

From: [Young, Daniel W.](#)
To: ["Greg Porreca"](#)
Cc: [Waller, Patrick](#)
Subject: RE: update with example
Date: Friday, May 15, 2009 4:17:38 PM
Attachments: [G0814.70003US00 - PROVISIONAL PT APP.DOC](#)
[G0814.70003US00.DOC](#)

Hi Greg,

Attached is the updated draft, which includes the new example below, and figures. Please notice claim 9, which includes the mutation detection reaction we discussed.

Thanks
Dan

-----Original Message-----

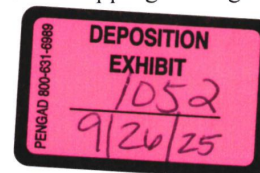
From: Greg Porreca [mailto:gporreca@gsgenetics.com]
Sent: Friday, May 15, 2009 2:55 PM
To: Young, Daniel W.
Subject: update with example

Dan-

Please add the following sentence after sentence 3 of paragraph 2 of the Summary:
Additionally, one technique might bias in favor of a locus, while another might bias against a locus, thus altering the overall coverage of that locus.

Example 1: Use of MIPs, hybridization, and mutation-detection MIPs to genotype a set of 1000 targets which could possibly include any of 50 specific known point mutations. The first step is to perform separate MIP, hybridization, and mutation-detection MIP reactions. Perform a MIP capture reaction essentially as described in Turner et al 2009, Nature methods 6:315-6; design a set of probes, where each probe flanks one of the 1000 targets. Separately, perform a hybridization enrichment reaction using the Agilent SureSelect procedure. Prior to selection, the genomic DNA to be enriched should be converted into a shotgun sequencing library using Illumina's 'Fragment Library' kit and protocols; use Agilent's web interface to design a set of probes which will hybridize to the targets. Separately, design a set of probes which will form MIPs only if mutations are present. Each probe will have its 3'-most base identity specific for a single known mutation. Perform the MIP reaction with this set of probes to selectively detect the presence of any mutant alleles.

Once all three reactions have been performed, the two MIP reactions are combined into a single tube, and run as one sample on the next-generation DNA sequencing instrument. The hybridization-enriched reaction is run as a separate sample on the next-generation DNA sequencing instrument. Reads from each 'sample' are combined by a software algorithm which forms a consensus diploid genotype at each position in the target set by evaluating the total coverage at each position, the origin of each read in that total coverage, the quality score of each individual read, and the presence (or absence) of any reads derived from mutation-specific MIPs overlapping the region.



ML-IPR-000789

Gregory Porreca, Ph.D.
Director of Technology
Good Start Genetics

8 St Mary's Street, Suite 621
Boston, MA 02215
P: 617-358-6004
F: 617-358-5677

ML-IPR-000790

METHODS AND COMPOSITIONS FOR ENHANCING THE SENSITIVITY OF NUCLEIC ACID ANALYSIS

Field of the Invention

The invention relates to methods for increasing the sensitivity of nucleic acid detection assays.

Background of the Invention

There are currently many genomic assays that utilize next-generation (polony-based) sequencing to generate data, including genome resequencing, RNA-seq for gene expression, bisulphite sequencing for methylation, and Immune-seq, among others. In order to make quantitative measurements (including genotype calling), these methods utilize the counts of sequencing reads of a given genomic locus as a proxy for the representation of that sequence in the original sample of nucleic acids. The majority of these techniques require a preparative step to construct a high-complexity library of DNA molecules that is representative of a sample of interest. Current assays use one of several alternative nucleic acid preparative techniques (e.g., amplification, for example PCR-based amplification; sequence-specific capture, for example, using immobilized capture probes; or target capture into a circularized probe) followed by a sequence analysis step. In order to reduce errors associated with the unpredictability (stochastic nature) of nucleic acid capture and sequence analysis techniques, current methods involve oversampling a target nucleic acid preparation in order to increase the likelihood that all sequences that are present in the original nucleic acid sample will be represented in the final sequence data. For example, due to sampling variations (stochastic variations) a genomic sequencing library may contain an over- or under-representation of particular sequences from a source nucleic acid sample (e.g., genome preparation) as a result of stochastic variations in the library construction process. Such variations can be particularly problematic when they result in target sequences from a genome being absent or undetectable in a sequencing library. For example, an under-representation of particular allelic sequences (e.g., heterozygotic alleles) from a genome in a sequencing library can result in an apparent homozygous representation in a sequencing library. As most downstream sequencing library quantification techniques depend on stochastic counting processes, these problems have typically been addressed by sampling enough (over-

ML-IPR-000791

sampling) to obtain a minimum number of observations necessary to make statistically significant decisions. However, the strategy of oversampling is generally limited to elimination of low-count Poisson noise, and the approach wastes resources and increases the expense required to perform such experiments. Moreover, oversampling can result in a reduced statistical confidence in certain conclusions (e.g., diagnostic calls) based on the data. Accordingly, new approaches are needed for overcoming variations in sequencing library preparatory methods.

Brief Summary of the Invention

Aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample. According to the invention, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases.

According to the invention, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity issue, because each technique has a different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

In some embodiments, the invention provides a method of increasing the sensitivity of a nucleic acid detection assay by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, assaying the sequences obtained in both first and second nucleic acid preparations, and using the sequence information from both first and second nucleic acid preparations to determine the genotype of the target nucleic acid in the biological sample, wherein the first and second preparative

methods have different systematic sequence biases. In some embodiments, the first and second nucleic acid preparations are combined prior to performing a sequence assay. In some embodiments, separate sequence assays are performed on the first and second nucleic acid preparations and the sequence information from both assays are combined to determine the genotype of the target nucleic acid in the biological sample. In some embodiments, the first preparative method is a PCR-based, a hybridization-based, or a circular probe-based preparative method.

In PCR-based (also referred to herein as amplification-based) preparative methods, genomic loci (target nucleic acids) are isolated directly by means of a polymerase chain reaction that selectively amplifies each locus using a pair of oligonucleotide primers. It is to be understood that primers will be sufficiently complementary to the target sequence to hybridize with and prime amplification of the target nucleic acid. Any one of a variety of art known methods may be utilized for primer design and synthesis. One or both of the primers may be perfectly complementary to the target sequence. Degenerate primers may also be used. Primers may also include additional nucleic acids that are not complementary to target sequences but that facilitate downstream applications, including for example restriction sites and identifier sequences (e.g., source sequences). PCR based methods include amplification of a single target nucleic acid and multiplex amplification (amplification of multiple target nucleic acids in parallel).

Hybridization-based preparative methods involve selectively immobilizing target nucleic acids for further manipulation. It is to be understood that one or more oligonucleotides (immobilization oligonucleotides), which in some embodiments may be from 15 to 170 nucleotides in length, are used which hybridize along the length of a target region of a genetic locus to immobilize it. In some embodiments, immobilization oligonucleotides are either immobilized before hybridization is performed (e.g., Roche/Nimblegen 'sequence capture'), or are prepared such that they include a moiety (e.g. biotin) which can be used to selectively immobilize the target nucleic acid after hybridization by binding to e.g., streptavidin-coated microbeads (e.g. Agilent 'SureSelect').

Circularization selection-based preparative methods selectively convert each region of interest into a covalently-closed circular molecule which is then isolated by removal (usually enzymatic, e.g. with exonuclease) of any non-circularized linear nucleic acid. In Oligonucleotide probes are designed which have ends that flank the region of interest. The probes are allowed to hybridize to the genomic target, and enzymes are used to first (optionally) fill in any gap between probe ends and second ligate the probe closed. Following

ML-IPR-000793

circularization, any remaining (non-target) linear nucleic acid is removed, resulting in isolation of target nucleic acid. Circularization selection-based preparative methods include molecular inversion probe capture reactions. In some embodiments, molecular inversion probe capture of a target nucleic acid of the nucleic acid preparation is indicative of the presence of a polymorphism in the target nucleic acid.

A variety of methods may be used to evaluate and compare bias profiles of each preparative technique. Next-generation sequencing may be used to quantitatively measure the abundance of each isolated target nucleic acid obtained from a certain preparative method. This abundance may be compared to a control abundance value (e.g., a known starting abundance of the target nucleic acid) and/or with an abundance determined through the use of an alternative preparative method. For example, a set of target nucleic acids may be isolated by one or more of the three preparative methods; the target nucleic acid may be observed x times using the amplification technique, y times using the hybridization enrichment technique, and z times using the circularization selection technique. A pairwise correlation coefficient may be computed between each abundance value (e.g. x and y, x and z, and y and z) to assess bias in nucleic acid isolation between pairs of preparative methods. Since the mechanisms of isolation are different in each approach, the abundances will usually be different and largely uncorrelated with each other.

In some embodiments, the second preparative method is a PCR-based, a hybridization-based, or a circular probe-based preparative method, provided that the second method is different from the first method.

In some embodiments, the invention provides a method of obtaining a nucleic acid preparation that is representative of a target nucleic acid in a biological sample by obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample, obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample, and combining the first and second nucleic acid preparations to obtain a combined preparation that is representative of the target nucleic acid in the biological sample.

In some embodiments of any of the methods described herein, a third preparation the target nucleic acid is obtained using a third preparative method that is different from the first and second preparative methods, wherein the first, second, and third preparative methods all have different systematic sequence biases. In some embodiments of any of the methods described herein, the different preparative methods are used for a plurality of different loci in

ML-IPR-000794

the biological sample to increase the sensitivity of a multiplex nucleic acid analysis. In some embodiments, the target nucleic acid has a sequence of a gene selected from Table 1.

Aspects of the invention also provide compositions, kits, devices, and analytical methods for increasing the sensitivity of nucleic acid assays. Aspects of the invention are particularly useful for increasing the confidence level of genotyping analyses. However, aspects of the invention may be used in the context of any suitable nucleic acid analysis, for example, but not limited to, a nucleic acid analysis that is designed to determine whether more than one sequence variant is present in a sample.

In some embodiments, a diagnosis, prognosis, or disease risk assessment is provided to a subject based on a genotype determined for that subject at one or more genetic loci (e.g., based on the analysis of a biological sample obtained from that subject). In some embodiments, an assessment is provided to a couple, based on their respective genotypes at one or more genetic loci, of the risk of their having one or more children having a genotype associated with a disease or condition (e.g., a homozygous or heterozygous genotype associated with a disease or condition). In some embodiments, a subject or a couple may seek genetic or reproductive counseling in connection with a genotype determined according to embodiments of the invention. The methods disclosed herein are useful for any application where sensitivity is important. For example, detection of cancer mutations in a heterogenous tissue sample, detection of mutations in maternally-circulating fetal DNA, and detection of mutations in cells isolated during a preimplantation genetic diagnostic procedure.

According to some aspects of the invention, methods of detecting a polymorphism in a nucleic acid in a biological sample are provided. In some embodiments, the methods comprise obtaining a nucleic acid preparation using a preparative method (e.g., any of the preparative methods disclosed herein) on a biological sample, and performing a molecular inversion probe capture reaction on the nucleic acid preparation, wherein a molecular inversion probe capture (e.g., using a mutation-detection MIP) of a target nucleic acid of the nucleic acid preparation is indicative of the presence of a mutation (polymorphism) in the target nucleic acid, optionally wherein the polymorphism is selected from Table 2.

In some embodiments of the methods disclosed herein, the target nucleic acid has a sequence of a gene selected from Table 1.

Brief Description of the Drawings

ML-IPR-000795

Figure 1 is a schematic example of an embodiment of the invention showing MIPs plus hybridization enrichment, plus optionally extra MIPs targeted to specific known, common disease-associated loci, e.g., to detect the presence of a polymorphism in a target nucleic acid.

Figure 2 is a graph of per-target abundance with MIP capture. In this graph, bias largely drives the heterozygote error rate, since targets which are less abundant here are less likely to be covered in sufficient depth during sequencing to adequately sample both alleles. This is from Turner et al 2009, Nature methods 6:315-6. Hybridization enrichment results in a qualitatively similar abundance distribution, but the abundance of a given target is likely not correlated between the two methods.

Figure 3 is a graph of correlation between two MIP capture reactions from Ball et al 2009, Nature biotechnology 27:361-8. Each point represents the target abundance in replicate 1 and replicate 2. Pearson correlation $r=0.956$. This indicates that MIP capture reproducibly biases targets to specific abundances. Hybridization enrichment is similarly correlated from one capture to the next.

Detailed Description of the Invention

Aspects of the invention relate to basing a nucleic acid sequence analysis on results from two or more different nucleic acid preparatory techniques that have different systematic biases in the types of nucleic acids that they sample. According to the invention, different techniques have different sequence biases that are systematic and not simply due to stochastic effects during nucleic acid capture or amplification. Accordingly, the degree of oversampling required to overcome variations in nucleic acid preparation needs to be sufficient to overcome the biases.

According to the invention, different techniques have different characteristic or systematic biases. For example, one technique may bias a sample analysis towards one particular allele at a genetic locus of interest, whereas a different technique would bias the sample analysis towards a different allele at the same locus. Accordingly, the same sample may be identified as being different depending on the type of technique that is used to prepare nucleic acid for sequence analysis. This effectively represents a sensitivity issue, because each technique has a different relative sensitivities for polymorphic sequences of interest.

According to aspects of the invention, the sensitivity of a nucleic acid analysis can be increased by combining the sequences from different nucleic acid preparative steps and using the combined sequence information for a diagnostic assay (e.g., for a making a call as to whether a subject is homozygous or heterozygous at a genetic locus of interest).

Currently, the ability of DNA sequencing to detect mutations is limited by the ability of the upstream sample isolation (amplification, enrichment, capture, etc) methods to reliably isolate the locus of interest. If one wishes to make heterozygote basecalls for a diploid genome (e.g. a human sample presented for molecular diagnostic sequencing), it is imperative that the isolation method produces near- or perfectly- uniform amounts of the two alleles to be sequenced.

Sample preparative methods may fall into three classes: 1) single- or several- target amplification (e.g. uniplex PCR, 'multiplex' PCR), 2) multi-target hybridization enrichment (e.g. Agilent SureSelect 'hybrid capture' [Gnirke et al 2009, *Nature methods* 27:182-9], Roche/Nimblegen 'sequence capture' [Hodges et al 2007, *Nature genetics* 39:1522-7], and 3) multi-target circularization selection (e.g. molecular inversion probes or padlock probes, [Porreca et al 2007, *Nature methods* 4:931-6, Turner et al 2009, *Nature methods* 6:315-6], 'selectors' [Dahl et al 2005, *Nucleic acids research* 33:e71]). Each of these methods can result in a pool of isolated product that does not adequately represent the input abundance distribution. For example, the two alleles at a heterozygous position can become skewed far from their input 50:50 ratio to something that results in a missed basecall during downstream sequencing. For example, if the ratio was skewed from 50:50 to 10:90, and the sample was sequenced to 10X average coverage, there is a high probability that one of the two alleles would not be observed once in the ten sequencing reads. This would reduce the sensitivity of the sequencing method by converting a heterozygous position to homozygous (where potentially the 'mutant' allele was the one not observed).

The methods disclosed herein are based, in part, on the discovery that certain classes of isolation methods, such as those described above, result in different modes of bias in the product. The disclosure provide methods for increasing the sensitivity of the downstream sequencing by using a combination of multiple isolation methods (one or more from at least two of the classes above) for a sample. This is particularly important in molecular diagnostics where high sensitivity is required to minimize the chances of 'missing' a disease-associated mutation. For example, given a nominal false-negative error rate of 1×10^{-3} for sequencing following circularization selection, and a false-negative error rate of 1×10^{-3} for sequencing following hybridization enrichment, one can achieve a final false-negative rate of

ML-IPR-000797

1×10^{-6} by performing both techniques on the sample (assuming failures in each method are fully independent). For a recessive disease with carrier frequency of 0.1, caused by a single fully-penetrant mutant allele, the number of missed carrier diagnoses would decrease from 1000 per million patients tested to 1 per million patients tested. Furthermore, if the testing was used in the context of prenatal carrier screening, the number of affected children born as a result of missing the carrier call in one parent would decrease from 25 per million to 25 per billion born.

Additionally, the disclosure provides combinations of preparative methods to effectively increase sequencing coverage in regions containing disease-associated alleles. Since heterozygote error rate is largely tied to both deviations from 50:50 allele representation, and in the case of next-generation DNA sequencing deviations from average abundance (such that less abundant isolated targets are more likely to be undersampled at one or both alleles), selectively increasing coverage in these regions will also selectively increase sensitivity. Furthermore, MIPs that detect presence or absence of specific known disease-associated mutations can be used to increase sensitivity selectively. These MIPs would have a targeting arm whose 3' -most region is complementary to the expected mutation, and has a fill-in length of 0 or more bp. Thus, the MIP will form only if the mutation is present, and its presence will be detected by sequencing.

Additionally, the algorithms disclosed herein may be used to determine base identity with varying levels of stringency depending on whether the given position has any known disease-associated alleles. Stringency can be reduced in such positions by decreasing the minimum number of observed mutant reads necessary to make a consensus basecall. This will effectively increase sensitivity for mutant allele detection at the cost of decreased specificity.

In some aspects, the disclosure provides methods for analyzing a plurality of target sequences which are genetic loci or portions of genetic loci (e.g., a genetic locus of Table 1). The genetic loci may be analyzed by sequencing to obtain a genotype at one or more polymorphisms (e.g., SNPs). Exemplary polymorphisms are disclosed in Table 2. The skilled artisan will appreciate that other polymorphisms are known in the art and may be identified, for example, by querying the Entrez Single Nucleotide Polymorphism database, for example, by searching with a GeneID from Table 1.

The mutations listed in Table 2 are documented polymorphisms in several disease-associated genes (CFTR is mutated in cystic fibrosis, GBA is mutated in Gaucher disease, ASPA is mutated in Canavan disease, HEXA is mutated in Tay Sachs disease). The

ML-IPR-000798

polymorphisms are of several types: insertion/deletion polymorphisms which will cause frameshifts (and thus generally interrupt protein function) unless the insertion/deletion length is a multiple of 3bp, and substitutions which can alter the amino acid sequence of the protein and in some cases cause complete inactivation by introduction of a stop codon.

The methods disclosed here are also useful for diagnosis of one or more heritable disorders. In some embodiments, a heritable disorder that may be diagnosed with the methods disclosed herein is a genetic disorder that is prevalent in the Ashkenazi Jewish population. In some embodiments, the heritable disorders are selected from: 21-Hydroxylase-Deficient Congenital Adrenal Hyperplasia; ABCC8-Related Hyperinsulinism; Alpha-Thalassemia, includes Constant Spring, & MR associated; Arylsulfatase A Deficiency-Metyachromatic Leukodystrophy; Biotinidase Deficiency-Holocarboxylase Synthetase Deficiency; Bloom's Syndrome; Canavan Disease; CFTR-Related Disorders-cystic fibrosis; Citrullinemia Type I; Combined MMA & Homocystinuria-dblC; Dystrophinopathies (DMD & BMD); Familial Dysautonomia; Fanconi Anemia-FANCC; Galactosemia-Classical: Galactokinase Deficiency & Galactose Epimerase Deficiency; Gaucher Disease; GJB2-Related DFNB 1 Nonsyndromic Hearing Loss and Deafness; Glutaric acidemia Type 1; Hemoglobinopathies beta-chain disorders; Glycogen Storage Disease Type 1A; Maple Syrup Urine Disease; Types 1A, 1B, 2, 3; Medium Chain Acyl-Coenzyme A; Dehydrogenase Deficiency-MCADD; Methylmalonic Acidemia; Mucopolidosis IV; Nemaline Myopathy; Nieman-Pick Type A-Acid Sphingomyelinase Deficiency; Non-Ketotic Hyperglycinemia-Glycine Encephalopathy; Ornithine Transcarbamylase Deficiency; PKU Phenylalanine Hydroxylase Deficiency; Propionic Acidemia; Short Chain Acyl-CoA Dehydrogenase Deficiency-SCADD; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy (SMN1)-SMA; Tay Sachs-HexA Deficiency; Usher Syndrome-Type I (Type IB, Type IC, Type ID, Type IF, Type IG) ; X-Linked Mental Retardation ARX-Related Disorders; X-Linked Mental Retardation with Cerebellar Cypoplasia and distinctive Facial Appearance; X-Linked Mental Retardation; includes 9, 21, 30, 46, 58, 63, 88, 89; X-linked mental retardation: FM1-Related Disorders-FRXA, Fragile X MR; X-linked SMR: Renpenning Syndrome 1; Zellweger Spectrum disorders - Peroxisomal Bifunctional Enzyme Deficiencies including Zellweger, NALD, infantile Refsums.

Table 1: Target Nucleic Acids

ML-IPR-000799

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
CYP21A2	1589	cytochrome P450, family 21, subfamily A, polypeptide 2	CAH1; CPS1; CA21H; CYP21; CYP21B; P450c21B; MGC150536; MGC150537; CYP21A2	201910	NG_008337.1	6p21.3
ABCC8	6833	ATP-binding cassette, sub-family C (CFTR/MRP), member 8	HI; SUR; HHF1; MRP8; PHH1; SUR1; ABC36; HRINS; TNDM2; ABCC8	600509	NG_008867.1	11p15.1
ATRX	546	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae)	SHS; XH2; XNP; ATR2; SFM1; RAD54; MRXHF1; RAD54L; ZNF-HX; MGC2094; ATRX	300032	NG_008838.1	Xq13.1-q21.1
ARSA	410	arylsulfatase A	MLD; ARSA	607574	NG_009260.1	22q13.31-qter; 22q13.33
PSAP	5660	Prosaposin	GLBA; SAPI; FLJ00245; MGC110993; PSAP	176801	NG_008835.1	10q21-q22
BTBD	686	Biotinidase	BTBD	609019	NG_008019.1	3p25
HLCS	3141	holocarboxylase synthetase (biotin-(propionyl-Coenzyme A-carboxylase (ATP-hydrolysing)) ligase)	HCS; HLCS	609018	NC_000021.7	21q22.1; 21q22.13
BLM	641	Bloom syndrome, RecQ helicase-like	BS; RECQ2; RECQL2; RECQL3; MGC126616; MGC131618; MGC131620; BLM	604610	NG_007272.1	15q26.1
ASPA	443	aspartoacylase (Canavan disease)	ASP; ACY2; ASPA	608034	NG_008399.1	17pter-p13
CFTR	1080	cystic fibrosis transmembrane conductance regulator (ATP-binding cassette sub-family C, member 7)	CF; MRP7; ABC35; ABCC7; CFTR/MRP; TNR-CFTR; dJ760C5.1; CFTR	602421	NC_000007.12	7q31.2
ASS1	445	argininosuccinate synthetase 1	ASS; CTLN1; ASS1	603470	NG_011542.1	9q34.1
MMACHC	25974	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria	cblC; FLJ25671; DKFZp564I122; RP11-291L19.3; MMACHC	609831	NC_000001.9	1p34.1

ML-IPR-000800

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
IKBKAP	8518	inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	FD; DYS; ELP1; IKAP; IKI3; TOT1; FLJ12497; DKFZp781H1425; IKBKAP	603722	NG_008788.1	9q31
FANCC	2176	Fanconi anemia, complementation group C	FA3; FAC; FACC; FLJ14675; FANCC	227645	NG_011707.1	9q22.3
GALK1	2584	galactokinase 1	GK1; GALK; GALK1	604313	NG_008079.1	17q24
GALT	2592	galactose-1-phosphate uridylyltransferase	GALT	606999	NC_000009.10	9p13
GALE	2582	UDP-galactose-4-epimerase	SDR1E1; FLJ95174; FLJ97302; GALE	606953	NG_007068.1	1p36-p35
GBA	2629	glucosidase, beta; acid (includes glucosylceramidase)	GCB; GBA1; GLUC; GBA	606463	NG_009783.1	1q21
GJB2	2706	gap junction protein, beta 2, 26kDa	HID; KID; PPK; CX26; DFNA3; DFNB1; NSRD1; DFNA3A; DFNB1A; GJB2	121011	NG_008358.1	13q11-q12
GCDH	2639	glutaryl-Coenzyme A dehydrogenase	GCD; ACAD5; GCDH	608801	NG_009292.1	19p13.2
G6PC	2538	glucose-6-phosphatase, catalytic subunit	G6PT; GSD1; GSD1a; MGC163350; G6PC	232200	NG_011808.1	17q21
HBB	3043	hemoglobin, beta	CD113t-C; beta-globin; HBB	141900	NG_000007.3	11p15.5
BCKDHA	593	branched chain keto acid dehydrogenase E1, alpha polypeptide	MSU; MSUD1; OVD1A; BCKDE1A; FLJ45695; BCKDHA	608348	NC_000019.8	19q13.1-q13.2
BCKDHB	594	branched chain keto acid dehydrogenase E1, beta polypeptide	E1B; FLJ17880; dJ279A18.1; BCKDHB	248611	NG_009775.1	6q13-q15
DBT	1629	dihydrolipoamide branched chain transacylase E2	E2; E2B; BCATE2; MGC9061; DBT	248610	NG_011852.1	1p31
DLD	1738	dihydrolipoamide dehydrogenase	E3; LAD; DLDH; GCSL; PHE3; DLD	238331	NG_008045.1	7q31-q32
ACADM	34	acyl-Coenzyme A dehydrogenase, C-4 to C-12 straight chain	MCAD; ACAD1; MCADH; FLJ18227; FLJ93013; FLJ99884;	607008	NG_007045.1	1p31

ML-IPR-000801

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
			ACADM			
MMAA	166785	methylmalonic aciduria (cobalamin deficiency) cblA type	cblA; MGC120010; MGC120011; MGC120012; MGC120013; MMAA	607481	NG_007536.1	4q31.22
MMAB	326625	methylmalonic aciduria (cobalamin deficiency) cblB type	ATR; cblB; MGC20496; MMAB	607568	NG_007096.1	12q24
MUT	4594	methylmalonyl Coenzyme A mutase	MCM; MUT	609058	NG_007100.1	6p12.3
MCOLN1	57192	mucolipin 1	ML4; MLIV; MST080; TRPML1; MSTP080; TRP-ML1; TRPM-L1; MCOLN1	605248	NC_000019.8	19p13.3-p13.2
ACTA1	58	actin, alpha 1, skeletal muscle	ACTA; ASMA; CFTD; MPFD; NEM1; NEM2; NEM3; CFTD1; CFTDM; ACTA1	102610	NG_006672.1	1q42.13
TPM3	7170	tropomyosin 3	TM3; TRK; NEM1; TM-5; TM30; TM30nm; TPMsk3; hscp30; MGC3261; FLJ41118; MGC14582; MGC72094; OK/SW-cl.5; TPM3	191030	NG_008621.1	1q21.2
TNNT1	7138	troponin T type 1 (skeletal, slow)	ANM; TNT; STNT; TNTS; FLJ98147; MGC104241; TNNT1	191041	NG_011829.1	19q13.4
NEB	4703	nebulin	NEM2; NEB177D; FLJ11505; FLJ36536; FLJ39568; FLJ39584; DKFZp686C1456; NEB	161650	NG_009382.1	2q22
SMPD1	6609	sphingomyelin phosphodiesterase 1, acid lysosomal	ASM; NPD; SMPD1	607608	NG_011780.1	11p15.4-p15.1
GLDC	2731	glycine dehydrogenase (decarboxylating)	GCE; NKH; GCSP; HYGN1; MGC138198; MGC138200; GLDC	238300	NC_000009.10	9p22
GCSH	2653	glycine cleavage system protein H (aminomethyl carrier)	GCE; NKH; GCSH	238330	NC_000016.8	16q23.2
AMT	275	aminomethyltransferase	GCE; NKH; GCST; AMT	238310	NC_000003.1	3p21.2-

ML-IPR-000802

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
					0	p21.1
OTC	5009	ornithine carbamoyltransferase	OCTD; MGC129967; MGC129968; MGC138856; OTC	300461	NG_008471.1	Xp21.1
PAH	5053	phenylalanine hydroxylase	PH; PKU; PKU1; PAH	612349	NG_008690.1	12q22-q24.2
DHPR	5860	quinoid dihydropteridine reductase	DHPR; PKU2; SDR33C1; FLJ42391; QDPR	612676	NG_008763.1	4p15.31
PTS	5805	6-pyruvoyltetrahydropterin synthase	PTPS; FLJ97081; PTS	261640	NG_008743.1	11q22.3-q23.3
PCCA	5095	propionyl Coenzyme A carboxylase, alpha polypeptide	PCCA	232000	NG_008768.1	13q32
PCCB	5096	propionyl Coenzyme A carboxylase, beta polypeptide	DKFZp451E113; PCCB	232050	NG_008939.1	3q21-q22
ACADS	35	acyl-Coenzyme A dehydrogenase, C-2 to C-3 short chain	SCAD; ACAD3; ACADS	606885	NG_007991.1	12q22-qter
DHCR7	1717	7-dehydrocholesterol reductase	SLOS; DHCR7	602858	NC_000011.8	11q13.2-q13.5
SMN1	6606	survival of motor neuron 1, telomeric	SMA; SMN; SMA1; SMA2; SMA3; SMA4; SMA@; SMNT; BCD541; T-BCD541; SMN1	600354	NG_008691.1	5q13
HEXA	3073	hexosaminidase A (alpha polypeptide)	TSD; MGC99608; HEXA	606869	NG_009017.1	15q23-q24
MYO7A	4647	myosin VIIA	DFNB2; MYU7A; NSRD2; USH1B; DFNA11; MYOVIIA; MYO7A	276903	NG_009086.1	11q13.5
USH1C	10083	Usher syndrome 1C (autosomal recessive, severe)	PDZ73; AIE-75; DFNB18; PDZ-45; PDZ-73; NY-CO-37; NY-CO-38; ush1cpst; PDZ-73/NY-CO-38; USH1C	605242	NC_000011.8	11p15.1-p14
CDH23	64072	cadherin-like 23	USH1D; DFNB12; FLJ00233; FLJ36499; KIAA1774; KIAA1812; MGC102761;	605516	NG_008835.1	10q21-q22

ML-IPR-000803

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
			DKFZp434P2350; CDH23			
PCDH15	65217	protocadherin 15	USH1F; DFNB23; DKFZp667A1711; PCDH15	605514	NG_009191.1	10q21.1
SANS	124590	Usher syndrome 1G (autosomal recessive)	SANS; ANKS4A; FLJ33924; USH1G	607696	NG_007882.1	17q25.1
ARX	170302	aristaless related homeobox	ISSX; PRIS; MRX29; MRX32; MRX33; MRX36; MRX38; MRX43; MRX54; MRX76; MRX87; MRXS1; ARX	300382	NG_008281.1	Xp21
OPHN1	4983	oligophrenin 1	OPN1; MRX60; OPHN1	300127	NG_008960.1	Xq12
JARID1C	8242	lysine (K)-specific demethylase 5C	MRXJ; SMCX; MRXSJ; XE169; JARID1C; DXS1272E; KDM5C	314690	NG_008085.1	Xp11.22-p11.21
FTSJ1	24140	FtsJ homolog 1 (E. coli)	JM23; MRX9; SPB1; TRM7; CDLIV; MRX44; FTSJ1	300499	NG_008879.1	Xp11.23
SLC6A8	6535	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	CRT; CT1; CRTR; MGC87396; SLC6A8	300036	NC_000023.9	Xq28
DLG3	1741	discs, large homolog 3 (Drosophila)	MRX; MRX90; NEDLG; NE-Dlg; SAP102; SAP-102; KIAA1232; DLG3	300189	NC_000023.9	Xq13.1
TM4SF2	7102	tetraspanin 7	A15; MXS1; CD231; MRX58; CCG-B7; TM4SF2; TALLA-1; TM4SF2b; DXS1692E; TSPAN7	300096	NG_009160.1	Xp11.4
ZNF41	7592	zinc finger protein 41	MRX89; MGC8941; ZNF41	314995	NG_008238.1	Xp11.23
FACL4	2182	acyl-CoA synthetase long-chain family member 4	ACS4; FACL4; LACS4; MRX63; MRX68; ACSL4	300157	NG_008053.1	Xq22.3-q23
PQBP1	10084	polyglutamine binding protein 1	SHS; MRX55; MRXS3; MRXS8; NPW38; RENS1; PQBP1	300463	NC_000023.9	Xp11.23

ML-IPR-000804

Gene name	Gene ID	Description	Gene aliases	OMIM	RefSeqGene	Chromosome map position
PEX1	5189	peroxisomal biogenesis factor 1	ZWS1; PEX1	602136	NG_008341.1	7q21.2
PXMP3	5828	peroxisomal membrane protein 3, 35kDa	PAF1; PEX2; PMP3; PAF-1; PMP35; RNF72; PXMP3	170993	NG_008371.1	8q21.1
PEX6	5190	peroxisomal biogenesis factor 6	PAF2; PAF-2; PXAAA1; PEX6	601498	NG_008370.1	6p21.1
PEX10	5192	peroxisomal biogenesis factor 10	NALD; RNF69; MGC1998; PEX10	602859	NG_008342.1	1p36.32
PEX12	5193	peroxisomal biogenesis factor 12	PAF-3; PEX12	601758	NG_008447.1	17q12
PEX5	5830	peroxisomal biogenesis factor 5	PXR1; PTS1R; PTS1-BP; FLJ50634; FLJ50721; FLJ51948; PEX5	600414	NG_008448.1	12p13.31
PEX26	55670	peroxisomal biogenesis factor 26	FLJ20695; PEX26M1T; Pex26pM1T; PEX26	608666	NG_008339.1	22q11.21

Table 2: Polymorphisms

Gene name	Gene ID	SNP ID	Mutation
CFTR	1080	rs63500661	TCACATCACCAAGTTAAAAAAAAAAAA[A/G]GGGGCGGGGGGCGAGAATGAAAATT
CFTR	1080	rs63107760	AAACAAGGATGAATTAAGTTTTTTT[-/T]AAAAAAGAAACATTTGGTAAGGGGA
CFTR	1080	rs62469443	ATCACCAAGTTAAAAAAAAAAAAAGGG[A/G]CGGGGGGGCAGAATGAAAATTGCAT
CFTR	1080	rs62469442	CTATTGAACCAGAACCAACAGGAAT[A/G]CCATAGCATTTTGTAACTAACTG
CFTR	1080	rs62469441	CAGGAGTTCAAGACCAGCCTACTAAA[A/C]CACACACACACACACACACACACAC
CFTR	1080	rs62469439	GATTAAATAATAGTGTATGTACCC[C/G]GCTTATAGGAGAAGAGGGTGTGTGT
CFTR	1080	rs62469438	ATTGTTATCTTTTCATATAAGGTAAC[A/T]GAGGCCAGAGAGATTAAATAACAT
CFTR	1080	rs62469437	TAATTTTAATTAAGTAAATTTAATTG[A/G]TAGATAAATAAGTAGATAAAATAAATA

ML-IPR-000805

Gene name	Gene ID	SNP ID	Mutation
CFTR	1080	rs62469436	GTATAAAAAAAAAAAAAAAAAAGTT[A/T]GAATGTTTTCTTGCATTCA GAGCCT
CFTR	1080	rs62469435	ATACTAAAAATTTAAAGTTCTTGC[A/G]ATATATTTCTTAATATCTTA CATC
CFTR	1080	rs62469434	TGCTGGGATTACAGGCGTGAGCCACC[A/G]CGCCTGGCCTGATGGGACA TATTTT
CFTR	1080	rs62469433	CTACAATATAAGTATAGTATTGCAAA[A/C]CCATCAGGAAGGGTGTTAA CTATTT
CFTR	1080	rs61763210	GTTGTCTCCAACTTTTTTTCAGGTG[- /AGA]AGGTGGCCAACCGAGCTTCGGAAG
CFTR	1080	rs61720488	TTTTTTCATAAAAGATTATATAAAGG[A/C]TATTGCTTTGAATCACAAA CACTA
CFTR	1080	rs61481156	ATCTAGTGAGCAGTCAGGAAAGAGAA[C/T]TTCCAGATCCTGGAAATCA GGGTTA
CFTR	1080	rs61443875	TAGAGTATAAAAAAAAAAAAAAAAAA[- /A]GTTGAATGTTTTCTTGCATTGAGA
CFTR	1080	rs61312222	TGCAAATGCCAACTATCAAAGATATT[C/G]GAGTATACTGTCAATAAACT TCATA
CFTR	1080	rs61159372	TCCTCAACAGTTAGAAACAATATTTT[C/G]AGTGATTCCCATGCCAACT TTAAT
CFTR	1080	rs61094145	TTTTTGGTATTGTTGTTAAATAAGTG[A/G]GAATTCAATACAGTATAATG TCTGT
CFTR	1080	rs61086387	CTTGAAATCGGATATATATATATATA[- /TGATATATATATATATATATATATATACATATATATATATA]GTA TTATCCCTGTTTCACAGTTT
CFTR	1080	rs60996744	AGAGGGGCTGTGAAGGACACCAAGGA[A/G]GAGACTAAGAGCCAGGAG GGAAAAC
CFTR	1080	rs60960860	TAGAGTTTATTAGCTTTTACTACTCT[A/G]CTTAGTTACTTTGTGTTACAG AATA
CFTR	1080	rs60923902	ACTAGTGATGATGAGCTTCTTTTCAT[- /AT]GTTTGTGGCTGCATAAATGTCTTC
CFTR	1080	rs60912824	GCAGAGAAAAGAGGGGCTGTGAAGGA[C/G]ACCAAGGAGGAGACTAAG AGCCAGG
CFTR	1080	rs60887846	TCAGAGGTCTACCACTGGTGCATAC[G/T]CTAATCACAGTGTGCGAAAAT TTTAC
CFTR	1080	rs60793174	AAGAAAGAGCAAAAGAGGGCAAACTT[C/T]TCATACATTTTGTATGTCG AAACCA
CFTR	1080	rs60788575	CCTAAAGTTTAAAAAGAAAAAAAAA[- /A]GGAAGAAGGAATTAATAATCCAAAG

ML-IPR-000806

Gene name	Gene ID	SNP ID	Mutation
CFTR	1080	rs60760741	GTGTGTGTGTATATATATATATAT[A/T]TATATATTTTTTTTCTGA GCCA
CFTR	1080	rs60456599	AAACTGTTGATGTTTTCATTTATTTA[C/G]ATCATTGGAAAACTTTAGATT CTAG
CFTR	1080	rs60363249	TTTATCCATTCTTAACCAGAACAGAC[A/G]TTTTTCAGAGCTGGTCCAG GAAAA
CFTR	1080	rs60355115	TTGAAATCGGATATATATATATAT[A/G]TATATATATATATATATATA TATAT
CFTR	1080	rs60308689	TAGTTTTTTATTCCTCATATTATTT[- /T]CAGTGGCTTTTTCTTCCACATCTTT
CFTR	1080	rs60271242	ACATAGTTCTCAGTGGTACAACACTACA[A/G]GTGATTTCTCTTTCTTATTT CTGG
CFTR	1080	rs60010318	AGAGCAATGGCATCCCTTGTCTTGTG[C/T]TATACAGGATGCAGCAATTT ATAGG
CFTR	1080	rs59961323	TTCTGTCTACATAAGATGTCATACTA[A/G]ATTATCTTTCCAGCATGCAT TCAG
CFTR	1080	rs59961270	CAGGGTGGCATGTTAGGCAGTGCTTA[A/G]AATAAATGAGTTGGTTATA CAAGTA
CFTR	1080	rs59837506	AGGACACACACACACACACACACA[- /CA]TGCACACACATTTAAATAGATGCAT
CFTR	1080	rs59572090	TAAAAAATTGGTATAATGAAATTGCA[C/T]TTGTAGTCTTTGGACATTTA AATCC
CFTR	1080	rs59548252	TTTCAATACTTAAGAGGTACGCAGAG[A/G]AAAGAGGGGCTGTGAAGGA CACCAA
CFTR	1080	rs59519859	CAGCAATGAATATTTTGGGCTGAGG[C/T]GCTGAGGGGTAAATTGCA GCCTGG
CFTR	1080	rs59509837	TTATGGTTTATATTTTGTGTCTTCT[- /CTTT]AACACATCTTTCTAGCAGAATTCA
CFTR	1080	rs59417037	GTATTTTAGTTTTTTTTTTGTTTG[- /T]TTTGTTTGTTTGTTTGTTTGT
CFTR	1080	rs59159458	TGGGTGACTCCATTTTACTTTTAGT[C/T]TGGTCTGTTGAGGCCTCGTGA GAGA
CFTR	1080	rs59048119	TATTTTCATGTATTTAGTTTTTTT[- /TTTT]GTTTGTGTTTGTGTTTGTGTTTGT
CFTR	1080	rs58970500	GTGTGTGTGTATATATATATATAT[A/T]TATATTTTTTTTCTGAGC CAAA
CFTR	1080	rs58942292	AACCTATTAGCATGTCTGGCAGAAAA[- /A]TAGATACTTAATAAATTCTTAAAT
CFTR	1080	rs58917054	GAGGCTTAGACAGTTTAAGTAACTCA[A/G]GCATGGTTACACAACCTAGC

ML-IPR-000807

Gene name	Gene ID	SNP ID	Mutation
GBA	2629	rs12752133	CTCTTCCGAGGTTCCACCCTGAACAC[C/T]TTCCTGCTCCCTCGTGGTGTA GAGT
GBA	2629	rs12747811	TTCTGACTGGCAACCAGCCCCACTCT[C/T]TGGGAGCCCTCAGGAATGAA CTTGC
GBA	2629	rs12743554	gctcagcctcccaggctggagtgcag[A/T]ggcgcgatctcggctcaccgcaacc
GBA	2629	rs12041778	CATGAACCACATCAAATGAGATTTAG[C/T]GGGAGTGGCACACACAGTC ATGACC
GBA	2629	rs12034326	AAGCAGCCCTGGGGAGTCGGGGCGGG[A/G]CCTGGATTGGAAAAGAGA CGGTAC
GBA	2629	rs11558184	CTCCAAGTTCTGGGAGCAGAGTGTGC[A/G]GCTAGGCTCCTGGGATCGA GGGATG
GBA	2629	rs11430678	GTTCTCTCAGTAAAtttttttttt[-/G/T]gttttgagacagagtcttgcctgt
GBA	2629	rs11264345	CTAGTACCTTACTTCCCTCAAGTTCA[A/T]TCATCTCACAGATATTTCTG AGCA
GBA	2629	rs10908459	aattagccgtgcgtggcgggtgc[C/T]tgtaatcccacgtacttggaggct
GBA	2629	rs10796940	CCATGGCCAGCCGGGAGGGGACGGG[A/C]ACACACAGACCCACACAG AGACTCA
GBA	2629	rs10668496	agcgagagagagagagagagagag[-/AG]gagCCGGCGCGAGAAGTACGCATGC
GBA	2629	rs7416991	CGTAGCAGTTAGCAGATGATAGGCGG[C/G/T]GAAATCTTATTTACAGG GCATTAA
GBA	2629	rs4024049	CTGGCCCTGGTGACAGTGGGGCTGTG[C/T]GTGGGGCCAGAGCCTTCTCA GAGGT
GBA	2629	rs4024048	CAGATACTGGCCCTGGTGACAGTGGG[A/G]CTGTGCGTGGGGCCAGAGC CTTCTC
GBA	2629	rs4024047	GACAGATACTGGCCCTGGTGACAGTG[G/T]GGCTGTGCGTGGGGCCAGA GCCTTC
GBA	2629	rs3841430	GGCTCctctctctctctctctct[-/TC]gctcgtctctctctctctctct
GBA	2629	rs3754485	GTTTCAGACCAGCCTGGCCAACATAG[C/T]GAAACCCCATCTCTACTAAA AATAA
GBA	2629	rs3205619	AGTGGGCGATTGGATGGAGCTGAGTA[C/T]GGGGCCCATCCAGGCTAAT CACACC
GBA	2629	rs2990227	CCGGGCTCCGTGAATGTTTGTACAT[C/G]TCTGAAGAACGTATGAATTA CATAA
GBA	2629	rs2990226	GAATCCCAACCCCGACGCTCGTCGCC[C/G]GGCTCCGTGAATGTTTGTCA CATGT
GBA	2629	rs2990225	GCGAATCCCAACCCCGACGCTCGTCG[C/T]CGGGCTCCGTGAATGTTTGT CACAT

ML-IPR-000809

Gene name	Gene ID	SNP ID	Mutation
GBA	2629	rs2990224	TGGGCAGAAGTCAGGGTCCAAAGAAA[G/T]GGCAAAGAAAAGTGTcagtggtca
ASPA	443	rs63751297	TAAGAAAGACGTTTTTGATTTTTTTC[A/G]GACTTCTCTGGCTCCACTACCCTGC
ASPA	443	rs62071301	CTGATTCTCTGGCCAGGAGCGGTGGCT[C/T]ACGCCTGTAATCCCAGCGCTTTGGG
ASPA	443	rs62071300	TAAAAATGCTGATTCTCTGGCCAGGAG[C/T]GGTGGCTCACGCCTGTAATCCAGC
ASPA	443	rs62071299	TTTAAAAATGCTGATTCTCTGGCCAGG[A/C]GCGGTGGCTCACGCCTGTAA TCCCA
ASPA	443	rs62071297	CAAGACCTGTCAAAGATCTGAGAAAT[A/T]TTACCCGACTTACAAGCTA ACCATT
ASPA	443	rs61697033	ACTGTAATAAGTGTGTGTAAGAAAAT[A/G]CACAAAATAATATAGCAGAGGGTAT
ASPA	443	rs60743592	CTTGAGGTCAGGAGTTCAAGACCAGT[C/T]TGGGCAACATGGGGAAAACCTTGTC
ASPA	443	rs60666840	AGGTTGCAGTGAGCCGAGATCATGCC[A/G]TTGCACTCCAGCCGGGGCAACAAAA
ASPA	443	rs60147514	ACAAGTGTCTTGAAATTATCTGTGAT[C/T]TGCTATAGAGCAATACTTTGTAAA
ASPA	443	rs59930743	GTGGGTATATGCAGCTCTATGCACTA[C/T]CTGCTCATTTATTTGGTAAATCTAA
ASPA	443	rs59690349	TGTGTGTGTGTGCGTGTGTGTGTGTG[-/TGTGTGTG]ATCATAAGAGTGGCTGCAGCAAAC
ASPA	443	rs59676360	AGTCTGGAGTGCAATGGTGCAATCTC[A/G]GCTCACTGCAGCCTCCACCTCCGGG
ASPA	443	rs59335404	CTCCTAATGGATATTCTCTAAATTT[G/T]CTGAACAGAATTTAACTTGAGCTGG
ASPA	443	rs58879097	ATTTAAAAATGGATTTCTAGAAAAAC[A/G]ATCACATACTGAATATTTTAGCAA
ASPA	443	rs58686774	CTATAAATGGGTAGCATGAGGGATTC[A/G]AGGAGGTGGCTGAAAGAAGCACGTA
ASPA	443	rs57511162	AAGAAACCAAGCATAGTAGAGTGTTA[A/G]AAAACCAAAGCAACTAAACAACGT
ASPA	443	rs55859596	CGGGGCTCAGAACTTGTAACAGAAAA[A/T]TAAAATATACTCCACTCAAGGGAAT
ASPA	443	rs55742972	TACTACACTTCACGGATACTGTACTT[-/GTACTT]TTTTCCAAATTGAAGGTTTTTGGC
ASPA	443	rs55640436	TTGTTTTTGTGTTTTGTTTTGTTTTT[-

ML-IPR-000810

Gene name	Gene ID	SNP ID	Mutation
			/GTTTTGTTTTT]TGAGATGGAGTCTCGCTCTGTCGCC
ASPA	443	rs36225687	TTGCCTTACTACACTTCACGGATAC[- /TGTA]TGTACTTTTTTTCCAAATTGAAGGT
ASPA	443	rs36051310	GAGGTGGCTGAAAGAAGCACGTATCC[- /C]TGATGGCATGGTTGCGGGTTATATG
ASPA	443	rs36034906	GAGAAAAGCAGTTCCTGGAACACCCC[- /C]ACCCCTTAACCCCTATCTCTGCTT
ASPA	443	rs36033666	TTACATATGTATACATGTGCCATGTT[- /T]GGTGTGCCGCACCCATTAACTCGTC
ASPA	443	rs35730123	CTTTTCCAGATTTTTTTTTTTTTTTT[- /T]GAGACAGAGTTTCACTCTTGTGTC
ASPA	443	rs35629100	TTTGAAATCTTAAGCTTTTATTTGG[- /G]TGTCACAGAGAAACAGGATCTGTAT
ASPA	443	rs35614631	TACTTTAAGTTTTAGGGTACATGTGC[- /A]CCATGTGCAGGTTTGTACATATGT
ASPA	443	rs35225782	ATTCATGACCAGCCACATAAATGCAC[- /A]GTATTACTTCGCAAGCATGCCAATG
ASPA	443	rs35178659	GTGCACTAGAATTAGCTAAAGTGGGG[- /G]AAAAAAGATGCATTTGATGGTCTA
ASPA	443	rs35095578	AACCTCCACCTCCCAGGTTCAAGAGA[- /A]TTCTCCTGCCTCAGCCTCCCAAGTA
ASPA	443	rs35002210	CCTCCCTGTGATCCGAAGTAGCAGAC[A/G]TACTTAACTTCCATGGTGGA TTGTT
ASPA	443	rs34744839	AAAACATTATTATATCTAGAAAAAA[- /A]TGTATCTTAACCATTTGTGGGAAGTG
ASPA	443	rs34680506	TTGAAGGTAAAAATCATAGGGAGTTGG[- /G]AGCTGTCCTCTTGCGCTGAATCAGT
ASPA	443	rs34365618	ACTTGTGGCCTTTTTGGAGAGGTTAG[- /CA]ACTCTGAAAACCTCTGTCCCTGGACC
ASPA	443	rs34275920	GAAGGAGAAAAAGAGAGGAAATAAGT[- /T]AAAAATAAACACAATTAATAAAG
ASPA	443	rs34109510	TGTATACATGTGCCATGTTGGTGTGC[C/T]GCACCCATTAACTCGTCATT AGCA
ASPA	443	rs34054576	TCACCTGTCACCTCCTATAGAACTTT[- /C]CCCTGACCCTCCTCTATAGCATTAA
ASPA	443	rs34015272	ATAAATGATCATCATTCACAGTAGGG[- /G]TTTGTGTTTGTGTTTTTCTGGAAA
ASPA	443	rs34002091	ACAGACATATCTACAAACACACTTTT[- /T]CACATATTTGTGTAAGTCATTTATG

ML-IPR-000811

Gene name	Gene ID	SNP ID	Mutation
ASPA	443	rs28940574	AAAGACAACATAAACTAACGCTCAATG[A/C]AAAAAGTATTCGCTGCTGT TTACAT
ASPA	443	rs28940279	TACCGTGTACCCCGTGTGTGAATG[A/C]GGCCGCATATTACGAAAAAGA AAGAA
ASPA	443	rs17850703	CAGGGCTGGAGGTAAAACCATTTATT[A/G]CTAACCCAGAGCAGTGAA GAAGTG
ASPA	443	rs17222495	TTCTTCATTGCCTATTGAAGAGAGAG[C/T]GGAATGCTTTGGTTGCCAGA TATGG
ASPA	443	rs17175228	CACAAGATCTCATTACTCAGGAGCTG[C/T]CCAAGTGTCTAATGTACTTA GTAA
ASPA	443	rs16953074	TTCTGTGTAACATTTCAATTAAGCAA[A/G]GGATTGCGCAAATCAAAAAT TGTC
ASPA	443	rs16953070	TAAAACGTATTGAAGGTATTATTGAC[G/T]CTGTTGAAGCAAAGAGAAC AAAACA
HEXA	3073	rs62022858	ATCTGCTCTTCCAGTTGGATGACAAG[C/T]CTTGCTGTCTAACACCTGCT GCAGA
HEXA	3073	rs62022857	CCATTTTTTGTGTATTTTTTTTTC[C/T]TGAATACTTTTATCGCAGTTG GTT
HEXA	3073	rs62017872	CCCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]AAAAAAAAAGAAACAA AACCCAA
HEXA	3073	rs62017871	AGTGGCTCCAAAAAGGTCATGGAACC[C/T]CTTGAGGATGATGCAAATT GACTCT
HEXA	3073	rs61662730	TAAAGTTACTTTTCTTTTATTGACTT[C/T]CCCTTATTTTTAACCTTATGC TTT
HEXA	3073	rs61329913	CAGAGTAAAAAAAAAAAAAAAAAAAA[- /A]GGAAGTAGCAGCAACAGCTTGGA
HEXA	3073	rs60920713	GTTGCCAGGGTTGAGTGCAGAGGCA[C/T]ATCTGGCTCACAGCAACCT CTGCC
HEXA	3073	rs60783213	AAGGCTTTTTTTTTTTTTTTTTTTT[- /TTTT]GAGACAGAGTCTTGCTGTGTACCC
HEXA	3073	rs60644867	GCCTACATTCTGCAAAGAGGAGGGAA[C/G]ATTCACAGCTCCATACTTG AACCT
HEXA	3073	rs60288568	CCAAAGGAGAAATAGCTCTAGGGGAGG[C/G]AGGTGGATGAGTATGCATG GGGAG
HEXA	3073	rs59888548	GACTCCATCTCAAAAAAAAAAAAAAAAAA[- /A]TGCAGTCTAATGGCAGAATTAGACT
HEXA	3073	rs59733856	TTATTTATTTATTTATTTTGA[A/G]ACAGGGTCTCTGTTGTCCAGG CTGG
HEXA	3073	rs59427837	TTTTGAGGCAGGGTCTCACTCTGTTG[C/T]CCAGGGTTGAGTGCAGAGGC

ML-IPR-000812

Gene name	Gene ID	SNP ID	Mutation
			ACATC
HEXA	3073	rs59171976	CGCCTTGCGAAGGCCCCACAGCTTGC[C/T]TGTGACAAACGTTTCATAGGC AAATG
HEXA	3073	rs58706602	GGAGGTCTGTACAAAGCACCTAC[C/T]TCATGGGTGAGTTCCACAG CAGAA
HEXA	3073	rs58696963	GAATCTTATAATTCAGTGTGTACCTC[- /CCTC]TGTTCATATTTTCGCAATTGAACT
HEXA	3073	rs58610850	AACATAGTATCTAATATAGCTTTACA[C/T]CCAAAGCCAAAATATGAATA CACTG
HEXA	3073	rs58016062	TTGTTTTGTTTTGTTTGGGGGGGGG[- /G]TTGTTTTCTGAGAGGGAGTCTTGC
HEXA	3073	rs57733983	CATACCAAAGGGCAGCTGGAGGGATA[C/T]AGACGGAAGTCATGTGGAG AGTGAA
HEXA	3073	rs57476645	CAGGTGTGAGCCACCACGACCACCAA[A/T]TTAGCTCTTTTACTCCTTC CCTTC
HEXA	3073	rs56870003	AGTGGTAGCTGATTTTGCTTCTGGAT[A/C]CTTTGCCACCTTCCCACTCTT TAAT
HEXA	3073	rs56338339	AAAGACCTGTTTCTTAAAAAAAAAAAA[- /AGAAAAAAAAAAAA]GAAAGAAAAGAAAAGAAAAAACAG
HEXA	3073	rs55995352	TAAAAAATCTTCAATGAGGAGATGT[C/T]CCCAGAGCAAGACAGCTGT AGGATG
HEXA	3073	rs55860138	AAAAGAAAAAAAAAAAAAAAAAAAAA[- /A]GAAAACAAAACCCAAACCCATAAAG
HEXA	3073	rs55743646	CCTGTCTCTAAAAGAAAAAAAAAAAA[A/G]AAAAAAAAAGAAAACAAA ACCCAAA
HEXA	3073	rs55665666	GTTATCATAGAAAAATATCACACTCT[- /GT]CTGTATCCCCACTTCCAGAACTGT
HEXA	3073	rs36106892	CAGGAGCTCATAGAATTACATACAAT[- /C]TTTTTTTTTTTTTTTGAGACAGCG
HEXA	3073	rs36091525	TTGAGAATCTTATAATTCAGTGTGA[- /CCTC]CCTCTGTTTCATATTTTCGCAATTG
HEXA	3073	rs35949555	CCACTACCACAGTGCCTAGAGAACAA[C/T]ATGTGTTTAATAATATTTAA ATAAT
HEXA	3073	rs35827424	CCCTGTCTCTAAAAGAAAAAAAAAAAA[- /A]AAAAAAAAAAGAAAACAAAACCCAA
HEXA	3073	rs35729578	CCATTATATCATTATTCCCACTCA[- /T]TTTCTTCATTCCAACCAAGATATAT
HEXA	3073	rs35649102	TCCGTCTCAAAAAAAAAAAAAAAAAAG[- /A]GAAAGGAATTATTCTCATGTATACA

ML-IPR-000813

Gene name	Gene ID	SNP ID	Mutation
HEXA	3073	rs35118677	CTGGGGCAGTTAAAAAGAAAAACAAA[-/C]CCCTGGTCCCTGCCCTTGAGGAGAT
HEXA	3073	rs35005352	CTCCAGGGTCCCATTCCAGGACCACA[-/C]GCCTGCTACCTCTGCAGCTCACTCA
HEXA	3073	rs34736306	GGATTGACATATACCAAGTTAGACGGA[-/T]TTTTTTTTCCATAAACCAAGGCTCA
HEXA	3073	rs34607939	ACAAATAATTACTACATATCTACAAC[A/G]TTCCAGATACAGAAGAAATGGCCAA
HEXA	3073	rs34496117	TAAACACACTTGAAACATCATATAAA[-/ATG]ATATTACTACAAGACTTAACCGTAA
HEXA	3073	rs34300017	ACACAGGTAATCCATGTTTATTATAG[-/A]AAAATGCCACATTACTCTTTATTGA
HEXA	3073	rs34206496	AGTTATCATAGAAAAATATCACACTC[-/TG]TCTGTATCCCCACTTCCAGAACTG
HEXA	3073	rs34110830	AATGAACTTACAGGAAGGTAATATAT[-/G]GGAAATAAACATCTTATTGAATTTA
HEXA	3073	rs34093438	GGACCCCTGAAAAGGCACAAGACACCC[-/T]TTCAGGTTACACTTCCTGAAAGCT
HEXA	3073	rs34085965	CCACCAATCACCAGAGCCTTCTGCTC[A/G]GGGGTACCTGAGGGAAAAC AAGCAA
HEXA	3073	rs34004907	AAAGACTGAAAAACATTCATAACTA[-/T]TTTCTTGTTATCCTCGGAAATGTC
HEXA	3073	rs28942072	TATCTTCATCTTGAGGAGATGAGGT[C/T]GATTCACCTGCTGGAAGTCAACC
HEXA	3073	rs28942071	TTGCCTATGAACGTTTGTACACTTC[C/T]GCTGTGAGTTGCTGAGGCGA GGTGT
HEXA	3073	rs28941771	GCTTGCTGTTGGATACATCTCGCCAT[C/T]ACCTGCCACTCTCTAGCATCC TGA
HEXA	3073	rs28941770	CCGGGGCTTGCTGTTGGATACATCTC[G/T]CCATTACCTGCCACTCTCTA GCATC

Examples

Example 1: MIP-capture reaction of a set of exon target nucleic acids

MIP probes are synthesized using standard column-based oligonucleotide synthesis by any number of vendors (e.g. IDT).

ML-IPR-000814

Hybridize capture probes to human genomic sample

Assemble hybridization reaction:

- 1.0 ul capture probe mix (~2.5 pmol)
- 2.0 ul 10X Ampligase buffer (Epicentre)
- 6.0 ul 500 ng/ul human genomic DNA (~16.7 fmol)
- 11 ul dH₂O

In a thermal cycler, heat reaction to 95C for 5 min to denature genomic DNA, then cool to 60C. Allow to incubate at 60C for 40 hours.

Convert hybridized probes into covalently-closed circular products containing target nucleic acids

Prepare fill-in / ligation reaction mixture:

- 0.25 ul 2 mM dNTP mix (Invitrogen)
- 2.5 ul 10X Ampligase buffer (Epicentre)
- 5.0 ul 5 U/ul Taq Stoffel fragment (Applied Biosystems)
- 12.5 ul 5 U/ul Ampligase (Epicentre)
- 4.75 ul dH₂O

Add 1.0 ul of this mix to the hybridized probe reaction, and incubate at 60C for 10 hours.

Purify circularized probe/target nucleic acid products from un-reacted probes and genomic DNA

Prepare exonuclease reaction mixture:

- 21 ul fill-in / ligation reaction product
- 2.0 ul 10X exonuclease I buffer (New England Biolabs)
- 2.0 ul 20 U/ul exonuclease I (New England Biolabs)
- 2.0 ul 100 U/ul exonuclease III (New England Biolabs)

Incubate at 37C for 60 min, then heat-inactivate by incubating at 80C for 15 min. Immediately cool to 4C for storage.

ML-IPR-000815

Amplify circular material by PCR using primers specific to the 'constant region' of the probes

Prepare PCR mixture:

- 5.0 ul 10X Accuprime reaction buffer (Invitrogen)
- 1.5 ul 10 uM CP-2-FA (5'-GCACGATCCGACGGTAGTGT-3')
- 1.5 ul 10 uM CP-2-RA (5'-CCGTAATCGGGAAGCTGAAG-3')
- 0.4 ul 25 mM dNTP mix (Invitrogen)
- 2.0 ul heat-inactivated exonuclease reaction mix
- 1.5 ul 10X SybrGreen (Invitrogen)
- 0.4 ul 2.5 U/ul Accuprime Pfx polymerase (Invitrogen)
- 37.7 ul dH₂O

Thermal cycle in real-time thermal cycler according to the following protocol, but stop cycling before amplification yield plateaus (generally 8-12 cycles):

1. 95C for 5 min
2. 95C for 30 sec
3. 58C for 60 sec
4. 72C for 60 sec
5. goto 2, N more times

Prepare a shotgun next-generation sequencing library for analysis

- Purify desired amplicon population from non-specific amplification products by gel extraction.
- Concatemerize amplicons into high-molecular weight products suitable for shearing
- Mechanically shear, using either a nebulizer, BioRuptor, Hydroshear, Covaris, or similar instrument. DNA should be sheared into fragments several hundred basepairs in length.
- Ligate adapters required for amplification by the sequencing platform used. If necessary, purify ligated product from unligated product and adapters.

Perform sequencing of library according to manufacturer's directions (e.g. Illumina, ABI, etc), reading the target sequence to determine abundance of the target nucleic acid..

ML-IPR-000816

Example 2: Use of MIPs, hybridization, and mutation-detection MIPs to genotype a set of 1000 targets

MIPs, hybridization, and mutation-detection MIPs are used to genotype a set of 1000 targets. The protocol permits detection of any of 50 specific known point mutations

First, separate MIP, hybridization, and mutation-detection MIP reactions are performed. A MIP capture reaction is performed essentially as described in Turner et al 2009, Nature methods 6:315-6. A set of MIPs is designed such that each probe in the set flanks one of the 1000 targets. Separately, a hybridization enrichment reaction is performed using the Agilent SureSelect procedure. Prior to selection, the genomic DNA to be enriched is converted into a shotgun sequencing library using Illumina's 'Fragment Library' kit and protocol. Agilent's web interface is used to design a set of probes which will hybridize to the target nucleic acids. Separately, a set of probes are designed (mutation-detection MIPs) which will form MIPs only if mutations (e.g., specific polymorphisms) are present. Each mutation-detection MIP has a 3'-most base identity that is specific for a single known mutation. A reaction with this set of mutation-detection MIPs is performed to selectively detect the presence of any mutant alleles.

Once all three reactions have been performed, the two MIP reactions are combined into a single tube, and run as one sample on the next-generation DNA sequencing instrument. The hybridization-enriched reaction is run as a separate sample on the next-generation DNA sequencing instrument. Reads from each 'sample' are combined by a software algorithm which forms a consensus diploid genotype at each position in the target set by evaluating the total coverage at each position, the origin of each read in that total coverage, the quality score of each individual read, and the presence (or absence) of any reads derived from mutation-specific MIPs overlapping the region.

It will be clear that the methods may be practiced other than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present disclosure are possible in light of the above teachings and, therefore, are within the scope of the claims. Preferred features of each aspect of the disclosure are as for each of the other aspects mutatis mutandis. The documents including patents, patent applications, journal articles, or other disclosures mentioned herein are hereby incorporated by reference in their entirety. In the event of conflict, the disclosure of present application controls, other than in the event of clear error.

What is claimed is:

ML-IPR-000817

Claims

1. A method of increasing the sensitivity of a nucleic acid detection assay, the method comprising:
 - obtaining a first preparation of a target nucleic acid using a first preparative method on a biological sample,
 - obtaining a second preparation of a target nucleic acid using a second preparative method on the biological sample,
 - assaying the sequences obtained in both first and second nucleic acid preparations,
 - using the sequence information from both first and second nucleic acid preparations to determine the genotype of the target nucleic acid in the biological sample, wherein the first and second preparative methods have different systematic sequence biases.
2. The method of claim 1, wherein the first and second nucleic acid preparations are combined prior to performing a sequence assay.
3. The method of claim 1, wherein separate sequence assays are performed on the first and second nucleic acid preparations and the sequence information from both assays are combined to determine the genotype of the target nucleic acid in the biological sample.
4. The method of claim 1, wherein the first preparative method is a PCR-based, a hybridization-based, or a circular probe-based preparative method.
5. The method of claim 4, wherein the second preparative method is a PCR-based, a hybridization-based, or a circular probe-based preparative method, provided that the second method is different from the first method.
6. A method of obtaining a nucleic acid preparation that is representative of the nucleic acid in a biological sample, the method comprising
 - obtaining a first nucleic acid preparation using a first preparative method on a biological sample,
 - obtaining a second nucleic acid preparation using a second preparative method on the biological sample, and

ML-IPR-000818

combining the first and second nucleic acid preparations to obtain a combined preparation that is representative of nucleic acid in the biological sample.

7. The method of claim 1 or 6, further comprising performing a third nucleic acid preparation using a third preparative method that is different from the first and second preparative methods, wherein the first, second, and third preparative methods all have different systematic sequence biases.

8. The method of claim any of the prior claims, wherein the different preparative methods are used for a plurality of different loci in the biological sample to increase the sensitivity of a multiplex nucleic acid analysis.

9. A method of detecting a polymorphism in a nucleic acid in a biological sample, the method comprising
obtaining a nucleic acid preparation using a preparative method on a biological sample,
and performing a molecular inversion probe capture reaction on the nucleic acid preparation, wherein a molecular inversion probe capture of a target nucleic acid of the nucleic acid preparation is indicative of the presence of a polymorphism in the target nucleic acid, optionally wherein the polymorphism is selected from Table 2.

10. The method of any of the prior claims, wherein the target nucleic acid has a sequence of a gene selected from Table 1.

Abstract

The invention relates to methods and compositions for increasing the representativeness and/or sensitivity of nucleic acid analyses by combining results from preparative techniques that have different systematic biases. Aspects of the invention are particularly useful for increasing the confidence of genotype determinations made in multiplex analyses.

FIGURE 1

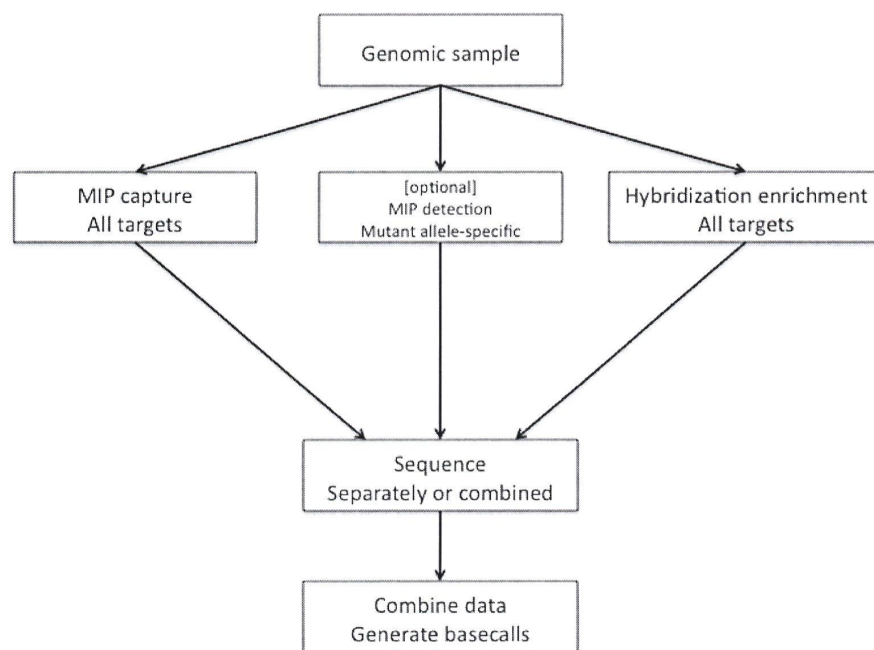
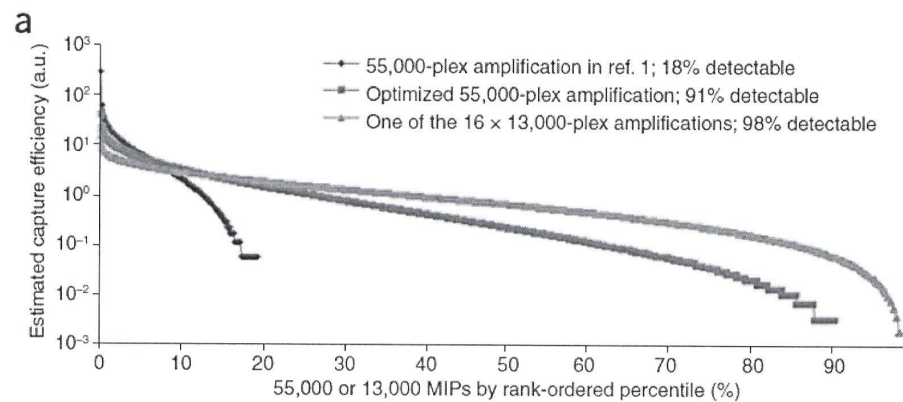


FIGURE 2



1666361.1

ML-IPR-000822

FIGURE 3

