



(12) 发明专利申请

(10) 申请公布号 CN 104946773 A

(43) 申请公布日 2015. 09. 30

(21) 申请号 201510390373. 9

(22) 申请日 2015. 07. 06

(71) 申请人 厦门万基生物科技有限公司

地址 361100 福建省厦门市翔安区火炬园翔
星路 98 号强业楼

(72) 发明人 陈洪亮 郑海灵 段利朋 祝兴强

(74) 专利代理机构 北京远大卓悦知识产权代理
事务所(普通合伙) 11369

代理人 李强

(51) Int. Cl.

C12Q 1/68(2006. 01)

权利要求书1页 说明书8页

(54) 发明名称

一种利用 SNP 进行产前亲权关系判定的方法

(57) 摘要

本发明提供了一种利用 SNP 进行产前亲权关系判定的方法,利用 SNP 作为遗传标记,结合高通量测序技术进行产前亲权关系的判定,只需提供母亲外周血 10ml,在母亲外周血浆中提取的游离 DNA 已经包含了胎儿的游离 DNA,所以母亲和胎儿只需一份样品即可。由于只需抽取孕妇的静脉血,操作简便,因此不会对孕妇和胎儿造成创伤,且孕 10 周后即可鉴定。

1. 一种利用 SNP 进行产前亲权关系判定的方法,其特征在于:利用 SNP 作为遗传标记,结合高通量测序技术进行产前亲权关系的判定。

2. 如权利要求 1 所述的一种利用 SNP 进行产前亲权关系判定的方法,其特征在于包括如下步骤:

步骤一,设计 SNP 位点及引物;

步骤二,提取孕妇样本和待定男性样本的样品 DNA;

步骤三,样品 DNA 的高通量测序前处理;

步骤四,高通量测序;

步骤五,高通量测序数据后处理,并筛选出孕妇样本和待定男性样本的纯合 SNP 位点;

步骤六,将孕妇样本和待定男性样本在每个相同的纯合 SNP 位点进行对比,最高碱基类型相同的位点定义为一个一致位点,最高碱基类型不同则将该位点定义为一个否定位点,统计出一致位点和否定位点的数目;

步骤七,当否定位点数目大于等于 5 个即可否定胎儿与待定男性的亲缘关系,当一致位点数目大于等于 35 个则不能否定胎儿与待定男性的亲缘关系。

3. 如权利要求 2 所述的一种利用 SNP 进行产前亲权关系判定的方法,其特征在于:所述步骤一中选择最小等位基因频率为 0.4-0.5 的 SNP 位点。

4. 如权利要求 2 所述的一种利用 SNP 进行产前亲权关系判定的方法,其特征在于所述步骤三包括:扩增含有 SNP 位点的片段,纯化扩增到的 PCR 产物,并将 PCR 产物末端修复,筛选得到平末端 DNA 片段,接头连接,得加接头的 DNA 片段,PCR 扩增及纯化得到小片段文库,检测小片段文库的文库浓度和片段大小。

5. 如权利要求 2 所述的一种利用 SNP 进行产前亲权关系判定的方法,其特征在于所述步骤五包括:

1) 将测序得到的序列进行初步过滤,并与人类基因组序列进行比对,筛选出错配碱基 < 3% 的唯一比对序列,然后根据比对结果统计出每个 SNP 位点的覆盖深度、碱基种类和每种碱基对应数目;

2) 根据统计结果,对每个 SNP 位点中四种碱基出现次数进行排序,四种碱基种类按照次数从多到少的顺序分别称作 Major_allele、Minor_allele、Third_allele 及 Fourth_allele,每种类型碱基出现的次数依次为 Major_num、Minor_num、Third_num 及 Fourth_num,并得出 SNP 位点的覆盖深度 Depth, Depth 等于四种碱基数目之和,即

$$Depth = Major_num + Minor_num + Third_num + Fourth_num;$$

3) 计算出在 SNP 位点中出现次数最多的碱基占该位点总覆盖深度的比例,即最高碱基比例 Major_percent = Major_num/Depth,筛选出覆盖深度 Depth 大于 200 层,且最高碱基比大于 99% 的位点。

6. 如权利要求 2 所述的一种利用 SNP 进行产前亲权关系判定的方法,其特征在于:所述孕妇样本采自孕妇外周血。

一种利用 SNP 进行产前亲权关系判定的方法

技术领域

[0001] 本发明涉及一种利用 SNP 进行产前亲权关系判定的方法。

背景技术

[0002] 亲子鉴定是根据人类遗传学的理论和实践,从子代和亲代的形态构造或者生理机能方面的相似特点,分析遗传特征,对可疑的父子关系或母子关系进行判断,并作出肯定或者否定的结论。

[0003] 判断亲子关系的理论依据是孟德尔遗传的分离律。按照这一规律,在配子细胞形成时,成对的等位基因彼此分离,分别进入各自的配子细胞。精、卵细胞受精形成子代,孩子的两个基因组一个来自母亲,一个来自父亲;因此,同对的等位基因也就是一个来自母亲,一个来自父亲。如果鉴定结果符合这一规律,则不排除亲生关系;若不符合,则排除亲生关系(基因变异情况除外)。最早用于检测遗传多态性的方法是使用限制性内切酶对人类基因组中可变数目的串联重复(VNTR)做限制性片段长度多态性分析。随着技术的进步,DNA聚合酶链式反应(PCR),使得较短的核酸片段也能用于分析,将分析目标集中到VNTR中的较短的短串联重复序列(STR),加上多重PCR技术的应用,迅速使STR基因座的分型检测在法医和刑侦中应用起来。STR也成为目前最常用的遗传标记。

[0004] 单核苷酸多态性(single nucleotide polymorphism,SNP)是第三代遗传学标记,这种遗传标记是由于单碱基突变使特定核苷酸位置上出现两种碱基,其中最少的一种在群体中的频率不少于1%。与第一代的RFLP及第二代的STR以长度的差异作为遗传标记的特点截然不同。SNP的分布密集,如果以1%的频率计算,在人基因组中就有300万个以上的SNP遗传标记,这可能达到了人类基因组多态位点数目的极限,因此被认为是应用前景最好的遗传标记物。在医学遗传学、群体遗传学和药物基因组学中有广泛的应用。在法医物证检验中,也由于SNP的含量丰富、遗传稳定,而引起了高度重视。

[0005] 产前亲子鉴定,也称胚胎期亲子鉴定、胎儿亲子鉴定,是指利用基因技术鉴定胎儿生物学意义上的父亲。现有产前亲子鉴定技术是从胎儿绒毛或者孕妇羊水中提取DNA物质,鉴定检测出胎儿的STR与疑似父亲的DNA进行比对,以确认亲子关系。

[0006] 现有技术中产前亲子鉴定通常需要利用羊水穿刺术抽取3-5毫升的羊水,但是抽取出来的羊水必须比较清澈透明,不能含有母亲的血液成分。虽然羊水穿刺术用于产前诊断至今已有30年的历史,准确性得到了医学界的公认。但羊水穿刺的难度较大,目前均需在具备三甲、大型医院采用B超可视监控的引导下完成,却仍有0.5%-1%的宫内感染和流产风险。

发明内容

[0007] 本发明的主要目的即在于克服现有技术的缺点,提供一种无创的利用SNP进行产前亲权关系判定的方法。

[0008] 本发明采用如下的技术方案:

[0009] 一种利用 SNP 进行产前亲权关系判定的方法,利用 SNP 作为遗传标记,结合高通量测序技术进行产前亲权关系的判定,包括如下步骤:

[0010] 步骤一,设计 SNP 位点及引物;

[0011] 步骤二,提取孕妇样本和待定男性样本的样品 DNA;

[0012] 步骤三,样品 DNA 的高通量测序前处理;

[0013] 步骤四,高通量测序;

[0014] 步骤五,高通量测序数据后处理,并筛选出孕妇样本和待定男性样本的纯合 SNP 位点;

[0015] 步骤六,将孕妇样本和待定男性样本在每个相同的纯合 SNP 位点进行对比,最高碱基类型相同的位点定义为一个一致位点,最高碱基类型不同则将该位点定义为一个否定位点,统计出一致位点和否定位点的数目;

[0016] 步骤七,当否定位点数目大于等于 5 个即可否定胎儿与待定男性的亲缘关系,当一致位点数目大于等于 35 个则不能否定胎儿与待定男性的亲缘关系。

[0017] 上述步骤一中选择最小等位基因频率为 0.4-0.5 的 SNP 位点。

[0018] 上述步骤三包括:扩增含有 SNP 位点的片段,纯化扩增到的 PCR 产物,并将 PCR 产物末端修复,筛选得到平末端 DNA 片段,接头连接,得加接头的 DNA 片段,PCR 扩增及纯化得到小片段文库,检测小片段文库的文库浓度和片段大小。

[0019] 上述步骤五包括:

[0020] 1) 将测序得到的序列进行初步过滤,并与人类基因组序列进行比对,筛选出错配碱基 < 3% 的唯一比对序列,然后根据比对结果统计出每个 SNP 位点的覆盖深度、碱基种类和每种碱基对应数目;

[0021] 2) 根据统计结果,对每个 SNP 位点中四种碱基出现次数进行排序,四种碱基种类按照次数从多到少的顺序分别称作 Major_allele、Minor_allele、Third_allele 及 Fourth_allele,每种类型碱基出现的次数依次为 Major_num、Minor_num、Third_num 及 Fourth_num,并得出 SNP 位点的覆盖深度 Depth, Depth 等于四种碱基数目之和,即 $Depth = Major_num + Minor_num + Third_num + Fourth_num$;

[0022] 3) 计算出在 SNP 位点中出现次数最多的碱基占该位点总覆盖深度的比例,即最高碱基比例 $Major_percent = Major_num / Depth$,筛选出覆盖深度 Depth 大于 200 层,且最高碱基比大于 99% 的位点。

[0023] 上述孕妇样本采自孕妇外周血。

[0024] 本发明的一种利用 SNP 进行产前亲权关系判定的方法与现有技术相比,利用 SNP 作为遗传标记,结合高通量测序技术进行产前亲权关系的判定,只需提供母亲外周血 10ml,在母亲外周血浆中提取的游离 DNA 已经包含了胎儿的游离 DNA,所以母亲和胎儿只需一份样品即可。由于只需抽取孕妇的静脉血,操作简便,因此不会对孕妇和胎儿造成创伤,且孕 10 周后即可鉴定。

具体实施方式

[0025] 以下举例对本发明的一种利用 SNP 进行产前亲权关系判定的方法进行详细说明。

[0026] 1、设计位点:在人类基因组中,找出最小等位基因频率 (MAF) 在 0.4-0.5 的 SNP 位

点,在 1-13、18、21 这 15 条染色体上共选出 1035 个位点,每条染色体上 SNP 数目相近。(设计过程中位点数目可变,染色体的分布可变)。

[0027] 2、设计引物:根据设计的 1035 个 SNP 位点,利用 ION AMPLISEQ DESIGNER 在线设计网站设计出覆盖目标 SNP 的引物。(设计引物可以用其他软件代替)。

[0028] 3、样品 DNA 提取。

[0029] 4、扩增含有 SNP 位点的片段:DNA 片段、多重 PCR 反应酶和 SNP Primers 混匀后,进行 PCR 反应。

[0030] 5、纯化扩增到的 PCR 产物。

[0031] 6、末端修复:将步骤 5 所得的混和 DNA 片段、End repair Enzyme 与 5X End repair Buffer 混合,室温下孵育进行反应。

[0032] 7、片段筛选:对步骤 6 所得混合 DNA 液进行片段筛选及纯化,得到片段中的平末端 DNA 片段。

[0033] 8、接头连接:将步骤 7 所得的 DNA 片段、ligase Enzyme、10X ligase Buffer、接头、BarcodeX 混合后,室温下孵育进行反应。

[0034] 9、片段筛选:对步骤 8 所得混合 DNA 液进行片段筛选及纯化,得加接头的 DNA 片段。

[0035] 10、PCR 扩增:对步骤 9 所得混合 DNA 片段、Platinum PCR Super Mix High Fidelity 与 Library Amplification Primer Mix 混合,进行 PCR 扩增及纯化,得到小片段文库。

[0036] 11、文库检测:将步骤 10 所得的扩增产物采用 Qubit 和 Agilent Bioanalyzer2100 检测文库浓度和片段大小。

[0037] 12、高通量测序:对步骤 11 所得到的合格文库进行高通量测序。

[0038] 13、数据预处理:将高通量测序得到的数据首先经过低质量过滤,同时筛选出长度大于 100bp 的序列,因为设计的引物加上目标序列长度都大于这个。(过滤的长度可根据设计的目标区域变化)。

[0039] 14、序列比对:将上步预处理后的序列与人类基因组序列 (hg19) 用 bowtie2 进行比对。(可用其他比对软件,参考基因组可用其他版本)。

[0040] 15、序列筛选:根据上步对比的结果,筛选出错配碱基在 4 个以下,只比对到一处的序列,称之为 unique read。(错配碱基数目可有其他标准,如长度的百分之三,或者目标区域的 SNP 数目加上 1)。

[0041] 16、SNP 数据统计:根据比对结果及 SNP 位点的位置统计出每个 SNP 位点的覆盖深度、碱基种类、每种碱基对应数目。

[0042] 17、根据统计结果,对每个 SNP 位点中四种碱基出现次数进行排序,四种碱基种类按照次数从多到少的顺序分别称作 Major_alle、Minor_alle、Third_alle 及 Fourth_alle,每种类型碱基出现的次数依次为 Major_num、Minor_num、Third_num 及 Fourth_num,并得出 SNP 位点的覆盖深度 Depth, Depth 等于四种碱基数目之和,即 $Depth = Major_num + Minor_num + Third_num + Fourth_num$ 。

[0043] 18、计算最高碱基比例:计算出在 SNP 位点中出现次数最多的碱基占该位点总覆盖深度的比例,即最高碱基比例 $Major_percent = Major_num / Depth$ 。

[0044] 19、筛选位点：筛选出覆盖深度 Detph 大于 200 层，且最高碱基比大于 99% 的位点，对于待定男性样本筛选条件可设为 Detph 大于 200 层，最高碱基比大于 98%。

[0045] 20、对于孕妇样本，由于母亲和胎儿在一个 SNP 位点上可能出现 4 种情况，即具有四种 SNP 组合类型：类型 1：母亲和胎儿都是纯合 SNP；类型 2：母亲纯合 SNP，胎儿杂合 SNP；类型 3：母亲杂合 SNP，胎儿纯合 SNP；类型 4：母亲和胎儿都是杂合 SNP。由于类型 2、3、4 中都包含有杂合的情况，所以本方法目前仅选用母亲胎儿均是纯合的类型 1 来判定亲子关系，具体为当某一 SNP 位点最高碱基比大于 99% 时，该位点即为类型 1，母亲胎儿均为纯合，该最高碱基即为母亲和胎儿的基因型。

[0046] 21、对于待定男性样本，在一个 SNP 上可能出现的类型有两种，纯合或者杂合，我们使用纯合的情况来判断他将要遗传给胎儿的基因型，从而判断亲子关系，具体为当某一 SNP 位点最高碱基比大于 99% 时，该位点即为纯合，该最高碱基即为待定男性的基因型。

[0047] 22、样本对比：位点筛选后，将孕妇样本和待定男性样本在每个相同的 SNP 位点之间进行对比，最高碱基类型 Major_allele 相同则将该位点定义为一个一致位点，最高碱基类型 Major_allele 不同则将该位点定义为一个否定位点，并且统计一致位点和否定位点的个数。

[0048] 23、关系判定：当否定位点个数大于等于 5 个即可以否定胎儿与待定男性的亲缘关系；当一致位点大于等于 35 个则不能否定胎儿与待定男性的亲缘关系。

[0049] 上述方法的具体应用过程的流程结合以下实例作进一步描述。

[0050] 1、材料

[0051] 供试共计 2 对，4 个样本，样本类型为血液，样本来源为一对夫妻并确认胎儿亲缘关系，另一对为毫无关系的样本。

[0052] 高通量测序法试剂主要由 3 个试剂盒组成，文库构建试剂盒（扩增文库构建试剂盒和连接文库构建试剂盒）、测序模板制备试剂盒以及测序试剂盒。

[0053] 2、方法

[0054] 将 4 个样本按照上述实验步骤经过建库和上机处理，得到样本数据，然后进行数据分析。

[0055] 1) 数据处理

[0056] 对上步拿到的数据进行预处理，即过滤掉长度少于 100bp，或者序列中测序质量在 Q20 的比例小于 50% 的序列，二者中任一条件满足就去掉。2 个样本过滤情况如下：

[0057]

	原始 reads 数目	过滤后 reads 数目
孕妇样本 1	5, 064, 168	3, 920, 852
待定男性 1	1, 314, 876	1, 237, 172
孕妇样本 2	2, 798, 659	2, 424, 338
待定男性 2	1, 844, 651	1, 742, 523

[0058] 2) 将过滤后的序列用 bowtie2 比对到参考序列人类基因组 HG19。

[0059] 根据上步比对的结果,筛选出错配数目小于等于 4,且比对到人类基因组一个位置的 reads。

[0060] 3) 利用软件 samtools 的命令 mpileup,统计出 4 个样本每个 SNP 的碱基类型,碱基比例及覆盖深度。

[0061] 根据上步统计结果,确定每个 SNP 的覆盖深度 Depth、最高碱基比 Major_percent 和最高碱基类型 Major_allele。根据筛选条件留下覆盖深度大于 200 层,最高碱基比大于 99%的 SNP。

[0062] 4) 在上步得到的数据中,将孕妇样本和待定男性样本在每一个相同 SNP 上之间进行比较,统计一致位点和否定位点,判定亲缘关系。

[0063] a. 没有亲缘关系的样本否定位点如下:

[0064]

孕妇样本 2			待定男性 2		
depth	major_allele	major_percent	depth	major_allele	major_percent
512	C	0.994	2710	A	0.984
301	G	1	2665	C	0.992
439	T	0.991	784	C	0.983
227	T	0.996	576	A	0.995
293	A	0.997	857	G	0.998
548	C	0.993	2118	G	0.994
678	G	0.99	1747	A	0.999
403	A	0.995	1452	G	0.997
644	C	0.991	1292	T	0.998
215	C	1	616	T	0.998

[0065] 这对样本否定位点为 10 个,应否定其亲缘关系。

[0066] b. 已知亲缘关系肯定位点如下:

[0067]

孕妇样本 1			待定男性 1		
depth	major_alle	major_percent	depth	major_alle	major_percent
1044	T	0.998	986	T	1
857	T	0.994	273	T	1
335	T	1	362	T	0.994
467	G	1	689	G	0.999
257	T	0.996	1971	T	0.993
865	A	0.994	1369	A	0.996
360	A	0.994	224	A	1

[0068]

421	T	0.995	707	T	1
438	T	0.995	1142	T	0.997
393	G	0.997	672	G	1
302	A	1	612	A	0.997
447	A	0.993	1295	A	0.998
2983	C	0.995	845	C	0.993
1213	A	0.995	5102	A	0.988
304	G	1	306	G	0.991
289	A	0.997	937	A	0.997
389	G	1	920	G	1
254	C	0.992	1554	C	0.997
809	A	0.999	7668	A	0.998
328	T	1	1206	T	0.992
666	C	0.995	3424	C	0.999
1146	T	0.993	1151	T	0.998
360	T	0.992	263	T	1
298	C	0.993	519	C	0.994
466	C	0.996	1520	C	1
252	G	1	276	G	0.989
518	C	0.994	1233	C	0.994
346	T	0.993	319	T	1
421	G	0.995	226	G	0.996
446	G	0.993	2771	G	0.998
256	G	0.994	605	G	1

[0069]

549	G	0.991	1016	G	0.998
236	G	0.992	403	G	0.998
232	C	0.991	550	C	0.996
941	G	1	4336	G	0.997
1400	G	0.99	1045	G	0.998
311	G	1	302	G	1

[0070] 这对样本一致位点共 37 个,不能否定其亲缘关系。

[0071] 上述仅为本发明的一个具体实施方式,但本发明的设计构思并不局限于此,凡利用此构思对本发明进行非实质性的改动,均应属于侵犯本发明保护范围的行为。