCAGATACTCATAACCTGCTAacacacacacatatacacacacatacacatacacacata
tacacacaccacacacatacacagacaccacacacatacccatacacagacacatacacat
gcacacacatatacacacacacctcaaatacatacacacacacacacatacatgtatac
acacatacacacacccacacatacaccacaaaaaccccacacacatacacatatacacacc
acacacaccacatacacacacgtatacacacatatatacacacacatgcatacat
acacaccacacatacatacagacacaccacacacacgtacacacaacacacaacacagac
acgtacacacactacagacatgtatgcacacatacacacacacacacatacatacacac
agacacatatacactacacacaccattacatacacacgtacacatacacccacacacacca
cacatacacacacccacacacacatacgccacacacacacccacaaaaaccgcacacacata
caaacatatacacactacaccacacatacacacacacacccacacacccacacacacata

cacacaccacacaccacacatacacgcaccacacatacacacacgtagacacaccaca
cacaccacagaaacacacattaacacaccacatacacatatgtatgtgcatatacacacc
cacaccccacacacacatgtataaagatttagatatatataaaacatatgttatatatat
gttgatgtaatatctaatatctatatatctaatatgtagtttattagctatctaatatct
atgtcatatatatcaaaatctttatatataaaaatatgtagaaatctttatacatatgtt
atatgtatataaagatttagatatataacatatgtaagttatatatatGTTAGTGTAATA
TCTAATATATAGTTTATTGGCTATCTAATATAatataaacagattatcaatattataagc
tattagaaaaatgcaagttaaggcagatgatatacctctttacacaccaactaCACACAC
CAACTacacacacacatacacacagacacacacgacacacaccatacacatgtacacaca
caccacatatacacaaacgtacacacacaccacacacacatacacaccacacacacacca
cacatacatacacatccacacaccacacatgtacacacgccacacacacacatacacaCC
ACATACACATATGTATGcacacatacacaccaacaccacacagacaccacacatgcataa
acatatagacatatacacaccacacaccatatgtacacatgtacacacacaccacatata
cacacaacacacacaaatacacacaccacacacacacacacaaaaaccccacacacacaca
aacatatacaccccacacatacgcatatatacacacacacatacacaccacacacataca
caccacacacacaccacacatacacacacgtacacacaccacacacacacaccacagacaca
caccacacatacatacacatacacacaccacacacacgtacacacaccacacacaccaca
gacacacatagacacaccacatacacacccacaccacacacacacaactcataccacaca
tacatacacaatagacacatacacaccacacacaccatacatacacacgtatacacacac
cacatatacacacacgtacacacacaccacacacacccacatgcacacaccacacacaca
tacaaatatacaccacacacacatacaccacacacacacggtgcacatacacacacatatac
acacaccagacacacataccacatacacatcacacatatatgtatacatgcatacacata
cacacacacatacacacacTCTCCTCAAGGCAGTTTATCCTCTGAGAACTTTAAATTTAC
AAAAGACACATATGTCCATTACTTTGAGAAGGACAGGAAAGAACCCACTTTCTTTTGCAG
CAACAGCAAGAGGGCCCTCCCGAGGCTCCTGCTCCCTGTCATAAGTCTCCTTGTTGAGGA
CATTCACAGGGTTCAGAACCCagggatcctgcatgggatggtgctttgctgattacttca
cctctgatttctttccactttcagaATACATAAACCAGTCCGTTCCCAAAAGGCCCGCTG **exon 28**
GCTCTGTGCAGAATCCTGTCTATCACAATCAGCCTCTGAACCCCGCGCCCAGCAGAGACC
CACACTACCAGGACCCCCACAGCACTGCAGTGGGCAACCCCGAGTATCTCAACACTGTCC **EGFR_e28**
AGCCCACCTGTGTCAACAGCACATTCGACAGCCCTGCCCACTGGGCCCAGAAAGGCAGCC
ACCAAATTAGCCTGGACAACCCTGACTACCAGCAGGACTTCTTTCCCAAGGAAGCCAAGC
CAAATGGCATCTTTAAGGGCTCCACAGCTGAAAATGCAGAATACCTAAGGGTCGCGCCAC
AAAGCAGTGAATTTATTGGAGCATGACCACGGAGGATAGTATGAGCCCTAAAAATCCAGA
CTCTTTCGATACCCAGGACCAAGCCACAGCAGGTCCTCCATCCCAACAGCCATGCCCGCA
TTAGCTCTTAGACCCACAGACTGGTTTTGCAACGTTTACACCGACTAGCCAGGAAGTACT
TCCACCTCGGGCACATTTTGGGAAGTTGCATTCCTTTGTCTTCAAACTGTGAAGCATTTA
CAGAAACGCATCCAGCAAGAATATTGTCCCTTTGAGCAGAAATTTATCTTTCAAAGAGGT
ATATTTGAAAAAAAAAAAAAAGTATATGTGAGGATTTTTATTGATTGGGGATCTTGGAGTT
TTTCATTGTCGCTATTGATTTTTACTTCAATGGGCTCTTCCAACAAGGAAGAAGCTTGCT
GGTAGCACTTGCTACCCTGAGTTCATCCAGGCCCAACTGTGAGCAAGGAGCACAAGCCAC
AAGTCTTCCAGAGGATGCTTGATTCCAGTGGTTCTGCTTCAAGGCTTCCACTGCAAAACA
CTAAAGATCCAAGAAGGCCTTCATGGCCCCAGCAGGCCGGATCGGTACTGTATCAAGTCA
TGGCAGGTACAGTAGGATAAGCCACTCTGTCCCTTCCTGGGCAAAGAAGAAACGGAGGGG
ATGGAATTCTTCCTTAGACTTACTTTTGTAAAAATGTCCCCACGGTACTTACTCCCCACT
GATGGACCAGTGGTTTCCAGTCATGAGCGTTAGACTGACTTGTTTGTCTTCCATTCCATT
GTTTTGAAACTCAGTATGCTGCCCCTGTCTTGCTGTCATGAAATCAGCAAGAGAGGATGA
CACATCAAATAATAACTCGGATTCCAGCCCACATTGGATTCATCAGCATTTGGACCAATA
GCCCACAGCTGAGAATGTGGAATACCTAAGGATAGCACCGCTTTTGTTCTCGCAAAAACG
TATCTCCTAATTTGAGGCTCAGATGAAATGCATCAGGTCCTTTGGGGCATAGATCAGAAG
ACTACAAAAATGAAGCTGCTCTGAAATCTCCTTTAGCCATCACCCCAACCCCCCAAAATT
AGTTTGTGTTACTTATGGAAGATAGTTTTCTCCTTTTACTTCACTTCAAAAGCTTTTTAC
TCAAAGAGTATATGTTCCCTCCAGGTCAGCTGCCCCCAAACCCCCTCCTTACGCTTTGTC
ACACAAAAAGTGTCTCTGCCTTGAGTCATCTATTCAAGCACTTACAGCTCTGGCCACAAC
AGGGCATTTTACAGGTGCGAATGACAGTAGCATTATGAGTAGTGTGGAATTCAGGTAGTA
AATATGAAACTAGGGTTTGAAATTGATAATGCTTTCACAACATTTGCAGATGTTTTAGAA
GGAAAAAAGTTCCTTCCTAAAATAATTTCTCTACAATTGGAAGATTGGAAGATTCAGCTA
GTTAGGAGCCCACCTTTTTTCCTAATCTGTGTGTGCCCTGTAACCTGACTGGTTAACAGC
AGTCCTTTGTAAACAGTGTTTTAAACTCTCCTAGTCAATATCCACCCCATCCAATTTATC
AAGGAAGAAATGGTTCAGAAAATATTTTCAGCCTACAGTTATGTTCAGTCACACACACAT
ACAAAATGTTCCTTTTGCTTTTAAAGTAATTTTTGACTCCCAGATCAGTCAGAGCCCCTA
CAGCATTGTTAAGAAAGTATTTGATTTTTGTCTCAATGAAAATAAAACTATATTCATTTC
CACTCTATTATGCTCTCAAATACCCCTAAGCATCTATACTAGCCTGGTATGGGTATGAAA
GATACAAAGATAAATAAAACATAGTCCCTGATTCTAAGAAATTCACA

Supplementary Table 1

Probe-set Assembly Table

probe	probe type	comm ents	5'PSS	len gth	5'SS	len gth	5'TGS	len gth	Tm	3'TGS	len gth	Tm	3'SS	len gth	3'PSS	len gth	5'half-probe	5'H PL	3'half-probe	3'H PL	PCR product	TPL
EGFR_e1	DS		GGGTTC	19	cgcta	5	GAGCTC	21	75,2	TGCGAC	21	83,5	c	1	TCTAGA	23	GGGTTCC	45	TGCGACC	45	GGGTTCCC	90
ctrl_1	contro	chr22	GGGTTC	19	cgctac	6	GGCCCA	21	75,6	GGCAAA	22	71,0	ac	2	TCTAGA	23	GGGTTCC	46	GGCAAAA	47	GGGTTCCC	93
EGFR_e7	DS		GGGTTC	19	cgctac	7	CCCGAG	22	70,9	GTGCCA	21	70,3	ctac	4	TCTAGA	23	GGGTTCC	48	GTGCCAC	48	GGGTTCCC	96
EGFR_e28	DS		GGGTTC	19	cgctac	9	TGGGCA	22	70,7	CACTGT	21	72,2	tctac	5	TCTAGA	23	GGGTTCC	50	CACTGTC	49	GGGTTCCC	99
EGFR_e21	MS-	speci	GGGTTC	19	cgcta	5	GCATGT	27	72,0	GGCCAA	21	77,5	atctac	7	TCTAGA	23	GGGTTCC	51	GGCCAAA	51	GGGTTCCC	102
EGFR_e8	DS		GGGTTC	19	cgctac	11	ATGTGG	23	71,9	TGCGTC	21	80,6	atctac	8	TCTAGA	23	GGGTTCC	53	TGCGTCC	52	GGGTTCCC	105
EGFR_e21+	MS+	speci	GGGTTC	19	cgctac	11	GCATGT	27	74,3								GGGTTCC	57			GGGTTCCC	108
ctrl_2	contro	chr1	GGGTTC	19	cgctac	12	CAGCTG	24	72,8	CTGGAC	21	72,7	atctac	12	TCTAGA	23	GGGTTCC	55	CTGGACA	56	GGGTTCCC	111
EGFR_e5	DS		GGGTTC	19	cgctac	13	CAAAAG	25	72,0	ATGGGA	21	74,6	atctac	13	TCTAGA	23	GGGTTCC	57	ATGGGAG	57	GGGTTCCC	114
ERBB2_e2	DS		GGGTTC	19	cgctac	19	CACCTC	21	73,1	GTGGTG	21	71,7	atctac	14	TCTAGA	23	GGGTTCC	59	GTGGTGC	58	GGGTTCCC	117
MET_e2	DS		GGGTTC	19	cgctac	16	GAGGAA	25	71,3	AGTACA	24	72,3	atctac	13	TCTAGA	23	GGGTTCC	60	AGTACAA	60	GGGTTCCC	120
ctrl_5	contro	chr2	GGGTTC	19	cgctac	21	AGTCCT	22	72,8	AGACGA	22	71,8	atctac	17	TCTAGA	23	GGGTTCC	62	AGACGAG	62	GGGTTCCC	124
EGFR_e2	DS		GGGTTC	19	cgctac	17	TTCTCA	28	70,2	CTGTGA	23	70,7	atctac	18	TCTAGA	23	GGGTTCC	64	CTGTGAG	64	GGGTTCCC	128
EGFR_e19	MS-	speci	GGGTTC	19	cgc	3	CCGTCG	28	71,3	AACATC	26	70,2	atctac	33	TCTAGA	23	GGGTTCC	50	AACATCT	82	GGGTTCCC	132
EGFR_e4	DS		GGGTTC	19	cgctac	21	AGTCAG	28	71,1	TCGATG	24	72,1	atctac	21	TCTAGA	23	GGGTTCC	68	TCGATGG	68	GGGTTCCC	136
ERBB2_e26	DS		GGGTTC	19	cgctac	30	CCCAGC	21	72,1	AGTGAG	22	71,8	atctac	25	TCTAGA	23	GGGTTCC	70	AGTGAGG	70	GGGTTCCC	140
ctrl_3	contro	chr17	GGGTTC	19	cgctac	26	TCCCTG	27	70,6	TATAGA	29	70,9	atctac	20	TCTAGA	23	GGGTTCC	72	TATAGAG	72	GGGTTCCC	144
MET_e21	DS		GGGTTC	19	cgctac	28	CAGAAG	27	71,7	ACACAC	22	70,7	atctac	29	TCTAGA	23	GGGTTCC	74	ACACACG	74	GGGTTCCC	148
EGFR_e20-2+	MS+	speci	GGGTTC	19	cgctac	22	CCTCAC	27	79,0								GGGTTCC	68			GGGTTCCC	152
EGFR_e20	MS-	speci	GGGTTC	19	cgctac	38	GGAAGC	21	71,7	CGTGGA	21	76,3	atctac	34	TCTAGA	23	GGGTTCC	78	CGTGGAC	78	GGGTTCCC	156
EGFR_e18	MS-	speci	GGGTTC	19	cgctac	30	GAAACT	31	70,6	GCTCCG	21	81,4	atctac	36	TCTAGA	23	GGGTTCC	80	GCTCCGG	80	GGGTTCCC	160
EGFR_e19+	MS+	speci	GGGTTC	19	cgctac	32	GTGAGA	31	71,5								GGGTTCC	82			GGGTTCCC	164
EGFR_e20-2	MS-	speci	GGGTTC	19	cgctac	38	CCTCAC	27	80,3	GCAGCT	21	75,8	atctac	40	TCTAGA	23	GGGTTCC	84	GCAGCTC	84	GGGTTCCC	168
ctrl_4	contro	chr11	GGGTTC	19	cgctac	44	TGCATG	23	70,4	GCTATG	29	70,5	atctac	34	TCTAGA	23	GGGTTCC	86	GCTATGT	86	GGGTTCCC	172

SALSA PCR Forward primer (Labeled): *GGGTTCCCTAAGGGTTGGA
SALSA PCR Reverse primer (Unlabeled): GTGCCAGCAAGATCCAATCTAGA

AC# V00604 Phage M13 genome
position: 3-99
5'-cgctactactattagtagaattgatgccacctttcagctcgcgccccaaatgaaaatatagctaaacaggttattgaccatttgcgaaatgtatctaatggtcaaactaaatctac-3'

Supplementary Table 2

Ordered oligonucleotide half-probes

ID	Sequence 5'-3'	Modification	Purification	Scale of synthesis	length
EGFR_e1_5'	GGGTTCCCTAAGGGTTGGAcgctaGAGCTCTTCGGGGAGCAGCGA	No	PGA	100nmole	45
ctrl_1_5'	GGGTTCCCTAAGGGTTGGAcgctacGGCCCAGATCACCGAGGAGGA	No	PGA	100nmole	46
EGFR_e7_5'	GGGTTCCCTAAGGGTTGGAcgctactCCCGAGGGCAAATACAGCTTTG	No	PGA	100nmole	48
EGFR_e28_5'	GGGTTCCCTAAGGGTTGGAcgctactacTGGGCAACCCCGAGTATCTCAA	No	PGA	100nmole	50
EGFR_e21_5'	GGGTTCCCTAAGGGTTGGAcgctaGCATGTCAAGATCACAGATTTTGGGCT	No	PGA	100nmole	51
EGFR_e8_5'	GGGTTCCCTAAGGGTTGGAcgctactactaATGTGGTGACAGATCACGGCTCG	No	PGA	100nmole	53
EGFR_e21+_5'	GGGTTCCCTAAGGGTTGGAcgctactactaGCATGTCAAGATCACAGATTTTGGGCG	No	PGA	100nmole	57
ctrl_2_5'	GGGTTCCCTAAGGGTTGGAcgctactactatCAGCTGGACGAGTACCAGGAGCTT	No	PGA	100nmole	55
EGFR_e5_5'	GGGTTCCCTAAGGGTTGGAcgctactactattCAAAAGTGTGATCCAAGCTGTCCCA	No	PGA	100nmole	57
ERBB2_e2_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaCACCTCTACCAGGGCTGCCAG	No	PGA	100nmole	59
MET_e2_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtGAGGAAGACCTTCAGAAGGTTGCTG	No	PGA	100nmole	60
ctrl_5_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaatAGTCCTGTGGCTACGGCACCAA	No	PGA	100nmole	62
EGFR_e2_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtaTTCTCAGCCTCCAGAGGATGTTCAATAA	No	PGA	100nmole	64
EGFR_e19_5'	GGGTTCCCTAAGGGTTGGAcgcCCGTCGCTATCAAGGAATTAAGAGAAGC	No	PGA	100nmole	50
EGFR_e4_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaatAGTCAGCAGTGACTTTCTCAGCAACATG	No	PGA	100nmole	68
ERBB2_e26_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacCCCAGCCCTCTACAGCGGTAC	No	PGA	100nmole	70
ctrl_3_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgTCCCTGCGCCATTGAGGTCTATAAAAT	No	PGA	100nmole	72
MET_e21_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccCAGAAGATAACGCTGATGATGAGGTGG	No	PGA	100nmole	74
EGFR_e20-2+_5	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattCCTCACCTCCACCGTGCAGCTCATCAT	No	PGA	100nmole	68
EGFR_e20_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttcagGGAAGCCTACGTGATGGCCAG	No	PGA	100nmole	78
EGFR_e18_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacGAAACTGAATTCAAAAAGATCAAAGTGCTGG	No	PGA	100nmole	80
EGFR_e19+_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctGTGAGAAAGTTAAAATTCCCGTCGCTATCAA	No	PGA	100nmole	82
EGFR_e20-2_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttcagCCTCACCTCCACCGTGCAGCTCATCAC	No	PGA	100nmole	84
ctrl_4_5'	GGGTTCCCTAAGGGTTGGAcgctactactattagtagaattgatgccacctttcagctcgcgTGCATGTTTGGAGCATCGACACA	No	PGA	100nmole	86
EGFR_e1_3'	TGCGACCCTCCGGGACGGCCGcTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	45
ctrl_1_3'	GGCAAAACTTCTGGCCCAGAAGacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	47
EGFR_e7_3'	GTGCCACCTGCGTGAAGAAGTctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	48
EGFR_e28_3'	CACTGTCCAGCCCACCTGTGTtctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	49
EGFR_e21_3'	GGCCAAACTGCTGGGTGCGGAaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	51
EGFR_e8_3'	TGCGTCCGAGCCTGTGGGGCCaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	52
ctrl_2_3'	CTGGACATCAAGCTGGCCCTGaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	56
EGFR_e5_3'	ATGGGAGCTGCTGGGGTGCAGaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	57
ERBB2_e2_3'	GTGGTGCAGGGAAACCTGGAAcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	58
MET_e2_3'	AGTACAAGACTGGGCCTGTGCTGGaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	60
ctrl_5_3'	AGACGAGGACTACGGCTGCGTCggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	62
EGFR_e2_3'	CTGTGAGGTGGTCCTTGGGAATTtggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	64
EGFR_e19_3'	AACATCTCCGAAAGCCAACAAGGAAAgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	82
EGFR_e4_3'	TCGATGGACTTCCAGAACCACCTGtaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	68
ERBB2_e26_3'	AGTGAGGACCCCACAGTACCCCtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	70
ctrl_3_3'	TATAGAGAAAGTTGATTACCCCCGGGATGaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	72
MET_e21_3'	ACACACGACCAGCCTCCTTCTGaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	74
EGFR_e20_3'	CGTGGACAACCCCCACGTGTGtgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	78
EGFR_e18_3'	GCTCCGGTGCGTTCGGCACGGtttgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	80
EGFR_e20-2_3'	GCAGCTCATGCCCTTCGGCTGaccatttgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	84
ctrl_4_3'	GCTATGTTAGAAGAAATGCTGTTTTGGCCtgcgaaatgtatctaatggtcaaactaaatctacTCTAGATTGGATCTTGCTGGCGC	5'-phosphate	PGA	100nmole	86

(continued on the next page)

Suppelmentary Table 2 (continued)

Positions of 0.2uM primers in 96-plate

	1	2	3	4	5	6	7	8	9	10	11	12
A	ctrl_1_5'	ctrl_1_3'	EGFR_e19+_5'	Blank	EGFR_e4_5'	EGFR_e4_3'	Blank	Blank	Blank	Blank	Blank	Blank
B	ctrl_2_5'	ctrl_2_3'	EGFR_e2_5'	EGFR_e2_3'	EGFR_e5_5'	EGFR_e5_3'	Blank	Blank	Blank	Blank	Blank	Blank
C	ctrl_3_5'	ctrl_3_3'	EGFR_e20_5'	EGFR_e20_3'	EGFR_e7_5'	EGFR_e7_3'	Blank	Blank	Blank	Blank	Blank	Blank
D	ctrl_4_5'	ctrl_4_3'	EGFR_e20-2_5'	EGFR_e20-2_3'	EGFR_e8_5'	EGFR_e8_3'	Blank	Blank	Blank	Blank	Blank	Blank
E	ctrl_5_5'	ctrl_5_3'	EGFR_e20-2+_5'	Blank	ERBB2_e2_5'	ERBB2_e2_3'	Blank	Blank	Blank	Blank	Blank	Blank
F	EGFR_e1_5'	EGFR_e1_3'	EGFR_e21_5'	EGFR_e21_3'	ERBB2_e26_5'	ERBB2_e26_3'	Blank	Blank	Blank	Blank	Blank	Blank
G	EGFR_e18_5'	EGFR_e18_3'	EGFR_e21+_5'	Blank	MET_e2_5'	MET_e2_3'	Blank	Blank	Blank	Blank	Blank	Blank
H	EGFR_e19_5'	EGFR_e19_3'	EGFR_e28_5'	EGFR_e28_3'	MET_e21_5'	MET_e21_3'	Blank	Blank	Blank	Blank	Blank	Blank

2

Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozlowski P

"Copy number variation of microRNA genes in the human genome"

BMC Genomics 2011, 12:183

Marcinkowska et al. BMC Genomics 2011, 12:183
http://www.biomedcentral.com/1471-2164/12/183

BMC Genomics

RESEARCH ARTICLE Open Access

Copy number variation of microRNA genes in the human genome

Malgorzata Marcinkowska[1], Maciej Szymanski[2], Wlodzimierz J Krzyzosiak[1] and Piotr Kozlowski[1*]

Abstract

Background: MicroRNAs (miRNAs) are important genetic elements that regulate the expression of thousands of human genes. Polymorphisms affecting miRNA biogenesis, dosage and target recognition may represent potentially functional variants. The functional consequences of single nucleotide polymorphisms (SNPs) within critical miRNA sequences and outside of miRNA genes were previously demonstrated using both experimental and computational methods. However, little is known about how copy number variations (CNVs) affect miRNA genes.

Results: In this study, we analyzed the co-localization of all miRNA *loci* with known CNV regions. Using bioinformatic tools we identified and validated 209 copy number variable miRNA genes (CNV-miRNAs) in CNV regions deposited in Database of Genomic Variations (DGV) and 11 CNV-miRNAs in two sets of CNVs defined as highly polymorphic. We propose potential mechanisms of CNV-mediated variation of functional copies of miRNAs (dosage) for different types of CNVs overlapping miRNA genes. We also showed that, consistent with their essential biological functions, miRNA *loci* are underrepresented in highly polymorphic and well-validated CNV regions.

Conclusion: We postulate that CNV-miRNAs are potential functional variants and should be considered high priority candidate variants in genotype-phenotype association studies.

Background

MicroRNAs (miRNAs) are a family of short (~20 nt), single-stranded, noncoding RNAs that are primarily involved in post-transcriptional down-regulation of gene expression in most eukaryotes [1]. Specific miRNAs are engaged in a variety of processes, including development, cell proliferation, differentiation and apoptosis [2]. Numerous studies have demonstrated that aberrant over-expression or down-regulation of certain miRNAs contribute to carcinogenesis and that these miRNAs can therefore be classified as either oncogenes (oncomirs) or tumor suppressors, respectively [3].

Mature, functional miRNAs are generated from primary precursors (pri-miRNA) encoded either by independent transcriptional units or within protein- or RNA-coding genes. In mammals, maturation of miRNAs involves two subsequent RNA cleavage steps. The first step takes place in the nucleus and is carried out by the Drosha nuclease to produce the secondary precursor (pre-miRNA) [4]. The pre-miRNAs (~60 nt) possess a hairpin structure, with the double-stranded portion interrupted by one or more mismatched nucleotides. Upon export to the cytoplasm, the pre-miRNA is further processed into an miRNA duplex by the RNAse III Dicer; [5] one of the duplex strands (passenger) is released, and the other serves as the mature miRNA [6]. The miRNA-induced silencing complex (miRISC) interacts with complementary target sequences, which are usually located within the 3' untranslated regions (3'UTRs) of mRNAs, causing mRNA degradation or inhibition of translation [7-9].

It is estimated that, in humans and other mammals, the expression of at least one-third of protein-coding genes is fine-tuned by approximately 1,000 miRNAs [10,11]. Currently, over 700 human miRNAs have been identified, and their sequences are deposited in miRBase (the microRNA database; http://www.mirbase.org).

Polymorphisms in miRNA genes can affect the expression of many downstream-regulated genes [12,13]. The most common form of polymorphism that affects the function of an miRNA (e.g., the structure of miRNA precursors, the efficiency of miRNA biogenesis and

* Correspondence: kozlowp@yahoo.com
[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland
Full list of author information is available at the end of the article

BioMed Central
© 2011 Marcinkowska et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

miRNA-target recognition) is the single nucleotide polymorphism (SNP). Computational and experimental studies have revealed many SNPs located in different parts of pre-miRNA sequences [14-16]. The occurrence of SNPs (including INDELs) in pre-miRNA regions is significantly lower than that in the surrounding reference sequences [16]. While sequences of mature miRNAs are the most conserved, the sequences of anti-miRNAs and the stems (outside miRNA and anti-miRNA) and loops of pre-miRNAs are somewhat less conserved [16]. SNPs naturally occurring within pre-miRNA sequences may affect miRNA biogenesis and impair miRNA-mediated gene silencing, as demonstrated by functional assays [15,17]. Recently, large genome-wide association study has demonstrated that also SNPs located outside (>14 kb) of pre-miRNA sequences can modulate miRNA expression both as *cis*- and *trans*-regulators (miRNA-eQTLs). One of identified miRNA-eQTLs (rs1522653) was shown to correlate with expression of 5 different miRNAs [18].

MiRNA target sites are also conserved genetic elements. Bioinformatic analyses show that SNPs are underrepresented in both experimentally validated and computationally predicted miRNA target sites, [16,19] and SNPs have the potential to either disrupt or create new miRNA target sites [19]. It has also been proposed that target site polymorphisms may play a role in evolution by altering miRNA specificity and function.

However, little is known about copy number variation (CNV) of miRNA genes. CNVs are segments of genomic DNA (roughly 1 kb to 1 Mb in length) that show variable numbers of copies in the genome due to deletions or duplications. CNVs recurrently occurring in a population are often called copy number polymorphisms (CNPs). Only a few CNV discovery studies report the presence of miRNAs in detected CNV regions and recognize their potential consequences [20-22]. Indeed, it was suggested that a comprehensive analysis of the co-localization of miRNAs and CNVs is needed [12].

Numerous studies show that CNVs can influence the expression of protein-coding genes in a copy number-dependent manner [23-25]. Recent results of genome-wide association study has confirmed such association for dozens of protein-coding genes and showed that CNVs capture at least 18% of the total detected genetic variation in gene expression [26]. It seems obvious that the expression of miRNA genes can also be modified by CNVs. This notion is supported by results from cancer genetics studies. For instance, there is a correlation between somatic copy number variation and the expression of miRNA genes, and miRNA genes recurrently amplified or lost in cancer genomes can serve as oncogenes or cancer suppressor genes, respectively [27-31].

In this study, by comparing the coordinates of human miRNAs with different sets of CNV regions (DGV-deposited and highly polymorphic), we identified over 200 human copy number variable miRNA *loci*. By comparing fractions of miRNAs and the genome that are covered by differentially validated CNV regions, we showed that miRNA *loci* are underrepresented in highly polymorphic CNVs, but not in CNVs deposited in the DGV database. We discuss the potential functional relevance of identified copy number variable miRNAs and propose models of how different types of CNVs can affect miRNA dosage.

Results and Discussion

Prior to bioinformatic identification of copy number variable miRNA genes (CNV-miRNAs), we compared the frequency of SNPs in annotated pre-miRNA sequences (3.7 SNPs/1,000 bp) and in reference human genome (4.8 SNPs/1,000 bp). Significantly lower number of SNPs in the pre-miRNA sequences (Fisher's exact test; $p < 0.0001$) most likely results from SNP purification effect and confirms general conservation of the analyzed pre-miRNA sequences. These analyses confirmed a SNP purification effect in pre-miRNA sequences reported previously [16]. The much higher number of SNPs identified in annotated pre-miRNA sequences in our study (N = 229; Additional file 1) versus N = 65 reported previously [16] results from the increased number of both SNPs (dbSNP - build 130; Apr 30, 2009; only annotated as 'single'; ~14 million SNPs) and miRNAs (miRBase - v 13.0), available in versions of databases used in this study.

To identify CNV-miRNAs, we compared the positions of miRNA *loci* with three sets of CNVs: 'DGV-deposited' (N = 29133; 30% genome coverage), 'polymorphic-SMC' (N = 1319; 1.2% genome coverage) [32] and 'polymorphic-DC' (N = 5037; 2.3% genome coverage) [22] CNVs. 'DGV-deposited' CNVs include all 29133 CNVs deposited in the Database of Genomic Variants (DGV update Aug 05, 2009 - http://projects.tcag.ca/variation). Two sets of 'polymorphic' CNVs ('polymorphic-SMC' [32] and 'polymorphic-DC' [22]) include highly polymorphic CNVs (minor allele frequency >0.01) validated by high-quality genotyping in two recent CNV-discovery studies using CNV-dedicated high-density hybrid arrays (combining traditional SNP probes and probes targeting CNVs) [22,32]. In both of these studies, precise breakpoints and unambiguous copy numbers were determined for each analyzed sample. All 'DGV-deposited' CNV-miRNA regions were further characterized by the following validation factors: (i) number of publications reporting CNVs (references), (ii) number of overlapping CNVs (DGV records) and (iii) number of observations in discovery studies (frequency) (Additional file 2). Since

the exact boundaries of miRNA genes (including regulatory elements) are difficult to determine, we used the genomic coordinates of all pre-miRNA *loci* deposited in miRBase (v 13.0; N = 715) as a proxy of miRNA gene sequences (three pre-miRNA *loci* located in the mitochondrial genome were excluded from our analysis) [33,34]. We realize, however, that CNVs overlapping other functional regions of miRNA coding genes (e.g., promoters) can also affect miRNA biogenesis and functionality, and those CNVs will be missed in our analysis.

The CNV-miRNAs identified in 'DGV-deposited' CNVs (N = 209) and in two sets of 'polymorphic' CNVs (N = 4 and N = 8) are shown in Additional file 2 and Table 1, respectively. Top-validated 'DGV-deposited' CNV-miRNAs are also shown in Table 2. Most miRNA *loci* identified in 'polymorphic' CNVs also overlapped with top-validated 'DGV-deposited' CNV regions (Table 1 and Table 2). All 'polymorphic' CNV-miRNAs were relatively frequent (combined minor genotype frequency >0.1 in at least one HapMap population). Among the identified miRNA-CNVs, we found deletions (e.g., hsa-mir-384 and hsa-mir-1324), duplications (e.g., hsa-mir-1972 and hsa-mir-1977), and multiple duplications (multiallelic polymorphisms; e.g., hsa-mir-1233 and hsa-mir-1268). The number of observed copies ranged from 0 (e.g., hsa-mir-384 and hsa-mir-650) to 6 (e.g., hsa-mir-1268).

The sequences of miRNA deposited in miRBase are derived from discovery studies in which many strict miRNA verification criteria were applied (e.g. hairpin forming potential, evolutionary conservation, presence in multiple clones/sequence reads or homogeneity of the 5'end). The SNP frequency analysis presented in this study also confirmed global conservation of annotated pre-miRNA sequences. However, there is still a possibility that some of the miRNAs in the miRBase represent experimental artifacts of false positive discoveries [35]. To provide additional data that can further validate miRNAs identified in CNVs we have conducted bioinformatic analysis of their expression and conservation. Table 1 and Table 2 show that according to different miRNA expression resources summarized in mimiRNA

Table 1 miRNA *loci* localized in polymorphic CNV regions

miRNA ID	miRNA position	dupl.	CNV region position	genotypes	CNV ID	functional relevance	expression (mimiRNA/[18])	conservation
miRNAs localized in 'polymorphic-SMC' CNV regions								
mir-1268	chr15:20014593-20014644		chr15:19803370-20089386	2,3,4,5,6	2057	1) recurrently deleted in classical Hodgkin's lymphoma [47]	not reported/NA	primates
mir-1233	chr15:32607783-32607864	chr15	chr15:32487975-32617680	0,1,2,3	2082	1)	not reported/NA	primates
mir-1972	chr16:15011679-15011755	chr16	chr16:14897364-15016088	2,3,4	2141		not reported/NA	primates
mir-384	chrX:76056092-76056179		chrX:76053855-76057477	0,1,2	2648		in several tissues/NA	mammals
miRNAs localized in 'polymorphic-DC' CNV regions								
miRNA ID	**miRNA position**	**dupl.**	**CNV region position**	**genotypes**	**CNV ID**	**functional relevance**	**expression (mimiRNA/[18])**	**conservation**
mir-1977	chr1:556050-556128	chrM	chr1:554403-560267	2,3,4	3.1		not reported/NA	primates
mir-1324	chr3:75762604-75762699		chr3:75464498-75782745	1,2	1432.2		not reported/NA	primates
mir-548i-2	chr4:9166887-9167035		chr4:9117494-9354801	1,2	1815.3		not reported/NA	primates
mir-1275	chr6:34075727-34075806		chr6:34071086-34077139	1,2	2853.1	2) upregulated in blood cells of MS patients [41]	not reported/NA	primates
mir-1302-2	chr9:20144-20281	chr1, 15,19	chr9:485-38531	2,3	4134_full		not reported/NA	primates
mir-1233	chr15:32461562-32461643	chr15	chr15:32450046-32662643	2,3,4,5	6351.3	1)	not reported/NA	primates
mir-1233	chr15:32607783-32607864	chr15	chr15:32450046-32662643	2,3,4,5	6351.3	1)	not reported/NA	primates
mir-650	chr22:21495270-21495365		chr22:20711019-21578950	0,1,2	8103_full	1)	in several tissues (mostly ovary and ovary-derived cancers)/ high	primates

dupl. - localization of duplicated copies; mimiRNA/[18] - miRNA expression according to database mimiRNA/and according to resent result of expression analysis in primary fibroblast cells (high - high expression, absent - low or undetectable expression in fibroblast cells, NA - not analyzed).

Table 2 miRNA *loci* localized in CNV regions validated by multiple overlapping CNVs

miRNAs localized in 'DGV-deposited' CNV regions validated by multiple overlapping CNVs							
miRNA ID	**miRNA position**	**dupl.**	**minimal CNV region**	**# CNVs**	**functional relevance**	**expression (mimiRNA/[18])**	**conservation**
mir-1977	chr1:556050-556128	chrM	chr1:554340-569354	6		not reported/NA	primates
mir-149	chr2:241044091-241044179		chr2:241039698-241051687	6	3) downregulated in squamous cell carcinoma of the tongue [44]	in multiple tissues/ high	vertebrates
mir-566	chr3:50185763-50185856		chr3:50173490-50214015	7		in several tissues/ absent	primates
mir-1324	chr3:75762604-75762699		chr3:75761737-75839337	6		not reported/NA	primates
mir-570	chr3:196911452-196911548		chr3:196905807-196918722	9		in several tissues/ absent	primates
mir-548i-2	chr4:9166887-9167035		chr4:9152768-9182838	9		not reported/NA	primates
mir-548i-3	chr8:7983873-7984021		chr8:7965981-8024983	14		not reported/NA	primates
mir-383	chr8:14755318-14755390		chr8:14741501-14763659	8	4) downregulated in non-obstructive azoospermia [39]	in multiple tissues/ absent	vertebrates
mir-661	chr8:145091347-145091435		chr8:145090343-145104971	8	5) downregulates the expression of metastatic tumor antigen 1 (MTA1), inhibits the motility, invasiveness, anchorage-independent growth, and tumorigenicity of cancer cells [48]	in several tissues (mostly ovary and ovary-derived cancers)/absent	primates
mir-1299	chr9:68292059-68292141		chr9:68291272-68298205	7		not reported/NA	primates
mir-126	chr9:138684875-138684959		chr9:138680837-138688363	14	6) suppresses cell growth in colon cancer [43]; downregulates HOXA9, playing a role in the development of many organs and often upregulated in myeloid leukemias [37]; regulates angiogenic signaling and vascular integrity [38]; overexpressed in ALL and AML [42]	high, in multiple tissues/high	vertebrates
mir-202	chr10:134911006-134911115		chr10:134903011-134918923	10		in several tissues/ absent	vertebrates
mir-1268	chr15:20014593-20014644		chr15:19975453-20046356	37	1) see Table 1	not reported/NA	primates
mir-1233	chr15:32461562-32461643	chr15	chr15:32461525-32469857	9	1) see Table 1	not reported/NA	primates
mir-1233	chr15:32607783-32607864	chr15	chr15:32599966-32615283	17	1) see Table 1	not reported/NA	primates
mir-662	chr16:760184-760278		chr16:750040-764098	6		in several tissues/ absent	primates
mir-1972	chr16:68621750-68621826	chr11	chr16:68621490-68653097	6		not reported/NA	primates
mir-142	chr17:53763592-53763678		chr17:53751608-53767652	11	7) increased expression correlates with rejection of organ transplants [40]; overexpressed in pre-B-ALL patients [46]; potentially involved in the development of blood cancer or brain tumors [45]	high, in multiple tissues/absent	vertebrates
mir-1270	chr19:20371080-20371162		chr19:20370872-20383238	9		not reported/NA	primates
mir-663	chr20:26136822-26136914		chr20:26136626-26139184	6		in several tissues/NA	primates
mir-650	chr22:21495270-21495365		chr22:21494381-21502189	38	1) see Table 1	in several tissues/ high	primates
mir-514-2	chrX:146171153-146171240		chrX:146168796-146174575	6		in several tissues/NA	mammals
mir-514-3	chrX:146173851-146173938		chrX:146168796-146174575	6		in several tissues/NA	mammals

dupl. - localization of duplicated copies; mimiRNA/[18] - miRNA expression according to database mimiRNA/and according to resent result of expression analysis in primary fibroblast cells (high - high expression, absent - low or undetectable expression in fibroblast cells, NA - not analyzed).

database [36] over half (14/26) of top-validated CNV-miRNAs (Table 1 and Table 2) were shown to be expressed in at least several tissues/cell lines (detailed expression profiles are shown in Additional file 3). MiRNA whose expression is not reported in mimiRNA were either not analyzed for expression or did not show expression in the analyzed tissues. Additionally, three out of ten (30%) top-validated CNV-miRNAs (Table 1 and Table 2) which expression in primary fibroblast cell lines was analyzed by the micro-fluidics-based TaqMan Human MiRNA Array show high level of expression [18]. Based on the currently available sequence data for miRNAs deposited in miRBase and blast searches of the vertebrate genomic sequences we also determined evolutionary conservation of the miRNAs found in top-validated CNV regions. Most of these miRNAs seem to be specific only for primates. There are, however, 8 miRNAs that are conserved across mammals or vertebrates (Table 1 and Table 2).

The functional relevance of several of the CNV-miRNAs identified in this survey was previously reported in the literature (manual screening; Table 1 and Table 2). CNV-miRNAs are involved in many processes and phenotypes (diseases), including organ development [37], angiogenesis [38], male infertility [39], transplant rejection [40], multiple sclerosis [41] and cancer. Many CNV-miRNAs are specifically deleted, amplified or expressed in different types of cancers [42-47] and can regulate the expression of important cancer-related genes [37,48]. The copy number variation of those functionally relevant miRNAs can modulate or predispose one to the aforementioned phenotypes.

In the next step, we determined whether the overlap of CNVs and miRNA *loci* was random (null hypothesis) or whether the CNVs were underrepresented at these *loci* (alternative hypothesis). To test this hypothesis, we compared fractions of miRNA *loci* and fractions of the genome covered by differentially defined CNV regions. Figure 1A shows that the fraction of miRNA *loci* covered by two sets of 'polymorphic' CNVs is approximately two times lower than expected (fraction of the covered genome). Although this effect was only marginally significant (Figure 1A), it suggested that at least highly polymorphic CNVs are under negative (purifying) selection at miRNA genes. Conversely, the fraction of miRNAs (0.292) covered by 'DGV-deposited' CNVs corresponded almost exactly to the fraction of the genome covered by those CNVs (0.299). The CNV purification effect was not observed, even after narrowing 'DGV-deposited' CNV regions by different validation factors defined above (Figure 1B and 1C). The fact that the purifying effect did not apply to the 'DGV-deposited' CNVs suggested that a significant portion of these CNVs are very rare, private, or significantly oversized or represents false positive artifacts. This observation is consistent with the conclusions from other recently published results [32,49].

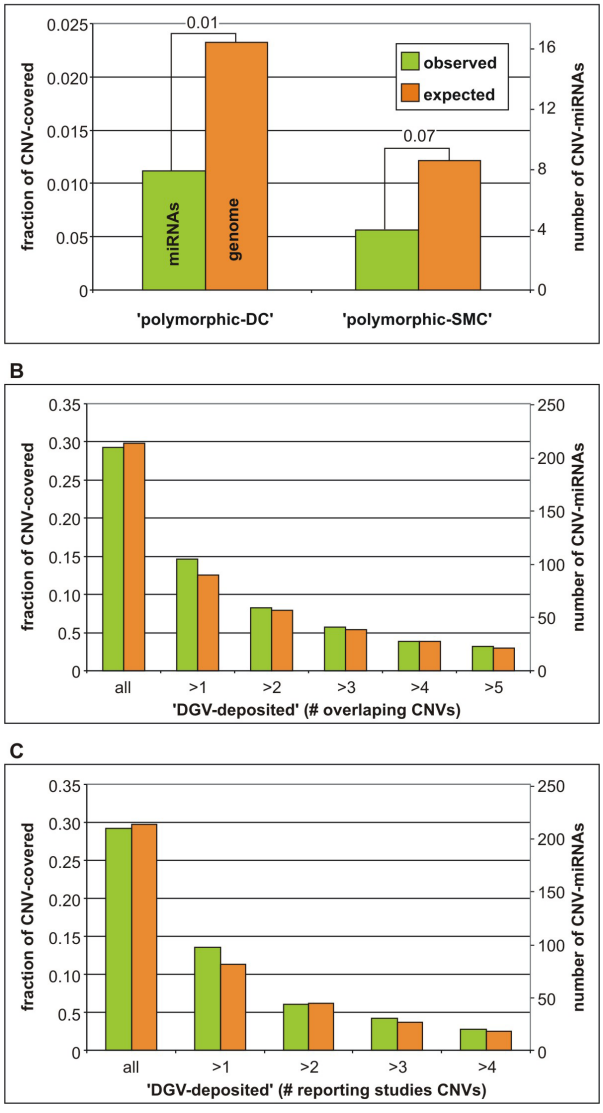


Figure 1 Comparison of observed and expected number (fraction) of miRNA *loci* located in different CNV regions. Expected values were estimated based on the fraction of the genome covered by CNVs. A) Graph showing the fractions of miRNA *loci* (observed number of CNV-miRNAs; green bars) and the genome (expected number of CNV-miRNAs; orange bars) covered by two sets of 'polymorphic' CNVs. Binomial probabilities of equal or lower than the observed number of miRNA *loci* covered by CNVs are indicated over the bars. B) and C) The fractions of miRNA *loci* and the genome covered by 'DGV-deposited' CNV regions gradually narrowed by the increasing number of overlapping CNVs (DGV records) (B) and the increasing number of reporting references (C).

Although copy number variation can influence gene expression through different mechanisms (e.g., position effect and deletion or duplication of regulatory elements

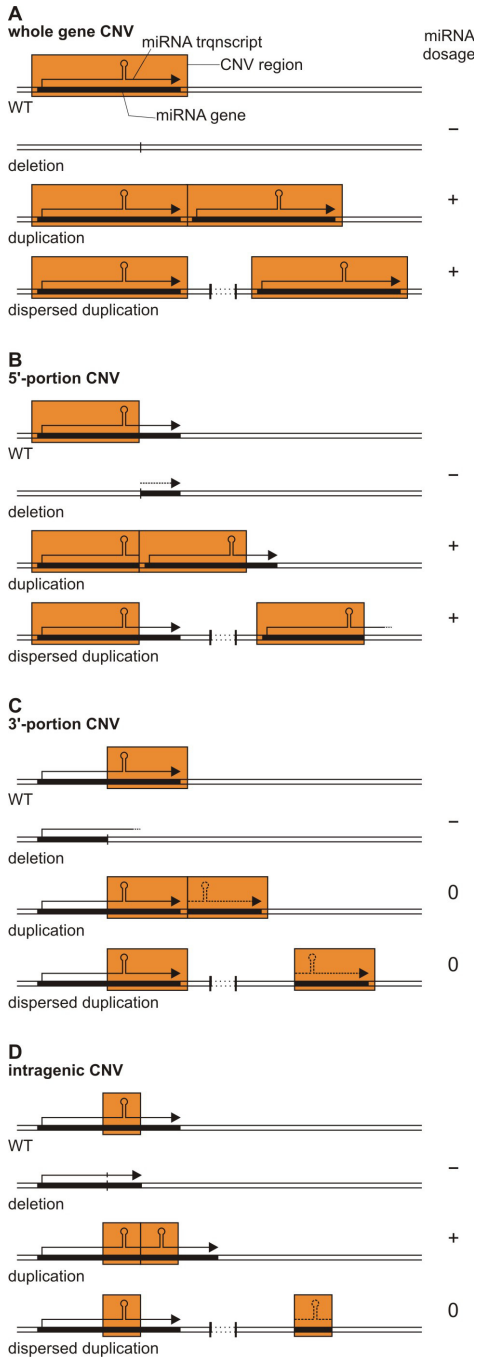


Figure 2 Potential mechanism of CNV-mediated variation of miRNA dosage. Schematic representation of an miRNA gene and its primary transcript (solid or dotted arrow-lines). The position of the pre-miRNA sequence is indicated as a hairpin-loop structure in the miRNA primary transcript. Dotted lines represent transcripts unlikely to be produced due to the lack of promoter and transcriptional start sequences. Orange boxes represent CNV regions (deletions, duplications and dispersed duplications). The following panels show a CNV spanning different parts of the miRNA gene: (A) whole gene, (B) 5'-portion, (C) 3'-portion and (D) intragenic region of the gene. +, - and 0 indicate potential increase, decrease and no change of miRNA dosage, respectively.

that control transcription or splicing), the most obvious mechanism is in the variability of dosage (number of functional copies). All of these mechanisms can affect both protein-coding and miRNA genes. However, mechanisms of dosage variation may be different for protein-coding and miRNA genes. In Figure 2, potential consequences of different CNV types overlapping different parts of miRNA genes are proposed. Not only whole gene amplification but also certain partial gene duplications (multiple duplications) can increase the dosage of miRNAs. Conversely, partial gene deletions may not always result in decreased miRNA dosage. This contrasts with the situation observed for protein-coding genes, in which only duplication of the entire gene (including the promoter and regulatory sequences) can lead to an increased number of functional copies, and almost every (even partial) gene deletion is deleterious.

Analysis of 11 miRNAs located in CNVs with well defined breakpoints (Table 1) showed that (i) 3 of these miRNAs are located in the protein coding genes which are entirely positioned within CNVs, (ii) 4 of the miRNAs are located in intergenic regions and are flanked by at least 20 kb of CNV sequences, (iii) 3 miRNAs are located in intergenic regions flanked by short CNV sequences (< 5 kb) and (iv) 1 miRNA is located in a gene of which the 3'end extends beyond CNV (Additional file 4). Taking into account the average size of a human gene (~30 kb) one can expect that miRNAs located in large CNVs (groups (i) and (ii)) will be expressed from genes entirely embedded within the CNV regions. According to the model presented in Figure 2A the expression of such miRNAs very likely will correlate with expression (number of copies) of genes from which these miRNAs are generated (no matter whether generated from protein-coding or non-coding transcripts). MiRNA located in short CNVs (group (iii)) most likely will form the tandem copies transcribed from one promoter. A number of such copies may modulate the number of miRNA precursors (pre-miRNAs) present in one primary transcript (pri-miRNA) and thus may modulate expression of miRNA (Figure 2D). Expression of miRNA whose gene only partially is embedded in CNV (iii) may be modified according to the model shown in Figure 2B and will depend on expression and stability of the transcript truncated at the 3'end. Moreover, it should be noted that some pre-miRNA sequences occur in the genome in multiple copies. Although the functionality of such copies is still mostly unknown, the duplicated copies of miRNA genes may mask the effect of copy number variations that usually affect only one copy.

Finally, not only common CNVs, but also CNVs implicated in specific diseases can affect miRNA *loci* and thus can play important role in pathogenesis. We

have identified 38 *loci* of miRNAs located in chromosomal regions implicated in microdeletion/microduplication syndromes (DECYPHER v5.0 [50]) (Additional file 5). For example, six miRNA *loci* (hsa-mir-185, hsa-mir-1306, hsa-mir-1286, hsa-mir-649, hsa-mir-301b and hsa-mir-130b) are located within genomic region implicated in DiGeorge syndrome. The role of somatic copy number variation of miRNA genes in cancer is extensively investigated in multiple studies (e.g. [27-31]) and was recently summarized in several review articles [51-53].

Conclusions

Although 'polymorphic' CNVs showed some purifying effects at miRNA *loci*, there were still many miRNA *loci* that overlapped with known CNV regions (Additional file 2 and Table 2), including those that are highly validated and confirmed by high-quality genotyping (Table 1). Taking into account the CNV genome coverage (1.2% 'polymorphic-SMC' and 2.3% 'polymorphic-DC') and the relatively small overlapping fractions (0.39 and 0.20, respectively) between the two sets of 'polymorphic' CNVs analyzed in this study, we estimated that up to 10% of the human genome is covered by highly polymorphic CNVs. This fraction corresponds to approximately 30 highly polymorphic CNV-miRNAs in the human genome (extrapolation of the fraction of miRNA *loci* covered by highly polymorphic CNVs analyzed in this study). It is likely that at least some of these *loci* are among the CNV-miRNAs identified from the top-validated 'DGV-deposited' CNVs (Table 2 and Additional file 2).

CNV-miRNAs are potential functional variants and should be considered high priority candidate variants in genotype-phenotype association studies, especially when they are located in regions implicated by linkage or association studies. As indicated in Table 1, only a small fraction of CNV-miRNAs were genotyped in three HapMap populations, which provides precise information about their polymorphisms. This is mostly due to the lack of appropriate methods for precise characterization of CNV polymorphisms. Although several genome-wide approaches that substantially fulfill the above requirement were proposed recently, a simple and inexpensive method that enables accurate characterization of several CNVs of interest in a large number of samples is still needed. The lack of such a method significantly hampers the analyses of CNVs and their correlation with the phenotype. To verify and characterize the polymorphisms of all CNV-miRNAs, we are developing several medium-throughput assays suited for large scale population studies that are focused on selected CNVs of potential functional effect. These assays will take advantage of the MLPA-based strategy proposed previously [54-56].

Methods

Genomic coordinates (hg18) of 718 human miRNA *loci*, 13 600 093 SNPs (only annotated as 'single'), 29 133 CNVs (only annotated as 'Copy Number') and 58 *loci* implicated in microdeletion syndromes were downloaded from miRBase v13.0 http://www.mirbase.org, dbSNP build 130; Apr 30, 2009, Database of Genomic Variants update Aug 05, 2009 http://projects.tcag.ca/variation and DECIPHER database v5.0 [50]http://decipher.sanger.ac.uk, respectively. The coordinates of 1319 CNVs described as 'polymorphic-SMC' and 5037 CNVs described as 'polymorphic-DC' were extracted from supplementary materials of references [32] and [22], respectively. The number of miRNA *loci* and fraction of genome covered by CNV regions were calculated using 'feature coverage' and 'base coverage' tools available on the Galaxy, web portal for large-scale interactive data analyses [57].

The expression profiles of CNV-miRNAs were generated with the use of mimiRNA database [36] that summarizes expression data from miRNA Atlas [58], quantitative real-time PCR [59,60] as well as microarray and deep sequencing data from GEO (Gene Expression Omnibus) [61]. The assessment of evolutionary conservation of microRNAs was done based on the data available at the miRBase and blast searches of the vertebrate genomic sequences with human pre-microRNAs.

All statistical analyses were performed using Statistica (StatSoft, Tulsa, OK). The Fisher's exact test for comparison of SNPs frequency in the annotated miRNA sequences and in the total genome sequence was calculated as described in [62], with the use of the online tool available on webpage http://www.langsrud.com/fisher.htm.

Additional material

Additional file 1: SNPs identified in pre-miRNA sequences. Excel table containing list of SNPs identified in annotated pre-miRNA sequences.

Additional file 2: miRNA identified in CNV regions. Excel table containing list of pre-miRNA annotated sequences identified in 'DGV-deposited' CNVs.

Additional file 3: Expression profiles of selected CNV-miRNAs. Expression profiles of selected CNV-miRNAs generated with the use of mimiRNA database [36]. The expression of all miRNAs was normalized in each tissue to a standard score spanning 1-1,000 (1,000 represents highest expression observed in tissue). The bars represent mean expression measured in multiple experiments and the error bars represent standard error of the mean. The variability of the expression level is indicated by colors (red - lowest variability; yellow - highest variability). Details can be found on mimiRNA webpage http://mimirna.centenary.org.au and in [36].

Additional file 4: miRNAs located in CNVs with well defined breakpoints. Excel table showing characteristics of miRNAs located in CNVs with well defined breakpoints.

Additional file 5: miRNAs located in chromosomal regions implicated in microdeletion/microduplication syndromes. Excel table containing list of miRNAs located in chromosomal regions implicated in microdeletion/microduplication syndromes (DECYPHER v5.0 [50]).

Acknowledgements
This work was supported by the Ministry of Science and Higher Education [N N302 278937, N N302 260938].
The authors have declared no conflict of interest.

Author details
[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland. [2]Computational Genomics Laboratory, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznan, Poland.

Authors' contributions
MM performed the computational analysis, literature screening, participated in the manuscript preparation. MS participated in the computational analysis (sequence conservation analysis) and the manuscript preparation. WJK participated in the design of the study and in the manuscript preparation. PK performed the statistical analysis, conceived of the study, and participated in its design and coordination. All authors have read and approved the final manuscript.

Received: 24 March 2010 Accepted: 12 April 2011
Published: 12 April 2011

References

1. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function.** *Cell* 2004, **116**: 281-297.
2. Kim VN, Nam JW: **Genomics of microRNA.** *Trends Genet* 2006, **22**: 165-173.
3. Esquela-Kerscher A, Slack FJ: **Oncomirs - microRNAs with a role in cancer.** *Nat Rev Cancer* 2006, **6**: 259-269.
4. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN: **The nuclear RNase III Drosha initiates microRNA processing.** *Nature* 2003, **425**: 415-419.
5. Bernstein E, Caudy AA, Hammond SM, Hannon GJ: **Role for a bidentate ribonuclease in the initiation step of RNA interference.** *Nature* 2001, **409**: 363-366.
6. Hammond SM, Bernstein E, Beach D, Hannon GJ: **An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells.** *Nature* 2000, **404**: 293-296.
7. Guo H, Ingolia NT, Weissman JS, Bartel DP: **Mammalian microRNAs predominantly act to decrease target mRNA levels.** *Nature* 2010, **466**: 835-840.
8. Pillai RS, Bhattacharyya SN, Artus CG, Zoller T, Cougot N, Basyuk E, Bertrand E, Filipowicz W: **Inhibition of translational initiation by Let-7 MicroRNA in human cells.** *Science* 2005, **309**: 1573-1576.
9. Yekta S, Shih IH, Bartel DP: **MicroRNA-directed cleavage of HOXB8 mRNA.** *Science* 2004, **304**: 594-596.
10. Lewis BP, Burge CB, Bartel DP: **Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.** *Cell* 2005, **120**: 15-20.
11. Rajewsky N: **microRNA target predictions in animals.** *Nat Genet* 2006, **38(Suppl)**: S8-13.
12. Borel C, Antonarakis SE: **Functional genetic variation of human miRNAs and phenotypic consequences.** *Mamm Genome* 2008, **19**: 503-509.
13. Georges M, Coppieters W, Charlier C: **Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease.** *Curr Opin Genet Dev* 2007, **17**: 166-176.
14. Iwai N, Naraba H: **Polymorphisms in human pre-miRNAs.** *Biochem Biophys Res Commun* 2005, **331**: 1439-1444.
15. Duan R, Pak C, Jin P: **Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA.** *Hum Mol Genet* 2007, **16**: 1124-1131.
16. Saunders MA, Liang H, Li WH: **Human polymorphism at microRNAs and microRNA target sites.** *Proc Natl Acad Sci USA* 2007, **104**: 3300-3305.
17. Sun G, Yan J, Noltner K, Feng J, Li H, Sarkis DA, Sommer SS, Rossi JJ: **SNPs in human miRNA genes affect biogenesis and function.** *RNA* 2009, **15**: 1640-1651.
18. Borel C, Deutsch S, Letourneau A, Migliavacca E, Montgomery SB, Dimas AS, Vejnar CE, Attar H, Gagnebin M, Gehrig C, *et al*: **Identification of cis- and trans-regulatory variation modulating microRNA expression levels in human fibroblasts.** *Genome Res* 2011, **21**: 68-73.
19. Chen K, Rajewsky N: **Natural selection on human microRNA binding sites inferred from SNP data.** *Nat Genet* 2006, **38**: 1452-1456.
20. Wong KK, deLeeuw RJ, Dosanjh NS, Kimm LR, Cheng Z, Horsman DE, MacAulay C, Ng RT, Brown CJ, Eichler EE, Lam WL: **A comprehensive analysis of common copy-number variations in the human genome.** *Am J Hum Genet* 2007, **80**: 91-104.
21. Lin CH, Li LH, Ho SF, Chuang TP, Wu JY, Chen YT, Fann CS: **A large-scale survey of genetic copy number variations among Han Chinese residing in Taiwan.** *BMC Genet* 2008, **9**: 92.
22. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, *et al*: **Origins and functional impact of copy number variation in the human genome.** *Nature* 2010.
23. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, *et al*: **Diet and the evolution of human amylase gene copy number variation.** *Nat Genet* 2007, **39**: 1256-1260.
24. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, *et al*: **The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility.** *Science* 2005, **307**: 1434-1440.
25. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36**: 949-951.
26. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, *et al*: **Relative impact of nucleotide and copy number variation on gene expression phenotypes.** *Science* 2007, **315**: 848-853.
27. Bottoni A, Piccin D, Tagliati F, Luchin A, Zatelli MC, degli Uberti EC: **miR-15a and miR-16-1 down-regulation in pituitary adenomas.** *J Cell Physiol* 2005, **204**: 280-285.
28. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, Aldler H, Rattan S, Keating M, Rai K, *et al*: **Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia.** *Proc Natl Acad Sci USA* 2002, **99**: 15524-15529.
29. Zhang L, Huang J, Yang N, Greshock J, Megraw MS, Giannakakis A, Liang S, Naylor TL, Barchetti A, Ward MR, *et al*: **microRNAs exhibit high frequency genomic alterations in human cancer.** *Proc Natl Acad Sci USA* 2006, **103**: 9136-9141.
30. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, Yoshida Y, Seto M: **Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma.** *Cancer Res* 2004, **64**: 3087-3095.
31. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM: **A microRNA polycistron as a potential human oncogene.** *Nature* 2005, **435**: 828-833.
32. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, *et al*: **Integrated detection and population-genetic analysis of SNPs and copy number variation.** *Nat Genet* 2008, **40**: 1166-1174.
33. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: **miRBase: tools for microRNA genomics.** *Nucleic Acids Res* 2008, **36**: D154-158.
34. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature.** *Nucleic Acids Res* 2006, **34**: D140-144.
35. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, *et al*: **Mammalian microRNAs: experimental evaluation of novel and previously annotated genes.** *Genes Dev* 2010, **24**: 992-1009.
36. Ritchie W, Flamant S, Rasko JE: **mimiRNA: a microRNA expression profiler and classification resource designed to identify functional correlations between microRNAs and their targets.** *Bioinformatics* 2010, **26**: 223-227.
37. Shen WF, Hu YL, Uttarwar L, Passegue E, Largman C: **MicroRNA-126 regulates HOXA9 by binding to the homeobox.** *Mol Cell Biol* 2008, **28**: 4609-4619.

38. Fish JE, Santoro MM, Morton SU, Yu S, Yeh RF, Wythe JD, Ivey KN, Bruneau BG, Stainier DY, Srivastava D: **miR-126 regulates angiogenic signaling and vascular integrity.** *Dev Cell* 2008, **15**: 272-284.
39. Lian J, Zhang X, Tian H, Liang N, Wang Y, Liang C, Li X, Sun F: **Altered microRNA expression in patients with non-obstructive azoospermia.** *Reprod Biol Endocrinol* 2009, **7**: 13.
40. Anglicheau D, Sharma VK, Ding R, Hummel A, Snopkowski C, Dadhania D, Seshan SV, Suthanthiran M: **MicroRNA expression profiles predictive of human renal allograft status.** *Proc Natl Acad Sci USA* 2009, **106**: 5330-5335.
41. Keller A, Leidinger P, Lange J, Borries A, Schroers H, Scheffler M, Lenhof HP, Ruprecht K, Meese E: **Multiple sclerosis: microRNA expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls.** *PLoS One* 2009, **4**: e7440.
42. Zhang H, Luo XQ, Zhang P, Huang LB, Zheng YS, Wu J, Zhou H, Qu LH, Xu L, Chen YQ: **MicroRNA patterns associated with clinical prognostic parameters and CNS relapse prediction in pediatric acute leukemia.** *PLoS One* 2009, **4**: e7826.
43. Guo C, Sah JF, Beard L, Willson JK, Markowitz SD, Guda K: **The noncoding RNA, miR-126, suppresses the growth of neoplastic cells by targeting phosphatidylinositol 3-kinase signaling and is frequently lost in colon cancers.** *Genes Chromosomes Cancer* 2008, **47**: 939-946.
44. Wong TS, Liu XB, Wong BY, Ng RW, Yuen AP, Wei WI: **Mature miR-184 as Potential Oncogenic microRNA of Squamous Cell Carcinoma of Tongue.** *Clin Cancer Res* 2008, **14**: 2588-2592.
45. Rossi S, Sevignani C, Nnadi SC, Siracusa LD, Calin GA: **Cancer-associated genomic regions (CAGRs) and noncoding RNAs: bioinformatics and therapeutic implications.** *Mamm Genome* 2008, **19**: 526-540.
46. Ju X, Li D, Shi Q, Hou H, Sun N, Shen B: **Differential microRNA expression in childhood B-cell precursor acute lymphoblastic leukemia.** *Pediatr Hematol Oncol* 2009, **26**: 1-10.
47. Hartmann S, Martin-Subero JI, Gesk S, Husken J, Giefing M, Nagel I, Riemke J, Chott A, Klapper W, Parrens M, *et al*: **Detection of genomic imbalances in microdissected Hodgkin and Reed-Sternberg cells of classical Hodgkin's lymphoma by array-based comparative genomic hybridization.** *Haematologica* 2008, **93**: 1318-1326.
48. Reddy SD, Pakala SB, Ohshiro K, Rayala SK, Kumar R: **MicroRNA-661, a c/EBPalpha target, inhibits metastatic tumor antigen 1 and regulates its functions.** *Cancer Res* 2009, **69**: 5639-5642.
49. Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, Krauss RM, Myers RM, Ridker PM, Chasman DI, *et al*: **Population analysis of large copy number variants and hotspots of human genetic disease.** *Am J Hum Genet* 2009, **84**: 148-161.
50. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP: **DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources.** *Am J Hum Genet* 2009, **84**: 524-533.
51. Deng S, Calin GA, Croce CM, Coukos G, Zhang L: **Mechanisms of microRNA deregulation in human cancer.** *Cell Cycle* 2008, **7**: 2643-2646.
52. Di Leva G, Croce CM: **Roles of small RNAs in tumor formation.** *Trends Mol Med* 2010, **16**: 257-267.
53. Ruan K, Fang X, Ouyang G: **MicroRNAs: novel regulators in the hallmarks of human cancer.** *Cancer Lett* 2009, **285**: 116-126.
54. Kozlowski P, Jasinska AJ, Kwiatkowski DJ: **New applications and developments in the use of multiplex ligation-dependent probe amplification.** *Electrophoresis* 2008, **29**: 4627-4636.
55. Kozlowski P, Roberts P, Dabora S, Franz D, Bissler J, Northrup H, Au KS, Lazarus R, Domanska-Pakiela D, Kotulska K, *et al*: **Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations.** *Hum Genet* 2007, **121**: 389-400.
56. Marcinkowska M, Wong KK, Kwiatkowski DJ, Kozlowski P: **Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example.** *ScientificWorldJournal* 2010, **10**: 2003-2018.
57. Taylor J, Schenck I, Blankenberg D, Nekrutenko A: **Using galaxy to perform large-scale interactive data analyses.** *Curr Protoc Bioinformatics* 2007, **Chapter 10**: Unit 10 15.
58. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, *et al*: **A mammalian microRNA expression atlas based on small RNA library sequencing.** *Cell* 2007, **129**: 1401-1414.
59. Gaur A, Jewell DA, Liang Y, Ridzon D, Moore JH, Chen C, Ambros VR, Israel MA: **Characterization of microRNA expression levels and their biological correlates in human cancer cell lines.** *Cancer Res* 2007, **67**: 2456-2468.
60. Lee EJ, Baek M, Gusev Y, Brackett DJ, Nuovo GJ, Schmittgen TD: **Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors.** *RNA* 2008, **14**: 35-42.
61. Barrett T, Edgar R: **Gene expression omnibus: microarray data storage, submission, retrieval, and analysis.** *Methods Enzymol* 2006, **411**: 352-369.
62. Agresti A: **A Survey of Exact Inference for Contingency Tables.** *Statist Sci* 1992, **7**: 131-153.

doi:10.1186/1471-2164-12-183
Cite this article as: Marcinkowska *et al.*: Copy number variation of microRNA genes in the human genome. *BMC Genomics* 2011 **12**:183.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

MATERIAŁY UZUPEŁNIAJĄCE DO PUBLIKACJI

Marcinkowska i wsp., *BMC Genomics* 2011

Additional file 1

SNPs identified in pre-miRNA sequences

miRNA ID	miRNA chromosomal localization	nuber of SNPs in anotated pre-miRNA sequence	SNP ID
hsa-mir-1302-2	chr1:20229-20366	5	rs11266858, rs4248191, rs11266859, rs422582, rs422363
hsa-mir-1977	chr1:556050-556128	4	rs9783068, rs41453547, rs9701099, rs2854138
hsa-mir-200b	chr1:1092347-1092441	1	rs72563729
hsa-mir-34a	chr1:9134314-9134423	2	rs72631823, rs35301225
hsa-mir-92b	chr1:153431592-153431687	1	rs12759620
hsa-mir-215	chr1:218357818-218357927	1	rs72631834
hsa-mir-559	chr2:47458318-47458413	1	rs58450758
hsa-mir-217	chr2:56063606-56063715	1	rs41291173
hsa-mir-216a	chr2:56069589-56069698	1	rs41291179
hsa-mir-1285-2	chr2:70333554-70333641	1	rs72904307
hsa-mir-1302-3	chr2:114057006-114057143	4	rs2441622, rs2441621, rs7589328, rs6542147
hsa-mir-663b	chr2:132731009-132731123	1	rs62165009
hsa-mir-1978	chr2:149355835-149355887	2	rs55723650, rs56489998
hsa-mir-1244	chr2:232286268-232286352	1	rs1804520
hsa-mir-149	chr2:241044091-241044179	2	rs71428439, rs2292832
hsa-mir-564	chr3:44878384-44878477	1	rs2292181
hsa-mir-1324	chr3:75762604-75762699	6	rs28620398, rs7614638, rs3008994, rs10155043, rs3008993, rs58827088
hsa-mir-568	chr3:115518012-115518106	1	rs28632138
hsa-mir-1248	chr3:187987155-187987260	1	rs73063489
hsa-mir-570	chr3:196911452-196911548	1	rs9860655
hsa-mir-943	chr4:1957909-1958002	1	rs1077020
hsa-mir-1255b-1	chr4:36104383-36104445	1	rs6841938
hsa-mir-1269	chr4:66825137-66825241	1	rs73239138
hsa-mir-1255a	chr4:102470482-102470594	1	rs28664200
hsa-mir-576	chr4:110629303-110629400	1	rs71603032
hsa-mir-577	chr4:115797364-115797459	1	rs34115976
hsa-mir-580	chr5:36183751-36183847	1	rs73080005
hsa-mir-1274a	chr5:41511491-41511561	1	rs318039
hsa-mir-581	chr5:53283091-53283186	3	rs788517, rs1694089, rs810917
hsa-mir-449b	chr5:54502231-54502327	1	rs10061133
hsa-mir-9-2	chr5:87998427-87998513	1	rs41265488
hsa-mir-1244	chr5:118338180-118338264	1	rs1804520
hsa-mir-1289-2	chr5:132791187-132791297	1	rs35296450
hsa-mir-1294	chr5:153706859-153707000	1	rs13186787
hsa-mir-146a	chr5:159844937-159845035	1	rs2910164
hsa-mir-585	chr5:168623183-168623276	2	rs62376934, rs62376935
hsa-mir-1229	chr5:179157884-179157952	1	rs2291418
hsa-mir-548a-1	chr6:18679994-18680090	1	rs12197631
hsa-mir-586	chr6:45273389-45273485	1	rs73735310
hsa-mir-339	chr7:1029095-1029188	3	rs13232101, rs72631820, rs72631831
hsa-mir-550-1	chr7:30295935-30296031	1	rs71528599
hsa-mir-590	chr7:73243464-73243560	1	rs6971711
hsa-mir-25	chr7:99529119-99529202	1	rs41274221
hsa-mir-93	chr7:99529327-99529406	1	rs72631824
hsa-mir-106b	chr7:99529552-99529633	1	rs72631827
hsa-mir-593	chr7:127509149-127509248	1	rs73721294
hsa-mir-96	chr7:129201768-129201845	2	rs73159662, rs41274239
hsa-mir-183	chr7:129201981-129202090	2	rs72631833, rs41281222
hsa-mir-595	chr7:158018171-158018266	1	rs4909237
hsa-mir-596	chr8:1752804-1752880	1	rs61388742
hsa-mir-548i-3	chr8:7983873-7984021	2	rs71313680, rs71313679
hsa-mir-1322	chr8:10720293-10720363	1	rs59878596
hsa-mir-548h-4	chr8:26962287-26962397	2	rs73235381, rs73235382
hsa-mir-124-2	chr8:65454260-65454368	1	rs72631829
hsa-mir-2053	chr8:113724898-113724988	1	rs10505168
hsa-mir-1206	chr8:129090326-129090384	1	rs2114358

hsa-mir-1208	chr8:129231544-129231616	2	rs56863230, rs2648841
hsa-mir-1234	chr8:145596284-145596367	1	rs2291134
hsa-mir-1302-2	chr9:20144-20281	5	rs11266858, rs4248191, rs11266859, rs422582, rs422363
hsa-mir-1299	chr9:68292059-68292141	1	rs62555121
hsa-mir-199b	chr9:130046821-130046930	1	rs72631835
hsa-mir-1265	chr10:14518581-14518666	1	rs11259096
hsa-mir-603	chr10:24604620-24604716	1	rs11014002
hsa-mir-604	chr10:29873939-29874032	2	rs2368393, rs2368392
hsa-mir-938	chr10:29931199-29931281	1	rs12416605
hsa-mir-605	chr10:52729339-52729421	1	rs2043556
hsa-mir-607	chr10:98578416-98578511	2	rs12778876, rs12780546
hsa-mir-608	chr10:102724732-102724831	1	rs4919510
hsa-mir-1307	chr10:105144000-105144148	1	rs7911488
hsa-mir-609	chr10:105968537-105968631	1	rs74154754
hsa-mir-2110	chr10:115923854-115923928	1	rs17091403
hsa-mir-202	chr10:134911006-134911115	1	rs12355840
hsa-mir-1908	chr11:61339209-61339288	1	rs174561
hsa-mir-194-2	chr11:64415403-64415487	1	rs11231898
hsa-mir-612	chr11:64968505-64968604	2	rs550894, rs12803915
hsa-mir-326	chr11:74723784-74723878	1	rs72561778
hsa-mir-1304	chr11:93106488-93106578	1	rs2155248
hsa-mir-548l	chr11:93839309-93839394	2	rs11020790, rs13447640
hsa-mir-1244	chr12:9283330-9283414	1	rs1804520
hsa-mir-1244	chr12:12156153-12156237	1	rs1804520
hsa-mir-196a-2	chr12:52671789-52671898	1	rs11614913
hsa-mir-548c	chr12:63302556-63302652	1	rs17120527
hsa-mir-617	chr12:79750443-79750539	1	rs12815353
hsa-mir-618	chr12:79853646-79853743	1	rs2682818
hsa-mir-492	chr12:93752305-93752420	1	rs2289030
hsa-mir-1178	chr12:118635822-118635912	1	rs7311975
hsa-mir-16-1	chr13:49521110-49521198	1	rs72631826
hsa-mir-622	chr13:89681437-89681532	1	rs59274393
hsa-mir-18a	chr13:90801006-90801076	1	rs41275866
hsa-mir-92a-1	chr13:90801569-90801646	2	rs72631821, rs9589207
hsa-mir-208b	chr14:22957036-22957112	1	rs2754157
hsa-mir-624	chr14:30553603-30553699	1	rs73251987
hsa-mir-625	chr14:65007573-65007657	1	rs12894182
hsa-mir-345	chr14:99843949-99844046	1	rs72631832
hsa-mir-431	chr14:100417097-100417210	1	rs12884005
hsa-mir-379	chr14:100558156-100558222	2	rs61991156, rs72631818
hsa-mir-299	chr14:100559884-100559946	1	rs41286566
hsa-mir-300	chr14:100577453-100577535	1	rs12894467
hsa-mir-1185-2	chr14:100580288-100580373	1	rs11844707
hsa-mir-453	chr14:100592280-100592359	1	rs56103835
hsa-mir-154	chr14:100595845-100595928	1	rs41286570
hsa-mir-412	chr14:100601537-100601627	1	rs61992671
hsa-mir-656	chr14:100602814-100602891	1	rs58834075
hsa-mir-1268	chr15:20014593-20014644	1	rs28599926
hsa-mir-1233	chr15:32461562-32461643	2	rs347881, rs347882
hsa-mir-1233	chr15:32607783-32607864	2	rs347881, rs347882
hsa-mir-627	chr15:40279060-40279156	1	rs2620381
hsa-mir-1282	chr15:41873149-41873249	1	rs11269
hsa-mir-147b	chr15:43512540-43512619	1	rs56073218
hsa-mir-184	chr15:77289185-77289268	1	rs41280052
hsa-mir-7-2	chr15:86956060-86956169	1	rs41276930
hsa-mir-1302-2	chr15:100318185-100318322	5	rs422363, rs422582, rs11266859, rs4248191, rs11266858
hsa-mir-662	chr16:760184-760278	1	rs9745376
hsa-mir-1826	chr16:33873009-33873093	2	rs1987294, rs62030476
hsa-mir-140	chr16:68524485-68524584	1	rs7205289
hsa-mir-1972	chr16:68621750-68621826	1	rs57629257
hsa-mir-1253	chr17:2598122-2598226	1	rs7217038
hsa-mir-548h-3	chr17:13387571-13387688	1	rs9913045

hsa-mir-423	chr17:25468223-25468316	1	rs6505162
hsa-mir-193a	chr17:26911128-26911215	1	rs60406007
hsa-mir-10a	chr17:44012199-44012308	1	rs72631828
hsa-mir-633	chr17:58375308-58375405	1	rs17759989
hsa-mir-187	chr18:31738779-31738887	1	rs41274312
hsa-mir-122	chr18:54269286-54269370	1	rs41292412
hsa-mir-1302-2	chr19:22973-23110	5	rs11266858, rs4248191, rs11266859, rs422582, rs422363
hsa-mir-220b	chr19:6446959-6447045	1	rs1053262
hsa-mir-1181	chr19:10375134-10375214	1	rs2569788
hsa-mir-27a	chr19:13808254-13808331	2	rs895819, rs11671784
hsa-mir-639	chr19:14501355-14501452	2	rs45556632, rs35149836
hsa-mir-125a	chr19:56888319-56888404	1	rs12975333
hsa-mir-1283-1	chr19:58883547-58883633	1	rs57111412
hsa-mir-520c	chr19:58902519-58902605	1	rs7255628
hsa-mir-521-2	chr19:58911660-58911746	1	rs13382089
hsa-mir-518d	chr19:58929943-58930029	1	rs73602910
hsa-mir-520h	chr19:58937578-58937665	1	rs56013413
hsa-mir-521-1	chr19:58943702-58943788	1	rs2561251
hsa-mir-516a-1	chr19:58951807-58951896	1	rs2569389
hsa-mir-1283-2	chr19:58953298-58953384	1	rs71363366
hsa-mir-1292	chr20:2581423-2581488	1	rs73576045
hsa-mir-663	chr20:26136822-26136914	3	rs28670321, rs2019798, rs7266947
hsa-mir-499	chr20:33041840-33041961	2	rs3746444, rs7267163
hsa-mir-646	chr20:58316927-58317020	2	rs6513496, rs6513497
hsa-mir-1-1	chr20:60561958-60562028	1	rs6122014
hsa-mir-124-3	chr20:61280297-61280383	1	rs34059726
hsa-mir-941-1	chr20:62021238-62021326	4	rs7268785, rs2427556, rs55795631, rs6089780
hsa-mir-941-2	chr20:62021545-62021633	1	rs34604519
hsa-mir-941-3	chr20:62021657-62021745	3	rs12625445, rs35544770, rs12625454
hsa-mir-647	chr20:62044428-62044523	1	rs73147065
hsa-mir-1286	chr22:18616657-18616734	1	rs71312743
hsa-mir-130b	chr22:20337593-20337674	1	rs72631822
hsa-mir-650	chr22:21495270-21495365	2	rs11558654, rs5996397
hsa-mir-548j	chr22:25281178-25281289	2	rs4822739, rs12161068
hsa-mir-1308	chrX:21990180-21990233	1	rs7051072
hsa-mir-548f-5	chrX:32569512-32569597	1	rs60180387
hsa-mir-222	chrX:45491365-45491474	1	rs72631825
hsa-mir-532	chrX:49654494-49654584	2	rs456615, rs456617
hsa-mir-325	chrX:76142220-76142317	1	rs72631830
hsa-mir-548i-4	chrX:83367416-83367492	1	rs72632467
hsa-mir-220a	chrX:122523627-122523736	2	rs72631819, rs72631817
hsa-mir-934	chrX:135460703-135460785	1	rs73558572
hsa-mir-891a	chrX:144917004-144917082	1	rs5965990
hsa-mir-105-2	chrX:151313540-151313620	1	rs72631816
hsa-mir-1184	chrX:153768829-153768927	1	rs56191956
hsa-mir-1184	chrX:154265943-154266041	1	rs56191956
hsa-mir-1184	chrX:154340372-154340470	1	rs56191956

Additional file 2

miRNA identified in CNV regions (continued on the next pages)

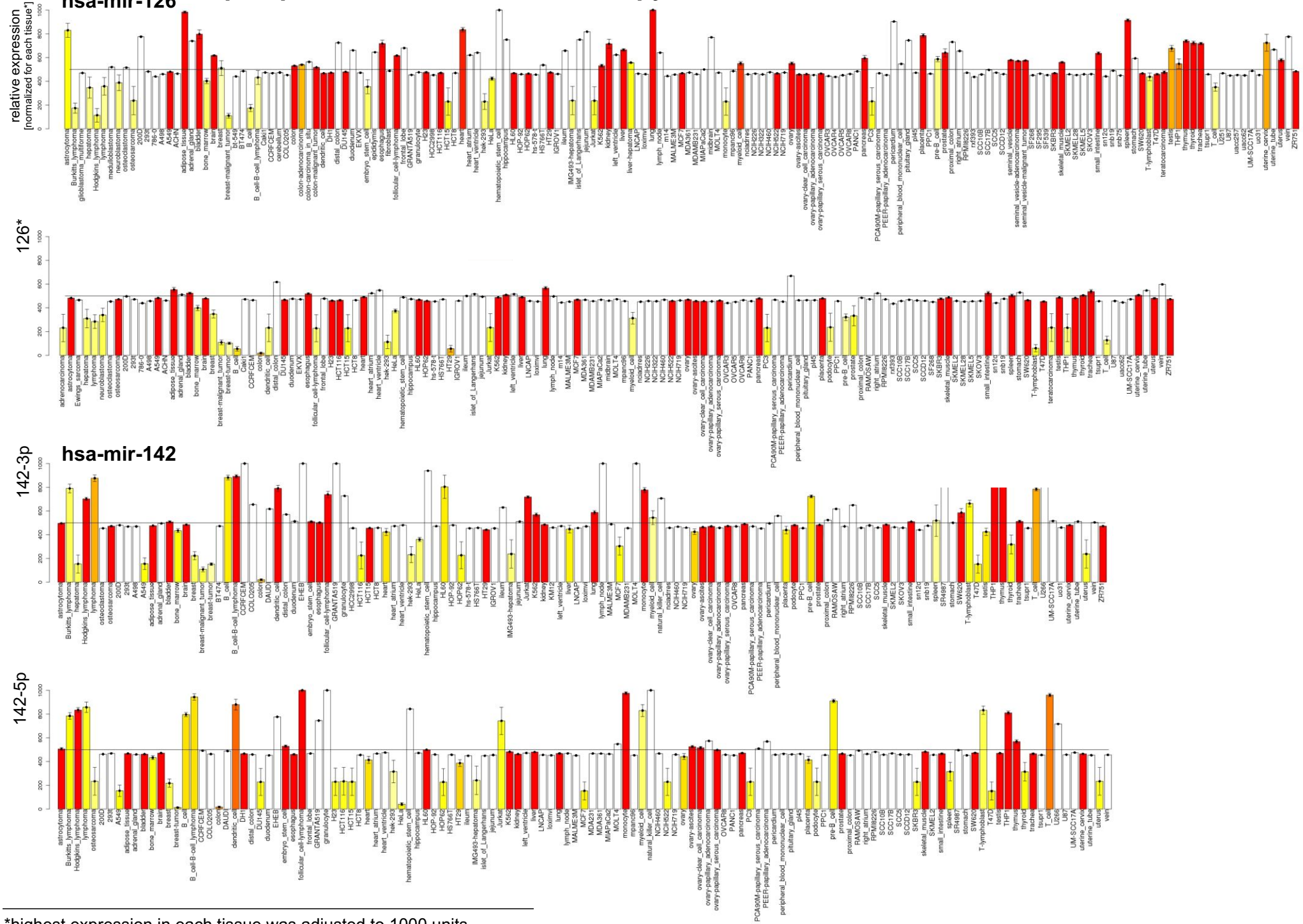
miRNA ID	miRNA chromosomal localization	miRNA ori	nuber of CNVs (DGV records) overlaping with miRNA location	number of discovery-studies reporting CNVs overlaping with miRNA location	references of discovery-studies	highest number of observations (Total Gain/Loss) reported in discovery-study
hsa-mir-1302-2	chr1:20229-20366	+	4	4	Redon (2006); Perry (2008); Locke (2006); Sharp (2005)	175
hsa-mir-1977	chr1:556050-556128	-	6	4	Redon (2006); Perry (2008); Wong (2007); Cooper (2008)	175
hsa-mir-200b	chr1:1092347-1092441	+	2	2	Perry (2008); Iafrate (2004)	11
hsa-mir-200a	chr1:1093106-1093195	+	2	2	Perry (2008); Iafrate (2004)	11
hsa-mir-429	chr1:1094248-1094330	+	2	2	Perry (2008); Iafrate (2004)	11
hsa-mir-551a	chr1:3467119-3467214	-	3	2	Jakobsson (2008); Redon (2006)	1
hsa-mir-34a	chr1:9134314-9134423	-	1	1	Simon-Sanchez (2007)	1
hsa-mir-320b-1	chr1:117015894-117015972	+	2	2	Wong (2007); Redon (2006)	56
hsa-mir-92b	chr1:153431592-153431687	+	2	2	Redon (2006); Wong (2007)	6
hsa-mir-555	chr1:153582765-153582860	-	1	1	de Smith (2007)	39
hsa-mir-556	chr1:160578960-160579054	+	5	4	Redon (2006); Wang (2007); Pinto (2007); Shaikh (2009)	3
hsa-mir-1255b-2	chr1:166234522-166234588	+	1	1	de Smith (2007)	5
hsa-mir-557	chr1:166611386-166611483	+	1	1	Wong (2007)	3
hsa-mir-320b-2	chr1:222511329-222511466	-	1	1	Perry (2008)	1
hsa-mir-1301	chr2:25405013-25405094	-	1	1	de Smith (2007)	2
hsa-mir-558	chr2:32610724-32610817	+	5	4	Simon-Sanchez (2007); Pinto (2007); Itsara (2009); Shaikh (2009)	4
hsa-mir-217	chr2:56063606-56063715	-	1	1	Redon (2006)	1
hsa-mir-216a	chr2:56069589-56069698	-	1	1	Redon (2006)	1
hsa-mir-216b	chr2:56081353-56081434	-	1	1	Redon (2006)	1
hsa-mir-1302-3	chr2:114057006-114057143	-	2	2	Redon (2006); Perry (2008)	5
hsa-mir-663b	chr2:132731009-132731123	-	4	4	Redon (2006); de Smith (2007); Kim (2009); Perry (2008)	121
hsa-mir-128-1	chr2:136139437-136139518	+	1	1	Gusev (2009)__sq__	6
hsa-mir-1978	chr2:149355835-149355887	-	1	1	Pinto (2007)	1
hsa-mir-1244	chr2:232286268-232286352	+	1	1	Redon (2006)	138
hsa-mir-1471	chr2:232465196-232465252	-	1	1	Redon (2006)	138
hsa-mir-149	chr2:241044091-241044179	+	6	3	Redon (2006); Shaikh (2009); Kim (2009)	2
hsa-mir-26a-1	chr3:37985899-37985975	+	1	1	Redon (2006)	8
hsa-mir-566	chr3:50185763-50185856	+	7	5	Shaikh (2009); Jakobsson (2008); Itsara (2009); Redon (2006); Wong (2007)	18
hsa-let-7g	chr3:52277334-52277417	-	2	2	Shaikh (2009); Wong (2007)	3
hsa-mir-135a-1	chr3:52303275-52303364	-	3	2	Shaikh (2009); Wong (2007)	6
hsa-mir-1324	chr3:75762604-75762699	+	6	5	Simon-Sanchez (2007); Redon (2006); Pinto (2007); Perry (2008); Kim (2009)	16
hsa-mir-1280	chr3:129563698-129563791	+	2	2	Jakobsson (2008); Redon (2006)	7
hsa-mir-1263	chr3:165371953-165372038	-	2	2	Redon (2006); Iafrate (2004)	1
hsa-mir-1224	chr3:185441887-185441971	+	1	1	Wong (2007)	3
hsa-mir-1248	chr3:187987155-187987260	+	2	2	Redon (2006); Gusev (2009)__sq__	2
hsa-mir-570	chr3:196911452-196911548	+	9	7	Redon (2006); Perry (2008); Kim (2009); Locke (2006); Sharp (2005); de Smith (2007); Wong (2007)	188
hsa-mir-922	chr3:198885764-198885844	-	2	2	Redon (2006); Wong (2007)	79
hsa-mir-571	chr4:333946-334041	+	1	1	Locke (2006)	5
hsa-mir-95	chr4:8057928-8058008	-	3	2	Gusev (2009)__sq__; Wong (2007)	10
hsa-mir-548i-2	chr4:9166887-9167035	-	9	8	Redon (2006); Shaikh (2009); Itsara (2009); Wang (2007); Zogopoulos (2007); Locke (2006); Sharp (2005); Cooper (2008)	222
hsa-mir-218-1	chr4:20138996-20139105	+	1	1	Wong (2007)	27
hsa-mir-577	chr4:115797364-115797459	+	2	2	Pinto (2007); Redon (2006)	2
hsa-mir-579	chr5:32430241-32430338	-	1	1	Redon (2006)	1
hsa-mir-1974	chr5:93930928-93930997	-	3	3	Redon (2006); Kim (2009); Perry (2008)	139
hsa-mir-583	chr5:95440598-95440672	+	4	3	Redon (2006); Pinto (2007); Shaikh (2009)	2
hsa-mir-548f-3	chr5:109877429-109877515	-	4	3	Redon (2006); Wang (2007); Pinto (2007)	1
hsa-mir-886	chr5:135444076-135444196	-	1	1	Redon (2006)	1
hsa-mir-1229	chr5:179157884-179157952	-	1	1	Iafrate (2004)	1
hsa-mir-1236	chr6:32032595-32032696	-	2	2	Redon (2006); Wong (2007)	36
hsa-mir-589	chr7:5501976-5502074	-	3	2	Wong (2007); Shaikh (2009)	5
hsa-mir-1183	chr7:21477201-21477289	+	1	1	Wong (2007)	20
hsa-mir-25	chr7:99529119-99529202	-	1	1	Locke (2006)	4
hsa-mir-93	chr7:99529327-99529406	-	1	1	Locke (2006)	4
hsa-mir-106b	chr7:99529552-99529633	-	1	1	Locke (2006)	4
hsa-mir-548o	chr7:101833194-101833307	-	2	2	Redon (2006); Wong (2007)	42
hsa-mir-129-1	chr7:127635161-127635232	+	1	1	de Smith (2007)	1
hsa-mir-182	chr7:129197459-129197568	-	2	1	Shaikh (2009)	3
hsa-mir-96	chr7:129201768-129201845	-	2	1	Shaikh (2009)	3
hsa-mir-183	chr7:129201981-129202090	-	2	1	Shaikh (2009)	3
hsa-mir-153-2	chr7:157059789-157059875	-	1	1	Redon (2006)	2
hsa-mir-596	chr8:1752804-1752880	+	1	1	Itsara (2009)	2
hsa-mir-548i-3	chr8:7983873-7984021	-	14	7	Redon (2006); Sebat (2004); Shaikh (2009); Zogopoulos (2007); Pinto (2007); Locke (2006); Sharp (2005)	124
hsa-mir-383	chr8:14755318-14755390	-	8	7	Redon (2006); Wang (2007); Pinto (2007); Perry (2008); McCarroll (2005); Locke (2006); Conrad (2005)	15
hsa-mir-320a	chr8:22158420-22158501	-	2	1	Redon (2006)	2
hsa-mir-599	chr8:100618040-100618134	-	1	1	Redon (2006)	17
hsa-mir-875	chr8:100618190-100618265	-	1	1	Redon (2006)	17

hsa-mir-1204	chr8:128877390-128877456	+	2	2	Pinto (2007); Wong (2007)	4
hsa-mir-661	chr8:145091347-145091435	-	8	3	Jakobsson (2008); Itsara (2009); Shaikh (2009)	8
hsa-mir-939	chr8:145590172-145590253	-	3	3	Jakobsson (2008); Redon (2006); Wong (2007)	16
hsa-mir-1234	chr8:145596284-145596367	-	3	3	Jakobsson (2008); Redon (2006); Wong (2007)	16
hsa-mir-1302-2	chr9:20144-20281	+	2	1	Perry (2008)	5
hsa-mir-31	chr9:21502114-21502184	-	1	1	Redon (2006)	1
hsa-mir-1299	chr9:68292059-68292141	-	7	4	Redon (2006); Locke (2006); Korbel (2007); de Smith (2007)	178
hsa-mir-7-1	chr9:85774483-85774592	-	2	2	Redon (2006); Perry (2008)	1
hsa-let-7a-1	chr9:95978060-95978139	+	1	1	Wong (2007)	18
hsa-let-7f-1	chr9:95978450-95978536	+	1	1	Wong (2007)	18
hsa-let-7d	chr9:95980937-95981023	+	1	1	Wong (2007)	18
hsa-mir-455	chr9:116011535-116011630	+	3	2	Simon-Sanchez (2007); Itsara (2009)	2
hsa-mir-126	chr9:138684875-138684959	+	14	6	Redon (2006); Jakobsson (2008); Perry (2008); Itsara (2009); Simon-Sanchez (2007); de Smith (2007)	9
hsa-mir-1265	chr10:14518581-14518666	+	1	1	Gusev (2009)__sq__	2
hsa-mir-511-1	chr10:17927113-17927199	+	1	1	de Smith (2007)	2
hsa-mir-1915	chr10:21825497-21825576	-	1	1	Shaikh (2009)	6
hsa-mir-604	chr10:29873939-29874032	-	1	1	Wong (2007)	3
hsa-mir-938	chr10:29931199-29931281	-	1	1	Wong (2007)	3
hsa-mir-548f-1	chr10:56037640-56037723	-	1	1	Pinto (2007)	1
hsa-mir-1254	chr10:70189081-70189177	+	1	1	Redon (2006)	3
hsa-mir-606	chr10:76982222-76982317	+	1	1	Redon (2006)	4
hsa-mir-1287	chr10:100144965-100145054	-	1	1	Shaikh (2009)	2
hsa-mir-608	chr10:102724732-102724831	+	1	1	Wong (2007)	7
hsa-mir-202	chr10:134911006-134911115	-	10	5	Redon (2006); Wong (2007); Jakobsson (2008); Simon-Sanchez (2007); Perry (2008)	39
hsa-mir-210	chr11:558089-558198	-	4	2	Jakobsson (2008); Redon (2006)	166
hsa-mir-675	chr11:1974565-1974637	-	2	1	Jakobsson (2008)	2
hsa-mir-130a	chr11:57165247-57165335	+	1	1	Wong (2007)	3
hsa-mir-612	chr11:64968505-64968604	+	1	1	Simon-Sanchez (2007)	1
hsa-mir-1244	chr12:9283330-9283414	-	1	1	Redon (2006)	93
hsa-mir-196a-2	chr12:52671789-52671898	+	1	1	Redon (2006)	3
hsa-mir-615	chr12:52714001-52714096	+	1	1	Redon (2006)	3
hsa-mir-616	chr12:56199213-56199309	-	1	1	Redon (2006)	37
hsa-let-7i	chr12:61283733-61283816	+	1	1	Sebat (2004)	1
hsa-mir-492	chr12:93752305-93752420	+	1	1	Mills (2006)	1
hsa-mir-1251	chr12:96409818-96409887	+	1	1	McCarroll (2005)	0
hsa-mir-619	chr12:107754813-107754911	-	1	1	Simon-Sanchez (2007)	1
hsa-mir-620	chr12:115070748-115070842	-	1	1	Wong (2007)	6
hsa-mir-622	chr13:89681437-89681532	+	1	1	Zogopoulos (2007)	2
hsa-mir-624	chr14:30553603-30553699	-	1	1	de Smith (2007)	2
hsa-mir-770	chr14:100388480-100388577	+	1	1	Pinto (2007)	1
hsa-mir-203	chr14:103653495-103653604	+	2	2	Jakobsson (2008); Shaikh (2009)	4
hsa-mir-1268	chr15:20014593-20014644	-	37	11	Redon (2006); Shaikh (2009); Zogopoulos (2007); Pinto (2007); Perry (2008); de Smith (2007); Sebat (2004); Kim (2009); Cooper (2008); McCarroll (2008); Wong (2007)	228
hsa-mir-211	chr15:29144527-29144636	-	2	2	de Smith (2007); Shaikh (2009)	5
hsa-mir-1233	chr15:32461562-32461643	-	9	7	de Smith (2007); Redon (2006); Locke (2006); Shaikh (2009); Jakobsson (2008); Kim (2009); Perry (2008)	213
hsa-mir-1233	chr15:32607783-32607864	-	17	14	de Smith (2007); Redon (2006); Locke (2006); Pinto (2007); Kidd (2008); Kim (2009); Perry (2008); Tuzun (2005); Sebat (2004); McCarroll (2008); Wong (2007); Itsara (2009); Sharp (2005); Zogopoulos (2007)	213
hsa-mir-627	chr15:40279060-40279156	-	1	1	Sharp (2005)	1
hsa-mir-1282	chr15:41873149-41873249	-	1	1	Redon (2006)	6
hsa-mir-630	chr15:70666612-70666708	+	2	2	Redon (2006); Iafrate (2004)	3
hsa-mir-184	chr15:77289185-77289268	+	1	1	Redon (2006)	1
hsa-mir-1302-2	chr15:100318185-100318322	-	3	2	Kidd (2008); Perry (2008)	15
hsa-mir-662	chr16:760184-760278	+	6	5	Shaikh (2009); Simon-Sanchez (2007); Redon (2006); Jakobsson (2008); Perry (2008)	64
hsa-mir-1225	chr16:2080197-2080286	-	2	2	Jakobsson (2008); Simon-Sanchez (2007)	3
hsa-mir-940	chr16:2261749-2261842	+	2	2	Jakobsson (2008); Simon-Sanchez (2007)	1
hsa-mir-1972	chr16:15011679-15011755	-	5	5	Perry (2008); Redon (2006); de Smith (2007); McCarroll (2008); Kim (2009)	222
hsa-mir-484	chr16:15644652-15644730	+	1	1	Locke (2006)	3
hsa-mir-1826	chr16:33873009-33873093	+	4	4	Redon (2006); Pinto (2007); de Smith (2007); Perry (2008)	213
hsa-mir-138-2	chr16:55449931-55450014	+	1	1	de Smith (2007)	1
hsa-mir-1538	chr16:68157212-68157272	-	1	1	Wheeler (2008)	1
hsa-mir-140	chr16:68524485-68524584	+	1	1	Redon (2006)	19
hsa-mir-1972	chr16:68621750-68621826	+	6	5	Redon (2006); Wong (2007); Perry (2008); Kim (2009); Shaikh (2009)	19
hsa-mir-22	chr17:1563947-1564031	-	4	3	Perry (2008); Jakobsson (2008); Shaikh (2009)	2
hsa-mir-1253	chr17:2598122-2598226	-	1	1	Wong (2007)	3
hsa-mir-548h-3	chr17:13387571-13387688	-	1	1	Gusev (2009)__sq__	6
hsa-mir-1180	chr17:19188412-19188480	-	1	1	Wong (2007)	3
hsa-mir-10a	chr17:44012199-44012308	-	2	2	Redon (2006); Itsara (2009)	2
hsa-mir-196a-1	chr17:44064851-44064920	-	1	1	Redon (2006)	2
hsa-mir-142	chr17:53763592-53763678	-	11	2	Wong (2007); Shaikh (2009)	21
hsa-mir-454	chr17:54569901-54570015	-	1	1	Redon (2006)	1
hsa-mir-301a	chr17:54583279-54583364	-	1	1	Redon (2006)	1
hsa-mir-548d-2	chr17:62898067-62898163	-	1	1	Sharp (2005)	1
hsa-mir-657	chr17:76713671-76713768	-	2	2	Wong (2007); Jakobsson (2008)	13
hsa-mir-338	chr17:76714278-76714344	-	2	2	Wong (2007); Jakobsson (2008)	13

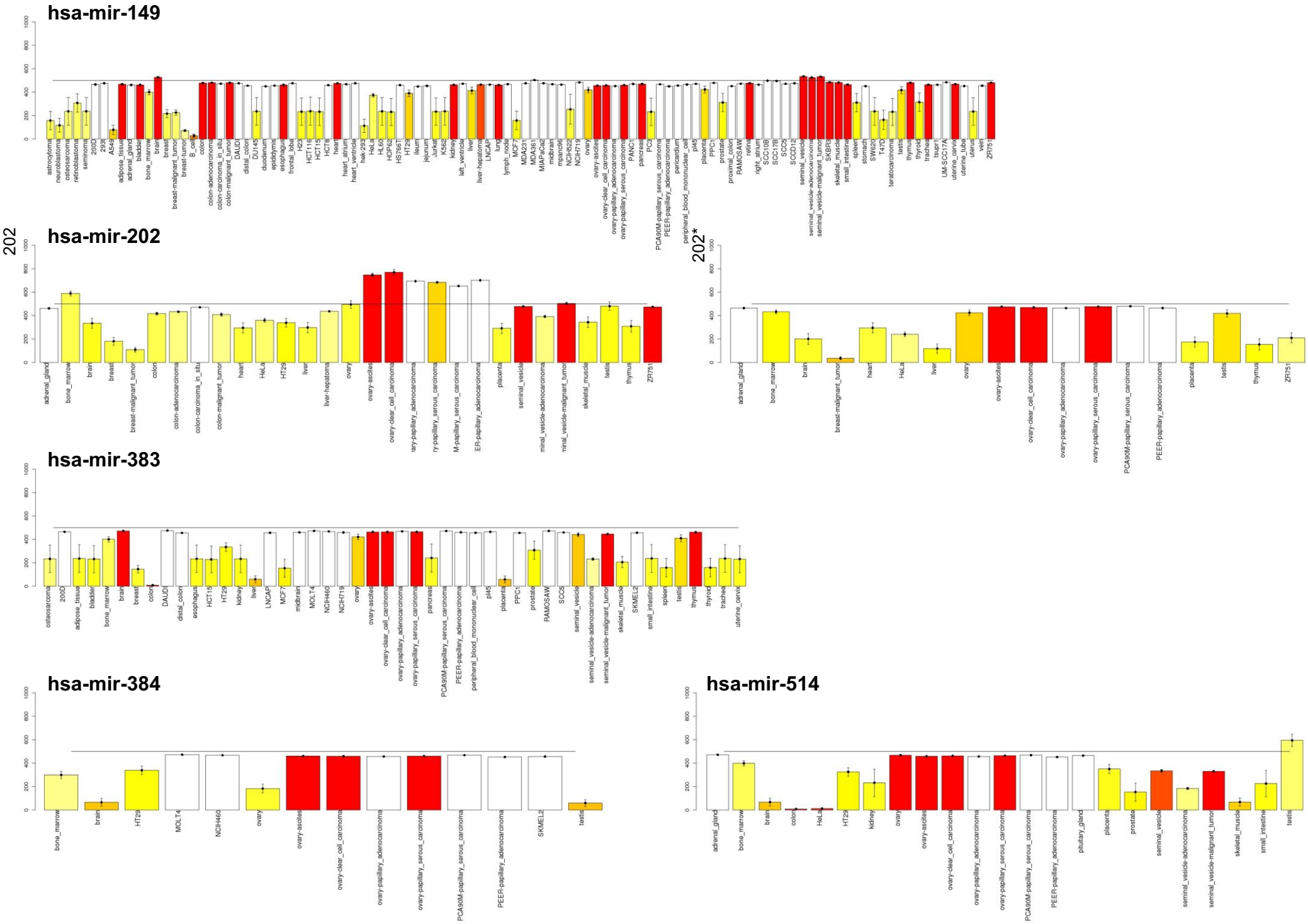
hsa-mir-1250	chr17:76721591-76721703	-	2	2	Wong (2007); Jakobsson (2008)	13
hsa-mir-1539	chr18:45267741-45267790	+	1	1	Kidd (2008)	1
hsa-mir-1302-2	chr19:22973-23110	+	1	1	Perry (2008)	20
hsa-mir-1909	chr19:1767158-1767237	-	3	2	de Smith (2007); Jakobsson (2008)	23
hsa-mir-1227	chr19:2185061-2185148	-	2	1	Wong (2007)	17
hsa-mir-638	chr19:10690080-10690179	+	1	1	Wong (2007)	3
hsa-mir-199a-1	chr19:10789102-10789172	-	2	1	Wong (2007)	11
hsa-mir-24-2	chr19:13808101-13808173	-	1	1	Wong (2007)	4
hsa-mir-27a	chr19:13808254-13808331	-	1	1	Wong (2007)	4
hsa-mir-23a	chr19:13808401-13808473	-	1	1	Wong (2007)	4
hsa-mir-181c	chr19:13846513-13846622	+	1	1	Wong (2007)	4
hsa-mir-181d	chr19:13846689-13846825	+	1	1	Wong (2007)	4
hsa-mir-1270	chr19:20371080-20371162	-	9	5	Simon-Sanchez (2007); Redon (2006); Itsara (2009); Kidd (2008); Shaikh (2009)	35
hsa-mir-641	chr19:45480290-45480388	-	1	1	Redon (2006)	1
hsa-mir-220c	chr19:53755341-53755423	-	2	2	Perry (2008); Wang (2007)	18
hsa-mir-150	chr19:54695854-54695937	-	3	2	Perry (2008); Wong (2007)	25
hsa-mir-99b	chr19:56887677-56887746	+	3	2	Redon (2006); Perry (2008)	15
hsa-let-7e	chr19:56887851-56887929	+	3	2	Redon (2006); Perry (2008)	15
hsa-mir-125a	chr19:56888319-56888404	+	3	2	Redon (2006); Perry (2008)	15
hsa-mir-512-1	chr19:58861745-58861828	+	1	1	Ahn (2009)	1
hsa-mir-935	chr19:59177373-59177463	+	1	1	Wong (2007)	5
hsa-mir-663	chr20:26136822-26136914	-	6	6	Redon (2006); Jakobsson (2008); Itsara (2009); Iafrate (2004); de Smith (2007); Perry (2008)	15
hsa-mir-1825	chr20:30289259-30289311	+	1	1	Wong (2007)	6
hsa-mir-499	chr20:33041840-33041961	+	1	1	Wong (2007)	3
hsa-mir-1257	chr20:59961997-59962113	-	3	2	Redon (2006); Kidd (2008)	10
hsa-mir-1-1	chr20:60561958-60562028	+	3	3	Simon-Sanchez (2007); Jakobsson (2008); Redon (2006)	32
hsa-mir-133a-2	chr20:60572564-60572665	+	3	3	Simon-Sanchez (2007); Jakobsson (2008); Redon (2006)	32
hsa-mir-124-3	chr20:61280297-61280383	+	4	3	Perry (2008); Wong (2007); Itsara (2009)	70
hsa-mir-941-1	chr20:62021238-62021326	+	1	1	Locke (2006)	3
hsa-mir-941-2	chr20:62021545-62021633	+	1	1	Locke (2006)	3
hsa-mir-941-3	chr20:62021657-62021745	+	1	1	Locke (2006)	3
hsa-mir-1914	chr20:62043262-62043341	-	1	1	Locke (2006)	3
hsa-mir-647	chr20:62044428-62044523	-	1	1	Locke (2006)	3
hsa-mir-185	chr22:18400662-18400743	+	5	4	Perry (2008); Locke (2006); Wong (2007); Jakobsson (2008)	31
hsa-mir-1306	chr22:18453581-18453665	+	2	2	Perry (2008); Redon (2006)	6
hsa-mir-1286	chr22:18616657-18616734	-	4	3	Perry (2008); Redon (2006); Jakobsson (2008)	6
hsa-mir-649	chr22:19718465-19718561	-	1	1	Wong (2007)	12
hsa-mir-650	chr22:21495270-21495365	+	38	14	Zogopoulos (2007); Itsara (2009); Kidd (2008); Wang (2007); Kim (2009); McCarroll (2005); Wong (2007); Sebat (2004); Locke (2006); Shaikh (2009); Sharp (2005); Levy (2007); Tuzun (2005); de Smith (2007)	67
hsa-mir-548j	chr22:25281178-25281289	-	1	1	Redon (2006)	1
hsa-mir-658	chr22:36570225-36570324	-	1	1	Redon (2006)	6
hsa-mir-659	chr22:36573631-36573727	-	2	2	Redon (2006); Wheeler (2008)	6
hsa-mir-1249	chr22:43975499-43975564	-	1	1	Redon (2006)	1
hsa-let-7a-3	chr22:44887293-44887366	+	4	2	Jakobsson (2008); Shaikh (2009)	7
hsa-let-7b	chr22:44888230-44888312	+	4	2	Jakobsson (2008); Shaikh (2009)	7
hsa-mir-651	chrX:8055006-8055102	+	3	3	Zogopoulos (2007); Pinto (2007); Wang (2007)	3
hsa-mir-98	chrX:53599909-53600027	-	1	1	Korbel (2007)	1
hsa-let-7f-2	chrX:53600878-53600960	-	1	1	Korbel (2007)	1
hsa-mir-384	chrX:76056092-76056179	-	4	4	Redon (2006); McCarroll (2005); Shaikh (2009); McCarroll (2008)	144
hsa-mir-1912	chrX:113792275-113792354	+	1	1	de Smith (2007)	13
hsa-mir-1264	chrX:113793386-113793454	+	1	1	de Smith (2007)	13
hsa-mir-1298	chrX:113855906-113856017	+	1	1	de Smith (2007)	13
hsa-mir-1911	chrX:113904000-113904079	+	1	1	de Smith (2007)	13
hsa-mir-448	chrX:113964273-113964383	+	1	1	de Smith (2007)	13
hsa-mir-320d-2	chrX:139836003-139836050	-	1	1	Redon (2006)	46
hsa-mir-513c	chrX:146078914-146078997	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-513b	chrX:146088254-146088337	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-513a-1	chrX:146102673-146102801	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-513a-2	chrX:146115036-146115162	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-506	chrX:146119930-146120053	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-507	chrX:146120194-146120287	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-508	chrX:146126123-146126237	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-509-2	chrX:146147970-146148060	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-509-3	chrX:146148862-146148936	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-509-1	chrX:146149742-146149835	-	2	2	Redon (2006); Pinto (2007)	1
hsa-mir-510	chrX:146161545-146161618	-	4	4	Redon (2006); Pinto (2007); Kidd (2008); Tuzun (2005)	1
hsa-mir-514-1	chrX:146168457-146168554	-	5	5	Redon (2006); Pinto (2007); Kidd (2008); Tuzun (2005); Perry (2008)	3
hsa-mir-514-2	chrX:146171153-146171240	-	6	5	Redon (2006); Pinto (2007); Kidd (2008); Tuzun (2005); Perry (2008)	3
hsa-mir-514-3	chrX:146173851-146173938	-	6	5	Redon (2006); Pinto (2007); Kidd (2008); Tuzun (2005); Perry (2008)	3
hsa-mir-1184	chrX:153768829-153768927	-	1	1	Levy (2007)	1
hsa-mir-1184	chrX:154265943-154266041	-	1	1	Levy (2007)	1
hsa-mir-1184	chrX:154340372-154340470	+	1	1	Levy (2007)	1

Additional File 3

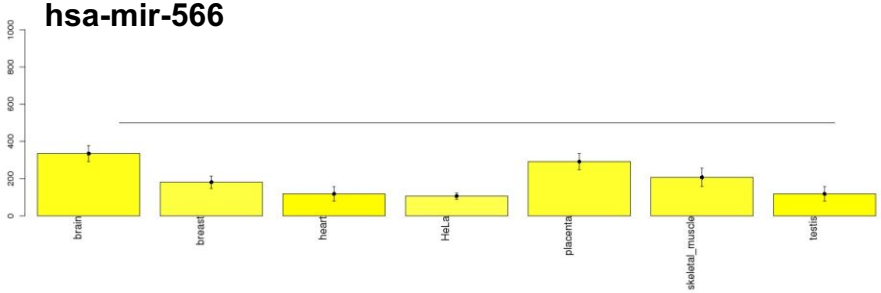
Expression profiles of selected CNV-miRNAs (continued on the next pages)



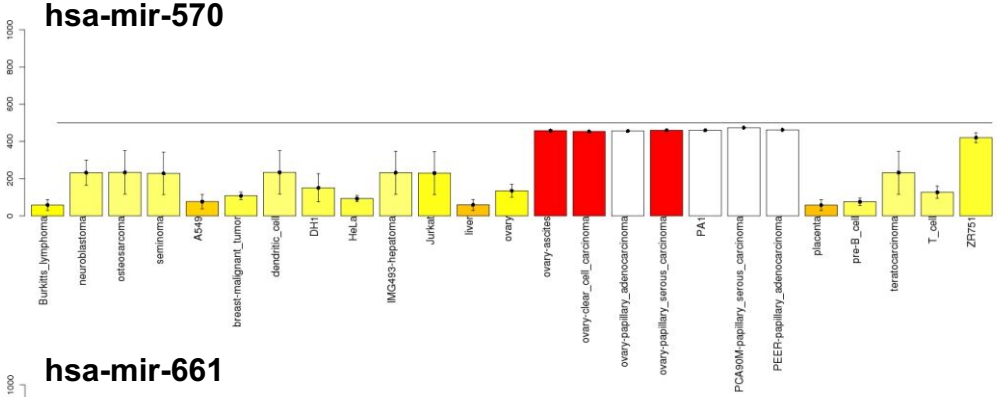
*highest expression in each tissue was adjusted to 1000 units


hsa-mir-149
hsa-mir-202
202
202*
hsa-mir-383
hsa-mir-384
hsa-mir-514

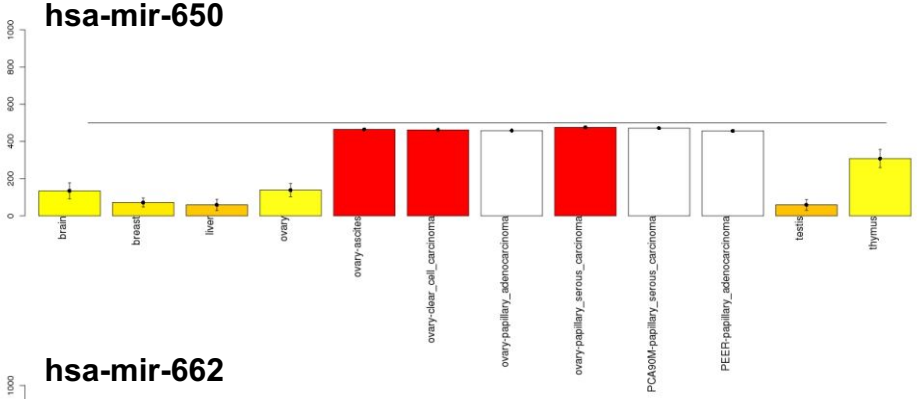
hsa-mir-566



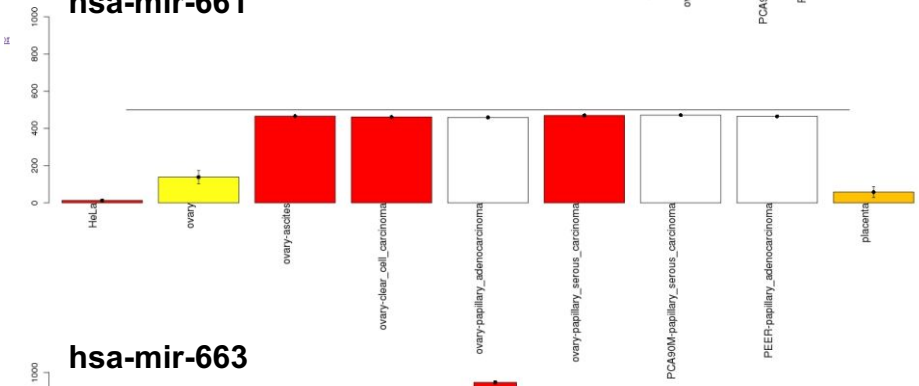
hsa-mir-570



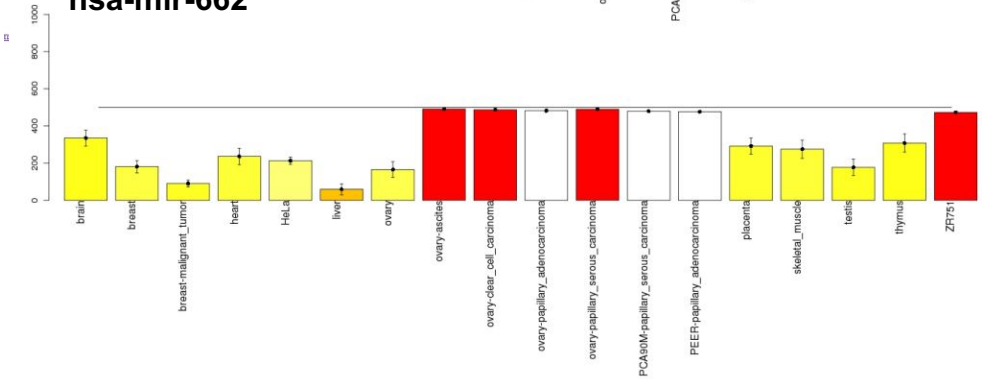
hsa-mir-650



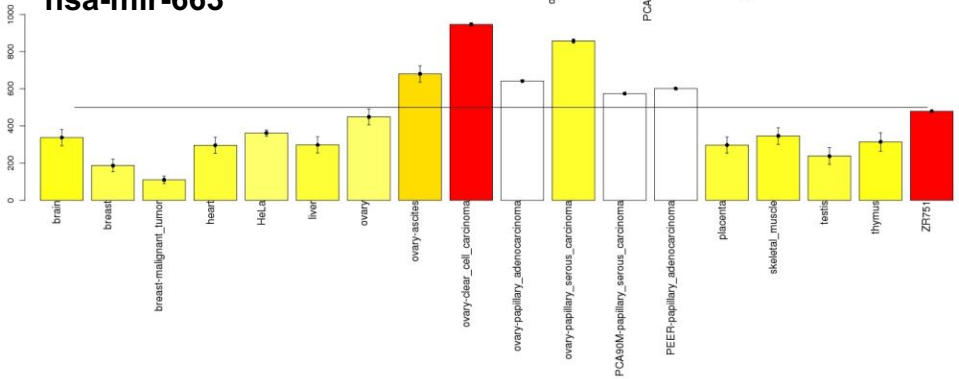
hsa-mir-661



hsa-mir-662



hsa-mir-663



Additional file 4

miRNAs located in CNVs with well defined breakpoints

miRNA ID	miRNA chromosomal localization	location	(distance to the 5'/3' breakpoint of CNV)
mir-1233	chr15:32461562-32461643	introm (mirtron) of GOLGA8A entirely spaned by CNV	(12kb/201kb)
mir-1233	chr15:32607783-32607864	introm (mirtron) of GOLGA8B entirely spaned by CNV	(158kb/55kb)
mir-1268	chr15:20014593-20014644	intergenic	(211kb/75kb)
mir-1275	chr6:34075727-34075806	intergenic	(5kb/1kb)
mir-1302-2	chr9:20144-20281	promoter region of WASH1 entirely spaned by CNV	(20kb/18kb)
mir-1324	chr3:75762604-75762699	intergenic	(298kb/20kb)
mir-1972	chr16:15011679-15011755	intron of PDXDC1 which 5' part extends beyond CNV	(114kb/4kb)
mir-1977	chr1:556050-556128	intergenic	(2kb/4kb)
mir-384	chrX:76056092-76056179	intergenic	(2kb/1kb)
mir-548i-2	chr4:9166887-9167035	intergenic	(49kb/188kb)
mir-650	chr22:21495270-21495365	intergenic	(784kb/84kb)

Additional file 5

miRNAs located in chromosomal regions implicated in microdeletion/microduplication syndromes

miRNA ID	miRNA chromosomal localization	microdeletion/microduplication syndrome*	syndrome locus chromosomal localization*
hsa-mir-429	chr1:1094248-1094330	1p36 microdeletion syndrome	chr1:1-5,308,621
hsa-mir-1977	chr1:556050-556128	1p36 microdeletion syndrome	chr1:1-5,308,621
hsa-mir-1302-2	chr1:20229-20366	1p36 microdeletion syndrome	chr1:1-5,308,621
hsa-mir-200a	chr1:1093106-1093195	1p36 microdeletion syndrome	chr1:1-5,308,621
hsa-mir-200b	chr1:1092347-1092441	1p36 microdeletion syndrome	chr1:1-5,308,621
hsa-mir-551a	chr1:3467119-3467214	1p36 microdeletion syndrome	chr1:1-5,308,621
hsa-mir-149	chr2:241044091-241044179	2q37 monosomy	chr2:239,619,630-242,951,149
hsa-mir-922	chr3:198885764-198885844	3q29 microdeletion / microduplication syndrome	chr3:197,156,626-198,982,266
hsa-mir-943	chr4:1957909-1958002	Wolf-Hirschhorn Syndrome	chr4:1-2,043,468
hsa-mir-571	chr4:333946-334041	Wolf-Hirschhorn Syndrome	chr4:1-2,043,468
hsa-mir-590	chr7:73243464-73243560	Williams-Beuren Syndrome (WBS)	chr7:71,970,679-74,254,837
hsa-mir-591	chr7:95686910-95687004	Split hand/foot malformation 1 (SHFM1)	chr7:95,371,796-96,617,422
hsa-mir-1322	chr8:10720293-10720363	8p23.1 deletion syndrome	chr8:8,156,705-11,803,128
hsa-mir-598	chr8:10930126-10930222	8p23.1 deletion syndrome	chr8:8,156,705-11,803,128
hsa-mir-597	chr8:9636592-9636688	8p23.1 deletion syndrome	chr8:8,156,705-11,803,128
hsa-mir-124-1	chr8:9798308-9798392	8p23.1 deletion syndrome	chr8:8,156,705-11,803,128
hsa-mir-602	chr9:139852692-139852789	9q subtelomeric deletion syndrome	chr9:139,523,184-140,273,252
hsa-mir-1302-2	chr15:100318185-100318322	15q26 overgrowth syndrome	chr15:97,175,493-100,338,915
hsa-mir-211	chr15:29144527-29144636	15q13.3 microdeletion syndrome	chr15:28,557,287-30,488,774
hsa-mir-631	chr15:73433005-73433079	15q24 recurrent microdeletion syndrome	chr15:72,164,227-73,949,332
hsa-mir-484	chr16:15644652-15644730	16p13.11 recurrent microdeletion / microduplication (neurocognitive disorder susceptibility locus)	chr16:15,411,955-16,191,749
hsa-mir-662	chr16:760184-760278	ATR-16 syndrome	chr16:1-774,373
hsa-mir-22	chr17:1563947-1564031	Miller-Dieker syndrome (MDS)	chr17:1-2,492,179
hsa-mir-33b	chr17:17657875-17657970	Potocki-Lupski syndrome (17p11.2 duplication syndrome) / Smith-Magenis Syndrome	chr17:16,646,746-20,422,653
hsa-mir-132	chr17:1899952-1900052	Miller-Dieker syndrome (MDS)	chr17:1-2,492,179
hsa-mir-212	chr17:1900315-1900424	Miller-Dieker syndrome (MDS)	chr17:1-2,492,179
hsa-mir-1180	chr17:19188412-19188480	Potocki-Lupski syndrome (17p11.2 duplication syndrome) / Smith-Magenis Syndrome	chr17:16,646,746-20,422,653
hsa-mir-193a	chr17:26911128-26911215	NF1-microdeletion syndrome	chr17:26,186,948-27,242,780
hsa-mir-365-2	chr17:26926543-26926653	NF1-microdeletion syndrome	chr17:26,186,948-27,242,780
hsa-mir-648	chr22:16843634-16843727	Cat-Eye Syndrome (Type I)	chr22:1-16,971,860
hsa-mir-185	chr22:18400662-18400743	22q11 deletion syndrome (Velocardiofacial/DiGeorge syndrome) / 22q11 duplication syndrome	chr22:16,926,349-20,666,469
hsa-mir-1306	chr22:18453581-18453665	22q11 deletion syndrome (Velocardiofacial/DiGeorge syndrome) / 22q11 duplication syndrome	chr22:16,926,349-20,666,469
hsa-mir-1286	chr22:18616657-18616734	22q11 deletion syndrome (Velocardiofacial/DiGeorge syndrome) / 22q11 duplication syndrome	chr22:16,926,349-20,666,469
hsa-mir-649	chr22:19718465-19718561	22q11 deletion syndrome (Velocardiofacial/DiGeorge syndrome) / 22q11 duplication syndrome	chr22:16,926,349-20,666,469
hsa-mir-301b	chr22:20337270-20337347	22q11 deletion syndrome (Velocardiofacial/DiGeorge syndrome) / 22q11 duplication syndrome	chr22:16,926,349-20,666,469
hsa-mir-130b	chr22:20337593-20337674	22q11 deletion syndrome (Velocardiofacial/DiGeorge syndrome) / 22q11 duplication syndrome	chr22:16,926,349-20,666,469
hsa-mir-650	chr22:21495270-21495365	22q11.2 distal deletion syndrome	chr22:20,445,848-22,026,229
hsa-mir-651	chrX:8055006-8055102	Steroid sulphatase deficiency (STS)	chrX:6,451,957-8,127,697

*according to DECYPHER v5.0 (https://decipher.sanger.ac.uk)

3

Marcinkowska M, Kozłowski P

„Wpływ polimorfizmu liczby kopii na zmienność fenotypową człowieka”

Postępy Biochemii 2011, 57:240-248

Wpływ polimorfizmu liczby kopii na zmienność fenotypową człowieka

Małgorzata Marcinkowska

Piotr Kozłowski✉

Instytut Chemii Bioorganicznej, Polska Akademia Nauk, Poznań

✉Instytut Chemii Bioorganicznej, Polska Akademia Nauk, ul. Z. Noskowskiego 12/14, 61-704 Poznań; e-mail: kozlowp@yahoo.com

Artykuł otrzymano 13 września 2010 r.
Artykuł zaakceptowano 4 grudnia 2010 r.

Słowa kluczowe: zmienność liczby kopii (CNV), NAHR, gen *AMY1*, osteoporoza, łuszczyca, HIV/AIDS

Wykaz skrótów: aCGH (ang. *array comparative genome hybridization*) – porównawcza hybrydyzacja genomowa do macierzy; CNV (ang. *copy number variation/variants*) – zmienność/warianty liczby kopii; MLPA (ang. *multiplex ligation-dependent probe amplification*) – zależna od ligacji multipleksowa amplifikacja sond; NAHR (ang. *non-allelic homologous recombination*) – niealleliczna rekombinacja homologiczna; OR (ang. *odd ratio*) – iloraz szans SNP (ang. *single nucleotide polymorphism*) – polimorfizm pojedynczego nukleotydu

Podziękowanie: Publikacja została przygotowana w trakcie realizacji projektu badawczego Ministerstwa Nauki i Szkolnictwa Wyższego Nr N N302 278937.

STRESZCZENIE

Zmienność fenotypowa populacji człowieka determinowana jest w większości przez dwa uzupełniające się czynniki: wpływ środowiska oraz wpływ informacji genetycznej. Za zmienność genetyczną odpowiedzialne są różnice (polimorfizmy, mutacje) występujące w genomie człowieka. Do niedawna uważano, że większość tych różnic stanowią niewielkie zmiany jednego lub kilku nukleotydów (SNP), których miliony występują w genomie człowieka. Najnowsze badania całych genomów pokazały jednak, że w genomie człowieka występują również polimorfizmy, obejmujące setki tysięcy par zasad DNA. Takie warianty sekwencji, określane mianem polimorfizmu liczby kopii (CNV), często obejmują geny i inne funkcjonalne elementy genomu. W artykule tym, na tle ogólnej charakterystyki polimorfizmu liczby kopii, przedstawiamy kilka przykładów wpływu tego polimorfizmu na fenotyp człowieka.

WPROWADZENIE

Genom człowieka obejmuje blisko 3 miliardy nukleotydów. Ich charakterystyczny układ zawiera informację genetyczną, będącą wspólną cechą genomów wszystkich ludzi. Mimo tego podobieństwa, porównanie genomów reprezentujących różne ludzkie populacje, jak również bezpośrednie porównanie indywidualnych genomów nawet blisko spokrewnionych osób, wykazuje istnienie szeregu różnic, czyli polimorfizmu. To właśnie polimorfizm genetyczny w znacznym stopniu odpowiedzialny jest za zróżnicowanie w obrębie naszej populacji. Zróżnicowanie to dotyczy większości cech fenotypowych, takich jak wygląd zewnętrzny, poziom markerów biochemicznych czy stan zdrowia. Polimorfizm genetyczny może determinować występowanie chorób, modyfikować ich ryzyko, ostrość objawów, przebieg oraz reakcje na stosowane terapie. Polimorfizmy o bardzo niskiej częstości w populacji lub polimorfizmy o bardzo silnym oddziaływaniu na fenotyp nazywa się mutacjami.

Do niedawna sądzono, że główną przyczyną genetycznej zmienności w populacji ludzkiej jest polimorfizm pojedynczych nukleotydów (SNP), który stanowi najpowszechniejszą formę polimorfizmu w genomie człowieka. Z tego powodu podjęto szereg wieloośrodkowych projektów (np. International HapMap Project, Programs for Genomic Applications NHLBI-PGA), zmierzających do dokładnego scharakteryzowania tego polimorfizmu w genomie człowieka [1,2]. Obecnie baza danych dbSNP zawiera ponad 11 milionów SNP w genomie człowieka [3], co odpowiada częstości ponad 1 SNP na 300 pz.

Analiza asocjacji setek tysięcy markerów SNP, doprowadziła w ostatnich latach do identyfikacji szeregu miejsc w genomie (*loci*), których polimorfizm związany jest z różnymi, powszechnie występującymi chorobami lub ich fenotypami składowymi (ang. *subfenotypes*). W przypadku fenotypów o charakterze ilościowym (np. wysokość, czy masa ciała), takie *loci* określa się mianem QTL (ang. *quantitative trait loci*). Przykładami największych osiągnięć w tym zakresie jest identyfikacja *loci*, których związek z takimi chorobami jak cukrzyca, choroby krążenia, astma czy rak płuc, piersi i prostaty, został potwierdzony w kilku różnych populacjach [4-8]. Pomimo tych osiągnięć, dotychczas wykryte sygnały asocjacji (markery SNP korelujące z chorobą) tłumaczą zaledwie niewielki procent (<5%) zmienności genetycznej badanych chorób, a w wielu przypadkach, pomimo wykrycia sygnału asocjacji, nie udało się zidentyfikować funkcjonalnych wariantów sekwencji, faktycznie modyfikujących ryzyko choroby.

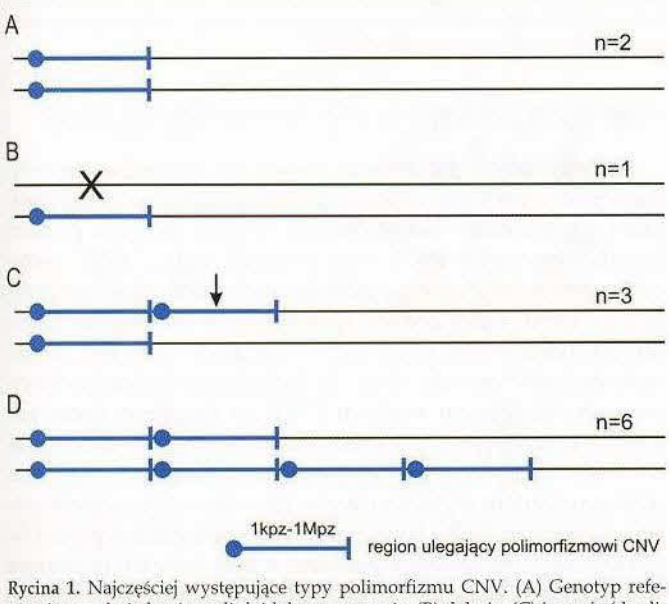
Nowe spojrzenie na polimorfizm genomu człowieka pojawiło się w 2004 roku, kiedy to w dwóch niezależnych pracach wykazano, że duże zmiany strukturalne mogą powszechnie występować w genomie człowieka [9,10]. Takie zmiany nazywane są zmiennością liczby kopii lub wariantami liczby kopii (CNV). Wcześniej, tak zdefiniowane warianty sekwencji znane były głównie,

jako mutacje utraty funkcji w genach związanych z chorobami człowieka.

STRUKTURA I CHARAKTERYSTYKA POLIMORFIZMU LICZBY KOPII

Warianty liczby kopii definiowane są, jako segmenty DNA o wielkości od 1 kpz do kilku Mpz, w których zaobserwowano relatywne zwiększenie lub zmniejszenie liczby kopii w porównywanych genomach [11]. CNV mogą występować zarówno w postaci delecji (Ryc. 1B), jak i insercji (Ryc. 1C). Większość insercji to duplikacje znanych fragmentów genomu. Takie duplikacje zwykle mają charakter bezpośrednich, tandemowych powtórzeń i często występują w formie powtórzeń wielokrotnych (Ryc. 1D). Część CNV ma charakter bardziej złożonych rearanżacji, często będących wynikiem wielokrotnych insercji, delecji i inwersji. Dodatkowym typem zmian strukturalnych są inwersje i translokacje, które jednak nie prowadzą do zmiany liczby kopii.

Jako że miejsca pęknięć chromosomu (końce CNV) często występują w obrębie segmentowych duplikacji, zwanych również powtórzeniami o niskiej liczbie kopii LCR (ang. *low-copy repeat*), najczęściej dyskutowanym mechanizmem powstawania CNV jest niealleliczna rekombinacja homologiczna (NAHR) [12-14]. Segmentowe duplikacje definiuje się, jako długie (>10 kpz) odcinki DNA, występujące w genomie w kilku kopiach charakteryzujących się wysokim stopniem homologii (>95%). Odpowiednie umiejscowienie tych kopii może indukować NAHR, a tym samym prowadzić do powstawania delecji lub duplikacji flankowanego przez segmentowe duplikacje odcinka DNA [15]. NAHR najczęściej zachodzi pomiędzy segmentowymi duplikacjami leżącymi w ramieniu tego samego chromosomu (zwykle w odległości <1 Mpz). Systematyczne badania przeprowadzone w skali całego genomu wykazały, że CNV występują około 4-krotnie częściej w rejonach o wielkości 50 kpz–10 Mpz, otoczonych przez segmentowe duplikacje [16]. Nie wszystkie znane CNV można jednak wytłumaczyć indukowaną przez segmentowe duplikacje NAHR.



Rycina 1. Najczęściej występujące typy polimorfizmu CNV. (A) Genotyp referencyjny – dwie kopie w diploidalnym genomie, (B) delecja, (C) insercja (duplikacja), (D) polimorfizm liczby tandemowych powtórzeń. Niebieska pogrubiona linia reprezentuje polimorficzny region o zmiennej liczbie kopii.

Więcej światła na mechanizm powstawania CNV rzuciły wyniki najnowszych badań, w których z wykorzystaniem masowego sekwencjonowania precyzyjnie (z nukleotydową rozdzielczością) zidentyfikowano pozycje setek CNV [17]. Badania te pokazały, że tylko 10–15% CNV może być wytłumaczonych mechanizmem NAHR. Końce pozostałych CNV (i) w około 50% wykazywały mikrohomologie (homologia odcinków DNA o długości 1–10 nt), (ii) w 10–15% zlokalizowane były w obrębie regionów o zmiennej liczbie tandemowych powtórzeń (VNTR, ang. *variable number tandem repeats*), (iii) w 30% zawierały krótsze lub dłuższe insercje, a (iv) w 5% były zakończone tępo (ang. *blunt ends*). Częste występowanie mikrohomologii w obrębie końców CNV sugeruje, że znacząca część CNV może powstawać w wyniku błędów replikacji polegających na uwolnieniu syntetyzowanej nici i przeniesieniu jej do innych widełek replikacyjnych, gdzie od miejsca wykazującego niewielką homologię kontynuowana jest replikacja. Proces ten określany jest mianem FoSTeS (ang. *replication fork stalling and template switching*) [18].

Prowadzone w wielu ośrodkach badania, pozwoliły na dokładniejsze poznanie strukturalnego polimorfizmu genomu człowieka i wykazały, że przynajmniej część polimorfizmów CNV ma charakter założycielski i cechuje się podobnymi właściwościami jak powszechne polimorfizmy SNP: dziedziczenie mendlowskie w rodzinie, dystrybucja w populacji zgodna z zasadą Hardy'ego-Weinberga, podobny jak SNP rozkład częstości, zakres nierównowagi sprzężeń (ang. *linkage disequilibrium*, LD) oraz występowanie w tych samych haplotypach w różnych populacjach ludzkich [12,16,19,20].

POLIMORFIZM LICZBY KOPII W GENOMIE CZŁOWIEKA

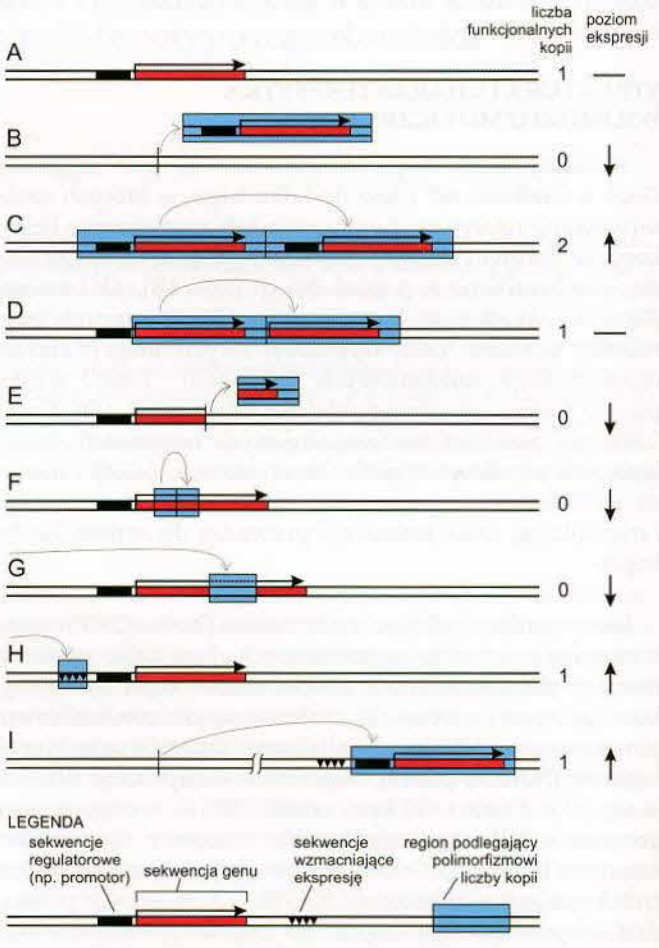
Zidentyfikowanie pojedynczych przypadków CNV spowodowało wzrost zainteresowania poznaniem struktury i funkcjonalnego znaczenia CNV w genomie człowieka. Dotychczas do identyfikacji CNV w skali całego genomu najczęściej wykorzystywano takie narzędzia jak: porównawcza hybrydyzacja genomowa do macierzy (aCGH) [13], mikromacierze SNP [19] czy analiza błędów dziedziczenia markerów SNP (w regionach CNV polimorfizmy SNP częściej wykazują niezgodność z równowagą Hardy'ego-Weinberga lub brak zgodności z zasadami dziedziczenia mendlowskiego) [21,22]. W przypadku zastosowania aCGH i mikromacierzy SNP, CNV identyfikowane są na podstawie porównania sygnałów hybrydyzacyjnych dwóch genomów, z których jeden traktowany jest jako genom referencyjny. Dodatkowo mikromacierze SNP pozwalają na identyfikację CNV w oparciu o wzory genotypów sąsiadujących ze sobą SNP. Przykładowo, nadreprezentacja (zgrupowanie) w określonym regionie genomu markerów SNP o genotypie homozygotycznym, sugeruje występowanie delecji jednego allelu. Ostatnio do identyfikacji CNV zastosowano również technologię masowego sekwencjonowania [17,23]. Metodami, stosowanymi do charakterystyki pojedynczych CNV są: zależna od ligacji multipleksowa amplifikacja sond (MLPA) [24], ilościowa reakcja łańcuchowa polimerazy (qPCR, ang. *quantitative PCR*) [25], fluorescencyjna hybrydyzacja *in situ*

(FISH, ang. *fluorescense in situ hybridization*) oraz PRT (ang. *paralogue ratio test*) [26]. W metodzie MLPA liczba kopii badanego regionu genomu określana jest na podstawie intensywności sygnałów amplifikacji specjalnie przygotowanych sond, z których każda składa się z dwóch oligonukleotydów (pół-sond), rozpoznających bezpośrednio sąsiadujące sekwencje w genomie. Tylko pół-sondy prawidłowo rozpoznające sekwencje docelowe mogą podlegać ligacji, a sygnał ich amplifikacji jest proporcjonalny do liczby kopii badanego regionu w genomie [27,28]. FISH jest metodą cytogenetyczną, która umożliwia bezpośrednią obserwację badanych regionów genomu, rozpoznawanych przez wyznakowane fluorescencyjnie sondy. Chociaż metoda ta charakteryzuje się niską rozdzielczością, pozwala ona nie tylko na określenie liczby kopii badanego regionu, ale również na określenie jego orientacji i pozycji w genomie. Metoda PRT, która została zastosowana do analizy genu *DEFB4*, polega na porównaniu intensywności sygnałów amplifikacji fragmentów genu, podlegającego zmienności liczby kopii oraz niepolimorficznych paralogów tego genu [26].

Zastosowanie wyżej wymienionych metod pozwoliło na identyfikację tysięcy wariantów CNV w różnych populacjach ludzkich. Przykładowo, Redon i wsp., z wykorzystaniem aCGH i mikromacierzy SNP, zidentyfikowali 1447 regionów CNV w czterech populacjach z Europy, Afryki i Azji [19]. Sumaryczna długość tych regionów wynosi 360 Mpz, co stanowi około 12% całkowitej długości genomu człowieka. Do tej pory w największej bazie danych polimorfizmu strukturalnego (Database of Genomic Variants, DGV) zgromadzono informacje o ponad 14 tys. regionów CNV, obejmujących około 30% genomu człowieka. Prawdopodobnie jednak, przynajmniej część z opisanych polimorfizmów CNV jest fałszywie pozytywnymi artefaktami lub reprezentuje bardzo rzadkie, „prywatne" warianty sekwencji. Wyniki ostatnich prac uzyskane za pomocą metod umożliwiających dokładne mapowanie końców CNV i precyzyjne genotypowanie poszczególnych CNV [13,29], pozwalają oszacować, że częsty (>1%) polimorfizm typu CNV obejmuje około 10% genomu człowieka.

ZWIĄZEK CNV Z FENOTYPEM CZŁOWIEKA

Chociaż wielokrotnie wykazano, że CNV (szczególnie delecje) częściej występują w regionach genomu o małym lub nieznanym znaczeniu funkcjonalnym, to jednak w obrębie znanych regionów CNV znajdują się setki ważnych, funkcjonalnych elementów genomu (np. geny kodujące białka). Przykładowo, w obrębie CNV zidentyfikowanych przez Redon i wsp. znajduje się blisko 3 tys. genów kodujących białka, w tym 285 genów związanych z chorobami człowieka, zaklasyfikowanych do bazy danych OMIM (Online Mendelian Inheritance in Man). Ponadto w regionach tych zlokalizowanych jest wiele innych, funkcjonalnych (lub potencjalnie funkcjonalnych) elementów genomu, w tym 50 sekwencji ultrakonserwatywnych, ponad 130 tys. zachowawczych elementów niekodujących oraz 67 sekwencji niekodujących RNA [19]. W innych badaniach, w wysoce polimorficznych obszarach CNV zidentyfikowano ponad 600 genów [13]. Takie CNV mogą wpływać na poziom ekspresji oraz funkcje genów i innych elementów funkcjonalnych, znajdujących się w ich obrębie.



Rycina 2. Wpływ polimorfizmu CNV na strukturę i funkcję genu. (A) Allel referencyjny z jedną funkcjonalną kopią genu. (B) Delecja genu powoduje zmniejszenie liczby funkcjonalnych kopii genu (zmniejszenie dawki genu). (C) Insercja (duplikacja) genu wraz z regionem regulatorowym, powoduje zwiększenie liczby funkcjonalnych kopii genu (zwiększenie dawki genu). (D) Duplikacja genu bez regionu regulatorowego nie powoduje zmiany liczby funkcjonalnych kopii genu (dawka genu pozostaje bez zmian). (E, F i G) Delecje lub insercje fragmentów genu prowadzące do uszkodzenia struktury genu. Większość tego typu wariantów sekwencji prowadzi do utraty funkcji genu (zmniejszenie dawki genu). (H) Translokacja/duplikacja sekwencji regulatorowej wzmacniającej ekspresję w pobliże genu (wzrost poziomu ekspresji bez zmiany dawki genu). (I) Translokacja genu w pobliże sekwencji wzmacniającej ekspresję - efekt pozycji (wzrost poziomu ekspresji bez zmiany dawki genu).

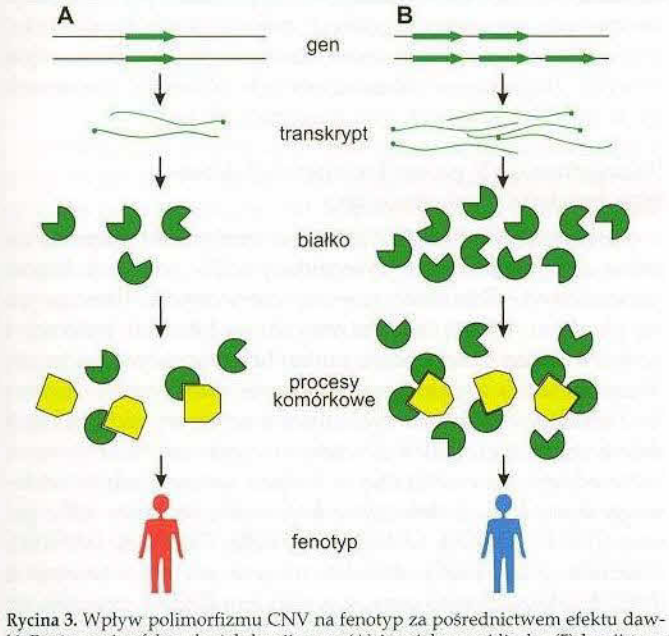
WPŁYW POLIMORFIZMU CNV NA EKSPRESJĘ GENÓW

Podstawowym fenotypem, mogącym podlegać modyfikacji pod wpływem zmiany struktury genu jest ekspresja, która na poziomie molekularnym wyraża się, jako poziom transkrybowanego RNA oraz poziom białka. CNV może oddziaływać na ekspresję genu poprzez: uszkodzenie/przerwanie genu, efekt pozycji (przeniesienie fragmentu genu) lub modyfikowanie elementów regulatorowych (np. regionów promotorowych) (Ryc. 2). Jednak najbardziej oczywistym mechanizmem wpływu CNV na ekspresję genu, jest zmiana liczby kopii, czyli efekt dawki, do którego zaistnienia niezbędne jest, aby region polimorficzny obejmował cały gen wraz z regionem regulatorowym. Zjawisko efektu dawki polega na tym, że liczba funkcjonalnych kopii genu wpływa na poziom jego transkryptu (mRNA), a tym samym na poziom kodowanego przez ten gen białka (Ryc. 3). Efekt dawki został potwierdzony dla wielu CNV, zarówno na poziomie transkryptu, jak i na poziomie białka [30-34].

Liczba kopii koreluje z poziomem ekspresji wielu genów

W celu kompleksowego zbadania, w jaki sposób CNV wpływa na ekspresję genów kodujących białka, porównano ekspresję (poziom transkryptu mRNA) ponad 14000 genów z liczbą kopii ponad 1300 najlepiej potwierdzonych regionów CNV w czterech populacjach ludzkich. Znaczącą korelację między CNV a poziomem mRNA zaobserwowano dla 99 genów. Dla blisko połowy tych genów, leżących w obszarach polimorficznych, wpływ CNV na ekspresję wynikał najprawdopodobniej z efektu dawki (pozytywna korelacja). Pozostałe geny albo leżały poza obszarem polimorficznym, z którym korelował poziom ich ekspresji albo ich ekspresja wykazywała negatywną korelację z liczbą kopii zasocjowanego polimorfizmu. Sugeruje to, że zmienne poziomy ekspresji tych genów nie były skutkiem efektu dawki, ale wynikały z innych form oddziaływania CNV na ekspresję, np. uszkodzenia struktury genów, efektu pozycji, czy oddziaływania na sekwencje regulujące ekspresję genów [35]. Bardziej szczegółowe badania, w których porównano poziomy transkrypcji kilku genów, leżących w obrębie delecyjnych polimorfizmów CNV, wykazały, że obserwowane różnice w ekspresji tych genów w znacznym stopniu korelowały z liczbą kopii. Zmienność poziomu mRNA w 26–88% wynikała ze zmienności dawki (liczby kopii) poszczególnych genów. Zmniejszenie liczby kopii badanych genów o jeden powodowało obniżenie poziomu mRNA o 30–38% [21]. W innych badaniach oszacowano, że zróżnicowanie ekspresji (poziomu mRNA) genów leżących w obrębie CNV w około 50% wynika ze zmienności liczby kopii tych genów [36].

Wpływ CNV na poziom mRNA sugeruje, że zmiana liczby kopii może również modyfikować poziom powstającego w procesie translacji białka. W literaturze przedstawiono kilka przykładów pozytywnej korelacji między liczbą kopii genów a poziomem kodowanych przez te geny białek [31,33,36-38]. Przykładem takiej korelacji jest CNV genu *FCGR3B*, znajdującego się na chromosomie 1 (1q23.3). Gen *FCGR3B* koduje receptor FcγRIIIb występujący na powierzchni komórek układu odpornościowego człowieka (głównie neutrofili, czyli granulocytów obojętnochłonnych), który umożliwia ich przyleganie do powierzchni pokrytych immunoglobulinami G (IgG) podczas wtórnej odpowiedzi immunologicznej. Przeprowadzone badania wykazały, że liczba receptorów FcγRIIIb na powierzchni neutrofili pozytywnie koreluje z liczbą kopii genu *FCGR3B*. Ponadto u osób z wyższą liczbą kopii genu *FCGR3B* obserwowano cztery razy wyższy poziom adhezji neutrofili do powierzchni pokrytych przeciwciałami IgG [38]. W innych badaniach obserwowano różnice w poziomach hormonów steroidowych, które wynikały z liczby kopii genu *UGT2B17*, kodującego enzym UDP-glukuronozylotransferazę, biorący udział w inaktywacji hormonów steroidowych [33].



Rycina 3. Wpływ polimorfizmu CNV na fenotyp za pośrednictwem efektu dawki. Porównanie efektu dwóch kopii genu (A) i zwiększonej liczby (5) kopii tego genu (B). Polimorfizm CNV zmienia liczbę funkcjonalnych kopii genu, co z kolei ma wpływ na poziom transkryptu oraz poziom kodowanego przez ten gen białka. Zmiana poziomu białka wpływa na regulowane przez to białko procesy komórkowe, co w konsekwencji może prowadzić do modyfikacji fenotypu człowieka.

Jednak nie zawsze zmiana liczby kopii genu musi prowadzić do zmiany poziomu jego ekspresji [35]. Przykładem takiego polimorfizmu jest CNV, w obrębie którego znajduje się gen *OPN1MW*. Gen ten koduje światłoczuły barwnik – opsynę, obecną w czopkach siatkówki oka, które odpowiedzialne są za widzenie barw. Pojedynczy allel (chromosom X) zawiera zwykle od 1 do 3 kopii genu *OPN1MW*. Obserwowany brak korelacji między liczbą kopii a poziomem ekspresji genu *OPN1MW* może wynikać z faktu, iż region polimorficzny obejmuje wyłącznie sekwencję genu, natomiast nie obejmuje regionu regulatorowego, zlokalizowanego 40 kpz powyżej tego genu. W związku z tym, bez względu na całkowitą liczbę kopii, tylko pierwsza kopia genu *OPN1MW* jest kopią funkcjonalną, kolejne, z powodu zbyt dużego dystansu od regionu regulatorowego, nie ulegają ekspresji [39,40].

CVN genów mikroRNA może mieć istotny udział w modyfikacji fenotypu człowieka

Jak już wcześniej wspomniano, polimorfizm liczby kopii może modyfikować nie tylko ekspresję genów kodujących białka, ale również sekwencje regulatorowe, m.in. niekodujące RNA, w tym mikroRNA. MikroRNA (miRNA) są to krótkie, jednoniciowe cząsteczki RNA (~21 nt), regulujące ekspresję genów na poziomie translacji, poprzez, nie w pełni komplementarne, oddziaływanie miRNA z regionem 3'UTR cząsteczek mRNA. Szacuje się, że u człowieka występuje około 1000 różnych miRNA, regulujących ekspresję przynajmniej 30% genów.

Autorzy niniejszego artykułu przeprowadzili badania, mające na celu identyfikację genów miRNA, leżących w polimorficznych obszarach CNV (CNV-miRNA, ang. *copy number variable microRNA genes*). W badaniach tych porównano lokalizację wszystkich znanych genów miRNA z lokalizacją znanych CNV [60]. Porównanie to pozwoliło na identyfikację 209 miRNA, z których 11 zlokalizowanych jest w grupie wysokopolimorficznych CNV. Wśród zidentyfikowanych CNV-miRNA występują: delecje (np. hsa-mir-384, hsa-mir-1324), duplikacje (np. hsa-mir-1972, hsa-mir-1977) i wielokrotne duplikacje (np. hsa-mir-1233, hsa-mir-1268), a liczba kopii tych miRNA waha się od 0 do 8. Dodatkowo, porównanie frakcji genomu objętej polimorfizmem CNV oraz frakcji miRNA zlokalizowanych w tych regionach, po-

zwoliło ustalić, że geny miRNA występują rzadziej w rejonach podlegających częstemu polimorfizmowi liczby kopii. Wynik ten sugeruje, że CNV podlegają negatywnej selekcji w regionach występowania genów miRNA, co potwierdza funkcjonalny charakter tych ostatnich.

Zmienność liczby kopii genów miRNA, poprzez efekt dawki, może modyfikować poziom dojrzałych miRNA, które z kolei mogą mniej lub bardziej efektywnie wyciszać translację białek z docelowych cząsteczek mRNA. Zmiana poziomu białka może mieć wpływ na szereg modyfikowanych przez te białka procesów/fenotypów. Funkcje wielu miRNA, zidentyfikowanych jako CNV-miRNA, zostały wcześniej powiązane z różnymi ważnymi fenotypami człowieka. Wśród fenotypów tych występują m.in.: męska niepłodność (hsa-mir-383) [41], odrzucanie przeszczepów (hsa-mir-142) [42], czy stwardnienie rozsiane (hsa-mir-1275) [43]. W szeregu prac wykazano również, że zidentyfikowane CNV-miRNA mogą odgrywać ważną rolę w kancerogenezie [44-47].

PRZYKŁADY WPŁYWU POLIMORFIZMU CNV NA FENOTYP CZŁOWIEKA

W związku z tym, że polimorfizm CNV może modyfikować poziom ekspresji genów, można założyć, że będzie on również ważnym czynnikiem modyfikującym fenotyp człowieka, zarówno w stanie normy i w stanach chorobowych. Poniżej zostanie opisanych kilka przykładów związku polimorfizmu CNV z fenotypem człowieka.

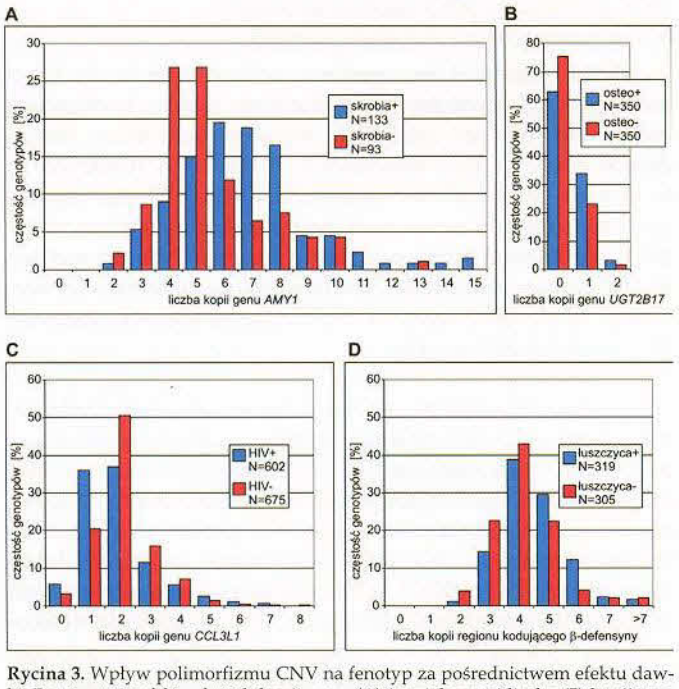
Polimorfizm CNV modyfikuje przystosowanie do zmiennych warunków środowiskowych

Przykładem genu, którego CNV modyfikuje fenotyp w sposób pozwalający na przystosowanie populacji do środowiska, jest gen *AMY1* (1p21.1) [31]. Gen *AMY1* koduje amylazę, enzym obecny w ślinie i soku trzustkowym, który odpowiedzialny jest za hydrolizę skrobi i innych wielocukrów. Gen *AMY1* znajduje się w obrębie CNV, którego liczba kopii różni się znacząco zarówno między poszczególnymi osobami jak i między różnymi populacjami. Liczba kopii genu *AMY1* w diploidalnym genomie człowieka może wynosić od dwóch do nawet 15. Analiza liczby kopii genu *AMY1* (analiza qPCR) oraz ilości amylazy w ślinie (analiza Western blot) u 50 osób pochodzenia europejskiego, ujawniła pozytywną korelację (R^2=0.351; P<0.0001) między liczbą kopii *AMY1* a poziomem amylazy w ślinie [31]. Jako że efektywność trawienia skrobi, na którą wpływa aktywność amylazy, jest czynnikiem mogącym podlegać zróżnicowanej selekcji w różnych populacjach, podjęto badania mające na celu porównanie liczby kopii genu *AMY1* między populacjami różniącymi się pod względem stosowanej diety. Badania przeprowadzono na grupie 133 osób z trzech populacji tradycyjnie stosujących dietę wysokoskrobiową (społeczeństwa rolnicze oraz zbieracze, których dieta oparta jest o korzenie i bulwy) oraz na grupie 93 osób z czterech populacji tradycyjnie stosujących dietę o niskiej zawartości skrobi (społeczeństwa zamieszkujące lasy deszczowe, jedzące dużo owoców oraz populacje północne, których podstawą żywienia są zwierzęta hodowlane i ryby). Przeprowadzona analiza wykazała, że osoby z populacji tradycyjnie spożywających dużo produktów skrobiowych miały średnio więcej (6,7) kopii genu *AMY1* w diploidalnym genomie niż osoby z populacji o niskiej konsumpcji skrobi (średnio 5,4) [31] (Ryc. 4A).

Istnienie różnicy w liczbie kopii *AMY1* i poziomie amylazy między poszczególnymi populacjami potwierdziło hipotezę, że w populacjach o diecie wysokoskrobiowej dobór naturalny faworyzował większą liczbę kopii genu *AMY1*, a tym samym wyższy poziom amylazy. W tym przypadku, zależna od CNV modyfikacja fenotypu ma wpływ na przystosowanie się poszczególnych populacji do środowiska, poprzez ułatwienie trawienia dostępnego pokarmu, a tym samym złagodzenie niekorzystnych objawów ewentualnych chorób jelitowych (np. biegunki) [31].

Polimorfizm CNV genów kodujących β defensyny zwiększa ryzyko łuszczycy

Łuszczyca (OMIM #177900) jest przewlekłą chorobą zapalną skóry człowieka, występującą u 2% populacji krajów rozwiniętych. Charakteryzuje się czerwonymi, łuszczącymi się plamami, występującymi zwykle na łokciach, kolanach i torsie. Badania histologiczne zmian łuszczycowych wykazują obecność stanu zapalnego i zakłócone różnicowanie naskórka. Ponadto w zmianach tych obserwuje się wysoki poziom β defensyn, małych białek o właściwościach antybakteryjnych, które odgrywają ważną rolę w inicjacji systemu odpornościowego skóry [36]. β defensyny kodowane są przez kilka genów (*DEFB1*, *DEFB4*, *SPAG11*, *DEFB103*, *DEFB104*, *DEFB105*, *DEFB106* i *DEFB107*) zlokalizowanych na chromosomie 8 (8p23.1). Wszystkie te geny, z wyjątkiem *DEFB1*, znajdują się w dużym, powtórzonym regionie, obejmującym około 300 kpz, który w różnych populacjach ludzkich występuje zwykle w 2 do 7 kopiach w diploidalnym genomie. Spotykane są jednak genotypy zawierające nawet 12 kopii tego regionu (Ryc. 4D). W związku z antybakteryjnymi i prozapalnymi



Rycina 3. Wpływ polimorfizmu CNV na fenotyp za pośrednictwem efektu dawki. Porównanie efektu dwóch kopii genu (A) i zwiększonej liczby (5) kopii tego genu (B). Polimorfizm CNV zmienia liczbę funkcjonalnych kopii genu, co z kolei ma wpływ na poziom transkryptu oraz poziom kodowanego przez ten gen białka. Zmiana poziomu białka wpływa na regulowane przez to białko procesy komórkowe, co w konsekwencji może prowadzić do modyfikacji fenotypu człowieka.

właściwościami β defensyn, wnioskuje się, że zmiana liczby kopii kodujących je genów może modyfikować ryzyko zapadalności na choroby infekcyjne i zapalne, m.in. łuszczycę. W przeprowadzonych badaniach asocjacji typu *case-control*[1] porównano liczbę kopii genów β defensyn u chorych na łuszczycę oraz w odpowiednio dobranych próbkach kontrolnych z Niemiec i Holandii [36]. Wykonane analizy pokazały, że liczba kopii genów β defensyn w grupach chorych na łuszczycę jest wyższa (średnio 4,51 i 4,54, odpowiednio w populacji niemieckiej i holenderskiej) w porównaniu z grupami osób zdrowych (4,14 i 4,18) (t-test, $p=2,95x10^{-5}$ i $p=1,65x10^{-6}$). Oszacowano, że każda dodatkowa kopia (powyżej 2) grupy genów β defensyn, zwiększa relatywne ryzyko zachorowania na łuszczycę o około 34%. Wynika to najprawdopodobniej z efektu dawki jednego, kilku lub wszystkich genów β defensyn w regionie polimorficznym. Wyższa liczba kopii powoduje zwiększoną ekspresję β defensyn, a te z kolei intensywniej stymulują keratynocyty do uwalniania szeregu interleukin, które mają właściwości prozapalne. To prowadzi do podwyższenia podstawowego poziomu czynników zapalnych, zwiększając ryzyko wystąpienia łuszczycy. W związku z tym, że region polimorficzny obejmuje kilka genów β defensyn, trudno jednak jednoznacznie określić, efekt którego genu (lub genów) jest związany z ryzykiem łuszczycy [36].

Polimorfizm CNV genu *UGT2B17* modyfikuje ryzyko osteoporozy

Osteoporoza (OMIM #166710) to najpowszechniejsza choroba kości charakteryzująca się występowaniem szeregu fenotypów składowych: niską gęstością mineralną kości (BMD, ang. *bone mineral density*), osłabieniem struktury przestrzennej kości (zmniejszeniem grubości kory kości (CT, ang. *cortical thinckness*) i podwyższeniem wskaźnika odkształcenia (BR, ang. *buckling ratio*) oraz występowaniem złamań osteoporotycznych (OF, ang. *osteoporotic fracture*). Chociaż wszystkie powyższe fenotypy składowe w znacznym stopniu determinowane są przez czynniki genetyczne, do tej pory udało się zidentyfikować zaledwie kilka genów lub SNP, które mogą zwiększać ryzyko rozwoju osteoporozy.

W związku z powyższym, przeprowadzono analizę CNV w całym genomie w grupie 350 osób ze zdiagnozowaną osteoporozą oraz w grupie 350 odpowiednio dobranych osób kontrolnych (niewykazujących symptomów choroby) z populacji chińskiej. Analiza asocjacji w badaniach typu *case-control* wszystkich zidentyfikowanych CNV ujawniła jeden częsty polimorfizm delecyjny w *locus* 4q13.2, który wykazywał znaczące różnice w rozkładzie genotypów w badanych grupach. Częstość obserwowanych genotypów w *locus* 4q13.2 w grupie badanej oraz kontrolnej wynosiła odpowiednio: 0 kopii (delecje homozygotyczne) 62,9% i 75,5%, jedna kopia (delecje heterozygotyczne) 34% i 23,1% oraz 2 kopie 3,1% i 1,4% (Ryc. 4B). Wyniki te pokazują, że posiadanie jednej lub dwóch kopii tego regionu, zwiększa ryzyko wystąpienia osteoporozy w porównaniu do osób nieposiadających żadnej kopii tego regionu (iloraz szans OR=1,73, $p=2,0x10^{-4}$). Powyższe wyniki zostały potwierdzone również w innych badaniach typu *case-control*, zarówno w innych populacjach azjatyckich jak i w populacji europejskiej [33].

[1]W polskiej literaturze naukowej takie badania określa się często jako badania kliniczno-kontrolne.

Chociaż w *locus* 4q13.2 znajduje się pięć genów (*UGT2B17, YTHDC1, TMPRSS11E, TMPRSS11E2 i UGT2B15*), dokładna analiza wykazała, że polimorfizm CNV obejmuje region 150 kpz i obejmuje tylko gen *UGT2B17* [33]. W dalszych badaniach przeanalizowano więc asocjację polimorfizmu genu *UGT2B17* w *locus* 4q13.2 z poszczególnymi fenotypami składowymi w patogenezie osteoporozy. Wyniki analiz w populacji chińskiej i europejskiej wyraźnie wskazały, że liczba kopii genu negatywnie koreluje z fenotypami BMD oraz CT, natomiast pozytywnie koreluje z fenotypem BR [33]. Z powyższych analiz wynika, że wpływ liczby kopii genu *UGT2B17* na ryzyko osteoporozy odbywa się poprzez modyfikację fenotypów składowych związanych z metabolizmem i strukturą kości.

Analiza funkcji genu *UGT2B17* pozwoliła na ustalenie mechanistycznego związku między liczbą kopii tego genu a ryzykiem wystąpienia osteoporozy. Gen *UGT2B17* koduje enzym z grupy UDP-glukuronylotransferaz, który odgrywa kluczową rolę w utrzymywaniu homeostazy i metabolizmie wielu endogennych molekuł, w tym hormonów steroidowych [48]. Niektóre z tych hormonów, androgeny i ich pochodne estrogeny, odgrywają znaczącą rolę w utrzymywaniu integralności kości gąbczastej i mają właściwości stymulujące formowanie kości. Wykazano, że wyższa liczba kopii genu *UGT2B17* powoduje podwyższenie poziomu syntezy enzymu inaktywującego androgeny, a w konsekwencji obniżenie poziomu hormonów steroidowych (testosteron, estradiol). To z kolei zwiększa ryzyko osteoporozy w wyniku nieprawidłowego formowania lub wzmożonej resorpcji kości [33].

Wyższa liczba kopii genu chemokiny CCL3L1 obniża ryzyko infekcji wirusem HIV

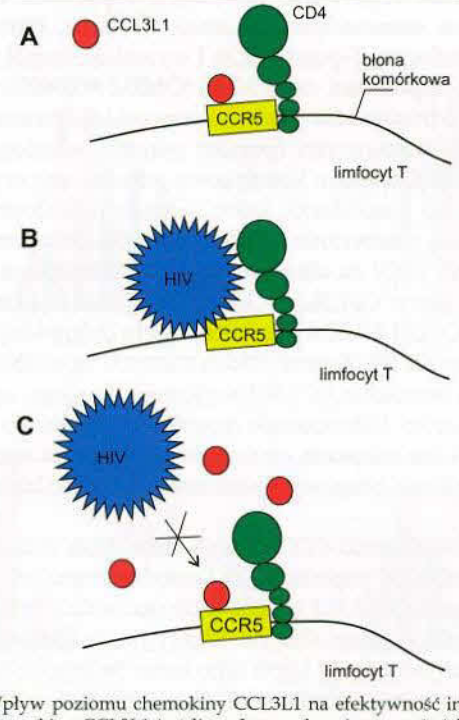
Wirus HIV czyli wirus niedoboru odporności człowieka (ang. *human immunodeficiency virus*), to wirus, który atakuje głównie limfocyty T-pomocnicze i wywołuje zespół nabytego niedoboru odporności, czyli AIDS (OMIM #609423). Przebieg zakażenia wirusem HIV jest wynikiem oddziaływania między czynnikami wirusowymi (poziom wiremii, wirulencja, zdolność transmisji, tropizm komórkowy wirusa) oraz czynnikami zależnymi od gospodarza, które w znacznym stopniu determinowane są genetycznie [49]. Jednym z takich czynników jest polimorfizm CNV na chromosomie 17. Polimorfizm ten obejmuje dwa geny: *CCL3L1* i *CCL4L1*, które kodują chemokiny CCL3L1 i CCL4L1. Głównym receptorem chemokiny CCL3L1 jest receptor CCR5. Ze wszystkich znanych ligandów receptora CCR5, chemokina CCL3L1 wykazuje do niego największe powinowactwo. Jednocześnie receptor CCR5 wraz z koreceptorem CD4 jest miejscem rozpoznawanym przez wirusa HIV, które umożliwia integrację i wnikanie wirusa do komórki [50].

Liczba kopii genu *CCL3L1* wykazuje dużą zmienność zarówno w obrębie populacji, jak i między populacjami. Analiza CNV genu *CCL3L1* u 1064 osób pochodzących z 57 różnych populacji pozwoliła na identyfikację genotypów zawierających od 0 do 14 kopii tego genu. Jednocześnie wykazano, że populacje afrykańskie charakteryzują się znacznie większą średnią liczbą kopii genu *CCL3L1* (średnio 6 kopii) niż populacje nieafrykańskie. Przykładowo, populacje europejskie (np. Hiszpanie, Francuzi) mają średnio dwie kopie, populacje centralno- i południowoazjatyckie mają średnio trzy kopie, a populacje wschodnioazjatyckie (np. Japończy-

cy, Chińczycy) i amerykańskie (np. Kolumbijczycy) mają średnio cztery kopie genu *CCL3L1*. Ustalono że, geograficzny region zamieszkiwania lub pochodzenia badanych osób tłumaczy blisko 35% zmienności rozkładu liczby kopii genu *CCL3L1* w populacjach ludzkich [37].

Podjęte badania typu *case-control* osób zakażonych HIV (HIV^+) oraz osób niezakażonych (HIV^-), pochodzących z czterech różnych populacji ludzkich wykazały, że we wszystkich tych populacjach średnia liczba kopii genu *CCL3L1* była wyższa u osób zdrowych niż u nosicieli wirusa HIV (Ryc. 4C). Wynik ten sugeruje, że wyższa liczba kopii genu *CCL3L1* może mieć działanie ochronne, zmniejszające ryzyko zakażenia wirusem HIV. Oszacowano, że każda dodatkowa kopia tego genu obniża ryzyko zakażenia wirusem HIV o 4,5–10%. Wynika to najprawdopodobniej z efektu dawki genu *CCL3L1*. Zwiększenie liczby kopii *CCL3L1*, wpływa na zwiększoną syntezę chemokiny CCL3L1, która jako ligand o wysokim powinowactwie, konkuruje z wirusem HIV o dostęp do receptora CCR5, tym samym utrudnia wiązanie wirusa do receptora, co w konsekwencji zmniejsza szanse infekcji i ryzyko zakażenia wirusem HIV (Ryc. 5). Dalsze badania wykazały, że liczba kopii genu *CCL3L1* wpływa nie tylko na obniżenie ryzyka infekcji wirusem HIV, ale również na przebieg choroby. Analiza liczby kopii genu *CCL3L1* u nosicieli wirusa HIV wykazała, że osoby z wyższą liczbą kopii tego genu, później rozwijały pełne objawy AIDS oraz wykazywały niższą śmiertelność w badanych przedziałach czasowych [37].

Wyższa liczba kopii genu *CCL3L1*, a tym samym wyższy poziom chemokiny CCL3L1, zmniejsza ryzyko infekcji wirusem HIV oraz prawdopodobnie innych chorób infekcyjnych. Jednak te same allele (wyższa liczba kopii genu *CCL3L1*) zwiększają ryzyko wystąpienia chorób zapalnych i autoimmunologicznych. Podobne zjawisko obserwuje się także w przypadku innych polimorfizmów, których allele obniżają ryzyko chorób infekcyjnych. Podjęte badania typu *case-control* wykazały, że osoby z wyższą liczbą kopii genu *CCL3L1* mają większe ryzyko wystąpienia ostrej choroby zapalnej Kawasaki [51] oraz dwóch chorób autoimmunologicznych: reumatoidalnego zapalenie stawów (RA, ang. *rheumatoid arthritis*) [52] i układowego toczenia rumieniowatego (SLE, ang. *systemic lupus erythematosus*) [53].



Rycina 5. Wpływ poziomu chemokiny CCL3L1 na efektywność infekcji wirusa HIV. (A) Chemokina CCL3L1 jest ligandem wykazującym największe powinowactwo do receptora CCR5, znajdującego się na powierzchni limfocytów pomocniczych T. (B) Receptor CCR5 wraz z koreceptorem CD4 jest miejscem rozpoznawanym przez wirusa HIV, które umożliwia integrację i wnikanie wirusa do komórki. (C) Zwiększony, w wyniku efektu dawki, poziom chemokiny CCL3L1 utrudnia dostęp wirusa HIV do receptora CCR5, a tym samym zmniejsza ryzyko infekcji.

Polimorfizm CNV genu *CYP2D6* wpływa na szybkość metabolizmu wielu leków

Innym przykładem wpływu polimorfizmu na fenotyp człowieka jest modyfikacja metabolizmu leków. Zmienność indywidualnej zdolności metabolizowania leków może mieć daleko idące konsekwencje, zarówno w kontekście toksyczności leków, jak również ich skuteczności w leczeniu. W metabolizmie leków znaczną rolę odgrywa cytochrom P450, katalizujący rozkład leków. P450 występuje w wielu formach izoenzymatycznych (szerzej opisanych w [54]), z których najważniejsze to CYP3A4 oraz CYP2D6 [55]. Enzym CYP2D6 jest zaangażowany w metabolizm 20-25% zatwierdzonych obecnie leków [56], w tym leków antydepresyjnych, przeciwbólowych, leków usuwających nudności, neuroleptycznych, leków przeciw arytmiom serca oraz leków przeciwnowotworowych (np. tamoxifen) [57].

Gen kodujący enzym CYP2D6 (*CYP2D6*) zlokalizowany jest na chromosomie 22 (22q13.2) i występuje u człowieka w różnej liczbie tandemowych powtórzeń [58]. Liczba kopii (powtórzeń) genu *CYP2D6* w diploidalnym genomie waha się od jednego do nawet 12. Wykazano, że liczba funkcjonalnych kopii genu *CYP2D6* wpływa na poziom jego ekspresji, a ta z kolei jest silnie skorelowana z szybkością metabolizowania leków. Osoby nieposiadające funkcjonalnych kopii genu *CYP2D6*, należą do grupy charakteryzującej się wolnym metabolizmem leków (PM, ang. *poor metabolizers*). W przeciwieństwie do nich, osoby posiadające dużą liczbę kopii genu *CYP2D6* (więcej niż 4 w diploidalnym genomie) charakteryzują się fenotypem bardzo szybkiego metabolizmu leków (UM, ang. *ultra-rapid metabolizers*). Fenotyp ten związany jest z brakiem odpowiedzi na terapie przy standardowych dawkach leków. Dodatkowym czynnikiem modyfikującym powyższe zależności, są mutacje punktowe, wpływające na poziom ekspresji i funkcjonalność poszczególnych kopii genu *CYP2D6* [59].

Poza dużym zróżnicowaniem osobniczym, średnia liczba kopii genu *CYP2D6*, a tym samym efektywność metabolizmu leków, różni się znacząco także między różnymi ludzkimi populacjami. Przykładowo, 30% populacji etiopskiej czy populacji saudyjskiej charakteryzuje się fenotypem UM, co koreluje z częstym występowaniem w tej populacji alleli zawierających 2, 3, 4 czy nawet 5 funkcjonalnych kopii genu *CYP2D6* i jednoczesnym brakiem w tej populacji osób nieposiadających ani jednej funkcjonalnej kopii tego genu. W przeciwieństwie do populacji wymienionych powyżej, fenotyp UM występuje bardzo rzadko w populacjach północno-europejskich i praktycznie nie występuje w populacjach azjatyckich, co koreluje z bardzo niską częstością alleli, zawierających więcej niż jedną kopię genu *CYP2D6* w tych populacjach [57,59].

PODSUMOWANIE

Polimorfizm liczby kopii w genomie człowieka jest intensywnie badanym zjawiskiem w wielu ośrodkach naukowych. Wymienione w niniejszej pracy przykłady związków tego polimorfizmu z fenotypem człowieka, to zaledwie kilka spośród obecnie znanych. Dotychczas prowadzone badania, identyfikujące coraz to nowsze CNV i ich wpływ na fenotyp człowieka, są jednak w znacznym stopniu ograniczone przez brak odpowiednich metod, pozwalających na precyzyjne określenie liczby kopii pojedynczych CNV w indywidualnych próbkach. Pewne nadzieje w tym zakresie można wiązać z zastosowaniem do genotypowania CNV metod masowego sekwencjonowania. Chociaż wiedza dotycząca udziału polimorfizmu CNV w zmienności fenotypu człowieka jest jeszcze niepełna, to jednak już teraz można stwierdzić, że polimorfizm liczby kopii jest istotnym czynnikiem modyfikującym nasz fenotyp.

PIŚMIENNICTWO

1. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM et al (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449: 851-861
2. Lee K, Kohane IS, Butte AJ (2003) PGAGENE: integrating quantitative gene-specific results from the NHLBI programs for genomic applications. Bioinformatics 19: 778-779
3. Sherry ST, Ward M, Sirotkin K (1999) dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res 9: 677-679
4. Weiss ST, Raby BA (2004) Asthma genetics 2003. Hum Mol Genet 13 Spec No 1: R83-89
5. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R et al (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087-1093
6. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Spitz MR, Eisen T, Amos CI, Houlston RS (2008) Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat Genet 40: 1407-1409
7. Doria A, Patti ME, Kahn CR (2008) The emerging genetic architecture of type 2 diabetes. Cell Metab 8: 186-200
8. Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC (2007) Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. Nat Genet 39: 513-516
9. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36: 949-951
10. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305: 525-528
11. Scherer SW, Lee C, Birney E, Altshuler DM, Eichler EE, Carter NP, Hurles ME, Feuk L (2007) Challenges and standards in integrating surveys of structural variation. Nat Genet 39: S7-15
12. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, Eichler EE (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am J Hum Genet 79: 275-290
13. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME (2009) Origins and functional impact of copy number variation in the human genome. Nature 464: 704-712
14. Hastings PJ, Lupski JR, Rosenberg SM, Ira G (2009) Mechanisms of change in gene copy number. Nat Rev Genet 10: 551-564
15. Shaw CJ, Lupski JR (2004) Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet 13 Spec No 1: R57-64
16. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE (2005) Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77: 78-88
17. Conrad DF, Bird C, Blackburne B, Lindsay S, Mamanova L, Lee C, Turner DJ, Hurles ME (2010) Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. Nat Genet 42: 385-391
18. Lee JA, Carvalho CM, Lupski JR (2007) A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 131: 1235-1247
19. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W et al (2006) Global variation in copy number in the human genome. Nature 444: 444-454
20. Eichler EE (2006) Widening the spectrum of human genetic variation. Nat Genet 38: 9-11
21. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM (2006) Common deletion polymorphisms in the human genome. Nat Genet 38: 86-92
22. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK (2006) A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38: 75-81
23. Chiang DY, Getz G, Jaffe DB, O'Kelly MJ, Zhao X, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES (2009) High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat Methods 6: 99-103
24. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G (2002) Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30: e57
25. Weaver S, Dube S, Mir A, Qin J, Sun G, Ramakrishnan R, Jones RC, Livak KJ (2010) Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. Methods 50: 271-276
26. Armour JA, Palla R, Zeeuwen PL, den Heijer M, Schalkwijk J, Hollox EJ (2007) Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Res 35: e19
27. Kozlowski P, Jasinska AJ, Kwiatkowski DJ (2008) New applications and developments in the use of multiplex ligation-dependent probe amplification. Electrophoresis 29: 4627-4636
28. Laczmanska I, Laczmanski L (2009) Metoda MLPA oraz jej zastosowanie w diagnostyce chorób uwarunkowanych genetycznie. Postepy Biol Kom 36: 555-563
29. McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A et al (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet 40: 1166-1174
30. Mileyko Y, Joh RI, Weitz JS (2008) Small-scale copy number variation and large-scale changes in gene expression. Proc Natl Acad Sci USA 105: 16659-16664
31. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC (2007) Diet and the evolution of human amylase gene copy number variation. Nat Genet 39: 1256-1260
32. Auer H (2008) Expression divergence and copy number variation in the human genome. Cytogenet Genome Res 123: 278-282
33. Yang TL, Chen XD, Guo Y, Lei SF, Wang JT, Zhou Q, Pan F, Chen Y, Zhang ZX, Dong SS et al (2008) Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. Am J Hum Genet 83: 663-674

34. Henrichsen CN, Chaignat E, Reymond A (2009) Copy number variants, diseases and gene expression. Hum Mol Genet 18: R1-8

35. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, Redon R, Bird CP, de Grassi A, Lee C, Tyler-Smith C, Carter N, Scherer SW, Tavare S, Deloukas P, Hurles ME, Dermitzakis ET (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315: 848-853

36. Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M, Reis A, Armour JA, Schalkwijk J (2008) Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet 40: 23-25

37. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ et al (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307: 1434-1440

38. Willcocks LC, Lyons PA, Clatworthy MR, Robinson JI, Yang W, Newland SA, Plagnol V, McGovern NN, Condliffe AM, Chilvers ER, Adu D, Jolly EC, Watts R, Lau YL, Morgan AW, Nash G, Smith KG (2008) Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. J Exp Med 205: 1573-1582

39. Yamaguchi T, Motulsky AG, Deeb SS (1997) Visual pigment gene structure and expression in human retinae. Hum Mol Genet 6: 981-990

40. Deeb SS (2006) Genetics of variation in human color vision and the retinal cone mosaic. Curr Opin Genet Dev 16: 301-307

41. Lian J, Zhang X, Tian H, Liang N, Wang Y, Liang C, Li X, Sun F (2009) Altered microRNA expression in patients with non-obstructive azoospermia. Reprod Biol Endocrinol 7: 13

42. Anglicheau D, Sharma VK, Ding R, Hummel A, Snopkowski C, Dadhania D, Seshan SV, Suthanthiran M (2009) MicroRNA expression profiles predictive of human renal allograft status. Proc Natl Acad Sci USA 106: 5330-5335

43. Keller A, Leidinger P, Lange J, Borries A, Schroers H, Scheffler M, Lenhof HP, Ruprecht K, Meese E (2009) Multiple sclerosis: microRNA expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls. PLoS One 4: e7440

44. Guo C, Sah JF, Beard L, Willson JK, Markowitz SD, Guda K (2008) The noncoding RNA, miR-126, suppresses the growth of neoplastic cells by targeting phosphatidylinositol 3-kinase signaling and is frequently lost in colon cancers. Genes Chromosomes Cancer 47: 939-946

45. Rossi S, Sevignani C, Nnadi SC, Siracusa LD, Calin GA (2008) Cancer-associated genomic regions (CAGRs) and noncoding RNAs: bioinformatics and therapeutic implications. Mamm Genome 19: 526-540

46. Ju X, Li D, Shi Q, Hou H, Sun N, Shen B (2009) Differential microRNA expression in childhood B-cell precursor acute lymphoblastic leukemia. Pediatr Hematol Oncol 26: 1-10

47. Wong TS, Liu XB, Wong BY, Ng RW, Yuen AP, Wei WI (2008) Mature miR-184 as Potential Oncogenic microRNA of Squamous Cell Carcinoma of Tongue. Clin Cancer Res 14: 2588-2592

48. Xue Y, Sun D, Daly A, Yang F, Zhou X, Zhao M, Huang N, Zerjal T, Lee C, Carter NP, Hurles ME, Tyler-Smith C (2008) Adaptive evolution of UGT2B17 copy-number variation. Am J Hum Genet 83: 337-346

49. Fellay J, Ge D, Shianna KV, Colombo S, Ledergerber B, Cirulli ET, Urban TJ, Zhang K, Gumbs CE, Smith JP et al (2009) Common genetic variation and the control of HIV-1 in humans. PLoS Genet 5: e1000791

50. Arenzana-Seisdedos F, Parmentier M (2006) Genetics of resistance to HIV infection: Role of co-receptors and co-receptor ligands. Semin Immunol 18: 387-403

51. Burns JC, Shimizu C, Gonzalez E, Kulkarni H, Patel S, Shike H, Sundel RS, Newburger JW, Ahuja SK (2005) Genetic variations in the receptor-ligand pair CCR5 and CCL3L1 are important determinants of susceptibility to Kawasaki disease. J Infect Dis 192: 344-349

52. McKinney C, Merriman ME, Chapman PT, Gow PJ, Harrison AA, Highton J, Jones PB, McLean L, O'Donnell JL, Pokorny V, Spellerberg M, Stamp LK, Willis J, Steer S, Merriman TR (2008) Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. Ann Rheum Dis 67: 409-413

53. Mamtani M, Rovin B, Brey R, Camargo JF, Kulkarni H, Herrera M, Correa P, Holliday S, Anaya JM, Ahuja SK (2008) CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythaematosus. Ann Rheum Dis 67: 1076-1083

54. Niemira M, Wisniewska A, Mazerska Z (2009) Rola polimorfizmu i zróżnicowanej ekspresji genów cytochromów P450 w metabolizmie ksenobiotyków. Postępy Biochem 55: 279-289

55. Maréchal JD, Kemp CA, Roberts GCK, Paine MJI, Wolf CR, Sutcliffe MJ (2008) Insights into drug metabolism by cytochromes P450 from modelling studies of CYP2D6-drug interactions. Br J Pharmacol 153: S82-S89

56. Ingelman-Sundberg M (2005) Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity. Pharmacogenomics J 5: 6-13

57. Knight JC (2009) Human genetic diversity. Functional consequences for health and disease. Oxford University Press Oxford, UK: 105-124

58. Steen VM, Molven A, Aarskog NK, Gulbrandsen AK (1995) Homologous unequal cross-over involving a 2.8 kb direct repeat as a mechanism for the generation of allelic variants of human cytochrome P450 CYP2D6 gene. Hum Mol Genet 4: 2251-2257

59. Johansson I, Lundqvist E, Bertilsson L, Dahl ML, Sjoqvist F, Ingelman-Sundberg M (1993) Inherited amplification of an active gene in the cytochrome P450 CYP2D locus as a cause of ultrarapid metabolism of debrisoquine. Proc Natl Acad Sci USA 90: 11825-11829

60. Marcinkowska *et al.* (2011) Copy number variation of microRNA genes in the human genome. BMC Genomics 12: 183

The influence of copy number polymorphism on the human phenotype

Malgorzata Marcinkowska, Piotr Kozlowski✉

Institute of Bioorganic Chemistry, Polish Academy of Sciences, 12/14 Z. Noskowskiego St., 61-704 Poznan, Poland

✉e-mail: kozlowp@yahoo.com

Key words: copy number variation (CNV), NAHR, *AMY1*, osteoporosis, psoriasis, HIV/AIDS

ABSTRACT

The variability of human populations in a large part is determined by two complementary factors: environment and genetic information. Genetic variation is caused by different genetic variants (polymorphisms and mutations) present in the human genome. Until recently it was thought that most of these variants are small changes of one or several nucleotides (SNPs) which in their millions are present in the human genome. However, it was recently shown that there are also polymorphisms that extend over hundreds of thousands of DNA base pairs in the human genome. Such alternations called copy number variation (CNV) often include genes and other functional genetic elements. In this article we present the general characteristics of copy number polymorphism and we discuss some examples of CNVs that influence human phenotypes.

4

Marcinkowska-Swojak M, Uszczynska B, Figlerowicz F, Kozlowski P

"An MLPA-based strategy for discrete CNV genotyping: CNV-miRNAs as an example"

Human Mutation 2013, 34:763-773

METHODS

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

An MLPA-Based Strategy for Discrete CNV Genotyping: CNV-miRNAs as an Example

Malgorzata Marcinkowska-Swojak, Barbara Uszczynska, Marek Figlerowicz, and Piotr Kozlowski*

European Centre for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland

Communicated by John McVey

Received 30 August 2012; accepted revised manuscript 24 January 2013.
Published online 5 February 2013 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22288

ABSTRACT: Copy number variation (CNV) has become well recognized in recent years. It has been estimated that common CNVs account for approximately 10% of the human genome and that they overlap hundreds of genes and other functional genetic elements. Although substantial progress in genome-wide CNV analysis has been made recently, there is still a need for a method that allows precise genotyping of selected CNVs. Here, we describe a novel strategy for CNV genotyping, taking advantage of the general principles of the multiplex ligation-dependent probe amplification (MLPA) method and short oligonucleotide probes, allowing easy custom design and generation of assays for almost any genomic region of interest. As a proof-of-concept, we developed two assays covering 17 candidate CNV regions that overlap human miRNA genes. Extensive quality control analysis demonstrated high reproducibility and reliability of the genotypes determined using our method. Detailed analysis of identified CNVs revealed that they are highly differentiated among the HapMap populations. The main advantages of the developed strategy include the simplicity of the assay design, its flexibility in terms of the selection of genomic regions, and its low cost (<$1–$10/genotype, depending on scale of experiment). These advantages make the presented strategy attractive for large-scale genetic analyses.
Hum Mutat 34:763–773, 2013. © 2013 Wiley Periodicals, Inc.

KEY WORDS: CNV; microRNA; MLPA; PRT; AluY insertion

Introduction

Copy number variation (CNV) in the human genome has become well recognized in recent years. CNVs are genomic regions (roughly 1 kb–1 Mb in length) that show a variable number of copies owing to deletions, duplications, or both. Common CNVs, often referred to as copy number polymorphisms (CNPs), account for approximately 10% of the human genome, overlapping hundreds of genes, regulatory sequences, and other functional genetic elements. The functional effects of CNVs are a subject of continuing investigation, and although the significance of the great majority of CNVs is uncertain, increasing numbers of CNVs are being associated with various human phenotypes, including diseases [Cantsilieris and White, 2013].

A number of methods have been developed to assess CNVs at the genome-wide level (reviewed in Carter (2007)), and major improvements to these methods (regarding the precision of CNV genotyping and breakpoint mapping) have recently been achieved [Chiang et al., 2009; Conrad et al., 2010a,b; McCarroll et al., 2008]. Despite these improvements, genome-wide approaches are still not feasible or practical for analyses of the large numbers of samples often required in genetic studies. Many of these methods also do not offer an adequate precision of CNV genotyping, and CNVs detected using genome-wide approaches usually require verification through alternative methods. Technical issues associated with CN genotyping have recently been reviewed [Cantsilieris and White, 2013]. Therefore, to follow up the analyses of individual CNVs of interest (e.g., located in regions implicated by linkage or association studies), a rapid, inexpensive, accurate, universal, and easy to set up locus-specific method is required [Cantsilieris and White, 2013; McCarroll and Altshuler, 2007].

The method that is currently most commonly used for CNV confirmation and copy number estimation is quantitative PCR (qPCR). However, in most cases, qPCR does not allow the identification of exact, discrete copy number genotypes (CN genotypes) [Fernandez-Jimenez et al., 2011; Fode et al., 2011]. Instead, the relative qPCR signal is usually used as a proxy of CN genotypes. The use of continuous PCR or hybridization signal instead of exact discrete CN genotypes hampers CNV analysis substantially (e.g., in allele inference and analysis of Mendelian inheritance, calculation of linkage disequilibrium, and investigation of the effect of individual CN genotypes) and decreases the power of CNV association studies [Ionita-Laza et al., 2009; McCarroll and Altshuler, 2007]. One method that overcomes most of the above limitations and allows identification of discrete CN genotypes is the paralog ratio test (PRT) [Armour et al., 2007; Hollox et al., 2008]. Other methods whose potential for CNV genotyping have been investigated previously include multiplex amplifiable probe hybridization (MAPH) as well as multiplex ligation-dependent probe amplification (MLPA) [den Dunnen and White, 2006]. A comprehensive review of the currently available locus-specific CNV genotyping methods was recently published [Ceulemans et al., 2012].

MLPA is a method that was first described in 2002 by Schouten and colleagues as a multiplex assay utilizing up to 45 probes specific for different genomic locations (often exons in genes of

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Piotr Kozlowski, Polish Academy of Sciences, Institute of Bioorganic Chemistry, Noskowskiego 12/14, Poznan, 61-704, Poland. E-mail: kozlowp@yahoo.com

Contract grant sponsors: Ministry of Science and Higher Education (N N302 278937); National Science Centre (2011/01/B/NZ5/02773, 2011/01/B/NZ2/04816, and 2012/05/N/ST6/03466).

© 2013 WILEY PERIODICALS, INC.

interest) [Schouten et al., 2002]. Each MLPA probe is composed of two half-probes whose target-specific sequences hybridize to directly adjacent target sequences, allowing the subsequent ligation and dosage-dependent amplification of probes specifically recognizing their targets. MLPA was originally designed as a tool to detect large mutations, and it has been successfully used to test and identify hundreds of large mutations in numerous disease-related genes, including *DMD*, *BRCA1*, *NF1*, *STK11*, and *TSC2* [Aretz et al., 2005; Bunyan et al., 2007; De Luca et al., 2007; Kozlowski et al., 2007b; Schouten et al., 2002]. Other applications of MLPA were subsequently proposed (reviewed in Kozlowski et al. (2008b)). The design and generation of the long probes utilized in the standard MLPA system is a complicated and time-consuming (and therefore expensive) process. In practice, this disadvantage seriously limits the applicability of MLPA to genes or sets of genes for which ready-to-use commercial kits are available. To overcome this limitation, several strategies for MLPA design that exclusively utilize short oligonucleotide probes that can easily be generated via chemical synthesis have been proposed [Kozlowski et al., 2007b; Sanchez-Mejias et al., 2010; White et al., 2007].

Here, we describe and evaluate a new MLPA-based method for discrete copy number genotyping of selected CNVs. This method takes advantage of a previously developed strategy for MLPA probe design [Kozlowski et al., 2007b; Marcinkowska et al., 2010]. The robustness of this strategy has been confirmed by its numerous applications, which include (1) large mutation detection in *TSC1* and *TSC2* [Kozlowski et al., 2007b]; (2) analysis of paralog sequences in *PKD1* [Kozlowski et al., 2008a]; (3) analysis of cancer genome [Liang et al., 2010; Marcinkowska et al., 2010]; (4) mouse transgene genotyping (e.g., *Cre* and *eGFP*) [Kozlowski et al., 2007a]; (5) analysis of conditional allele conversion [Liang et al., 2010]; and (6) analysis of strand-specific gene expression [Mykowska et al., 2011]. As a model for testing our method, we employed CNV regions overlapping human miRNA genes (CNV-miRNAs). The selected regions included both unique ("easier") and segmentally duplicated ("more difficult") sequences [Cantsilieris and White, 2013]. The developed assays allowed us to confirm CN polymorphism in almost 50% of the selected candidate CNV regions and to determine the exact copy number in all but one of the investigated regions. The proposed strategy allows assays targeting almost any region of the genome to be designed and discrete genotyping of both biallelic and multiallelic CNVs to be performed. The relatively low per-genotype cost makes this technique an attractive method for the genotyping of individual CNVs in large groups of samples, allowing it to be applied in genotype–phenotype association studies.

Materials and Methods

DNA Samples

DNA samples were purchased from the Coriell Institute (www.coriell.org). A total of 96 samples was obtained from three HapMap populations: 48 samples from a European population (CEU) from the Centre d'Etude du Polymorphisme Humain Collection, representing 16 family trios (two parents and one child); 24 unrelated samples from a Han Chinese population in Beijing, China (CHB); and 24 unrelated samples from a Yoruba population from Ibadan, Nigeria (YRI). According to information provided by the Coriell Institute, all samples were diluted in deionized water to a working concentration of 50 ng/μl.

Selection of CNVs

Seventeen CNV regions containing miRNA genes (CNV-miRNAs) were selected from top-validated CNV regions identified previously via bioinformatic comparison of CNV and miRNA localization [Marcinkowska et al., 2011]. Five of the selected miRNAs were localized in CNV regions validated through high accuracy genotyping reported in Conrad et al. (2010b) and McCarroll et al. (2008). The remaining 12 miRNAs were localized in CNV regions deposited in DGV (update Aug 05, 2009—http://projects.tcag.ca/variation) validated with multiple (at least 6) overlapping CNVs. In this case, as a CNV region, we considered the smallest region covered by all overlapping CNVs (Supp. Fig. S1).

MLPA Analysis

The MLPA probes and general probe set layout were designed according to a previously proposed strategy [Kozlowski et al., 2007b; Marcinkowska et al., 2010]. This strategy utilizes only short oligonucleotide probes that can easily be generated via standard chemical synthesis. Briefly, each probe was composed of two half-probes of equal size, and the total probe length ranged from 93 to 180 nt. The target sequences for the probes were selected to avoid SNPs, repeat elements and sequences of extremely high or low GC content (Supp. Table S1).

The MLPA reactions were run according to the manufacturer's general recommendations (MRC-Holland, Amsterdam, the Netherlands), as described earlier in Kozlowski et al. (2007b) and Schouten et al. (2002). The products of MLPA reaction were subsequently diluted 10× in HiDi formamide containing GS Liz600, which was used as a DNA sizing standard, and separated via capillary electrophoresis (POP7 polymer) in an ABI Prism 3130XL apparatus (Applied Biosystems, Carlsbad, CA, USA).

The obtained electropherograms were analyzed using GeneMarker software (v1.91). The signal intensities (peak heights) were retrieved and transferred to prepared Excel sheets (available upon request). For each individual sample, the signal intensity of each probe was divided by the average signal intensity of the control probes to normalize the obtained values and to equalize run-to-run variation. Then, the normalized signals of region-specific probe pairs for all samples were presented in signal scatter plots. As the signal of MLPA probes is proportional to the copy number, the signals of corresponding probes of multiple samples form distinct clusters representing CN genotypes and the distances between subsequent clusters are almost equal. We assumed that subsequent clusters correspond to genotypes differing by one copy. Taking the above into account, we calculated the CN genotype of the first cluster (lowest signal cluster) dividing its average signal value (distance from 0) by its distance to the subsequent cluster and rounding the obtained value to the closest integer. The subsequent integers were assigned as genotypes of subsequent clusters. Upon visual examination of signal scatter plots the upper and lower boundaries of each cluster were manually defined (Fig. 1C). Then with the use of specially prepared Excel sheets (available upon request), each sample in which MLPA probe signals were located within defined cluster borders was assigned to the particular CN genotype. Samples located outside of all of the defined clusters were called as N (no call).

qPCR and Statistical Analysis

The qPCR analysis was performed with the use of MESA GREEN qPCR MasterMix Plus for SYBR® Assay (Eurogentec, Seraing, Belgium) according to manufacturer's general recommendations.

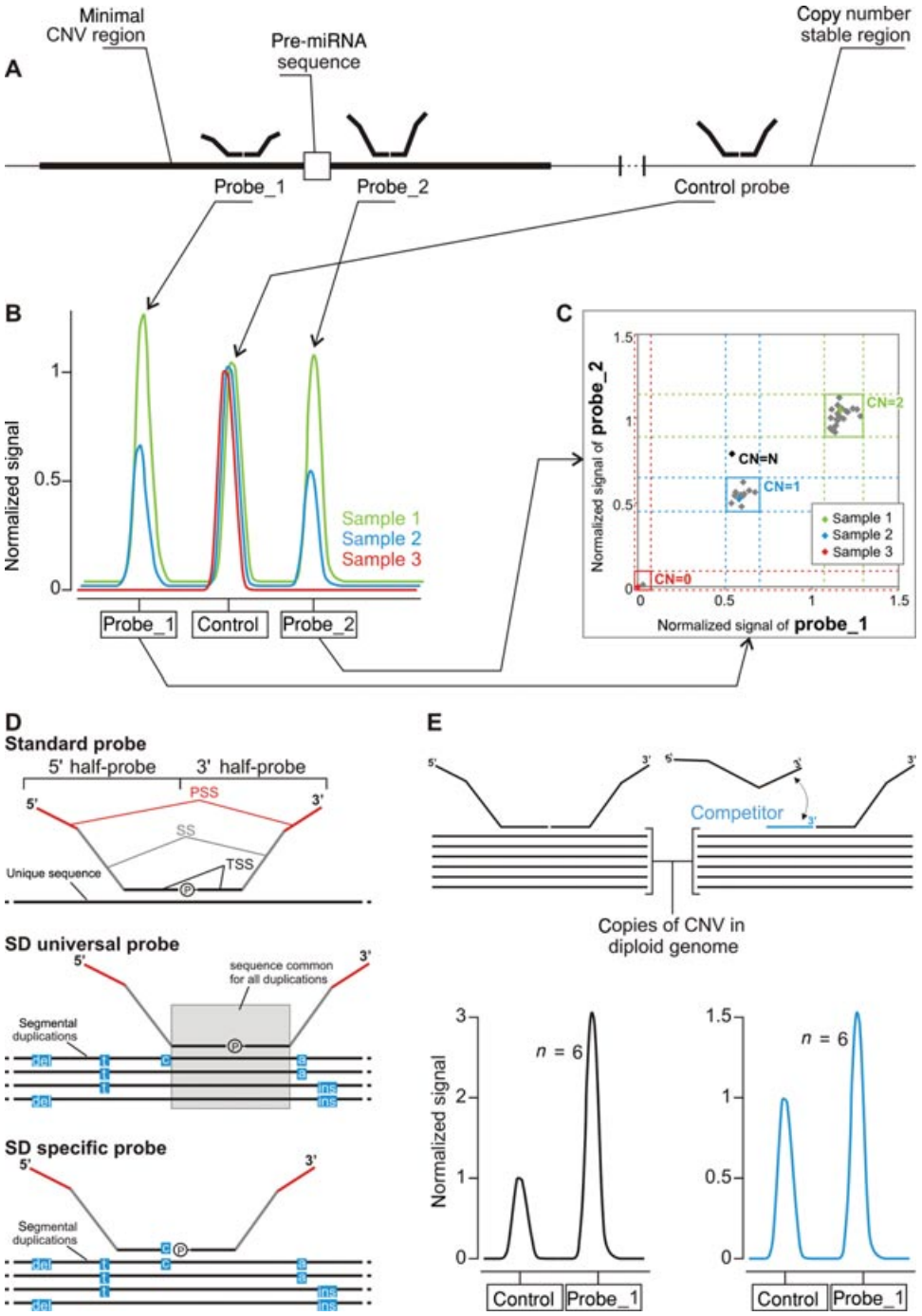


Figure 1. Strategy for MLPA-based assay design for discrete genotyping of selected CNV-miRNAs. **A**: Schematic representation map of a CNV-miRNA region with an indicated minimal CNV region (thick black line), a pre-miRNA sequence (white box) and the position of a pair of CNV-specific probes, as well as the position of one control probe. **B**: Overlapped electropherograms of three DNA samples normalized against the signal of the control probe. The electropherograms represent samples with two copies (sample 1), one copy (sample 2), and zero copies (sample 3) of the investigated region. **C**: Two-dimensional signal scatter plot depicting normalized signals of probe_1 (*x* axis) and probe_2 (*y* axis) for all of the samples analyzed in an hypothetical experiment. The normalized signals group into distinct clusters representing discrete CN genotypes. Samples 1, 2, and 3, whose electropherograms are shown in panel **B**, are indicated with colors. Doted lines indicate cluster boundaries defined based on visual examination of the signal scatter plot. **D**: The three types of MLPA probes used in this study, from top to bottom, are as follows. (1) Standard probe targeting a unique genomic sequence. Each MLPA probe is composed of two half-probes: a 5′half-probe and a 3′half-probe. Target-specific sequences TSS, stuffer sequences SS, and primer-specific sequences PSS are indicated on the graph. (2) SD universal probe—common to all SD copies present in the reference genome. Blue rectangles indicate nucleotides differentiating particular copies of SD from a consensus sequence. (3) SD-specific probe—specifically distinguishes one copy of SD from other copies present in the genome. **E**: Effect of a competitor on the absolute signal of a multicopy probe. Schematic representation of a multicopy probe used without (left side) and with a target-specific competitor (right side). Corresponding electropherograms are shown below. The sequence of a competitor oligonucleotide is identical to the target-specific sequence of the 5′half-probe but lacks the universal primer sequence for PCR. Thus, a ligated competitor will not be amplified in the PCR step, and the use of a competitor therefore decreases the absolute signal of the specific probe, leading to increased uniformity of the multiplexed probe signals.

PCR primers for the eight monomorphic regions and two control regions were designed to overlap target sequences of corresponding MLPA probes. The details of qPCR analysis and primer sequences are available upon request.

All statistical analyses were performed using Statistica (StatSoft, Tulsa, OK) or Prism v. 4.0 (GraphPad, San Diego, CA). All of the human genome positions indicated in this report refer to the February 2009 (GRCh37/hg19) human reference sequence.

Results

To develop and test the strategy for MLPA-based multiplex genotyping of CNVs, we selected 17 CNV-miRNAs. CNV-miRNAs are candidate copy number variable regions spanning sequences annotated as miRNA precursors (miRNA genes) [Marcinkowska et al., 2011]. CNV-miRNAs were recently identified in the human genome by computationally overlapping the coordinates of miRNA genes with either CNV regions deposited in the Database of Genomic Variants (DGV) or CNVs validated by high-quality genotyping [Conrad et al., 2010b; McCarroll et al., 2008]. For the purpose of the present study, we selected highly validated CNV-miRNAs that were either covered by at least six reports in DGV (n = 12) or identified by high-quality genotyping of HapMap samples (n = 5) [Conrad et al., 2010b; McCarroll et al., 2008] (Table 1).

General Strategy of Assay Design and Analysis

The general strategy for the design of the MLPA probes and assays is presented in Figure 1. For each candidate CNV region, we have designed and generated two MLPA probes located in close proximity to miRNA precursor sequences (Fig. 1A and Supp. Fig. S1), in most cases (12), on either side of an annotated miRNA precursor sequence. The probes were designed according to a strategy that was previously developed and described in detail [Kozlowski et al., 2007b; Marcinkowska et al., 2010]. All of the MLPA probes were divided into two groups (MLPA assays): CNVmiR1 and CNVmiR2 (Table 1 and Supp. Table S1). The CNVmiR1 assay involved 25 probes, including 20 probes specific for 10 different CNV regions. The CNVmiR2 assay involved 19 probes, including 14 probes specific for seven CNV regions. It was generally more difficult to design the MLPA probes for the CNV regions covered by the CNVmiR2 assay because of the complex structure of the segmental duplications (SDs) and excessive amount of repeat elements in these regions (Supp. Fig. S1). Each assay, with the exception of probes specific for selected CNV-miRNAs, also included five control probes specific for stable copy number regions that were used to normalize the run-to-run variation of the MLPA probe signals.

Most of the MLPA probes were designed to be specific for unique genomic sequences (Fig. 1D—Standard probe). However, in cases when CNVs could not be unambiguously mapped to a specific genomic position due to overlap with SDs, the MLPA probes were designed to recognize a target sequence common to all SD copies (Fig. 1D—SD universal probes). In fact, the number of SD copies present in a reference genome may represent just one of many alleles of the investigated CNVs. This allele is not necessarily the most frequent or ancestral. Therefore, in the present study, when calling the CN genotypes, we counted all of the copies of the investigated region, regardless of how many copies of the region are present in the reference genome. In one case (CNV-miRNA-663) in which CNV was mapped unambiguously to a specific SD copy (the other copy is located on a different chromosome and does not include the sequence of the miRNA-663 precursor), the MLPA probes were designed to be specific only for the SD copy identified as being copy number variable. In this case, the target sequences of the MLPA probes were selected to be specific for the copy of interest, and the ligation points of the MLPA probes were located directly adjacent to the nucleotides differentiating the two SD copies (Fig. 1D—SD-specific probe). The similar strategy was previously successfully used (by one of us P.K.) for detection of large mutations in the highly duplicated *PKD1* gene [Kozlowski et al., 2008a].

As some probes map to multiple positions in the genome (multicopy genotypes), their absolute signal is much stronger than the signals of other probes (detecting ~2 copies). This situation substantially increases the disparities between signal intensities (peak heights). Extremely high peaks may exceed the upper detection range and often generate artifacts in the separation of MLPA products (e.g., the occurrence of extra peaks or an aberrant peak shape). Therefore, to increase the uniformity of signals in the assays designed herein, we reduced the signals of high-signal probes through the use of probe-specific competitors designed according to a strategy whose usefulness was demonstrated previously [Kozlowski et al., 2007a]. Briefly, the competitors are short oligonucleotides that recognize the same target sequence as one of the MLPA half-probes. A competitor added to the MLPA probe mix competes with the corresponding probe and decreases its signal roughly proportionally to the ratio of the probe to its competitor (Fig. 1E and Supp. Fig. S2).

MLPA results are usually analyzed by comparison with a reference sample in which the copy number of all of the investigated regions is known (usually $2n$). This type of approach is not practical in the case of multiplex genotyping of common polymorphisms because each sample can exhibit a different combination of genotypes, and we do not have prior knowledge regarding the genotypes of the analyzed samples. Therefore, for analysis of common CNVs, we propose an alternative approach in which the normalized signals of two probes targeting a particular CNV region are presented in a 2D signal scatter plot (Fig. 1C). The signal of one probe is shown on the x axis and the other on the y axis. As the signal of an MLPA probe is proportional to the copy number, the signals of multiple samples form distinct clusters corresponding to particular CN genotypes. The CN genotype of each sample was called based on the location of its signals within the defined clusters (see *Materials and Methods*).

Results of CNV Genotyping using the Developed Assays

The prepared MLPA assays were used for analysis of two sets of DNA samples: (1) 48 European samples (CEU) representing 16 family trios; and (2) 48 unrelated non-European samples (CHB + YRI) consisting of 24 African samples (YRI) and 24 Asiatic samples (CHB). For the purposes of method validation, all experiments were performed twice.

Representative MLPA electropherograms of the CNVmiR1 and CNVmiR2 assays and signal scatter plots presenting the genotyping results of individual CNV-miRNAs are shown in Figure 2 and summarized in Table 1. The complete set of signal scatter plots and the complete list of identified genotypes are shown in Supp. Figure S3 and Supp. Table S2, respectively. As can be observed in the presented results, eight of the 17 (47%) tested CNV regions proved to be polymorphic in at least one of the analyzed populations. Three of these CNVs (miRNA-384, miRNA-383, and miRNA-1972) were classified as biallelic. The CN genotypes of these CNVs (0–2 copies or 4–6 copies) can be easily elucidated based on the presence of just two CN alleles with zero and one copy or two and three copies per allele, respectively. Five other polymorphic CNVs (miRNA-570,

Table 1. General Characteristics of the CNV-miRNAs Covered by CNVmiR1 and CNVmiR2 Assays

Assay ID	CNV-miRNA ID	CNV-miRNA Chromosomal Coordinates (hg18/hg19)	Overlap of CNV-miRNA with SD	Probe type[a]	Probe targets in reference genome[b]	Use of probe-specific Competitor	Type of CNV polymorphism	Observed genotypes
CNVmiR1	CNV-miRNA-126	chr9:138680837–138688363/chr9:139561016–139568542	No	Standard	Unique	No	Monomorphic	2
	CNV-miRNA-142	chr17:53751608–53767652/chr17:56396609–56412653	No	Standard	Unique	No	Monomorphic	2
	CNV-miRNA-149	chr2:241039698–241051687/chr2:241391025–241403014	No	Standard	Unique	No	Monomorphic	2
	CNV-miRNA-383	chr8:14741501–14763659/chr8:14697130–14719288	No	Standard	Unique	No	Biallelic	1, 2
	CNV-miRNA-384[d]	chrX:76053855–76057477/chrX:76137461–76141083	No	Standard	Unique	No	Biallelic	W: 1, 2; M: 0, 1
	CNV-miRNA-566	chr3:50173490–50214015/chr3:50198486–50239011	No	Standard	Unique	No	Monomorphic	2
	CNV-miRNA-570	chr3:196905807–196918722/chr3:195420627–195433542	No	Standard	Unique	Yes	Multiallelic	2–7
	CNV-miRNA-1233[d]	chr15:32450046–32662643/chr15:34662754–34875351	Yes	SD universal	Multiple (2)	Yes	Multiallelic	3–5
	CNV-miRNA-1268[d]	chr15:19975453–20046356/chr15:22474089–22544992	Yes	Standard	Unique	Yes	Multiallelic	2–8
	CNV-miRNA-1275[d]	chr6:34071086–34077139/chr6:33963108–33969161	No	Standard	Unique	No	Monomorphic	2
CNVmiR2	CNV-miRNA-202	chr10:134903011–134918923/chr10:135053021–135068932	No	Standard	Unique	No	Monomorphic	2
	CNV-miRNA-514	chrX:146167253–146174575/chrX:146359561–146366883	Yes	SD universal	Multiple (3)	Yes	Multiallelic	W: 5–9; M: 2–5
	CNV-miRNA-661	chr8:145090343–145104971/chr8:145018355–145032983	No	Standard	Unique	No	Monomorphic	2
	CNV-miRNA-662	chr16:750040–764098/chr16:810039–824097	No	Standard	Unique	No	Monomorphic	2
	CNV-miRNA-663	chr20:26136626–26139184/chr20:26188626–26191184	Yes	SD specific	Unique	No	Monomorphic[c]	2
	CNV-miRNA-650	chr22:21494381–21502189/chr22:23164381–23172189	Yes	Standard	Unique	No	Polymorphic	Genotypes not determined
	CNV-miRNA-1972[d]	chr16:14997420–15016088/chr16:15089919–15108587chr16:68621490–68653097/chr16:70063989–70095596	Yes	SD universal	Multiple (2)	Yes	Biallelic	4–6

[a]For explanation, see Figure 1;
[b]Details in Supp. Figure S1;
[c]AluY insertion polymorphism;
[d]CNV-miRNAs selected based on high-quality genotyping [Conrad et al., 2010b; McCarroll et al., 2008];
W, women; M, men.

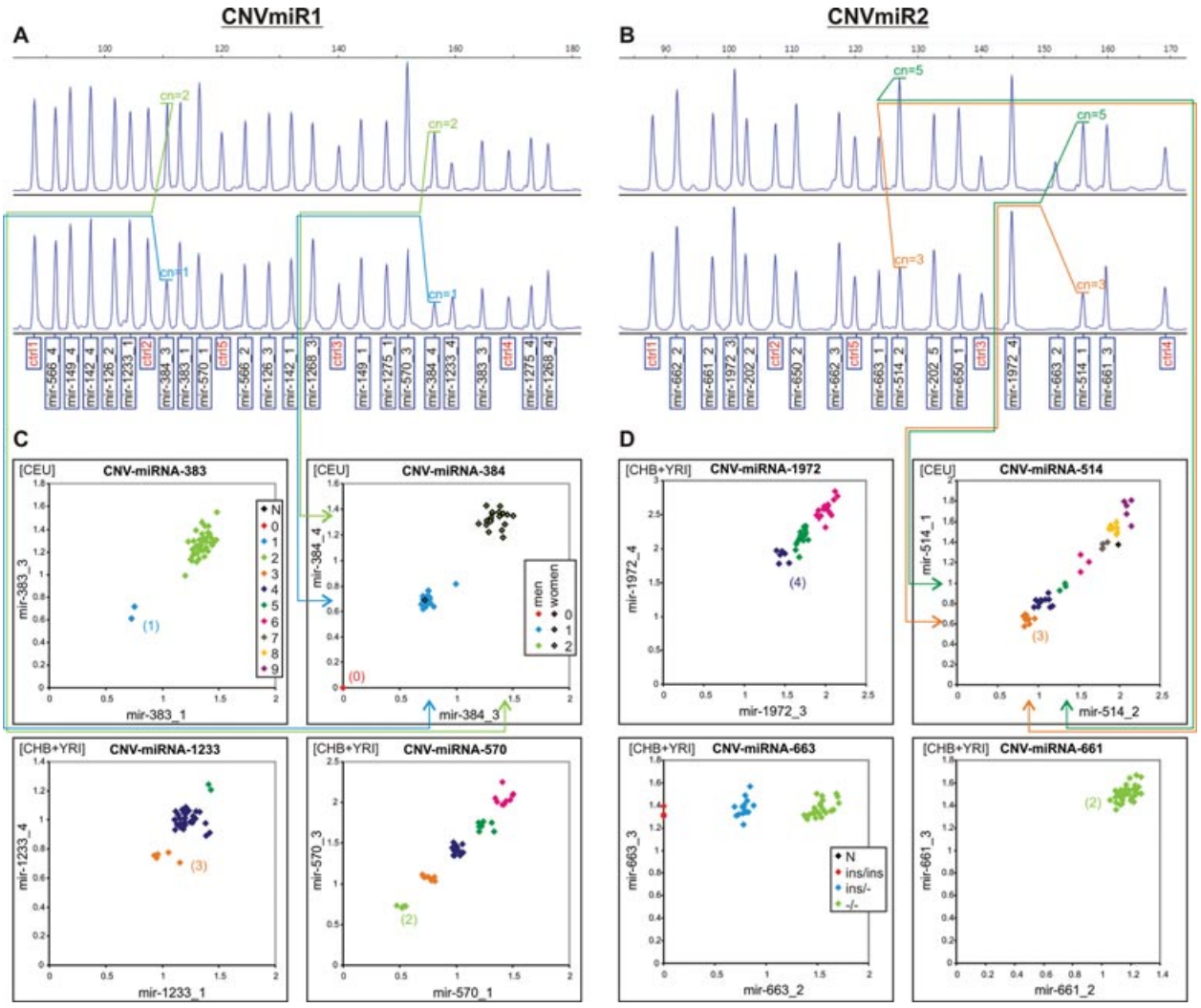


Figure 2. The results of MLPA-based CNV genotyping. **A** and **B**: Representative electropherograms of CNVmiR1 (A) and CNVmiR2 (B) assays. Probe IDs are indicated below the electropherograms. Note that the signals of paired region-specific probes are synchronized. Examples are indicated on the electropherograms. **C** and **D**: Selected signal scatter plots of CNV-miRNA regions covered by the CNVmiR1 and CNVmiR2 assays, respectively. Each sample is shown as a square colored according to the predicted copy number genotype. The *X*, *Y* coordinates represent the normalized signals of a pair of probes targeting an investigated region (probes IDs are indicated along the *x* and *y* axes). The analyzed sample sets CEU or CHB+YRI are indicated on the graphs. Number in parenthesis present on each graph indicates CN genotype of the lowest signal cluster. The selected signal scatter plots represent (1) biallelic CNVs with CN-alleles [0, 1] and [2, 3] (CNV-miRNA-383 and CNV-miRNA-1972, respectively); (2) a biallelic CNV with CN-alleles [0, 1] located on the X chromosome (CNV-miRNA-384). Note that in this case symbols representing women are bordered to distinguish them from symbols representing men (see the inset legend). (3) Multiallelic CNVs with different numbers of CN genotypes (CNV-miRNA-1233, CNV-miRNA-570, and CNV-miRNA-514; (4) CN-monomorphic region CNV-miRNA-661; and (5) a CN-monomorphic region with an AluY insertion affecting probe mir-663_2 (CNV-miRNA-663)—for details, see Supp. Figure S4.

miRNA-514, miRNA-650, miRNA-1233, and miRNA-1268) were classified as multiallelic. The genotypes of these CNVs range from two to nine copies. In one case of a multiallelic CNV (miRNA-650), the signals observed in the signal scatter plot do not allow individual genotype clusters to be distinguished. The CN genotypes of two CNVs located on the X chromosome (CNV-mir-384 and CNV-mir-514) clearly show a distinct distribution in men and women, which as expected, corresponds to the presence of one and two X chromosomes, respectively.

Nine of the 17 (53%) tested CNV regions were found to be copy number monomorphic in the analyzed samples (Fig. 2 and Supp. Fig. S3). The signal scatter plots of all but one of these regions show a clear single cluster. As it is shown in Supp. Figure S4 coefficient of variation (CV) of signal of probes representing monomorphic regions in most cases is below 0.1 corresponding to less than 10% of probe signal values. The average CV of "monomorphic" probes is similar to the average CV of control probes (0.07 vs. 0.06; *t*-test *P* val = 0.49; Supp. Fig. S4). We confirmed monomorphism of these regions with the use of qPCR. In all cases, the distribution of qPCR signals was unimodal, and signal variation of tested regions was similar to that observed in control regions. The interesting example of CN-monomorphic regions is CNV-miRNA-663, whose signal scatter plot shows an unusual pattern with three distinct clusters, resulting from the polymorphic signal of just one probe (mir-663_2). The second probe (mir-663_3) clearly shows monomorphic signal (Fig. 2D and Supp. Fig. S3). To elucidate the observed results further, we amplified and sequenced the target sequence of the mir-663_2 probe, and we found a common AluY insertion polymorphism

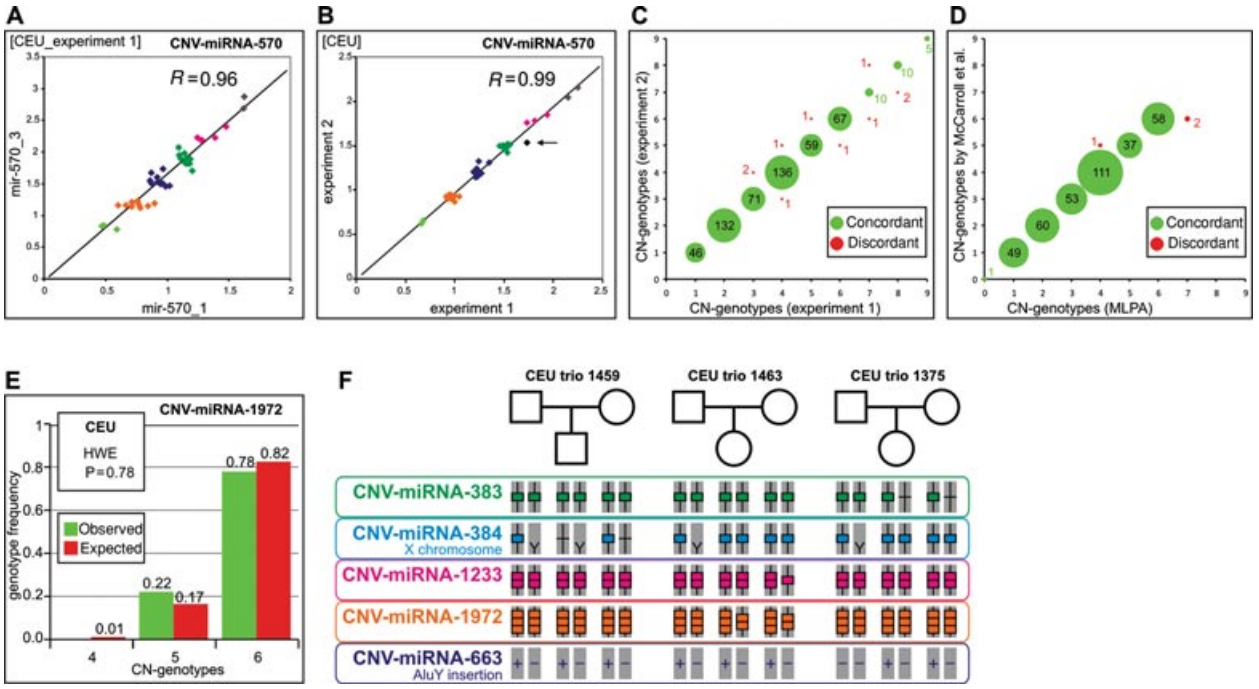


Figure 3. Quality control analysis of the obtained genotyping results. **A**: Representative signal scatter plot CNV-miRNA-570 showing the correlation between the signals of two probes targeting a single CNV region (probe-to-probe comparison). The selected example represents the result of experiment 1 performed on the CEU sample set. The trend line and correlation coefficient are indicated on the graph. **B**: Representative result of experiment-to-experiment comparisons showing the correlation between two subsequent genotyping experiments performed in the CEU sample set. The *x* axis and *y* axis show the average signals of CNV-miRNA-570-specific probes obtained in experiment 1 and experiment 2, respectively. The trend line and correlation coefficient are indicated on the graph. The arrowhead indicates a sample that was genotyped discordantly in two experiments. **C**: Comparison of the CN genotypes determined in two subsequent experiments. The overall reproducibility is >98%. Green and red circles indicate concordant and discordant results, respectively, with the number of results in or next to the circle. **D**: Comparison of 372 CN genotypes assigned in our MLPA-based study and a previous microarray-based study [McCarroll et al., 2008]. Color key as in C. **E**: Agreement with Hardy–Weinberg equilibrium of the CNV-miRNA-1972 genotypes in the CEU population. Green and red bars indicate the observed and expected CN-genotype frequencies, respectively. **F**: Mendelian inheritance of inferred CN-alleles in representative CEU parent-offspring trios. Copy number alleles are graphically depicted as a pile of rectangles in which each rectangle represents one copy of the investigated region. Alleles of the AluY insertion are depicted as either + (present) or – (absent).

whose genotypes correlate with the observed MLPA pattern (Supp. Fig. S5).

Validation of the Results of MLPA-Based CNV Genotyping

Although there is a growing interest in the identification and analysis of CNVs in human and other genomes, there is still no available method that can serve as the gold standard for CNV genotyping. Therefore, to evaluate the performance of the genotyping strategy proposed here, we carried out a stringent validation analysis using various technical, genetic, and computational criteria. As demonstrated in Figure 3, the signal of MLPA probes shows high probe-to-probe and experiment-to-experiment correlation (Fig. 3A and B, respectively). The determined discrete CN genotypes show high experiment-to-experiment reproducibility (Fig. 3C), good agreement with previous results [Conrad et al., 2010b; McCarroll et al., 2008] (Fig. 3D) and are consistent with Hardy–Weinberg equilibrium (Fig. 3E) and Mendelian inheritance (Fig. 3F). Finally, we showed that the genotype clusters distinguished based on visual examination were also distinguished by the expectation maximization algorithm (Supp. Fig. S6). Details of validation analyses are described in the Supporting Information and summarized in Supp. Table S3.

Comparison of CNV-miRNA Polymorphism in Three Human Populations

As noted above, we confirmed copy number polymorphism in eight out of 17 selected miRNA loci. At seven of these loci, we were able to distinguish integer CN genotypes. We characterized all of these polymorphisms in terms of the range and frequency of CN genotypes in the three ethnic populations. The minor allele frequency (MAF) and combined minor genotype frequency (cMGF—the combined frequency of all but the most frequent genotype) were determined for biallelic and multiallelic CNVs, respectively (Fig. 4 and Supp. Table S4). As is shown in Figure 4, in most cases (except for CNV-mir-383 and CNV-mir-384, which are noninformative because of a low MAF), the allele frequency and genotype distributions differ significantly between the analyzed populations. For example, a CN-allele containing two copies of mir-1972 that is a major allele (67%) in the CHB population is a minor allele in the YRI and CEU populations, in which it shows frequencies of 15% and 11%, respectively.

Discussion

As new CNV regions are still being discovered and characterized, there is a growing need for methods allowing for the precise

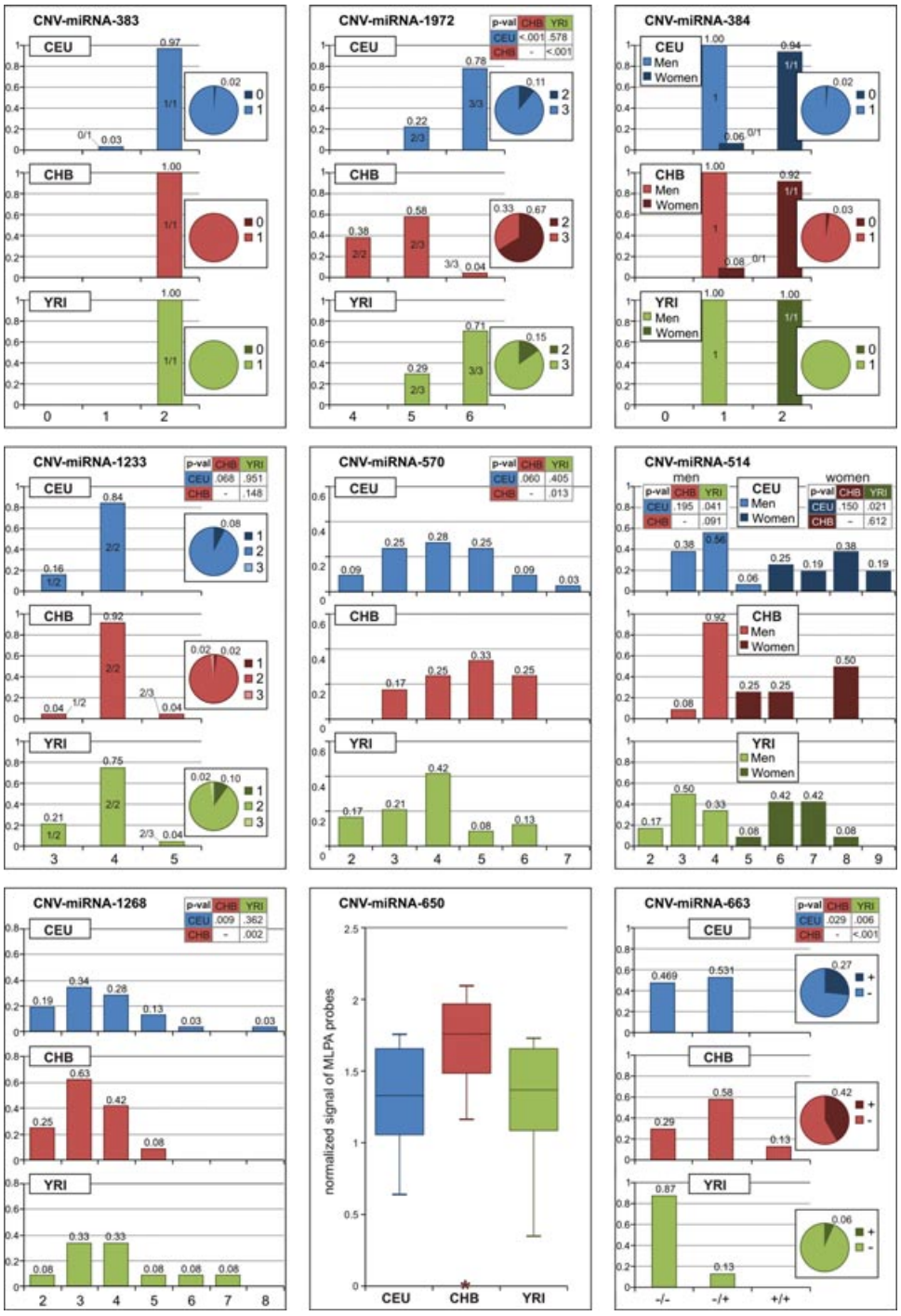


Figure 4. Comparison of the CN-genotype and CN-allele frequency distributions in the three tested populations. Blue, red, and green bar plots show the observed frequencies (*y* axis) of the copy number genotypes (*x* axis) in the CEU, CHB, and YRI samples, respectively. The CN-genotype and CN-allele frequencies in the CEU population were calculated based only on the genotypes observed in the parent samples. Alleles constituting particular genotypes of biallelic and three-allelic CNVs are indicated on the bars. For CNV-miRNA-384 and CNV-miRNA-514, the genotype frequencies were calculated separately for men and women because these CNV-miRNAs are localized on the X chromosome. For simple biallelic and triallelic CNVs, the CN-allele frequencies were calculated and are presented in pie charts. Minor allele frequencies are depicted next to the charts. The table insets indicate the *P* values for pairwise comparisons of the genotype distribution multiallelic CNVs or allele frequency biallelic CNVs performed using chi-squared or Fisher's exact tests, respectively. For CNV-miRNA-650, for which exact genotypes cannot be assigned, the distributions of the normalized signals of the MLPA probes in the analyzed populations are shown (box-and-whisker plots).

(discrete) genotyping of individual CNVs. Here, we proposed an MLPA-based strategy that allows multiplex CNV genotyping to be performed in almost all genomic regions of interest. Note, however, that the exact position of an MLPA probe has to be located in a sequence free of repetitive elements or SNPs present in the region of interest. Multiplexed CNV regions can be selected based either on their location (e.g., proximity to an association signal) or on prior knowledge about their relationship to an investigated problem.

Using this strategy, we developed two MLPA assays for genotyping seven and 10 highly validated CNV regions overlapping with human miRNA genes. The multiplexing factor used in the assays developed here, or an even higher factor, can easily be achieved through the generation of MLPA probes via standard chemical synthesis. A further increase in the multiplexing capacity can be achieved either through the use of two-color (or multiple-color) labeling on two distinct pairs of universal primers [White et al., 2004] or via the generation of longer MLPA probes composed of multiple short oligonucleotides [Serizawa et al., 2010].

Using the developed assays, we analyzed 96 HapMap samples from three human populations [CEU, CHB, and YRI] and confirmed polymorphism in eight out of 17 (47%) candidate CNV regions. Among the identified polymorphisms, there were both simple biallelic polymorphisms (3) and complex multiallelic polymorphisms associated with multiple genotypes containing up to nine copies of an investigated region (5). Although the candidate CNV regions were carefully selected based on previous data, nine of the selected regions were not polymorphic in the analyzed samples. We confirmed the polymorphic status of all of the candidate regions selected based on previous results of high-quality CNV genotyping obtained using CNV-dedicated microarrays [Conrad et al., 2010b; McCarroll et al., 2008]. However, many candidate regions, selected based on multiple reports in the DGV, turned out to be monomorphic. The high proportion of monomorphic regions may be explained by the fact that a significant portion of the CNVs deposited in DGV are very rare, private, oversized or represent false positive artifacts.

To evaluate the performance of the developed assays, we carried out a strict quality control analysis. All of the tests performed demonstrated that our results showed high reproducibility and were correlated with well-validated reference genotypes, in addition to the concordance of the determined genotypes with the predictions of Hardy–Weinberg equilibrium and Mendelian inheritance.

We believe that direct comparisons of competing methods are often strongly biased in favor of the new method. Among other reasons, this type of bias can result from the fact that researchers are usually experts in the proposed technology and are less familiar with alternative methods. Additionally, such comparisons can be affected by the selected target (here, genomic regions). Therefore, we chose not to perform a direct comparison of our MLPA-based strategy with the other methods of locus-specific CNV genotyping. Instead, we propose comparison of our results with CNV genotyping results obtained using alternative methods, such as qPCR, PRT, or MAPH, whose successful application in different genetic analyses, including association studies, has been reported previously in well-respected journals. This evaluation led us to the conclusion that our results are most comparable with those obtained via the PRT (see Fig. 3 in Armour et al. (2007)). The above conclusion is based on visual evaluation of the published PRT results and PRT characteristics discussed below. It was demonstrated that the PRT allows the CN genotypes of multiallelic CNVs to be distinguished discretely and reliably [Carpenter et al., 2012; Hollox et al., 2008]. It was also shown that the PRT may be multiplexed [Walker et al., 2009]. However, the multiplexing capacity of the PRT is lower than that of MLPA. Similar to our approach, the separation of low-CN genotypes is better than high-CN genotypes. Unfortunately, PRT assays cannot be designed for all CNVs. They are limited only to genomic regions containing specific paralog sequences used for designing locus-specific and reference probes. Currently, qPCR is the most commonly used method for CNV validation and genotyping. However, although several examples of excellent qPCR genotyping results can be found in the literature (see Fig. 1 in Hosono et al. (2009) or Supp. Fig. 2 in Pelak et al. (2011)), qPCR generally does not allow discrete CN genotypes to be distinguished (e.g., Supp. Fig. 13 in Gonzalez et al. (2005), Fig. 6 in Waszak et al. (2010), Fig. 5 in Fernandez-Jimenez et al. (2011) and Fig. 2 in Fode et al. (2011)), and the applicability of qPCR for the quantification of copy numbers has previously been questioned [Armour et al., 2007; Fernandez-Jimenez et al., 2011; Fode et al., 2011]. It also does not allow inference of alleles from observed genotypes. Another method that can be used for CNV genotyping is MAPH [den Dunnen and White, 2006; Sellner and Taylor, 2004]. MAPH may be considered as a sister method of MLPA [Schouten et al., 2002], and many aspects of these methods, including probe signal characteristics, are similar. However, MAPH assays are more difficult to develop, and MAPH is therefore much less popular than MLPA at present (PubMed). Nevertheless, as MAPH is a hybridization-based method, it can be more easily scaled up and presents the potential for the development of highly multiplexed assays [Kousoulidou et al., 2008; Tyson et al., 2009]. As MAPH takes advantage of similar principles as MLPA, some aspects of the strategy proposed here can be adopted to the MAPH platform.

Almost all of the analyzed polymorphic CNV-miRNAs showed substantial differences in terms of their genotype/allele frequencies and distributions in the three examined human populations. This finding may suggest that these CNVs, overlapping miRNA gene sequences, are functional polymorphisms that modify phenotypes that have been subjected to different selective pressures in human populations. Although the actual roles of particular polymorphisms must be proven in extensive functional and association studies, which are not within the scope of this project, a review of the literature indicates that some of the miRNAs identified here as being CN polymorphic are involved in various biological and physiological processes (Supp. Table S5). These miRNAs are mostly associated with the regulation of various genes and processes involved in cancer [Wang et al., 2011; Wulfken et al., 2011] but are also implicated in the regulation of drug activities [Tili et al., 2010] and apoptosis [Sudbery et al., 2010]. An interesting example of one of these miRNAs is miRNA-383, which is involved in the regulation of spermatogenesis, for which downregulation was observed in nonobstructive azoospermia and was associated with male infertility [Lian et al., 2009, 2010). The deletion polymorphism of miRNA-383 observed in the European population may be one of the factors involved in the downregulation of this miRNA.

Finally, as a byproduct of our study, we detected a common AluY insertion located 2,061 nt downstream of the miRNA-663 precursor (Supp. Fig. S5). The sequence of this 320 nt-long insertion indicates that it arose according to a typical retrotransposition mechanism [Comas et al., 2001; Cordaux et al., 2009]. It is located in a poly-(A)7 tract and is composed of the entire AluY sequence (303 nt) and 17 target-site duplication nucleotides. The presence of this insertion in all of the analyzed populations indicates that it arose before the divergence of these populations, but its different frequencies (Africans—6%, Europeans—27%, Asians—42%) may suggest that different selective pressures have acted on it in the analyzed populations. This common AluY insertion may be interesting both as a marker of evolutionary processes and as a potential phenotype modifying variant [Comas et al., 2001; Cordaux et al., 2009].

Concluding, we developed and validated a strategy for the discrete genotyping of CNVs in complex genomes. The main advantages of this strategy except the high reliability are the ease of assay design, its flexibility in terms of the selection of genomic regions, and its low cost (from even below $1 for large projects to about $10 for low-scale projects). These advantages make the presented strategy attractive for the large-scale genotyping of individual CNVs, as is required in association studies.

Acknowledgments

Thanks to David Kwiatkowski from Brigham and Women's Hospital, Harvard Medical School in whose laboratory one of us (P.K.) developed the original version of the MLPA probe design strategy. MMS and BU received the scholarship from "Scholarship support for PhD students specializing in majors strategic for Wielkopolska's development", European Social Fund.

Disclosure statement: The authors declare no conflict of interest.

References

1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.

Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res 21:974–984.

Aretz S, Stienen D, Uhlhaas S, Loff S, Back W, Pagenstecher C, McLeod DR, Graham GE, Mangold E, Santer R, Propping P, Friedl W. 2005. High proportion of large genomic STK11 deletions in Peutz–Jeghers syndrome. Hum Mutat 26:513–519.

Armour JA, Palla R, Zeeuwen PL, den Heijer M, Schalkwijk J, Hollox EJ. 2007. Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic Acids Res 35:e19.

Boguslawska J, Wojcicka A, Piekielko-Witkowska A, Master A, Nauman A. 2011. MiR-224 targets the 3′UTR of type 1 5′-iodothyronine deiodinase possibly contributing to tissue hypothyroidism in renal cancer. PLoS ONE 6:e24541.

Bunyan DJ, Skinner AC, Ashton EJ, Sillibourne J, Brown T, Collins AL, Cross NC, Harvey JF, Robinson DO. 2007. Simultaneous MLPA-based multiplex point mutation and deletion analysis of the dystrophin gene. Mol Biotechnol 35:135–140.

Cantsilieris S, White SJ. 2013. Correlating multiallelic copy number polymorphisms with disease susceptibility. Hum Mutat 34:1–13.

Carpenter D, Farnert A, Rooth I, Armour JA, Shaw MA. 2012. CCL3L1 copy number and susceptibility to malaria. Infect Genet Evol 12:1147–1154.

Carter NP. 2007. Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet 39:S16–S21.

Ceulemans S, van der Ven K, Del-Favero J. 2012. Targeted screening and validation of copy number variations. Methods Mol Biol 838:311–328.

Chan E, Patel R, Nallur S, Ratner E, Bacchiocchi A, Hoyt K, Szpakowski S, Godshalk S, Ariyan S, Sznol M, Halaban R, Krauthammer M, et al. 2011. MicroRNA signatures differentiate melanoma subtypes. Cell Cycle 10:1845–1852.

Chiang DY, Getz G, Jaffe DB, O'Kelly MJ, Zhao X, Carter SL, Russ C, Nusbaum C, Meyerson M, Lander ES. 2009. High-resolution mapping of copy-number alterations with massively parallel sequencing. Nat Methods 6:99–103.

Cicatiello L, Mutarelli M, Grober OM, Paris O, Ferraro L, Ravo M, Tarallo R, Luo S, Schroth GP, Seifert M, Zinser C, Chiusano ML, et al. 2010. Estrogen receptor alpha controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microRNAs. Am J Pathol 176:2113–2130.

Comas D, Plaza S, Calafell F, Sajantila A, Bertranpetit J. 2001. Recent insertion of an Alu element within a polymorphic human-specific Alu insertion. Mol Biol Evol 18:85–88.

Conrad DF, Bird C, Blackburne B, Lindsay S, Mamanova L, Lee C, Turner DJ, Hurles ME. 2010a. Mutation spectrum revealed by breakpoint sequencing of human germline CNVs. Nat Genet 42:385–391.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, et al. 2010b. Origins and functional impact of copy number variation in the human genome. Nature 464:704–712.

Cordaux R, Bouchon D, Greve P. 2009. The impact of endosymbionts on the evolution of host sex-determination mechanisms. Trends Genet 27:332–341.

Costa FF, Bischof JM, Vanin EF, Lulla RR, Wang M, Sredni ST, Rajaram V, Bonaldo Mde F, Wang D, Goldman S, Tomita T, Soares MB. 2011. Identification of microRNAs as potential prognostic markers in ependymoma. PLoS ONE 6:e25114.

De Luca A, Bottillo I, Dasdia MC, Morella A, Lanari V, Bernardini L, Divona L, Giustini S, Sinibaldi L, Novelli A, Torrente T, Schrinzi A, et al. 2007. Deletions of NF1 gene and exons detected by multiplex ligation-dependent probe amplification. J Med Genet 44:800–808.

den Dunnen JT, White SJ. 2006. MLPA and MAPH: sensitive detection of deletions and duplications. Curr Protoc Hum Genet Chapter 7:Unit 7.14.

Do CB, Batzoglou S. 2008. What is the expectation maximization algorithm? Nat Biotechnol 26:897–899.

Feng L, Xie Y, Zhang H, Wu Y. 2011. Down-regulation of NDRG2 gene expression in human colorectal cancer involves promoter methylation and microRNA-650. Biochem Biophys Res Commun 406:534–538.

Fernandez-Jimenez N, Castellanos-Rubio A, Plaza-Izurieta L, Gutierrez G, Irastorza I, Castano L, Vitoria JC, Bilbao JR. 2011. Accuracy in copy number calling by qPCR and PRT: a matter of DNA. PLoS ONE 6:e28910.

Fode P, Jespersgaard C, Hardwick RJ, Bogle H, Theisen M, Dodoo D, Lenicek M, Vitek L, Vieira A, Freitas J, Andersen PS, Hollox EJ. 2011. Determination of beta-defensin genomic copy number in different populations: a comparison of three methods. PLoS ONE 6:e16768.

Fraley C, Raftery, AE. 2006. MCLUST Version 3 for R: Normal Mixture Modeling and Model-Based Clustering. Technical Report no. 504. Seattle, WA: Department of Statistics, University of Washington.

Gillen AE, Gosalia N, Leir SH, Harris A. 2011. MicroRNA regulation of expression of the cystic fibrosis transmembrane conductance regulator gene. Biochem J 438:25–32.

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, et al. 2005. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307:1434–1440.

Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M, Reis A, Armour JA, et al. 2008. Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet 40:23–25.

Hosono N, Kato M, Kiyotani K, Mushiroda T, Takata S, Sato H, Amitani H, Tsuchiya Y, Yamazaki K, Tsunoda T, Zembutsu H, Nakamura Y, et al. 2009. CYP2D6 genotyping for functional-gene dosage analysis by allele copy number detection. Clin Chem 55:1546–1554.

Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. 2009. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics 93:22–26.

Jian P, Li ZW, Fang TY, Jian W, Zhuan Z, Mei LX, Yan WS, Jian N. 2011. Retinoic acid induces HL-60 cell differentiation via the upregulation of miR-663. J Hematol Oncol 4:20.

Jung M, Mollenkopf HJ, Grimm C, Wagner I, Albrecht M, Waller T, Pilarsky C, Johannsen M, Stephan C, Lehrach H, Nietfeld W, Rudel T, et al. 2009. MicroRNA profiling of clear cell renal cell cancer identifies a robust signature to define renal malignancy. J Cell Mol Med 13:3918–3928.

Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, et al. 2008. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet 40:1253–1260.

Kousoulidou L, Mannik K, Sismani C, Zilina O, Parkel S, Puusepp H, Tonisson N, Palta P, Remm M, Kurg A, Patsalis PC. 2008. Array-MAPH: a methodology for the detection of locus copy-number changes in complex genomes. Nat Protoc 3:849–865.

Kozlowski P, Bissler J, Pei Y, Kwiatkowski DJ. 2008a. Analysis of PKD1 for genomic deletion by multiplex ligation-dependent probe assay: absence of hot spots. Genomics 91:203–208.

Kozlowski P, Jasinska AJ, Kwiatkowski DJ. 2008b. New applications and developments in the use of multiplex ligation-dependent probe amplification. Electrophoresis 29:4627–4636.

Kozlowski P, Lin M, Meikle L, Kwiatkowski DJ. 2007a. Robust method for distinguishing heterozygous from homozygous transgenic alleles by multiplex ligation-dependent probe assay. Biotechniques 42:584, 586, 588.

Kozlowski P, Roberts P, Dabora S, Franz D, Bissler J, Northrup H, Au KS, Lazarus R, Domanska-Pakiela D, Kotulska K, Jozwiak S, Kwiatkowski DJ. 2007b. Identification of 54 large deletions/duplications in TSC1 and TSC2 using MLPA, and genotype-phenotype correlations. Hum Genet 121:389–400.

Lian J, Tian H, Liu L, Zhang XS, Li WQ, Deng YM, Yao GD, Yin MM, Sun F. 2010. Downregulation of microRNA-383 is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation by targeting IRF1. Cell Death Dis. 1:e94

Lian J, Zhang X, Tian H, Liang N, Wang Y, Liang C, Li X, Sun F. 2009. Altered microRNA expression in patients with non-obstructive azoospermia. Reprod Biol Endocrinol 7:13.

Liang MC, Ma J, Chen L, Kozlowski P, Qin W, Li D, Goto J, Shimamura T, Hayes DN, Meyerson M, Kwiatkowski DJ, Wong KK. 2010. TSC1 loss synergizes with KRAS activation in lung cancer development in the mouse and confers rapamycin sensitivity. Oncogene 29:1588–1597.

Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozlowski P. 2011. Copy number variation of microRNA genes in the human genome. BMC Genomics 12:183.

Marcinkowska M, Won K-K, Kwiatkowski DJ, Kozlowski P. 2010. Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example. TheScientificWorldJO 10:2003–2018.

McCarroll SA, Altshuler DM. 2007. Copy-number variation and association studies of human disease. Nat Genet 39:S37–S42.

McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, Shapero MH, de Bakker PI, Maller JB, Kirby A, Elliot AL, Parkin M, et al. 2008. Integrated detection

and population-genetic analysis of SNPs and copy number variation. Nat Genet 40:1166–1174.

Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, et al. 2011. Mapping copy number variation by population-scale genome sequencing. Nature 470:59–65.

Mykowska A, Sobczak K, Wojciechowska M, Kozlowski P, Krzyzosiak WJ. 2011. CAG repeats mimic CUG repeats in the misregulation of alternative splicing. Nucleic Acids Res 39:8938–8951.

Pan J, Hu H, Zhou Z, Sun L, Peng L, Yu L, Sun L, Liu J, Yang Z, Ran Y. 2010. Tumor-suppressive mir-663 gene induces mitotic catastrophe growth arrest in human gastric cancer cells. Oncol Rep 24:105–112.

Pelak K, Need AC, Fellay J, Shianna KV, Feng S, Urban TJ, Ge D, De Luca A, Martinez-Picado J, Wolinsky SM, Martinson JJ, Jamieson BD, et al. 2011. Copy number variation of KIR genes influences HIV-1 control. PLoS Biol 9:e1001208.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051-07–0. URL: http://www.R-project.org/.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, et al. 2006. Global variation in copy number in the human genome. Nature 444:444–454.

Sanchez-Mejias A, Nunez-Torres R, Fernandez RM, Antinolo G, Borrego S. 2010. Novel MLPA procedure using self-designed probes allows comprehensive analysis for CNVs of the genes involved in Hirschsprung disease. BMC Med Genet 11:71.

Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. 2002. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 30:e57.

Sellner LN, Taylor GR. 2004. MLPA and MAPH: new techniques for detection of gene deletions. Hum Mutat 23:413–419.

Serizawa RR, Ralfkiaer U, Dahl C, Lam GW, Hansen AB, Steven K, Horn T, Guldberg P. 2010. Custom-designed MLPA using multiple short synthetic probes: application to methylation analysis of five promoter CpG islands in tumor and urine specimens from patients with bladder cancer. J Mol Diagn 12:402–408.

Sudbery I, Enright AJ, Fraser AG, Dunham I. 2010. Systematic analysis of off-target effects in an RNAi screen reveals microRNAs affecting sensitivity to TRAIL-induced apoptosis. BMC Genomics 11:175.

Tili E, Michaille JJ, Adair B, Alder H, Limagne E, Taccioli C, Ferracin M, Delmas D, Latruffe N, Croce CM. 2010. Resveratrol decreases the levels of miR-155 by upregulating miR-663, a microRNA targeting JunB and JunD. Carcinogenesis 31:1561–1566.

Tyson J, Majerus TM, Walker S, Armour JA. 2009. Quadruplex MAPH: improvement of throughput in high-resolution copy number screening. BMC Genomics 10:453.

Walker S, Janyakhantikul S, Armour JA. 2009. Multiplex paralogue ratio tests for accurate measurement of multiallelic CNVs. Genomics 93:98–103.

Wang W, Sun J, Li F, Li R, Gu Y, Liu C, Yang P, Zhu M, Chen L, Tian W, Zhou H, Mao Y, et al. 2011. A frequent somatic mutation in CD274 3′-UTR leads to protein over-expression in gastric cancer by disrupting miR-570 binding. Hum Mutat 33:480–484.

Waszak SM, Hasin Y, Zichner T, Olender T, Keydar I, Khen M, Stutz AM, Schlattl A, Lancet D, Korbel JO. 2010. Systematic inference of copy-number genotypes from personal genome sequencing data reveals extensive olfactory receptor gene content diversity. PLoS Comput Biol 6:e1000988.

White SJ, Vink GR, Kriek M, Wuyts W, Schouten J, Bakker B, Breuning MH, den Dunnen JT. 2004. Two-color multiplex ligation-dependent probe amplification: detecting genomic rearrangements in hereditary multiple exostoses. Hum Mutat 24:86–92.

White SJ, Vissers LE, Geurts van Kessel A, de Menezes RX, Kalay E, Lehesjoki AE, Giordano PC, van de Vosse E, Breuning MH, Brunner HG, den Dunnen JT, Veltman JA. 2007. Variation of CNV distribution in five different ethnic populations. Cytogenet Genome Res 118:19–30.

Wulfken LM, Moritz R, Ohlmann C, Holdenrieder S, Jung V, Becker F, Herrmann E, Walgenbach-Brunagel G, von Ruecker A, Muller SC, Ellinger J. 2011. MicroRNAs in renal cell carcinoma: diagnostic implications of serum miR-1233 levels. PLoS ONE 6:e25787.

Zhang X, Zhu W, Zhang J, Huo S, Zhou L, Gu Z, Zhang M. 2010. MicroRNA-650 targets ING4 to promote gastric cancer tumorigenicity. Biochem Biophys Res Commun 395:275–280.

MATERIAŁY UZUPEŁNIAJĄCE DO PUBLIKACJI

Marcinkowska-Swojak i wsp., *Human Mutation* 2013

SUPPORTING INFORMATION

Validation of the results of MLPA-based CNV genotyping

First, to test the reproducibility of our results without any bias that may result from CN-genotype calling, we directly compared the signals of the CNV-specific probe pairs (probe-to-probe comparison) and the signals of these probes in repeated experiments (experiment-to-experiment comparison). As correlation analysis would not be informative in the case of monomorphic CNVs or CNVs with few genotype clusters we performed probe-to-probe and experiment-to-experiment comparisons only for the CNVs with three or more genotypes (each represented by at least 3 samples) observed in the analyzed sample sets. Representative correlation analyses are shown in Figures 3A and 3B and are summarized in Supp. Table S3. As shown in Supp. Table S3, in all cases, both the probe-to-probe and experiment-to-experiment correlations are high. Most of the correlation coefficient (R) values are well above 0.9. The somewhat lower R values for CNVs with fewer observed genotypes are not a result of lower correlations, but a result of the overall smaller copy number range.

Next, to determine the reproducibility and accuracy of the genotype calling, we compared the genotypes determined in two subsequent experiments. As shown in Supp. Table S3 for all but one (CNV-miRNA-514) of the genotyped CNVs, we obtained almost perfect reproducibility (only two discordant genotypes). In the case of CNV-miRNA-514, the reproducibility was substantially lower (91%) owing to the poorer separation of the genotype clusters in the signal scatter plot (see Figure 2). All of the observed genotype discordances are single copy number shifts, and most of them affect high copy number genotypes (Figure 3C).

As 4 of the CNVs determined in our study to be polymorphic were previously genotyped in the same group of samples, we decided to compare our results with two sets of CN-genotype data reported in two recent CNV discovery studies (Conrad, et al., 2010; McCarroll, et al., 2008). These sets of genotype data were obtained through high-quality genotyping using CNV-dedicated high-density hybrid arrays (combining traditional SNP probes and probes targeting CNVs). The results of these two studies have been used as reference data for validation of many other CNV discovery approaches (1000 Genomes Project Consortium, 2010; Abyzov, et al., 2011; Korn, et al., 2008; Mills, et al., 2011; Waszak, et al., 2010). As observed in Supp. Table S3 and Figure 3D, only 3 out of the 384 total analyzed sample genotypes (4 CNVs x 96 samples) were discordant. The overall concordance with previous results was >98%, and all of discordant results consisted of shifts of only a single copy. Moreover, it is worth noting that two of the discordant genotypes of CNV-miRNA-1268 (now called as 7, but previously as 6 copies) most likely result from a previously applied genotyping strategy that does not call genotypes

higher than 6 copies (McCarroll, et al., 2008) (genotypes with higher numbers were most likely rounded down to 6 copies) (see Figure 3D). Additionally, the genotypes of CNV-miRNA-1275 that were monomorphic in the samples analyzed in the present study show perfect concordance with previous results (CNV-miRNA-1275 were polymorphic only in three samples not analyzed in our study) (Conrad, et al., 2010). This finding suggests that our approach presents a low or no false discovery rate.

In the next step, the genotyping of bi-allelic CNVs was evaluated in way similar to what is commonly performed for the assessment of SNP genotypes. All of the bi-allelic CNVs were tested for agreement with Mendelian inheritance patterns, and the common CNV-miRNA-1972 was also tested for agreement with Hardy-Weinberg equilibrium (Supp. Table S3). Consistency with a Mendelian inheritance pattern was observed in all but one case, and CNV-miRNA-1972 shows good agreement with Hardy-Weinberg equilibrium in all of the tested populations. The only deviation from a Mendelian inheritance pattern (CEU trio 1349; CNV-miRNA-384) results from the occurrence in the offspring sample of a deletion allele that is not present in either parent. The same Mendelian inconsistency (same CNV, same trio) is present in genotypes determined previously via high-quality genotyping (McCarroll, et al., 2008). It was suggested that this inconsistency may be a result of cell line-specific artifacts (a deletion) (McCarroll, et al., 2008; Redon, et al., 2006). As such, it argues for, rather than against the quality of our assays. As in most cases of multi-allelic CNVs, the constituent alleles cannot be inferred from the genotypes to evaluate the genotyping results for multi-allelic CNVs, we compared the correlation of genotypes in parent-offspring and mother-father pairs. In all cases, the parent-offspring correlation was substantially higher than the mother-father correlation (Supp. Table S3). The observed correlation coefficient values approximate the values expected for perfect and unbiased heritability (0 and 0.5 for mother-father and parent-offspring pairs, respectively). Although CNV-miRNA-1233 is multi-allelic, its simple genotype distribution pattern allows for relatively confident prediction of its alleles (see Figure 4). We assumed that CNV-miRNA-1233 was three-allelic and that the observed CN-genotypes with 3, 4, and 5 copies resulted from the following CN-allele combinations: 1/2, 2/2 and 2/3, respectively. We disregarded genotype 1/3, which, if present, would be extremely rare (see Figure 4). Based on the above assumption, we calculated the frequency of all of the inferred alleles and showed that the genotyping results for CNV-miRNA-1233 are perfectly consistent with a Mendelian inheritance mode and in agreement with Hardy-Weinberg equilibrium in all three of the tested populations (Supp. Table S3 and Figure 3).

Finally, to verify that the genotype assignment based on our visual examination of signal scatter plots is not affected by subjective biases, we analyzed the signal data for the selected CNVs using the Expectation Maximization (EM) algorithm. As the EM algorithm is a model-

based clustering method that creates clusters based on the number and distribution of data points, we applied it only for the analysis of CNVs with relatively few genotypes (an appreciable number of samples per genotype). The application of EM showed that the genotype clusters distinguished based on visual examination were also distinguished by EM and that in most cases, the probability of sample assignment to a particular genotype cluster was high (Supp. Figure S6). The above conclusions are limited to CNVs with relatively simple genotype patterns and cannot be directly extrapolated to complex multi-allelic CNVs.

Cost and throughput of CNV genotyping

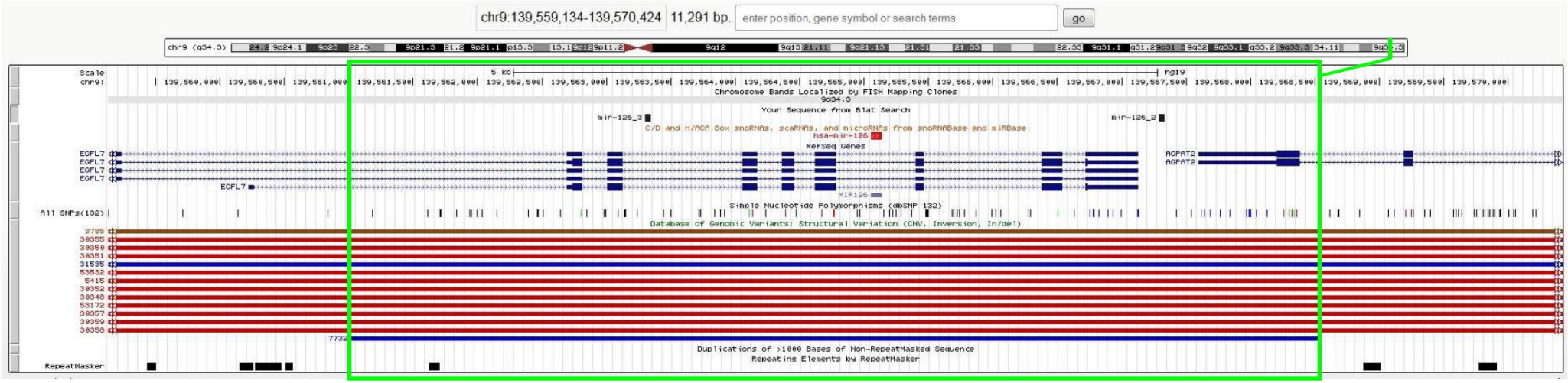
Taking into account the current cost of capillary running and MLPA reagents (both approximately $3 per sample), we estimate the minimal per-sample cost of the MLPA assays to be ~$6. The capacity of the assay to perform multiplex analysis of approximately 10 CNV regions means that the minimal cost per genotype may be reduced to approximately $0.5. Note, however, that the above figures depend on the multiplexing factor and the size of the experiment (number of samples to be analyzed) and due to the initial cost of probe synthesis the actual cost may increase substantially for low scale experiments. The initial cost of probe synthesis amounts to approximately $3000 per assay (once synthesized, the amount of synthesized probes is sufficient for as many as a million analyses). The contribution of probe synthesis to the overall cost of genotyping can be minimized or even disregarded in the case of large projects in which hundreds or thousands of samples are going to be analyzed. Additional cost and assay development problems can be caused by poorly performing probes that have to be replaced during preliminary experiments. Poor performance includes low or no signal (usually caused by poor quality probe synthesis) or unexpected signal variation that may be the result of unreported SNPs. Although in assays presented here no probe had to be replaced due to poor performance our previous experience indicates that about 5% of probes (1 probe per assay) have to be resynthesized or redesigned.

Designing a full set of MLPA probes takes 1-2 days (depending on the experimenter's skill and experience). Oligonucleotide dilution and probe mix preparation takes about 4 h.

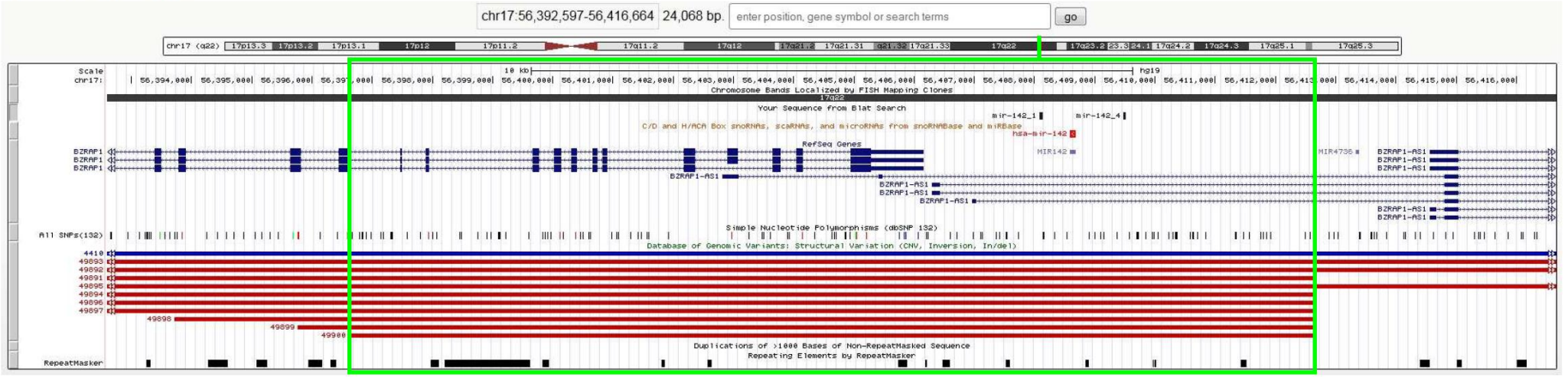
The use of a standard 96-well thermocycler and any multi-capillary DNA analyzer allows the analysis of 96 samples (960 genotypes) per day. This number can easily be scaled up through the use of additional thermal blocks or thermocyclers. It takes approximately 2 hours of experimenter time for the preparation of an MLPA reaction and approximately one hour to set up capillary electrophoresis.

CNVmiR1

CNV-miRNA-126

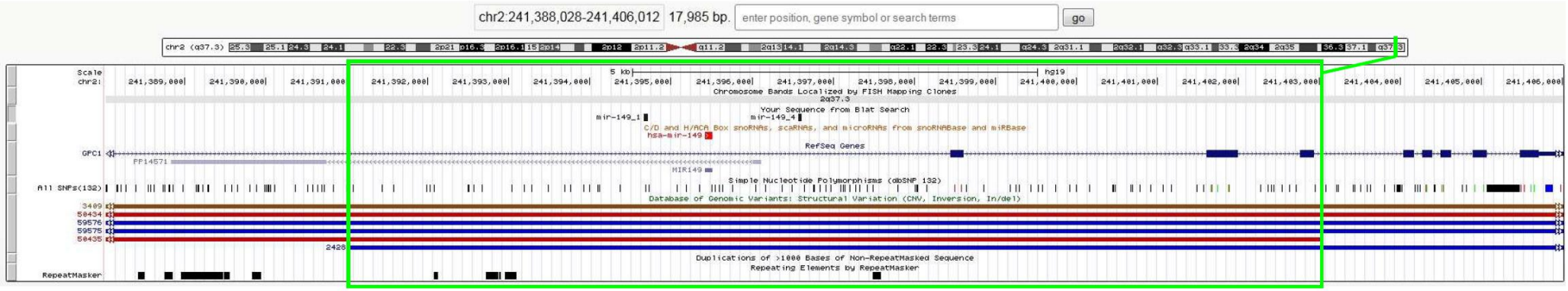


CNV-miRNA-142

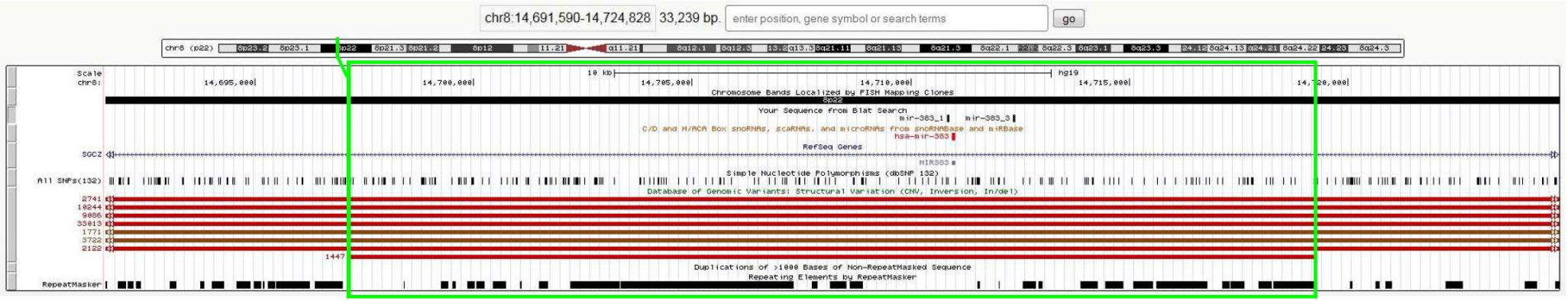


Supp. Figure S1. Screenshots from the UCSC Genome Browser (hg19) depicting the CNV-miRNA regions tested in our study. The visualized UCSC tracks include miRNA precursors, RefSeq genes, SNPs, CNV regions from DGV, segmental duplications and repeat elements. The positions of the MLPA probes are indicated in the track "Your Sequence from Blat Search". Note that in some cases, the MLPA probes map to more than one position in the reference genome. The green frame indicates the minimal CNV region defined previously (Marcinkowska, et al., 2010). Owing to the size of the investigated regions and the number of details included, most of the screenshots are only for computer review. Consequently, printouts of this image will be unreadable. (Continued on next page.)

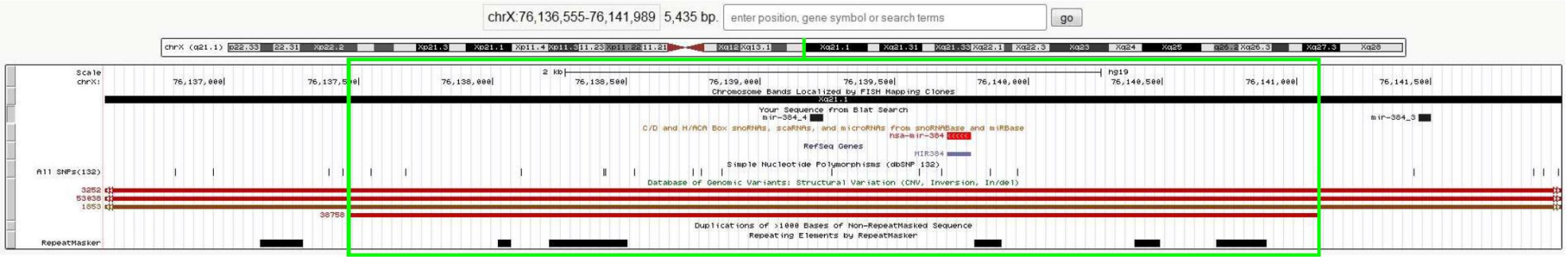
CNV-miRNA-149



CNV-miRNA-383

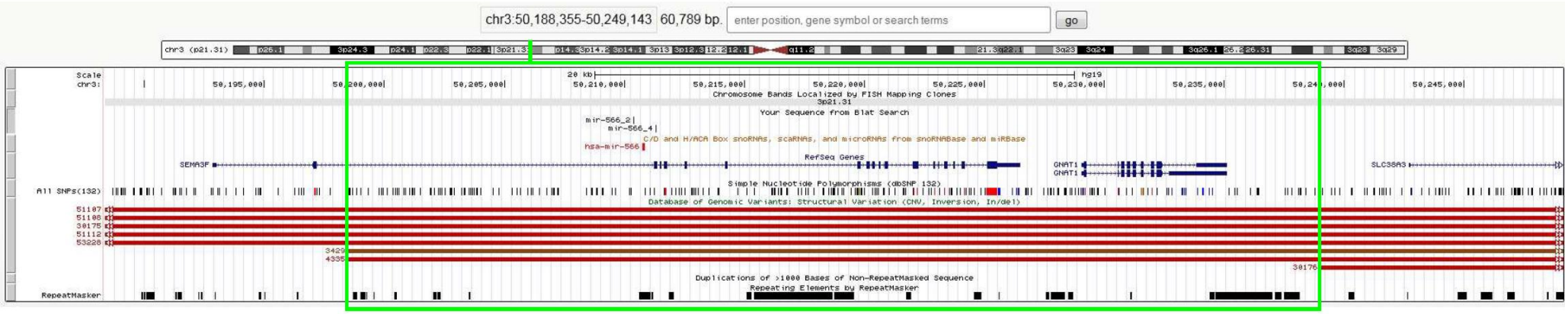


CNV-miRNA-384

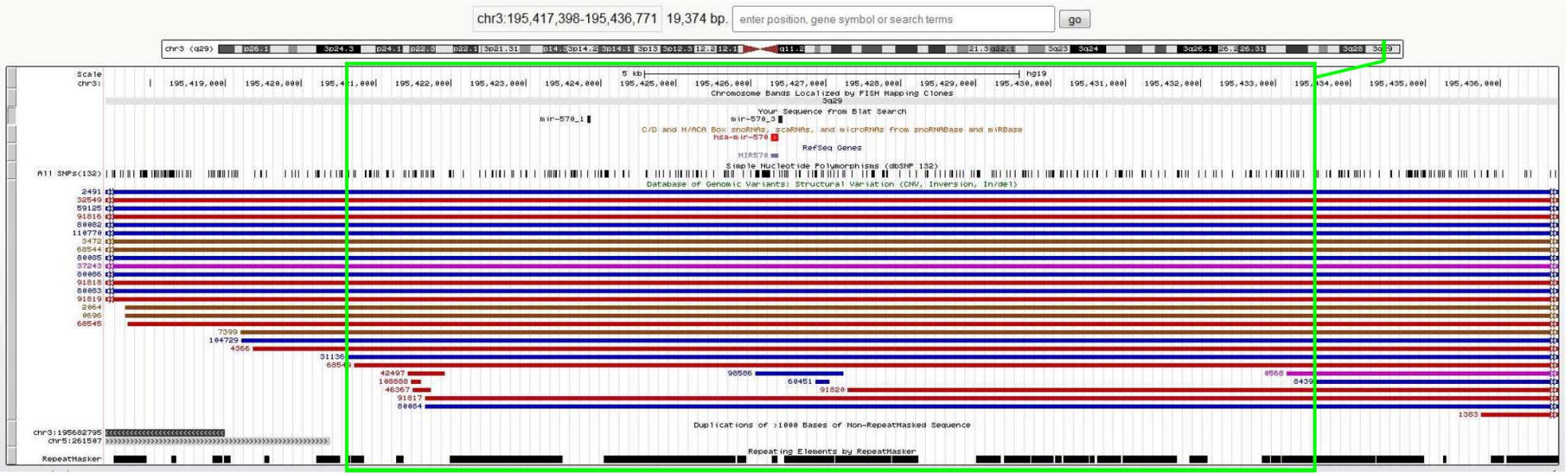


Supp. Figure S1. (continued)**.**

CNV-miRNA-566

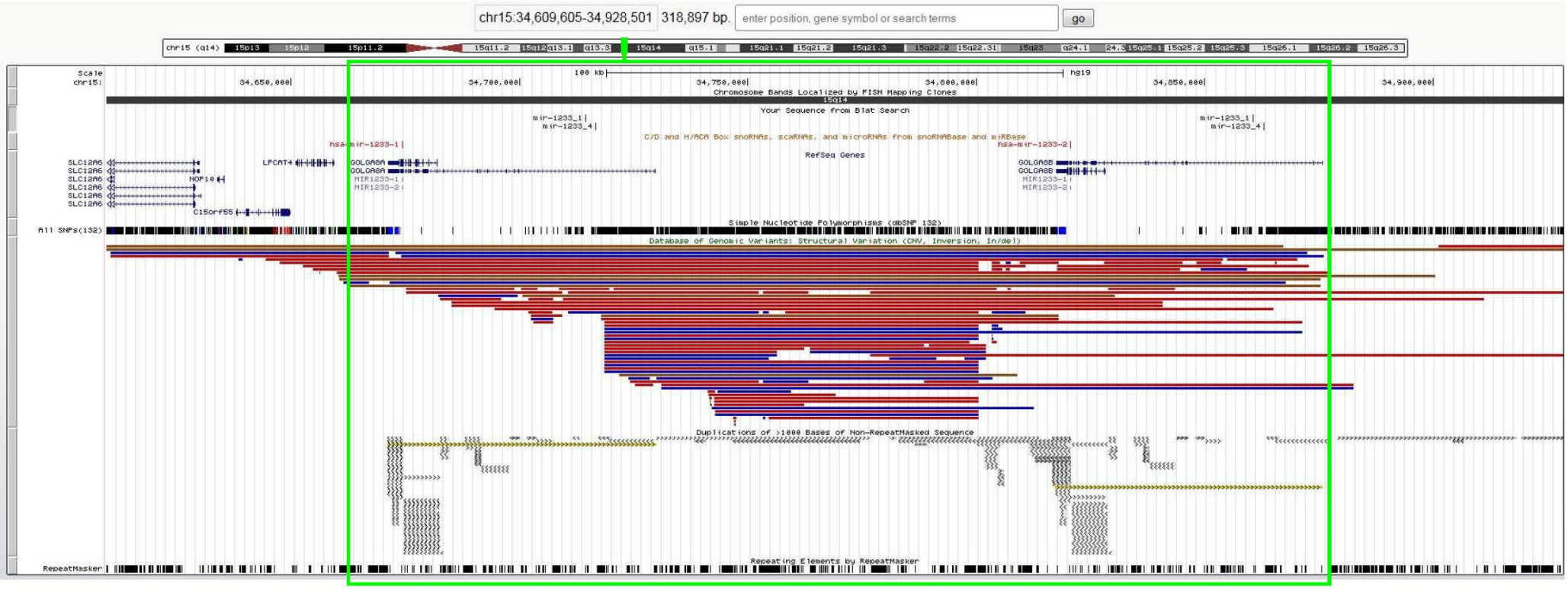


CNV-miRNA-570

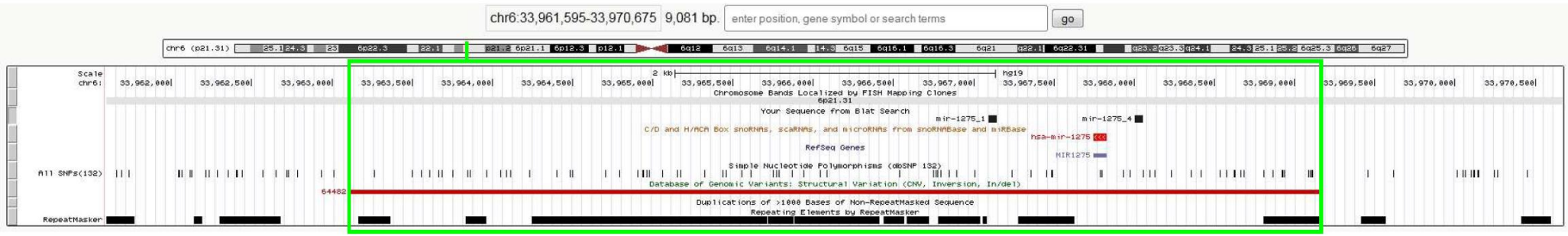


Supp. Figure S1. (continued)**.**

CNV-miRNA-1233

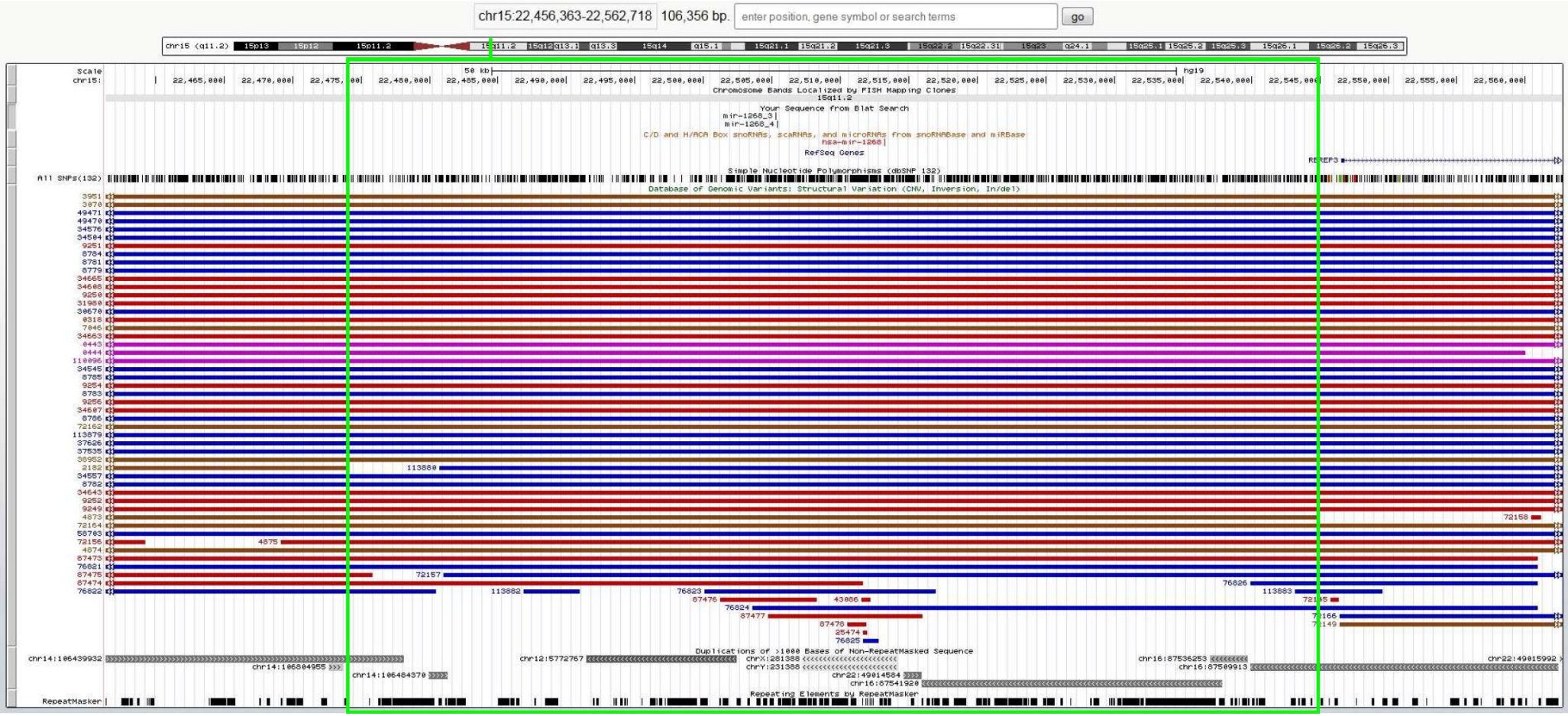


CNV-miRNA-1275



Supp. Figure S1. (continued)**.**

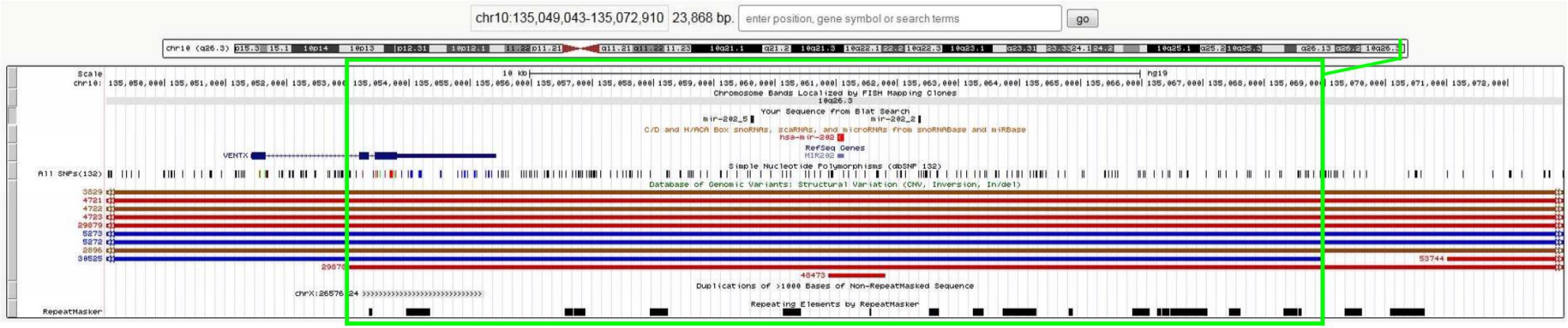
CNV-miRNA-1268



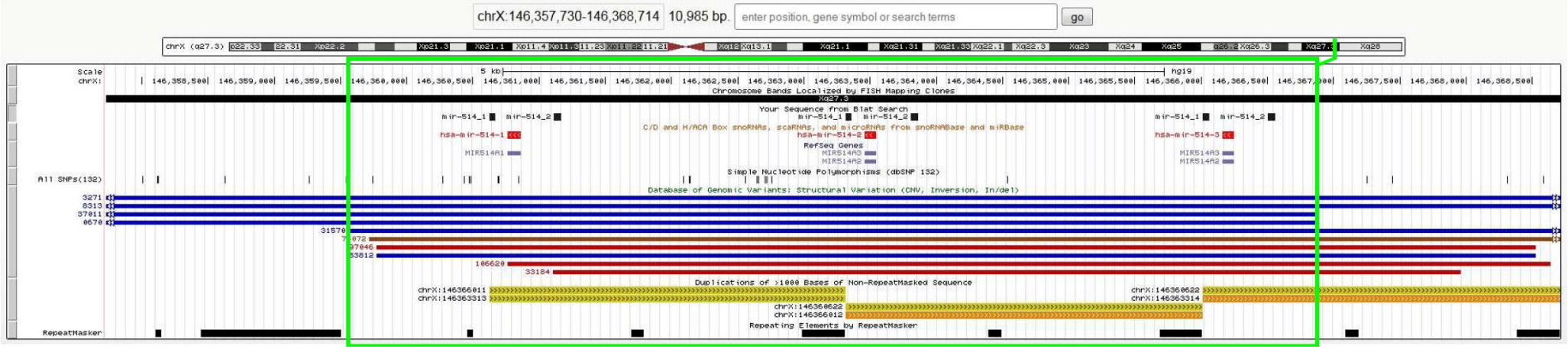
Supp. Figure S1. (continued)**.**

CNVmiR2

CNV-miRNA-202

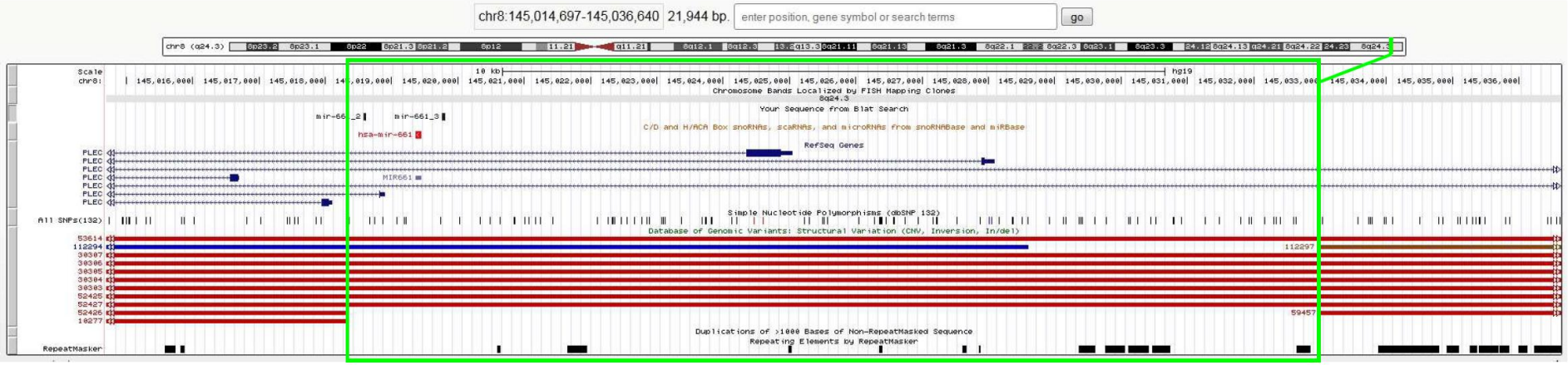


CNV-miRNA-514

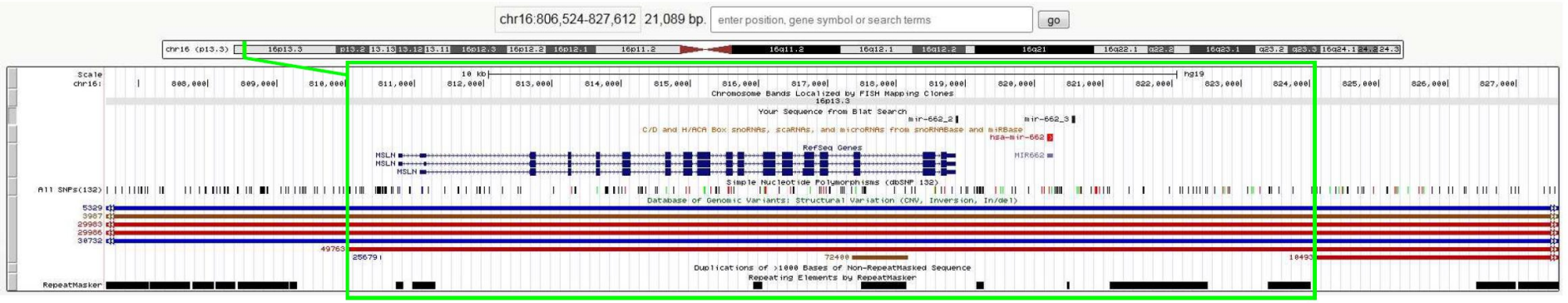


Supp. Figure S1. (continued)**.**

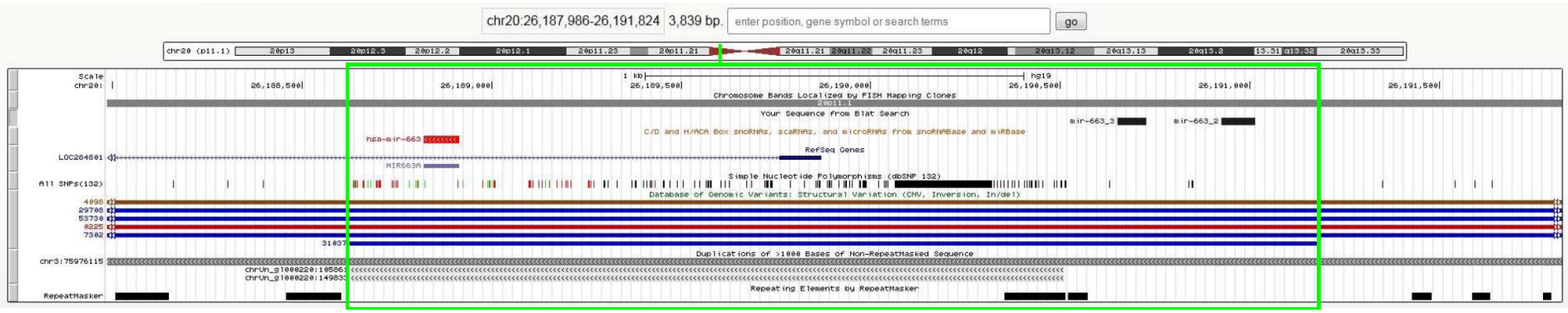
CNV-miRNA-661



CNV-miRNA-662

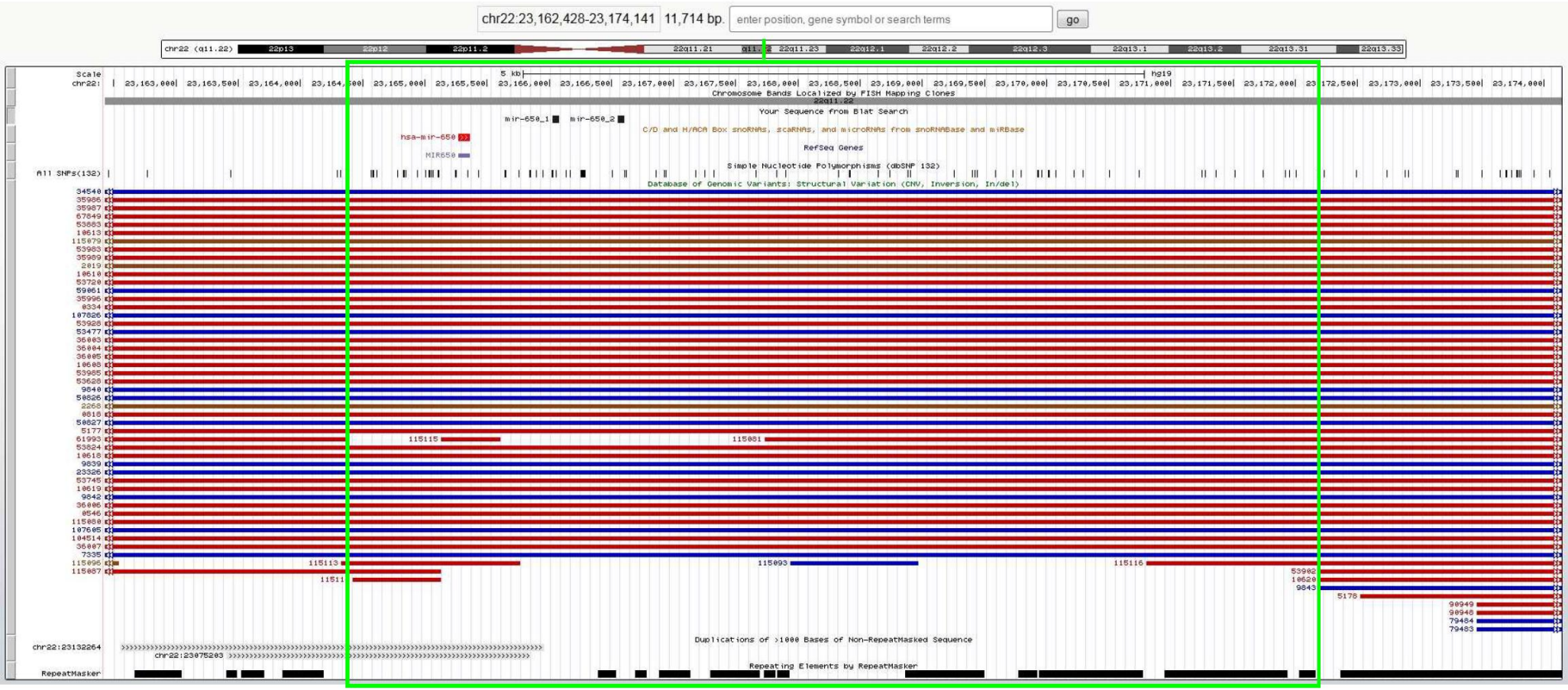


CNV-miRNA-663



Supp. Figure S1. (continued)**.**

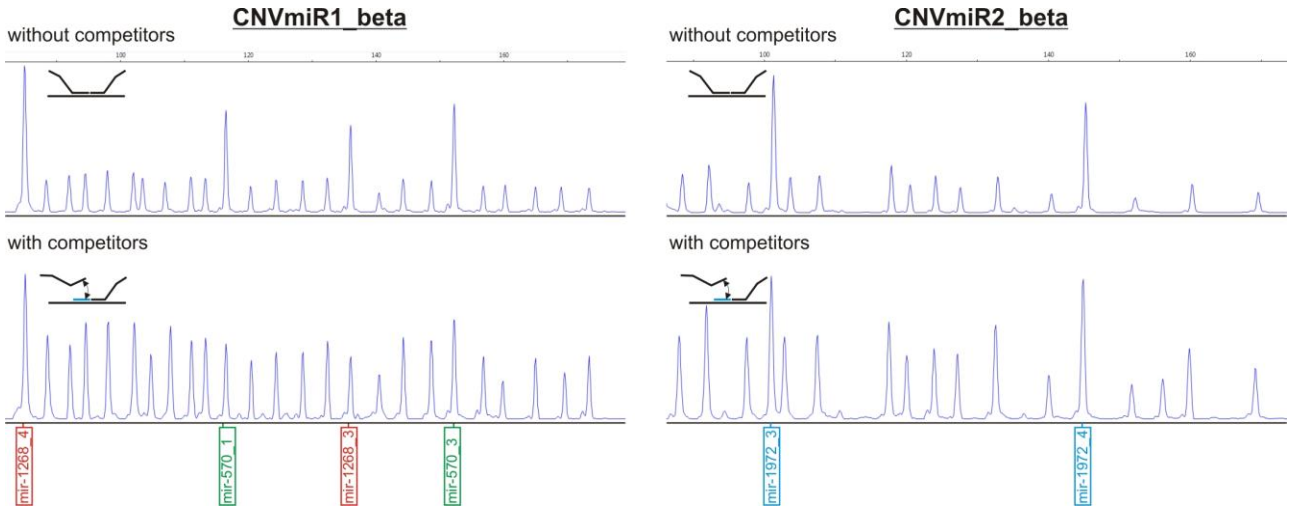
CNV-miRNA-650



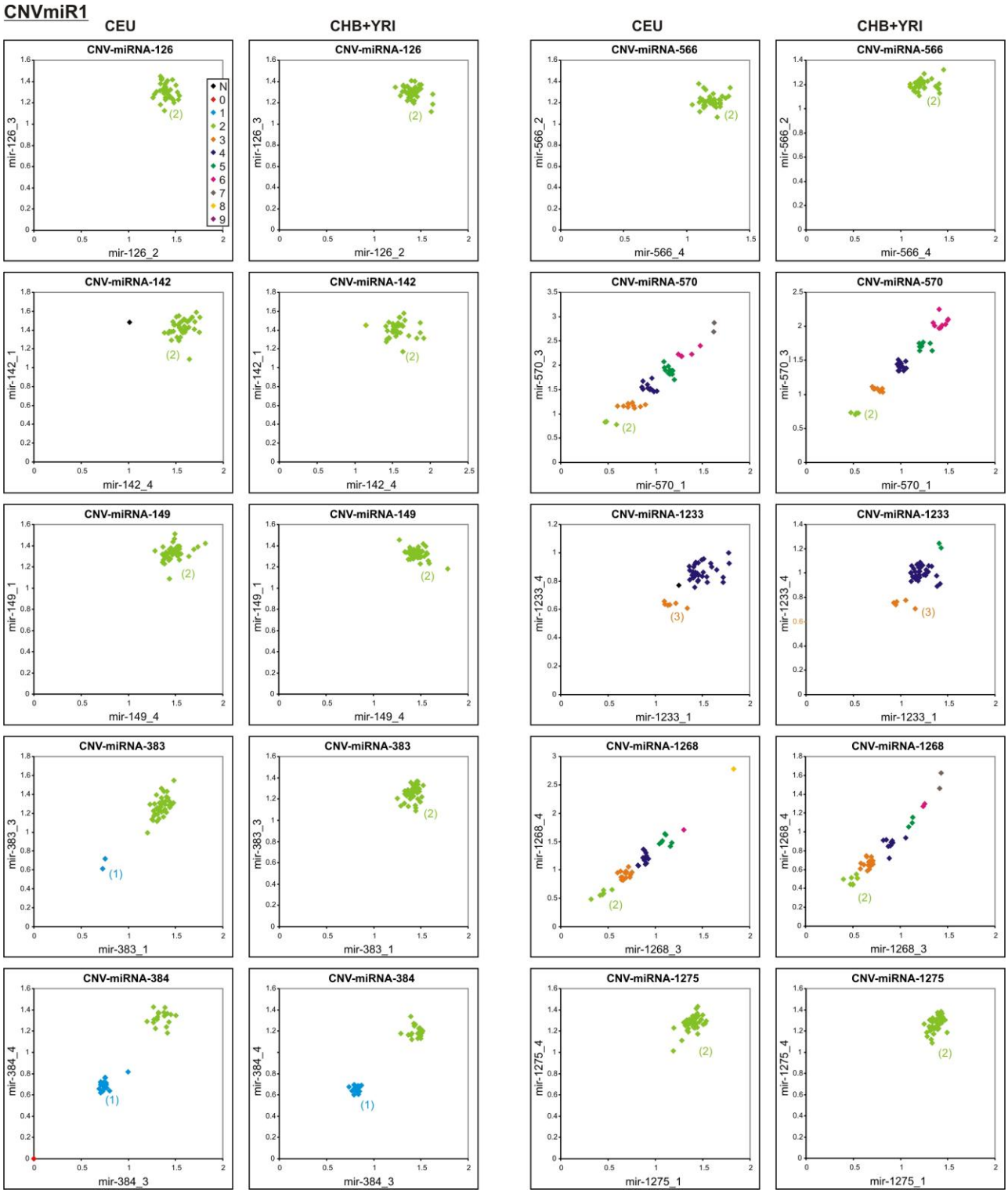
Supp. Figure S1. (continued)**.**

CNV-miRNA-1972

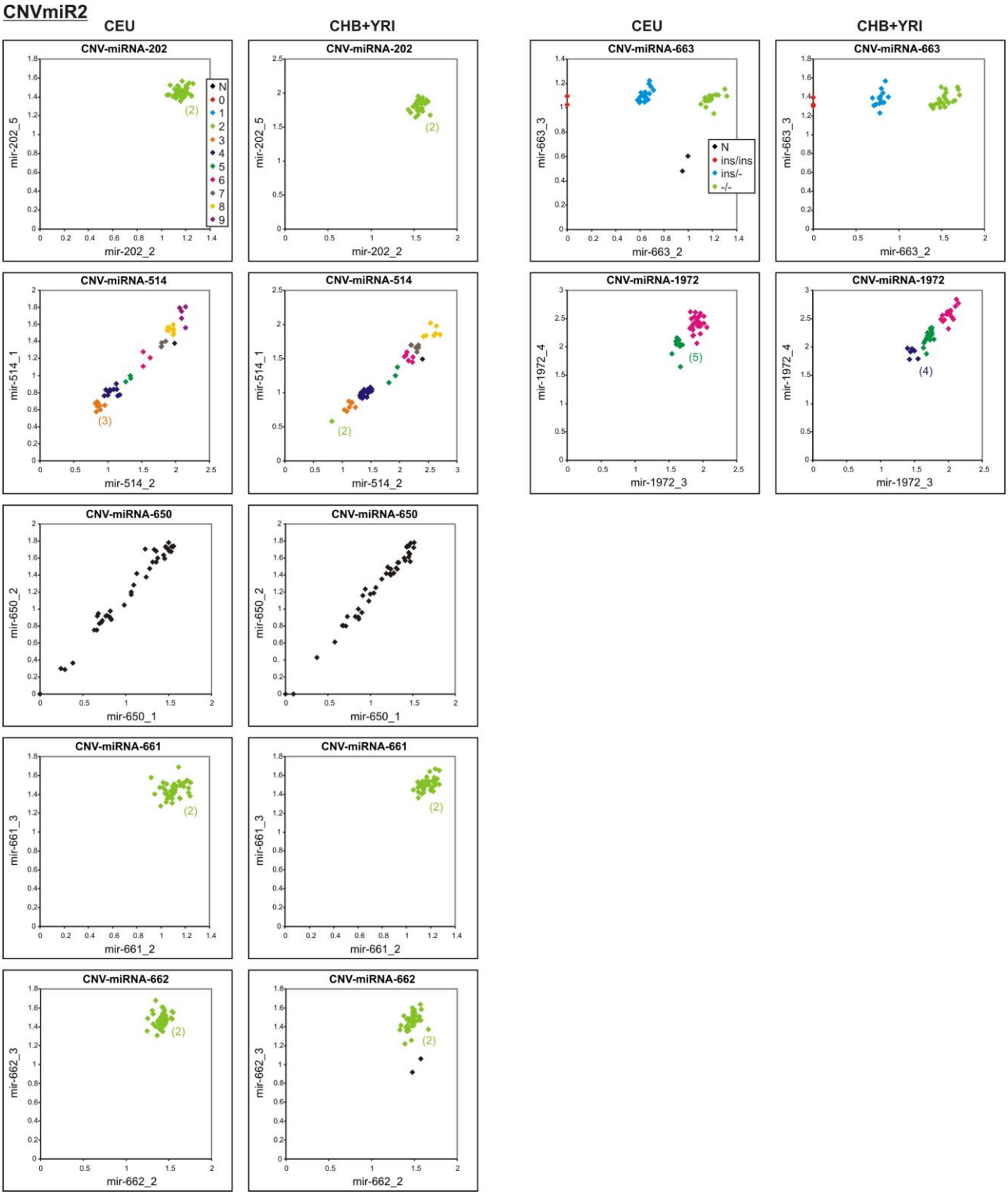
Supp. Figure S1. (continued)**.**



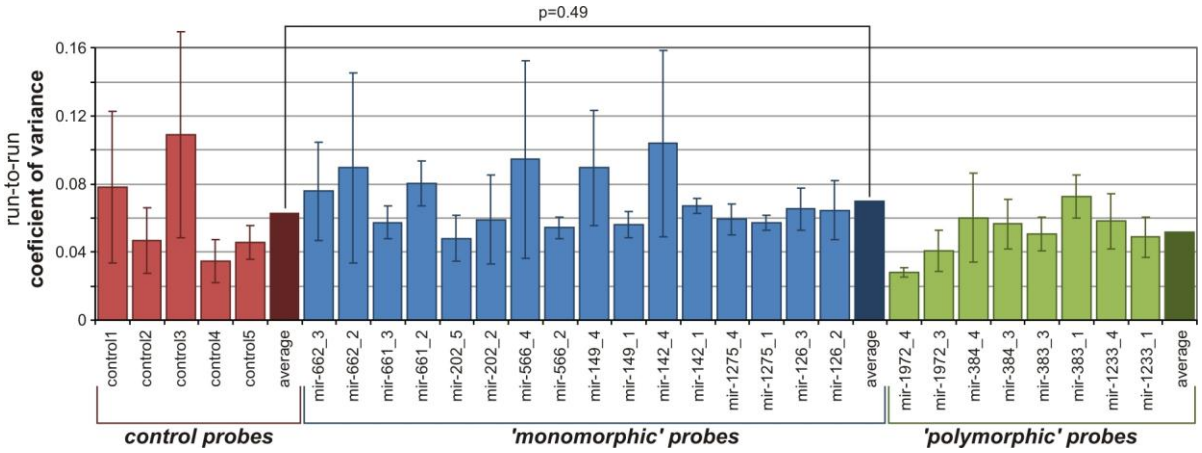
Supp. Figure S2. Effect of competitors on the signals of targeted MLPA probes. The presented electropherograms show probe signals obtained both without (upper electropherograms) and with (lower electropherograms) the addition of MLPA competitors specific for the mir-570_1, mir-570_2, mir-1268_3, mir-1268_4, mir-1972_3 and mir-1972_4 probes. For clarity, only the position of the competed probes is indicated in the figure. In all of our experiments, the competitors were added in the same concentration as the MLPA probes (i.e., at a 1:1 ratio with the competed 5'half probes). As the competitors were designed and tested during preliminary experiments the order of the probes differs somewhat from that in the final experiments (the positions of few probes were changed).



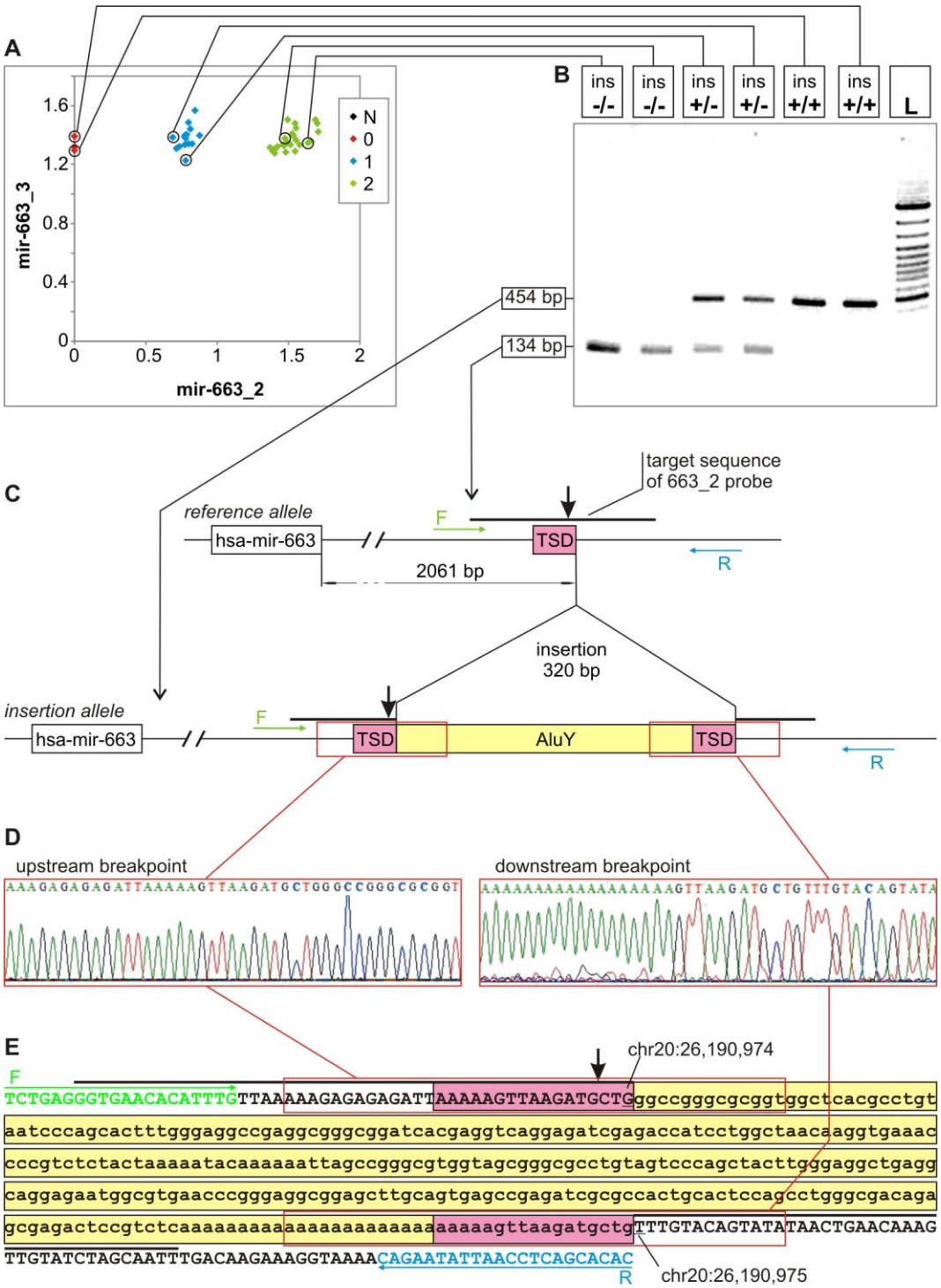
Supp. Figure S3. All of the signal scatter plots of the CNV-miRNA regions covered by the CNVmiR1 and CNVmiR2 assays. The corresponding results of experiment 1 performed on the CEU and CHB+YRI sample sets are shown next to each other on the left and right sides, respectively. Each sample is shown as a square, colored according to the predicted copy number genotype. The X, Y coordinates represent the normalized signals of probes targeting the investigated regions (probes IDs are indicated along the x and y axes). Number in parenthesis present on each graph indicates CN-genotype of the lowest signal-cluster. (Continued on next page.)



Supp. Figure S3. (continued)**.**



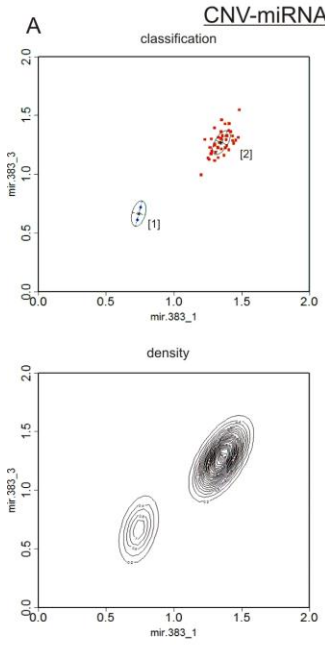
Supp. Figure S4. Comparison of the run-to-run (sample-to-sample) signal variation of control probes, probes representing monomorphic CNV-miRNAs ('monomorphic' probes) and probes representing simple polymorphism with few genotypes (in this case variation was calculated only for the biggest cluster, the most common genotype). The bars represent coefficient of variation (CV) calculated for individual probes as well as average CV calculated for selected group of probes. Presented CVs of "monomorphic' and 'polymorphic' probes are averaged values calculated for four independent experiments (experiment for samples CEU, experiment for samples CHB+YRI, each performed twice). Presented CVs of control probes are averaged values of 8 independent experiments (four experiments as described above were performed with two probe mixes, CNVmiR1 and CNVmiR2). Standard deviations of averaged CVs are shown as error bars.



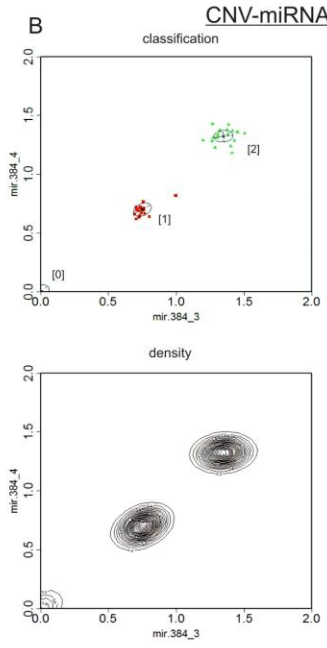
Supp. Figure S5. Common AluY insertion affecting the signal of the mir-663_2 probe. A) Signal scatter plot of the CNV-miRNA-663 assay. The signal of the mir-663_2 probe separates all of the samples into 3 distinct clusters. The circled samples were subjected to PCR amplification using F and R primers spanning the target sequence of the mir-663_2 probe. B) Agarose gel (negative) showing the separation of PCR products from selected DNA samples (see panel A). L – GeneRuler 100 bp DNA ladder (Fermentas Life Sciences). Depending on which cluster the DNA samples were derived from, the PCR products show (i) a single band of 134 bp corresponding to the reference human genome sequence (reference allele – derived from cluster "2"); (ii) a single band of 454 bp corresponding to an insertion of 320 bp (insertion allele – derived from cluster "0"); and (iii)

two bands, representing the heterozygous genotype (derived from cluster “1”). C) Schematic representation of the reference and insertion alleles revealed by sequence analysis of homozygous PCR products (see panel B). Sequence analysis showed that the 320 bp insertion is composed of 303 bp of the AluY sequence and 17 bp of a target site duplication (TSD) located 2,061 bp downstream of annotated miRNA-663 precursor (hsa-mir-663). The indicated target sequence of the mir-663_2 probe is disrupted in the insertion allele. D) Fragments of sequencing results spanning upstream and downstream insertion breakpoints. E) Reference (capital letters) and insertion (lowercase letters) sequences with annotated sequences of F and R primers (green and blue arrows, respectively), TSD (pink background), AluY (yellow background) and the target sequence (thick black line above the sequence) with the indicated half-probe ligation point (vertical arrowhead) of the mir-663_2 MLPA probe.

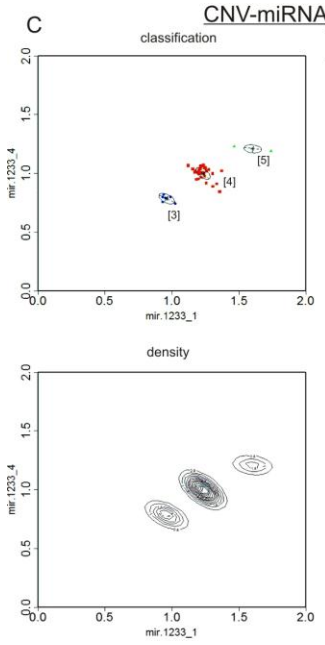
As it is shown in the figure this relatively small insertion does not directly affect sequence or copy number of miRNA-663 precursor. However, it rearranges the target sequence of the mir-663_2 probe in a way that prevents the ligation of half probes and consequently reduces the signal of the mir-663_2 probe by 50% or 100% (no signal) when present in a heterozygous or homozygous state, respectively. We found that in all of the tested populations, this AluY insertion is in Hardy-Weinberg equilibrium and in perfect agreement with a Mendelian mode of inheritance in the analyzed parent-offspring trios. However, the frequency of this insertion was found to be extremely differentiated between populations (rare in YRI (0.06), but frequent in CEU (0.27) and CHB (0.42)) (Figure 4).



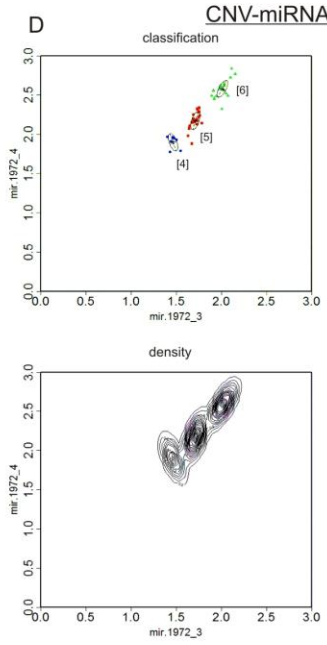
sample ID	cluster [1]	cluster [2]	CN-genotype
NA07034	2.89E-20	1.00E+00	2
NA07048	9.45E-32	1.00E+00	2
NA07055	3.85E-26	1.00E+00	2
NA07345	2.71E-29	1.00E+00	2
NA07348	2.52E-38	1.00E+00	2
NA07357	5.83E-34	1.00E+00	2
NA10830	1.50E-27	1.00E+00	2
NA10835	3.63E-23	1.00E+00	2
NA10851	3.45E-36	1.00E+00	2
NA10854	2.18E-31	1.00E+00	2
NA10857	9.97E-32	1.00E+00	2
NA10859	5.24E-28	1.00E+00	2
NA10861	1.06E-34	1.00E+00	2
NA10863	1.00E+00	2.05E-17	1
NA11839	2.47E-25	1.00E+00	2
NA11840	1.02E-26	1.00E+00	2
NA11881	6.12E-29	1.00E+00	2
NA11882	5.49E-42	1.00E+00	2
NA11994	3.85E-24	1.00E+00	2
NA12043	4.34E-22	1.00E+00	2
NA12044	5.49E-31	1.00E+00	2
NA12056	1.47E-38	1.00E+00	2
NA12057	4.82E-30	1.00E+00	2
NA12154	2.31E-24	1.00E+00	2
NA12234	1.00E+00	3.80E-19	1
NA12236	1.08E-30	1.00E+00	2
NA12249	1.84E-21	1.00E+00	2
NA12264	1.81E-37	1.00E+00	2
NA12707	2.12E-26	1.00E+00	2
NA12716	9.70E-34	1.00E+00	2
NA12717	5.58E-34	1.00E+00	2
NA12751	1.39E-36	1.00E+00	2
NA12762	5.06E-40	1.00E+00	2
NA12763	4.37E-28	1.00E+00	2
NA12801	3.44E-30	1.00E+00	2
NA12812	2.40E-16	1.00E+00	2
NA12813	1.76E-23	1.00E+00	2
NA12864	2.09E-26	1.00E+00	2
NA12873	1.88E-44	1.00E+00	2
NA12878	6.06E-38	1.00E+00	2
NA12891	9.18E-33	1.00E+00	2
NA12892	7.89E-35	1.00E+00	2



sample ID	cluster [0]	cluster [1]	cluster [2]	CN-genotype
NA07034	1.34E-44	1.00E+00	2.12E-49	1
NA07048	1.34E-48	1.00E+00	5.91E-46	1
NA07055	4.34E-147	9.68E-31	1.00E+00	2
NA07345	7.35E-159	1.87E-36	1.00E+00	2
NA07348	1.82E-136	3.13E-26	1.00E+00	2
NA07357	7.06E-52	1.00E+00	4.35E-40	1
NA10830	1.99E-47	1.00E+00	4.36E-46	1
NA10835	7.75E-42	1.00E+00	4.07E-55	1
NA10851	3.55E-49	1.00E+00	4.85E-45	1
NA10854	7.64E-71	1.00E+00	1.28E-23	1
NA10857	2.39E-47	1.00E+00	2.08E-47	1
NA10859	7.94E-138	4.25E-25	1.00E+00	2
NA10861	6.97E-154	7.36E-34	1.00E+00	2
NA10863	1.86E-166	1.49E-39	1.00E+00	2
NA11839	1.39E-46	1.00E+00	1.19E-49	1
NA11840	7.27E-163	2.16E-37	1.00E+00	2
NA11881	2.96E-46	1.00E+00	2.11E-46	1
NA11882	1.68E-177	3.65E-43	1.00E+00	2
NA11994	3.69E-43	1.00E+00	3.02E-53	1
NA12043	1.62E-45	1.00E+00	3.98E-48	1
NA12044	1.30E-150	4.62E-32	1.00E+00	2
NA12056	2.24E-48	1.00E+00	1.83E-46	1
NA12057	1.63E-169	1.73E-40	1.00E+00	2
NA12154	4.58E-46	1.00E+00	9.73E-47	1
NA12234	3.06E-160	3.39E-37	1.00E+00	2
NA12236	8.97E-173	6.84E-43	1.00E+00	2
NA12249	1.57E-146	1.51E-27	1.00E+00	2
NA12264	1.90E-41	1.00E+00	9.53E-54	1
NA12707	2.60E-52	1.00E+00	1.66E-39	1
NA12716	1.19E-39	1.00E+00	1.59E-57	1
NA12717	1.55E-172	1.93E-41	1.00E+00	2
NA12751	1.36E-150	4.13E-32	1.00E+00	2
NA12762	4.83E-47	1.00E+00	7.86E-51	1
NA12763	7.38E-161	9.08E-39	1.00E+00	2
NA12801	2.58E-45	1.00E+00	2.84E-50	1
NA12812	1.14E-51	1.00E+00	2.04E-40	1
NA12813	1.06E-142	2.29E-28	1.00E+00	2
NA12864	1.00E+00	2.14E-46	1.55E-214	0
NA12873	2.22E-45	1.00E+00	4.06E-49	1
NA12878	1.54E-152	3.69E-31	1.00E+00	2
NA12891	3.13E-48	1.00E+00	3.66E-45	1



sample ID	cluster [3]	cluster [4]	cluster [5]	CN-genotype
NA18501	1.00E+00	3.16E-23	9.79E-66	3
NA18507	3.07E-19	1.00E+00	1.01E-21	4
NA18516	3.14E-20	1.00E+00	4.97E-21	4
NA18522	1.00E+00	5.13E-20	3.98E-68	3
NA18526	8.39E-114	5.75E-45	1.00E+00	5
NA18529	2.55E-33	1.00E+00	3.28E-10	4
NA18532	1.80E-19	1.00E+00	1.96E-21	4
NA18561	3.37E-21	1.00E+00	6.16E-20	4
NA18562	8.72E-20	1.00E+00	2.71E-20	4
NA18563	1.40E-19	1.00E+00	8.52E-21	4
NA18570	1.62E-16	1.00E+00	5.96E-28	4
NA18572	3.63E-20	1.00E+00	1.79E-19	4
NA18573	7.60E-15	1.00E+00	1.06E-25	4
NA18576	1.91E-27	1.00E+00	3.52E-15	4
NA18577	2.86E-13	1.00E+00	1.03E-26	4
NA18579	2.80E-20	1.00E+00	8.84E-22	4
NA18603	4.71E-20	1.00E+00	4.37E-21	4
NA18605	4.44E-29	1.00E+00	1.06E-13	4
NA18612	7.25E-28	1.00E+00	8.78E-15	4
NA18620	1.00E+00	8.95E-21	1.45E-61	3
NA18621	3.31E-24	1.00E+00	1.51E-17	4
NA18622	6.83E-28	1.00E+00	5.35E-15	4
NA18623	1.63E-20	1.00E+00	3.89E-24	4
NA18624	1.27E-15	1.00E+00	5.10E-27	4
NA18856	1.00E+00	2.62E-22	1.15E-65	3
NA18859	2.75E-27	1.00E+00	3.89E-15	4
NA18863	1.00E+00	7.66E-27	6.71E-73	3
NA19100	1.41E-26	1.00E+00	1.31E-15	4
NA19103	1.48E-26	1.00E+00	1.25E-16	4
NA19127	2.14E-17	1.00E+00	6.92E-24	4
NA19132	4.82E-22	1.00E+00	1.06E-19	4
NA19138	6.93E-22	1.00E+00	3.97E-20	4
NA19141	1.91E-25	1.00E+00	4.54E-17	4
NA19152	6.06E-16	1.00E+00	5.78E-31	4
NA19153	7.58E-22	1.00E+00	5.31E-20	4
NA19159	4.21E-23	1.00E+00	2.20E-18	4
NA19172	4.84E-18	1.00E+00	5.08E-22	4
NA19192	8.19E-24	1.00E+00	6.65E-18	4
NA19193	1.03E-21	1.00E+00	6.40E-20	4
NA19202	2.06E-77	2.49E-18	1.00E+00	5
NA19204	1.00E+00	4.83E-18	7.07E-61	3
NA19222	1.15E-19	1.00E+00	4.64E-21	4
NA19240	1.97E-19	1.00E+00	2.07E-21	4



sample ID	cluster [4]	cluster [5]	cluster [6]	CN-genotype
NA18501	1.25E-78	9.32E-10	1.00E+00	6
NA18507	9.93E-53	8.37E-04	9.99E-01	6
NA18516	3.99E-21	1.00E+00	7.51E-05	5
NA18522	2.34E-46	5.86E-03	9.94E-01	6
NA18526	1.11E-04	1.00E+00	3.71E-09	5
NA18532	2.46E-14	1.00E+00	9.96E-06	5
NA18537	1.00E+00	4.88E-04	1.51E-15	4
NA18540	9.93E-01	6.57E-03	2.74E-14	4
NA18561	1.30E-18	9.99E-01	5.82E-04	5
NA18563	3.81E-21	1.00E+00	5.48E-05	5
NA18570	1.46E-04	1.00E+00	6.39E-09	5
NA18572	4.75E-16	1.00E+00	1.75E-05	5
NA18573	2.05E-13	1.00E+00	1.11E-06	5
NA18576	1.00E+00	5.05E-09	2.71E-22	4
NA18577	2.10E-14	1.00E+00	3.86E-06	5
NA18603	1.00E+00	1.41E-06	8.85E-20	4
NA18605	2.84E-21	1.00E+00	2.00E-04	5
NA18612	1.00E+00	2.69E-05	2.94E-17	4
NA18620	6.99E-19	1.00E+00	1.21E-04	5
NA18621	1.00E+00	6.84E-06	4.64E-18	4
NA18622	2.81E-11	1.00E+00	3.34E-07	5
NA18623	1.00E+00	1.36E-06	3.83E-19	4
NA18624	1.70E-13	1.00E+00	2.37E-06	5
NA18856	2.49E-72	2.04E-09	1.00E+00	6
NA18859	6.45E-15	1.00E+00	1.38E-06	5
NA18863	1.65E-69	1.46E-07	1.00E+00	6
NA19100	4.74E-12	1.00E+00	6.47E-07	5
NA19103	1.69E-109	1.48E-14	1.00E+00	6
NA19127	6.40E-77	1.17E-09	1.00E+00	6
NA19132	6.70E-73	2.78E-10	1.00E+00	6
NA19138	2.27E-97	1.17E-13	1.00E+00	6
NA19140	5.29E-70	8.81E-08	1.00E+00	6
NA19141	4.10E-66	3.32E-07	1.00E+00	6
NA19152	2.51E-20	1.00E+00	2.80E-04	5
NA19153	1.49E-53	3.68E-05	1.00E+00	6
NA19159	1.67E-08	1.00E+00	4.49E-08	5
NA19172	2.47E-23	1.00E+00	3.03E-04	5
NA19192	1.74E-48	4.45E-04	1.00E+00	6
NA19193	3.66E-74	3.24E-11	1.00E+00	6
NA19202	3.69E-111	1.29E-16	1.00E+00	6
NA19204	3.42E-07	1.00E+00	1.98E-08	5
NA19222	9.42E-52	6.51E-07	1.00E+00	6
NA19240	1.85E-64	7.39E-07	1.00E+00	6

Supp. Figure S6. CNV-miRNA genotyping using the EM algorithm. In panels A, B, C and D, four examples of genotype assignment using the EM algorithm for CNV-miRNA-383, CNV-miRNA-384, CNV-miRNA-1233 and CNV-miRNA-1272, respectively, are shown. Each panel includes a classification plot (upper), density plot (lower) and table showing the confidence of sample assignment to particular genotype clusters. The last column of each table indicates the CN-genotypes determined according to visual examination.

Unattended analysis of the signal data was performed using R, version 2.13.1 (R Development Core Team, 2011) and the Expectation Maximization algorithm provided in the mclust package, version 3.4.10 (Fraley, 2006). Clustering of the signal data was carried out by applying the Mclust() function with the equal shape and volume ellipsoidal model (EEV). The sensitivity of the EM algorithm was set using control values provided by emControl () function [eps~10^{-16}; tol; itmax=(Inf,Inf); equalPro=False]. The number of iterations was set based on tol=($1x10^{-12}$,$1.5x10^{-8}$) that was the same for all cluster analyses. The general concept of EM algorithm was recently described (Do and Batzoglou, 2008).

Supp. Table S1. Characteristics of designed MLPA probes
[Table only for computer review. Printouts of this table will be unreadable.]

Assay CNVmiR1

probe ID	probe location (hg19)	probe type[1]	5'PSS	length	5'SS	length	5'TSS	length	Tm	5' HPL	3'TSS	length	Tm	3'SS	length	3'PSS	length	3' HPL	TPL
control1	chr22:30039296-30069338	control	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75.6	46	GGCAAAACTTCTGGCCCAGAAG	22	71.0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	47	93
mir-566_4	chr3:50211301-50211343	standard	GGGTTCCCTAAGGGTTGGA	19	cgctact	7	acgaggaccacgaccgcatgta	22	72.1	48	cgtgggcagcaaggactacgt	21	72.0	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	48	96
mir-149_4	chr2:241396569-241396612	standard	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	tccccctctagctcaagccaga	22	70.8	49	gccccctgcatcagtacattcc	22	72.0	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	50	99
mir-142_4	chr17:56409474-56409517	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	caggagcccaaggctatcccaa	22	73.1	51	caaaaatggtggccatgttggg	22	71.3	atctac	6	TCTAGATTGGATCTTGCTGGCGC	23	51	102
mir-126_2	chr9:139567291-139567335	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	tggtttctgtgaggccacagaga	23	70.0	52	gctttgtcctggggacatgttg	22	70.9	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	53	105
mir-1233_1	chr15:34714381-34714427	SD universal	GGGTTCCCTAAGGGTTGGA	19	cgctactactatt	13	ctggttcattggcaaacaggca	22	70.7	54	gctctgagcagcacattcagaagttg	26	71.8	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	54	108
mir-1233_1K		competitor					ctggttcattggcaaacaggca	22											
control2	chr1:156105818-156105862	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72.8	55	CTGGACATCAAGCTGGCCCTG	21	72.7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	56	111
mir-384_3	chrX:76141456-76141502	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	tctttgggacaggtcccagaaatg	24	70.6	57	cccagtggactttttcttgccttg	23	70.9	actaaatctac	11	TCTAGATTGGATCTTGCTGGCGC	23	57	114
mir-383_1	chr8:14710837-14710885	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattag	15	ttcatctcacctgaagcatgctgg	24	70.3	58	ccactccagtccaccaaatgcagtt	25	73.5	actaaatctac	11	TCTAGATTGGATCTTGCTGGCGC	23	59	117
mir-570_1	chr3:195423823-195423871	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtaga	19	cagtccatccagctcctgcatt	22	70.5	60	gtccgtcctaccaataacctctcactt	27	70.3	ctaaatctac	10	TCTAGATTGGATCTTGCTGGCGC	23	60	120
mir-570_1K		competitor					cagtccatccagctcctgcatt	22											
control5	chr2:109545794-109545837	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaat	21	AGTCCTGTGGCTACGGCACCAA	22	72.8	62	AGACGAGGACTACGGCTGCGTC	22	71.8	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	62	124
mir-566_2	chr3:50210365-50210410	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattg	23	accctgagcgtgggagcattaa	22	70.0	64	cccccacctgagcaaaaactcttt	24	71.7	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	64	128
mir-126_3	chr9:139563299-139563344	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattg	23	ttccacgatcacctccagtgtcct	24	72.0	66	aatccttgctcttgccccgaca	22	71.5	taatggtcaaactaaatctac	21	TCTAGATTGGATCTTGCTGGCGC	23	66	132
mir-142_1	chr17:56408086-56408133	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattg	23	ccagagccctcttctgatctcatagg	26	72.0	68	cagggccaaacccgtatcttct	22	70.4	tctaatggtcaaactaaatctac	23	TCTAGATTGGATCTTGCTGGCGC	23	68	136
mir-1268_3	chr15:22505273-22505320	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgc	27	tgaccaggaattagtccccaccag	24	71.3	70	gtgacactcacagccttacccaca	24	71.4	tctaatggtcaaactaaatctac	23	TCTAGATTGGATCTTGCTGGCGC	23	70	140
mir-1268_3K		competitor					tgaccaggaattagtccccaccag	24											
control3	chr17:3397657-3397712	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70.6	72	TATAGAGAAAGTTGATTACCCCCGGGATG	29	70.9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	72	144
mir-149_1	chr2:241394664-241394713	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgcca	29	caaaaaaatcctcagagcactgggga	26	72.3	74	agggaagtgtcccagtgactgagg	24	72.3	tgtatctaatggtcaaactaaatctac	27	TCTAGATTGGATCTTGCTGGCGC	23	74	148
mir-1275_1	chr6:33967092-33967143	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacc	31	gaagctatagatgctaaggcaaccgg	26	70.3	76	ggctcaggttctcttcttcaaagctc	26	71.1	tgtatctaatggtcaaactaaatctac	27	TCTAGATTGGATCTTGCTGGCGC	23	76	152
mir-570_3	chr3:195426375-195426424	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttt	33	tgccatgcttctctctgtgaaaaacc	26	70.9	78	ccaagccgaagtcagcataacatc	24	70.1	gaaatgtatctaatggtcaaactaaatctac	31	TCTAGATTGGATCTTGCTGGCGC	23	78	156
mir-570_3K		competitor					tgccatgcttctctctgtgaaaaacc	26											
mir-384_4	chrX:76139186-76139233	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttttcagct	40	tgagcagggaaaggcaaggaa	21	69.0	80	gtttcatgttgacacagaggaggttg	27	72.5	aaatgtatctaatggtcaaactaaatctac	30	TCTAGATTGGATCTTGCTGGCGC	23	80	160
mir-1233_4	chr15:34716713-34716770	SD universal	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttt	34	aaaaggcaacaacttttctacagaccagc	29	70.0	82	catgatcagctagaaaacaaggcaagagg	29	72.0	aaatgtatctaatggtcaaactaaatctac	30	TCTAGATTGGATCTTGCTGGCGC	23	82	164
mir-1233_4K		competitor					aaaaggcaacaacttttctacagaccagc	29											
mir-383_3	chr8:14712337-14712388	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttttcagc	39	cagagagtggcccctctttaatctctg	26	70.2	84	ctaccctagacatcaaaggctgcact	26	70.2	ttgcgaaatgtatctaatggtcaaactaaatctac	35	TCTAGATTGGATCTTGCTGGCGC	23	84	168
control4	chr11:14515205-14515256	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttttcagctcgcg	44	TGCATGTTTGGAGCATCGACACA	23	70.4	86	GCTATGTTAGAAGAAATGCTGTTTTGGCC	29	70.5	tgcgaaatgtatctaatggtcaaactaaatctac	34	TCTAGATTGGATCTTGCTGGCGC	23	86	172
mir-1275_4	chr6:33968002-33968054	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttttcagctc	41	cgactcaccctgttcttggttcaaaagag	28	74.1	88	ctgggtgtgtggagacagaagaagt	25	71.1	accatttgcgaaatgtatctaatggtcaaactaaatctac	40	TCTAGATTGGATCTTGCTGGCGC	23	88	176
mir-1268_4	chr15:22505357-22505414	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccaccttttcagctc	41	acgagagacacacatctcaccccattctgt	30	75.1	90	gcagccaacacagtgcacattcagcata	28	75.6	ccatttgcgaaatgtatctaatggtcaaactaaatctac	39	TCTAGATTGGATCTTGCTGGCGC	23	90	180
mir-1268_4K		competitor					acgagagacacacatctcaccccattctgt	30											

Supp. Table S1. (continued)

Assay CNVmiR2

probe ID	probe location (hg19)	probe type[1]	5'PSS	length	5'SS	length	5'TGS	length	Tm	5' HPL	3'TGS	length	Tm	3'SS	length	3'PSS	length	3' HPL	TPL
control1	chr22:30039296-30069338	control	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	GGCCCAGATCACCGAGGAGGA	21	75.6	46	GGCAAAACTTCTGGCCCAGAAG	22	71.0	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	47	93
mir-662_2	chr16:818858-818901	standard	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	agacacgtcttgtggcctccg	21	71.3	48	aggactttctgtgaccccaccag	23	70.8	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	48	96
mir-661_2	chr8:145018575-145018619	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactact	10	acctcaagcctgcagctgaacc	22	71.2	51	cactgcactcttcaacttcgcc	23	72.4	tctac	5	TCTAGATTGGATCTTGCTGGCGC	23	51	102
mir-1972_3	chr16:15102560-15102612	SD universal	GGGTTCCCTAAGGGTTGGA	19	cgctacta	8	agagccacagcctttcacaatctga	25	70.2	52	agtgaatggtgcagagagctttcttgtc	28	71.1	ac	2	TCTAGATTGGATCTTGCTGGCGC	23	53	105
mir-1972_3K		competitor					agagccacagcctttcacaatctga	25											
mir-202_2	chr10:135062340-135062384	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactatt	13	aaggaacaaggctgaggcctca	22	70.0	54	gaggcctcttcagtgaccatgtc	23	70.8	aaatctac	8	TCTAGATTGGATCTTGCTGGCGC	23	54	108
control2	chr1:156105818-156105862	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactat	12	CAGCTGGACGAGTACCAGGAGCTT	24	72.8	55	CTGGACATCAAGCTGGCCCTG	21	72.7	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	56	111
mir-650_2	chr22:23166555-23166605	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagta	17	tccacctcccttcctcatgga	21	70.1	57	catcgagcatttctaattttcatggctgtc	30	71.7	ctac	4	TCTAGATTGGATCTTGCTGGCGC	23	57	114
mir-662_3	chr16:820548-820591	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtag	18	ttcaggctggtgaaggtgtcgat	23	70.4	60	gtccatggagatgttggcgtg	21	70.8	gtcaaactaaatctac	16	TCTAGATTGGATCTTGCTGGCGC	23	60	120
control5	chr2:109545794-109545837	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaat	21	AGTCCTGTGGCTACGGCACCAA	22	72.8	62	AGACGAGGACTACGGCTGCGTC	22	71.8	ggtcaaactaaatctac	17	TCTAGATTGGATCTTGCTGGCGC	23	62	124
mir-663_3	chr20:26190651-26190727	SD specific	GGGTTCCCTAAGGGTTGGA	19	cgctac	6	ccaaggcagttaatcacatattaacaactctttagatac	39	70.3	64	ataaggtttatagcaatgaggtacaaaattggttcaac	38	70.2	tac	3	TCTAGATTGGATCTTGCTGGCGC	23	64	128
mir-514_2	chrX:146361109-146361166	SD universal	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaa	20	gaatcagcagtatgcagtcctcagatc	27	70.3	66	caacagttaagcagaaaatgcttaactccag	31	70.1	aactaaatctac	12	TCTAGATTGGATCTTGCTGGCGC	23	66	132
mir-514_2K		competitor					gaatcagcagtatgcagtcctcagatc	27											
mir-202_5	chr10:135059594-135059645	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattga	24	cttttaagacagcagccgactgtca	25	70.2	68	ttctttaaacagggctgatgagggacc	27	71.9	tggtcaaactaaatctac	18	TCTAGATTGGATCTTGCTGGCGC	23	68	136
mir-650_1	chr22:23166029-23166081	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattga	24	ctattcctagctcttcaatgcagggac	27	70.2	70	gtagggacaaggagtttactgcttgg	26	70.1	taatggtcaaactaaatctac	21	TCTAGATTGGATCTTGCTGGCGC	23	70	140
control3	chr17:3397657-3397712	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatg	26	TCCCTGCGCCATTGAGGTCTATAAAAT	27	70.6	72	TATAGAGAAAGTTGATTACCCCCGGGATG	29	70.9	aatggtcaaactaaatctac	20	TCTAGATTGGATCTTGCTGGCGC	23	72	144
mir-1972_4	chr16:15102860-15102914	SD universal	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattga	24	gaatgtccatgaaggttggttggtatatagg	31	71.3	74	tcaggagcacctggcttttaagga	24	70.1	gcgaaatgtatctaatggtcaaacta	27	TCTAGATTGGATCTTGCTGGCGC	23	74	148
mir-1972_4K		competitor					gaatgtccatgaaggttggttggtatatagg	31											
mir-663_2	chr20:26190927-26191015	SD specific	GGGTTCCCTAAGGGTTGGA	19	cgctactactatta	14	ggtgaacacatttgttaaaaagagagagattaaaaagttaagatg	45	72.2	78	ctgtttgtacagtatataactgaacaaagttgtatctagcaatt	44	70.1	actaaatctac	11	TCTAGATTGGATCTTGCTGGCGC	23	78	156
mir-514_1	chrX:146360624-146360670	SD universal	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttt	35	cctctgaaaggcacttaagaggcttc	26	70.8	80	ccttgagggaggccagcacta	21	71.8	tttgcgaaatgtatctaatggtcaaactaaatctac	36	TCTAGATTGGATCTTGCTGGCGC	23	80	160
mir-514_1K		competitor					cctctgaaaggcacttaagaggcttc	26											
mir-661_3	chr8:145019762-145019813	standard	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttca	37	cagcaaaccccatgaagtcctataccc	26	71.3	82	atgaaggcctcatacagaccaggagc	26	72.3	gcgaaatgtatctaatggtcaaactaaatctac	33	TCTAGATTGGATCTTGCTGGCGC	23	82	164
control4	chr11:14515205-14515256	control	GGGTTCCCTAAGGGTTGGA	19	cgctactactattagtagaattgatgccacctttttcagctcgcg	44	TGCATGTTTGGAGCATCGACACA	23	70.4	86	GCTATGTTAGAAGAAATGCTGTTTTGGCC	29	70.5	tgcgaaatgtatctaatggtcaaactaaatctac	34	TCTAGATTGGATCTTGCTGGCGC	23	86	172

[1]for details see Figure 1D
5'PSS, 3'PSS - 5' and 3' primer-specific sequence, respectively
5'SS, 3'SS - 5' and 3' stuffer sequence, respectively
5'TSS, 3'TSS - 5' and 3' target-specific sequence, respectively
Tm- melting temperature
5' half-probe - complete sequence of 5' half-probe
3' half-probe - complete sequence of 3' half-probe
5'HPL, 3'HPL - 5' and 3' half-probe length

SALSA PCR Forward primer (Labeled): *GGGTTCCCTAAGGGTTGGA
SALSA PCR Reverse primer (Unlabeled): GTGCCAGCAAGATCCAATCTAGA

Sequence used for generation of all 5' and 3' stuffer sequences: AC#V00604, Phage M13 genome, position 3-99
5'-
cgctactactattagtagaattgatgccacctttttcagctcgcgccccaaatgaaaatatagctaaacaggttattgaccattgcgaaatgtatctaatggtcaaactaaatctac-3'

Supp. Table S2. The genotyping results of this study and their concordance with previous studies

					CNVmiR1 assay											CNVmiR2 assay					
					CNV-miRNA-383		CNV-miRNA-384		CNV-miRNA-1233			CNV-miRNA-570		CNV-miRNA-1268		CNV-miRNA-1972		CNV-miRNA-514		AluY insertion	
sample ID	sex	HapMap population	family ID	family relation	this study	previous study	this study	previous study MC	this study	previous study MC*	previous study C	this study	previous study	this study	previous study MC	this study	previous study MC*	this study	previous study	this study	previous study
NA12878	W	CEU	1463	daughter	2	NG	2	2	4	4	4	3	NG	2	2	5	5	7	NG	+/-	NG
NA12892	W	CEU	1463	mother	2	NG	2	2	3	3	3	3	NG	2	2	5	5	8	NG	+/-	NG
NA12891	M	CEU	1463	father	2	NG	1	1	4	4	4	4	NG	2	2	6	6	3	NG	+/-	NG
NA12740	W	CEU	1444	daughter	2	NG	2	2	4	4	4	2	NG	3	3	5	5	9	NG	+/-	NG
NA12751	W	CEU	1444	mother	2	NG	2	2	4	4	4	2	NG	2	2	6	6	9	NG	+/-	NG
NA12750	M	CEU	1444	father	2	NG	1	1	4	4	4	3	NG	4	4	5	5	4	NG	-/-	NG
NA10835	M	CEU	1416	son	2	NG	1	1	3	3	3	3	NG	3	3	6	6	4	NG	+/+	NG
NA12249	W	CEU	1416	mother	2	NG	2	2	4	4	4	4	NG	3	3	6	6	6	NG	+/-	NG
NA12248	M	CEU	1416	father	2	NG	1	1	3	3	3	3	NG	2	2	6	6	3	NG	+/-	NG
NA10863	W	CEU	1375	daughter	1	NG	2	2	4	4	4	6	NG	4	4	6	6	9	NG	+/-	NG
NA12234	W	CEU	1375	mother	1	NG	2	2	4	4	4	5	NG	3	3	6	6	6	NG	+/-	NG
NA12264	M	CEU	1375	father	2	NG	1	1	4	4	4	5	NG	3	3	6	6	5	NG	-/-	NG
NA12707	M	CEU	1358	son	2	NG	1	1	4	4	4	6	NG	3	3	6	6	4	NG	-/-	NG
NA12717	W	CEU	1358	mother	2	NG	2	2	4	4	4	6	NG	5	N	6	6	7	NG	+/-	NG
NA12716	M	CEU	1358	father	2	NG	1	1	3	3	3	5	NG	5	N	6	6	4	NG	-/-	NG
NA10854	W	CEU	1349	daughter	2	NG	1	1	4	4	4	3	NG	5	5	6	6	5	NG	-/-	NG
NA11840	W	CEU	1349	mother	2	NG	2	2	4	4	4	5	NG	8	N	6	N	9	NG	-/-	NG
NA11839	M	CEU	1349	father	2	NG	1	1	4	4	4	5	NG	3	3	5	5	3	NG	+/-	NG
NA10859	W	CEU	1347	daughter	2	NG	2	2	4	4	4	5	NG	4	4	6	6	5	NG	-/-	NG
NA11882	W	CEU	1347	mother	2	NG	2	2	4	4	4	6	NG	3	3	5	5	8	NG	+/-	NG
NA11881	M	CEU	1347	father	2	NG	1	1	4	4	4	4	NG	3	3	6	6	3	NG	-/-	NG
NA10857	M	CEU	1346	son	2	NG	1	1	4	4	4	5	NG	4	4	6	6	3	NG	+/-	NG
NA12044	W	CEU	1346	mother	2	NG	2	2	4	4	4	4	NG	4	4	6	6	7	NG	-/-	NG
NA12043	M	CEU	1346	father	2	NG	1	1	4	4	4	5	NG	4	4	6	6	4	NG	+/-	NG
NA07348	W	CEU	1345	daughter	2	NG	2	2	4	4	4	2	NG	4	4	6	6	8	NG	+/-	NG
NA07345	W	CEU	1345	mother	2	NG	2	2	4	4	4	4	NG	4	4	6	6	6	NG	-/-	NG
NA07357	M	CEU	1345	father	2	NG	1	1	4	4	4	3	NG	5	N	6	6	4	NG	+/-	NG
NA10851	M	CEU	1344	son	2	NG	1	1	3	3	3	5	NG	3	3	6	6	3	NG	-/-	NG
NA12057	W	CEU	1344	mother	2	NG	2	2	4	4	4	5	NG	2	2	6	6	8	NG	-/-	NG
NA12056	M	CEU	1344	father	2	NG	1	1	3	3	3	3	NG	4	4	6	6	4	NG	-/-	NG
NA12864	M	CEU	1459	son	2	NG	0	0	4	4	4	5	NG	4	4	6	6	3	NG	+/-	NG
NA12873	W	CEU	1459	mother	2	NG	1	1	4	4	4	5	NG	3	3	6	6	6	NG	+/-	NG
NA12872	M	CEU	1459	father	2	NG	1	1	4	4	4	6	NG	3	3	6	6	4	NG	+/-	NG
NA12801	M	CEU	1454	son	2	NG	1	1	4	4	4	4	NG	3	3	6	6	3	NG	-/-	NG
NA12813	W	CEU	1454	mother	2	NG	2	2	4	4	4	2	NG	2	2	6	6	7	NG	-/-	NG
NA12812	M	CEU	1454	father	2	NG	1	1	4	4	4	4	NG	4	4	6	6	4	NG	-/-	NG
NA12753	W	CEU	1447	daughter	2	NG	2	2	4	4	4	2	NG	5	N	6	6	8	NG	+/-	NG
NA12763	W	CEU	1447	mother	2	NG	2	2	4	4	4	3	NG	3	3	5	5	8	NG	+/-	NG
NA12762	M	CEU	1447	father	2	NG	1	1	4	4	4	4	NG	5	5	5	5	4	NG	+/-	NG
NA10830	M	CEU	1408	son	2	NG	1	1	4	4	4	4	NG	3	3	6	6	5	NG	+/+	NG
NA12236	W	CEU	1408	mother	2	NG	2	2	4	4	4	4	NG	4	4	6	6	9	NG	-/-	NG
NA12154	M	CEU	1408	father	2	NG	1	1	4	4	4	4	NG	3	3	5	5	3	NG	+/-	NG
NA10861	W	CEU	1362	daughter	2	NG	2	2	4	4	4	3	NG	5	N	6	6	7	NG	+/-	NG
NA11995	W	CEU	1362	mother	2	NG	2	2	4	4	4	2	NG	4	4	6	6	8	NG	-/-	NG
NA11994	M	CEU	1362	father	2	NG	1	1	4	4	4	3	NG	4	4	6	6	3	NG	+/-	NG
NA07048	M	CEU	1341	son	2	NG	1	1	3	3	3	7	NG	5	N	6	6	4	NG	-/-	NG
NA07055	W	CEU	1341	mother	2	NG	2	2	3	3	3	7	NG	6	6	6	6	8	NG	-/-	NG
NA07034	M	CEU	1341	father	2	NG	1	1	4	4	4	3	NG	3	3	6	6	4	NG	-/-	NG
NA18526	W	CHB	-	unrelated	2	NG	2	2	5	5	5	6	NG	2	2	5	5	6	NG	+/-	NG
NA18529	W	CHB	-	unrelated	2	NG	2	2	4	4	4	4	NG	3	N	6	N	7	NG	+/-	NG
NA18532	W	CHB	-	unrelated	2	NG	2	2	4	4	4	5	NG	5	N	5	5	5	NG	+/-	NG
NA18537	W	CHB	-	unrelated	2	NG	2	2	4	4	4	5	NG	2	2	4	4	6	NG	+/-	NG
NA18540	W	CHB	-	unrelated	2	NG	1	N	4	4	4	4	NG	3	3	4	4	5	NG	-/-	NG
NA18542	W	CHB	-	unrelated	2	NG	2	2	4	4	4	3	NG	3	3	5	5	7	NG	+/-	NG
NA18561	M	CHB	-	unrelated	2	NG	1	1	4	4	4	6	NG	2	2	5	5	4	NG	+/-	NG
NA18562	M	CHB	-	unrelated	2	NG	1	1	4	4	4	6	NG	3	3	5	5	4	NG	+/-	NG
NA18563	M	CHB	-	unrelated	2	NG	1	1	4	4	4	6	NG	4	4	5	5	4	NG	+/+	NG
NA18570	W	CHB	-	unrelated	2	NG	2	2	4	4	4	4	NG	3	3	5	5	7	NG	-/-	NG
NA18571	W	CHB	-	unrelated	2	NG	2	2	4	4	4	5	NG	3	3	4	4	6	NG	+/+	NG
NA18572	M	CHB	-	unrelated	2	NG	1	1	4	4	4	6	NG	2	2	5	5	4	NG	+/-	NG
NA18573	W	CHB	-	unrelated	2	NG	2	2	4	4	4	4	NG	3	3	5	5	8	NG	-/-	NG
NA18576	W	CHB	-	unrelated	2	NG	2	2	4	4	4	6	NG	3	3	4	4	7	NG	+/+	NG
NA18577	W	CHB	-	unrelated	2	NG	2	2	4	4	4	5	NG	3	3	5	5	7	NG	+/-	NG
NA18579	W	CHB	-	unrelated	2	NG	2	2	4	4	4	5	NG	2	2	4	5	7	NG	+/-	NG
NA18603	M	CHB	-	unrelated	2	NG	1	1	4	4	4	5	NG	3	3	4	4	4	NG	+/-	NG
NA18605	M	CHB	-	unrelated	2	NG	1	1	4	4	4	5	NG	3	3	5	5	4	NG	-/-	NG
NA18612	M	CHB	-	unrelated	2	NG	1	1	4	4	4	4	NG	3	3	4	4	3	NG	+/-	NG
NA18620	M	CHB	-	unrelated	2	NG	1	1	3	3	3	3	NG	5	N	5	5	4	NG	-/-	NG
NA18621	M	CHB	-	unrelated	2	NG	1	1	4	4	4	3	NG	2	2	4	4	4	NG	+/-	NG
NA18622	M	CHB	-	unrelated	2	NG	1	1	4	4	4	4	NG	3	3	5	5	4	NG	+/-	NG
NA18623	M	CHB	-	unrelated	2	NG	1	1	4	4	4	3	NG	3	3	4	4	4	NG	-/-	NG
NA18624	M	CHB	-	unrelated	2	NG	1	1	4	4	4	5	NG	3	3	5	5	4	NG	-/-	NG
NA18501	M	YRI	Y004	unrelated	2	NG	1	1	3	3	3	3	NG	7	6	6	6	2	NG	-/-	NG
NA18507	M	YRI	Y009	unrelated	2	NG	1	1	4	4	4	5	NG	2	2	X	6	2	NG	-/-	NG
NA18516	M	YRI	Y013	unrelated	2	NG	1	1	4	4	4	4	NG	3	3	6	5	3	NG	-/-	NG
NA18522	M	YRI	Y016	unrelated	2	NG	1	1	3	3	3	4	NG	4	4	6	6	3	NG	-/-	NG
NA18856	M	YRI	Y023	unrelated	2	NG	1	1	3	3	3	2	NG	7	6	6	6	4	NG	-/-	NG
NA18859	M	YRI	Y012	unrelated	2	NG	1	1	4	4	4	5	NG	3	3	5	5	4	NG	-/-	NG
NA18863	M	YRI	Y024	unrelated	2	NG	1	1	3	3	3	4	NG	4	4	6	6	4	NG	-/-	NG
NA19100	W	YRI	Y105	unrelated	2	NG	2	2	4	4	4	4	NG	6	6	5	5	6	NG	-/-	NG
NA19103	M	YRI	Y042	unrelated	2	NG	1	1	4	4	4	3	NG	3	3	6	6	3	NG	-/-	NG
NA19127	W	YRI	Y077	unrelated	2	NG	2	2	4	4	4	4	NG	3	3	6	6	7	NG	+/-	NG
NA19132	W	YRI	Y101	unrelated	2	NG	2	2	4	4	4	3	NG	4	4	6	6	5	NG	-/-	NG
NA19138	M	YRI	Y043	unrelated	2	NG	1	1	4	4	4	6	NG	3	3	6	6	4	NG	-/-	NG
NA19140	W	YRI	Y071	unrelated	2	NG	2	2	4	4	4	3	NG	3	3	6	6	6	NG	-/-	NG
NA19141	M	YRI	Y071	unrelated	2	NG	1	1	4	4	4	4	NG	5	5	6	6	3	NG	-/-	NG
NA19152	W	YRI	Y072	unrelated	2	NG	2	2	4	4	4	6	NG	4	4	5	5	6	NG	+/-	NG
NA19153	M	YRI	Y072	unrelated	2	NG	1	1	4	4	4	4	NG	4	4	6	6	3	NG	-/-	NG
NA19159	W	YRI	Y056	unrelated	2	NG	2	2	4	4	4	2	NG	2	2	5	5	6	NG	-/-	NG
NA19172	W	YRI	Y047	unrelated	2	NG	2	2	4	4	4	4	NG	3	3	5	5	6	NG	+/-	NG
NA19192	M	YRI	Y112	unrelated	2	NG	1	1	4	4	4	2	NG	3	3	6	6	3	NG	-/-	NG
NA19193	W	YRI	Y112	unrelated	2	NG	2	2	4	4	4	2	NG	4	4	6	6	6	NG	-/-	NG
NA19202	W	YRI	Y045	unrelated	2	NG	2	2	5	5	N	3	NG	6	6	6	6	6	NG	-/-	NG
NA19204	W	YRI	Y048	unrelated	2	NG	2	2	3	3	3	6	NG	4	4	5	5	6	NG	-/-	NG
NA19222	W	YRI	Y058	unrelated	2	NG	2	2	4	4	4	4	NG	4	4	6	6	7	NG	-/-	NG
NA19240	W	YRI	Y117	unrelated	2	NG	2	2	4	4	4	4	NG	5	5	6	6	6	NG	-/-	NG

W – woman; M – man; CEU – European population; CHB – Asiatic population, Han Chinese; YRI – African population; NG – not genotyped; C and MC – genotyped in previous study by Conrad et al. 2010 and McCarroll et al. 2008,

respectively; *Note that due to different strategy of CN-genotype calling (different CN-reference assumption) genotypes of two CNV regions (CNV-mir-1233 and CNV-mir-1972) called by McCarroll et al. were adjusted before comparison with CN-genotypes determined in our study and CN-genotypes determined by Conrad et al.; Red font indicates discordant genotypes.

Supp. Table S3. Quality control analyses of obtained CNV genotyping results

	CNV-miRNA ID	Reproducibility			Concordance [%] with previous studies		HWE p-value	Mendelian inheritance	Genotype correlation (R) parent-offspring/ mother-father
		probe-to-probe correlation (R) CEU/CHB+YRI	exp.-to-exp. correlation (R) CEU/CHB+YRI	genotyping reproducibility [%]	McCarroll et al. 2008	Conrad et al. 2010b	CEU/CHB/YRI	CEU trios	CEU trios
Bi-allelic	CNV-mir-383	NA/NA	NA/NA	100	NG	NG	NA due to low MAF	all passed	-
	CNV-mir-384 (chrX)	NA/NA	NA/NA	100	100	NG	NA due to low MAF	1 trio failed	-
	CNV-mir-1972	NA/0.94	NA/0.90	99[1]	99[4]	NG	0.78/0.31/0.71	all passed	-
Multi-allelic	CNV-mir-1233	NA/0.71	0.72/0.95	100	100	100	0.89/0.99/0.99	all passed	0.45/-0.18
	CNV-mir-570	0.96/0.98	0.99/0.99	99[2]	NG	NG	-	-	0.55/0.27
	CNV-mir-1268	0.98/0.98	0.98/0.99	100	98[5]	NG	-	-	0.46/0.07
	CNV-mir-514 (chrX)	0.98/0.98	0.99/0.99	91[3]	NG	NG	-	-	NA due to location on X chromosome
polymorphic (genotypes not determined)	CNV-mir-650	0.99/0.99	0.99/0.99	NA	NG	NA	-	-	-

NA – not analyzed; exp.-to-exp. – experiment-to-experiment; R – correlation coefficient; NG – not genotyped in previous experiment; [1-5]discordant genotypes: [1](5>6), [2](6>5), [3](2x3>4, 4>3, 4>5, 6>7, 7>6, 7>8, 2x8>7), [4](4>5), [5](2x7>6).

Supp. Table S4. Characteristics of the CNV-miRNA polymorphisms in the three human populations

	CNV-miRNA ID	Observed genotypes			Inferred alleles			MAF			cMGF		
		CEU	CHB	YRI	CEU	CHB	YRI	CEU	CHB	YRI	CEU	CHB	YRI
Bi-allelic	CNV-mir-383	1,2	2	2	0,1	1	1	0.02	-	-	0.03	-	-
	CNV-mir-384 (chrX)	M:0,1 W:1,2	M:1 W:1,2	M:1 W:2	0,1	1	1	0.02	0.03	-	W:0.06	W:0.08	-
	CNV-mir-663 (AluY ins)	+/+,+/-,-/-	+/+,+/-,-/-	+/-,-/-	+,-	+,-	+,-	0.27	0.42	0.06	0.47	0.42	0.13
	CNV-mir-1972	5,6	4,5,6	5,6	2,3	2,3	2,3	0.11	0.33*	0.15	0.22	0.42	0.29
Multi-allelic	CNV-mir-1233	3,4	3,4,5	3,4,5	1,2	1,2,3	1,2,3	0.08	0.4	0.12	0.16	0.08	0.25
	CNV-mir-570	2,3,4,5,6,7	3,4,5,6	2,3,4,5,6	-	-	-	-	-	-	0.72	0.67	0.58
	CNV-mir-1268	2,3,4,5,6,8	2,3,4,5	2,3,4,5,6,7	-	-	-	-	-	-	0.66	0.38	0.67
	CNV-mir-514 (chrX)	M:3,4,5 W:5,6,7,8,9	M:3,4 W:5,6,8	M:2,3,4 W:5,6,7,8	-	-	-	-	-	-	W:0.63	W:0.5	W:0.58

Underlining – major genotypes and major alleles; W – women; M – men; + presence of the AluY insertion; - absence of the AluY insertion; MAF – minor allele frequency; *note that the minor allele in CHB is the major allele in CEU and YRI; cMGF – combined minor genotype frequency.

Supp. Table S5. Functional relevance of miRNAs identified as CN-polymorphic

miRNA	Functional relevance of CN-polymorphic miRNAs	ref.
miRNA-384	**apoptosis** affects sensitivity to TNF-related apoptosis-inducing ligand (TRAIL)-induced apoptosis	(Sudbery, et al., 2010)
	chloride transport targets mRNA of *cystic fibrosis transmembrane conductance regulator* (*CFTR*); inhibits expression of *SLC12A2*; may play an important role in regulating chloride transport in epithelial cells	(Gillen, et al., 2011)
miRNA-1233	**cancer** renal cell carcinoma (RCC)-associated oncomir and a potential biomarker for RCC patients	(Wulfken, et al., 2011)
miRNA-514	**cancer** downregulated in clear cell RCC; differentiates malignant from non-malignant tissue	(Jung, et al., 2009)
miRNA-570	**cancer** frequent somatic disruption of the miRNA-570-binding site in the *CD274* 3'UTR leads to overexpression of CD274 protein in gastric cancer	(Wang, et al., 2011)
	cancer regulated by estrogen receptor alpha in luminal-like breast cancer cells	(Cicatiello, et al., 2010)
miRNA-650	**cancer** overexpressed in acral compared to non-acral melanoma	(Chan, et al., 2011)
	cancer represses the expression of *NDRG2*, a potential tumor suppressor gene	(Feng, et al., 2011)
	cancer overexpression may promote the growth of cancer cells by targeting inhibitor of growth 4 (ING4)	(Zhang, et al., 2010)
miRNA-663	**cancer** considered to be a tumor suppressor, induces mitotic catastrophe in gastric cancer cells; downregulation may lead to the development of gastric cancer.	(Pan, et al., 2010)
	cancer may play an important role in all trans-retinoic acid (ATRA)-induced differentiation of acute myeloid leukemia (AML) HL-60 cells; may be used in the treatment of hematological malignancies	(Jian, et al., 2011)
	cancer/inflammation targets multiple genes implicated in the immune response; upregulated by resveratrol (a natural antioxidant); may help to optimize the use of resveratrol as both an anti-inflammatory and anti-cancer agent against malignances associated with high levels of miRNA-155	(Tili, et al., 2010)
miRNA-383	**infertility** plays a potential role in regulating spermatogenesis in human males; downregulated in testicular tissues of patients with non-obstructive azoospermia (NOA)	(Lian, et al., 2009)
	infertility downregulation is associated with hyperactive proliferation of germ cells in infertile male patients, with maturation arrest (MA); overexpression of miR-383 results in suppression of proliferation, G1-phase arrest and induction of apoptosis, whereas silencing of miR-383 reverses these effects; targets tumor suppressor interferon regulatory factor-1 (IRF1)	(Lian, et al., 2010)
	cancer potential prognostic marker in ependymomas	(Costa, et al., 2011)
	cancer directly targets and downregulates *type 1 iodothyronine deiodinase* (*DIO1*), playing an important role in the regulation of cell proliferation and differentiation; overexpressed in ccRCC; regulates *DIO1* expression in ccRCC	(Boguslawska, et al., 2011)
miRNA-1268	No data	
miRNA-1972	No data	

References to Supp. Table S5

Boguslawska J, Wojcicka A, Piekielko-Witkowska A, Master A, Nauman A. 2011. MiR-224 targets the 3'UTR of type 1 5'-iodothyronine deiodinase possibly contributing to tissue hypothyroidism in renal cancer. PLoS One 6:e24541.

Chan E, Patel R, Nallur S, Ratner E, Bacchiocchi A, Hoyt K, Szpakowski S, Godshalk S, Ariyan S, Sznol M and others. 2011. MicroRNA signatures differentiate melanoma subtypes. Cell Cycle 10:1845-1852.

Cicatiello L, Mutarelli M, Grober OM, Paris O, Ferraro L, Ravo M, Tarallo R, Luo S, Schroth GP, Seifert M and others. 2010. Estrogen receptor alpha controls a gene network in luminal-like breast cancer cells comprising multiple transcription factors and microRNAs. Am J Pathol 176:2113-2130.

Costa FF, Bischof JM, Vanin EF, Lulla RR, Wang M, Sredni ST, Rajaram V, Bonaldo Mde F, Wang D, Goldman S and others. 2011. Identification of microRNAs as potential prognostic markers in ependymoma. PLoS One 6:e25114.

Feng L, Xie Y, Zhang H, Wu Y. 2011. Down-regulation of NDRG2 gene expression in human colorectal cancer involves promoter methylation and microRNA-650. Biochem Biophys Res Commun 406:534-538.

Gillen AE, Gosalia N, Leir SH, Harris A. 2011. MicroRNA regulation of expression of the cystic fibrosis transmembrane conductance regulator gene. Biochem J 438:25-32.

Jian P, Li ZW, Fang TY, Jian W, Zhuan Z, Mei LX, Yan WS, Jian N. 2011. Retinoic acid induces HL-60 cell differentiation via the upregulation of miR-663. J Hematol Oncol 4:20.

Jung M, Mollenkopf HJ, Grimm C, Wagner I, Albrecht M, Waller T, Pilarsky C, Johannsen M, Stephan C, Lehrach H and others. 2009. MicroRNA profiling of clear cell renal cell cancer identifies a robust signature to define renal malignancy. J Cell Mol Med 13:3918-3928.

Lian J, Tian H, Liu L, Zhang XS, Li WQ, Deng YM, Yao GD, Yin MM, Sun F. 2010. Downregulation of microRNA-383 is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation by targeting IRF1. Cell Death Dis 1:e94.

Lian J, Zhang X, Tian H, Liang N, Wang Y, Liang C, Li X, Sun F. 2009. Altered microRNA expression in patients with non-obstructive azoospermia. Reprod Biol Endocrinol 7:13.

Pan J, Hu H, Zhou Z, Sun L, Peng L, Yu L, Sun L, Liu J, Yang Z, Ran Y. 2010. Tumor-suppressive mir-663 gene induces mitotic catastrophe growth arrest in human gastric cancer cells. Oncol Rep 24:105-112.

Sudbery I, Enright AJ, Fraser AG, Dunham I. 2010. Systematic analysis of off-target effects in an RNAi screen reveals microRNAs affecting sensitivity to TRAIL-induced apoptosis. BMC Genomics 11:175.

Tili E, Michaille JJ, Adair B, Alder H, Limagne E, Taccioli C, Ferracin M, Delmas D, Latruffe N, Croce CM. 2010. Resveratrol decreases the levels of miR-155 by upregulating miR-663, a microRNA targeting JunB and JunD. Carcinogenesis 31:1561-1566.

Wang W, Sun J, Li F, Li R, Gu Y, Liu C, Yang P, Zhu M, Chen L, Tian W and others. 2011. A frequent somatic mutation in CD274 3'-UTR leads to protein over-expression in gastric cancer by disrupting miR-570 binding. Hum Mutat 33:480-484.

Wulfken LM, Moritz R, Ohlmann C, Holdenrieder S, Jung V, Becker F, Herrmann E, Walgenbach-Brunagel G, von Ruecker A, Muller SC and others. 2011. MicroRNAs in renal cell carcinoma: diagnostic implications of serum miR-1233 levels. PLoS One 6:e25787.

Zhang X, Zhu W, Zhang J, Huo S, Zhou L, Gu Z, Zhang M. 2010. MicroRNA-650 targets ING4 to promote gastric cancer tumorigenicity. Biochem Biophys Res Commun 395:275-280.

OŚWIADCZENIA WSPÓŁAUTORÓW

OKREŚLAJĄCE ICH UDZIAŁ W TWORZENIU PUBLIKACJI

ZAWARTYCH W NINIEJSZEJ PRACY DOKTORSKIEJ

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ
ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
fax: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327
NIP 777-00-02-062

Poznań, 10 czerwca 2013

Dr hab. Piotr Kozłowski, prof. IChB PAN

OŚWIADCZENIA

Dotyczy rozprawy doktorskiej mgr inż. Małgorzaty Marcinkowskiej-Swojak:

Mgr inż. Małgorzata Marcinkowska-Swojak wykonywała pracę doktorską w Instytucie Chemii Bioorganicznej, PAN od 2010 r. Jej praca doktorska jest znaczną częścią projektu badawczego (grantu) MNiSW pod tytułem „Opracowanie i zastosowanie nowej metody do genotypowania powszechnego polimorfizmu liczby kopii (CNP) w genomie człowieka".

Od początku swojej działalności mgr inż. Małgorzata Marcinkowska-Swojak wykazywała się dużą samodzielnością i przedsiębiorczością, a wraz z postępem czasu zdobyła dużą wiedzę i doświadczenie w zakresie realizowanego tematu oraz dojrzałość naukową, przejawiającą się umiejętnością umieszczenia wyników swoich badań w szerokim kontekście ogólnego stanu wiedzy w dziedzinie genetyki i genomiki.

Jako, że od początku byłem opiekunem naukowym mgr inż. Małgorzaty Marcinkowskiej-Swojak, a od początku 2013 roku również jej promotorem, jestem głównym autorem wszystkich publikacji przedkładanych przez doktorantkę w ramach rozprawy doktorskiej, i mogę dobrze ocenić jej udział w poszczególnych pracach. We wszystkich przypadkach rola mgr inż. Małgorzaty Marcinkowskiej-Swojak była znacząca (przynajmniej 50%), wykonywała ona wszystkie eksperymenty, wszystkie lub większość analiz oraz brała udział w przygotowaniu manuskryptów. Moja rola polegała na zaplanowaniu badań, pozyskaniu środków, koordynacji badań i przygotowaniu manuskryptów, wykonywałem też, równolegle z doktorantką, niektóre analizy.

Poniżej przedstawiam zakres prac wykonanych przez mgr inż. Małgorzatę Marcinkowską-Swojak oraz mój udział w poszczególnych publikacjach.

- Marcinkowska M, Wong KK, Kwiatkowski DJ, Kozlowski P.
  Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example.
  TheScientificWorldJournal. 2010; 10: 2003-2018.

  Mgr inż. Małgorzata Marcinkowska-Swojak wykonała w powyższej publikacji wszystkie eksperymenty oraz analizy, zaplanowała większość sond MLPA do testu EGFRmut+, przygotowała ilustracje, brała udział w przygotowaniu manuskryptu oraz przygotowała wszystkie materiały suplementarne. Wykonywane przez doktorantkę testy przyczyniły się do stworzenia standardów oraz walidacji poszczególnych kroków przedstawionego w publikacji protokołu.

  Mój udział w tej publikacji oceniam na około 40%. Polegał on na zaplanowaniu badań (wstępną koncepcję testu EGFR opracowałem wspólnie z prof. Davidem Kwiatkowskim z Harvard University w Bostonie), nadzorowaniu i koordynacji eksperymentów i analiz oraz wykonaniu większości prac związanych z przygotowaniem manuskryptu. Nadzorowałem pracę doktorantki oraz zapoznawałem ją z zagadnieniami genetyki będącymi tłem oraz bezpośrednim przedmiotem publikacji.

- Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozlowski P.
  Copy number variation of microRNA genes in the human genome.
  BMC Genomics. 2011; 12: 183.

  Mgr inż. Małgorzata Marcinkowska-Swojak wykonała w powyższej publikacji większość analiz genetycznych i genomicznych (z wyjątkiem analiz konserwatywności sekwencji miRNA wykonanych przez dr Macieja Szymańskiego z UAM w Poznaniu), wykonała przegląd literatury dotyczącej funkcji badanych miRNA, brała udział w przygotowaniu ilustracji i manuskryptu oraz przygotowała materiały suplementarne do publikacji.

  Mój udział w tej publikacji oceniam na około 30%. Polegał on na zaplanowaniu oraz egzekwowaniu badań oraz wyborze narzędzi i testów statystycznych, nadzorowaniu i koordynacji badań oraz przygotowaniu manuskryptu.

Nadzorowałem pracę doktorantki oraz zapoznawałem ją z zagadnieniami genetyki i genomiki będącymi tłem oraz bezpośrednim przedmiotem publikacji.

- Marcinkowska M, Kozłowski P.
  The influence of copy number polymorphism on the human phenotype.
  Postepy Biochem. 2011; 57: 240-248.

  Wspólnie z mgr inż. Małgorzatą Marcinkowską-Swojak przygotowaliśmy powyższy artykuł przeglądowy. Udział każdego z nas szacuję na około 50%. Praca polegała na gromadzeniu oraz przeglądaniu literatury, prezentowaniu licznych przykładów CNV oraz metod ich analizy w celu wybrania najbardziej reprezentatywnych przykładów oraz przygotowaniu ilustracji i manuskryptu. We wszystkich etapach przygotowania artykułu uczestniczyliśmy w mniej więcej porównywalnej części. Praca nad tym artykułem przeglądowym była doskonałą okazją dla doktorantki do zgłębienia szerokiego aspektu zjawiska zmienności liczby kopii oraz jego miejsca i znaczenia we współczesnej genetyce i genomice.

- Marcinkowska-Swojak M, Uszczynska B, Figlerowicz M, Kozlowski P.
  An MLPA-based strategy for discrete CNV genotyping: CNV-miRNAs as an example.
  Hum Mutat. 2013; 34: 763-773.

  Mgr inż. Małgorzata Marcinkowska-Swojak wykonała w niniejszej pracy wszystkie eksperymenty, zaplanowała wszystkie testy i sondy MLPA, przygotowała narzędzia i wykonała niemal wszystkie analizy, przygotowała materiały suplementarne i ryciny oraz brała udział w przygotowaniu manuskryptu. Jedyną analizą, której nie wykonała mgr inż. Małgorzata Marcinkowska, jest analiza automatycznego rozpoznawania genotypów liczby kopii z użyciem algorytmu EM (ang. Expectation Maximization), przedstawiona na Suplementarnej Rycinie S6. Analiza ta została wykonana przez mgr Barbarę Uszczyńską, doktorantkę IChB PAN.

  Mój udział w niniejszej publikacji oceniam na około 40%. Polegał on na zaplanowaniu oraz egzekwowaniu badań oraz wyborze narzędzi i testów statystycznych, nadzorowaniu i koordynacji badań oraz przygotowaniu manuskryptu. Nadzorowałem pracę doktorantki oraz zapoznawałem ją z zagadnieniami genetyki i genomiki będącymi tłem oraz bezpośrednim

przedmiotem publikacji. Zapoznawałem doktorantkę z narzędziami genetycznymi i genomicznymi stosowanymi w niniejszej pracy.

Proszę o kontakt, w przypadku jakichkolwiek pytań dotyczących powyżej przedstawionych oświadczeń.

Piotr Kozłowski

BRIGHAM & WOMEN'S HOSPITAL -- HARVARD MEDICAL SCHOOL

David J. Kwiatkowski, M.D. Ph. D.
Professor of Medicine, HMS
Senior Physician, BWH
Leader, Cancer Genetics Program, DFHCC
NIH NINDS Javits Neuroscience Investigator

5/25/13

MAILING ADDRESS:
1 Blackfan Circle, 6-216
Brigham & Women's Hospital
Boston, MA 02115
(617) 355-9005
Fax (617) 355-9016
Email: dk@rics.bwh.harvard.edu

To Whom It May Concern:

Regarding PhD thesis of Malgorzata Marcinkowska-Swojak.

This letter is to certify my contribution to the following publication of which I was a co-author.

Marcinkowska M, Wong KK, Kwiatkowski DJ, Kozlowski P.
Design and generation of MLPA probe sets for combined copy number and small-mutation analysis of human genes: EGFR as an example.
TheScientificWorldJournal. 2010 Oct 12;10:2003-18.

I participated in conceiving the idea of MLPA test for copy number (amplification) and small-mutation analysis of EGFR gene. I also provided DNA samples used in this study. My colleague in Boston, Dr. Kwok Wong made similar minimal contributions. I estimate my overall contribution to this paper at 15%, and Dr. Wong's at 5%.

Feel free to contact me if further information is required.

Sincerely,

Sincerely,
David J. Kwiatkowski

dr Maciej Szymański Poznań, 11.06.2013r.

OŚWIADCZENIE

Oświadczam, iż mój udział w publikacji

Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozlowski P. „Copy number variation of microRNA genes in the human genome" BMC Genomics 2011; 12:183.

polegał na wykonaniu obliczeniowej analizy regionów CNV w genomie człowieka. Brałem również udział w przygotowaniu manuskryptu. Swój całkowity udział w tej publikacji szacuję na 15 %.

Z poważaniem

Poznań, 12.06.2013r.

OŚWIADCZENIE

Dotyczy udziału w publikacji:

Marcinkowska M, Szymanski M, Krzyzosiak WJ, Kozlowski P. „Copy number variation of microRNA genes in the human genome" BMC Genomics 2011; 12:183.

Jako współautor powyższej publikacji, oświadczam iż brałem udział w przygotowaniu projektu badań, wchodzących w skład tej publikacji oraz uczestniczyłem w końcowym sprawdzaniu jej manuskryptu. Swój całkowity udział w powyższej publikacji szacuję na 3%.

Z poważaniem

prof. dr hab. Włodzimierz J. Krzyżosiak

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ
ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
fax: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327
NIP 777-00-02-062

Poznań, 10.06.2013r.

ZAŚWIADCZENIE

Dotyczy:

Marcinkowska-Swojak M, Uszczynska B, Figlerowicz F, Kozlowski P. „An MLPA-based strategy for discrete CNV genotyping: CNV-miRNAs as an example." Human Mutation 2013; 34, 763-773

Na potrzeby powyższej publikacji, której jestem współautorem, opracowałam algorytm EM (Expectation Maximization), umożliwiający analizę klastrowania i obiektywne genotypowanie zmienności liczby kopii w badanych próbkach. Swój całkowity udział w powyższej publikacji szacuję na około 5%.

Z poważaniem

Barbara Uszczyńska

POLSKA AKADEMIA NAUK

INSTYTUT CHEMII BIOORGANICZNEJ

ul. Noskowskiego 12/14, 61-704 Poznań, Polska
tel.: +48-61 centrala 852 85 03, sekretariat 852 89 19
fax: +48-61 852 05 32 e-mail: office@ibch.poznan.pl
Regon 000849327

Poznań, 12.06.2013r.

OŚWIADCZENIE

Dotyczy:

Marcinkowska-Swojak M, Uszczynska B, Figlerowicz F, Kozlowski P. "An MLPA-based strategy for discrete CNV genotyping: CNV-miRNAs as an example" Human Mutation 2013; 34, 763-773

Oświadczam, iż brałem udział w przygotowaniu koncepcji badań, stanowiących podstawę powyższej publikacji oraz uczestniczyłem w przygotowaniu jej tekstu do druku. Mój całkowity udział w publikacji szacuję na 5 %.

Z poważaniem

prof. dr hab. Marek Figlerowicz