



(21) 申请号 202410813874.2

G16B 20/30 (2019.01)

(22) 申请日 2024.06.24

G16H 50/30 (2018.01)

(71) 申请人 首都医科大学附属北京天坛医院

地址 100160 北京市丰台区南四环西路119号

申请人 深圳华大生命科学研究院

(72) 发明人 姜涛 李健康 韩洞明 刘海龙

陈璇 刘子嘉 金鑫

(74) 专利代理机构 北京高沃律师事务所 11569

专利代理师 宋春雨

(51) Int. Cl.

G16B 35/20 (2019.01)

G16B 20/40 (2019.01)

G16B 20/50 (2019.01)

G16B 20/20 (2019.01)

权利要求书1页 说明书8页

(54) 发明名称

一种筛选髓母细胞瘤新遗传易感基因的方法

(57) 摘要

本发明涉及肿瘤易感基因技术领域,特别是涉及一种筛选髓母细胞瘤新遗传易感基因的方法。本发明提供的方法筛选掉包含已知易感基因突变的家系后,留下的均为不包含已知易感基因突变的家系,在这些家系的突变数据里寻找候选易感基因,可以提高方法的可解释率(即髓母细胞瘤在分子机制上发病原因的解釋率);本发明通过筛选含有2次以上杂合性缺失变异位点的基因,并过滤掉GDI中已经报道了的正常人群基因损害值较高的基因,可以降低筛选结果的假阳性,提高筛选结果的准确性。

1. 一种筛选髓母细胞瘤新遗传易感基因的方法,其特征在于,包括以下步骤:

选择多个家系样本为待测样本;所述家系样本包括髓母细胞瘤患者的血液、髓母细胞瘤患者的肿瘤组织和髓母细胞瘤患者父母的血液;

将所述待测样本进行全基因组测序和变异位点检测,得到每个样本的VCF文件;按照家系关系将所述每个样本的VCF文件进行合并,得到家系VCF文件;

将所述家系VCF文件进行质控和注释,去除已知髓母细胞瘤易感基因发生突变的家系VCF文件,得到不含已知易感基因突变的家系VCF文件;

在所述不含已知易感基因突变的家系VCF文件中进行筛选,得到所述髓母细胞瘤新遗传易感基因;所述筛选的标准为:同时满足1)~3)的基因,具体为:

1) 出现2次以上杂合性缺失变异位点的基因;所述杂合性缺失变异位点为:变异位点在髓母细胞瘤患者的血液测序数据为杂合,肿瘤测序数据为野生型纯合或者突变型纯合,父亲和/或母亲的血液测序数据为杂合;

2) GDI为评级为Medium以下或无评级的基因;

3) 脑部平均表达量 $\geq 10$ TPM的基因。

2. 根据权利要求1所述的方法,其特征在于,所述质控的标准包括:去掉Qual By Depth < 2.0, RMS Mapping Quality < 40.0, Fisher Strand > 60.0, Strand Odds Ratio > 3.0, Mapping Quality Rank Sum Test < -12.5和Read Pos Rank Sum Test < -8.0的变异位点。

3. 根据权利要求1所述的方法,其特征在于,所述变异位点检测的变异位点包括点突变、插入突变和缺失突变。

4. 根据权利要求1所述的方法,其特征在于,所述注释的标准包括:(1) 利用已知的突变数据库,过滤掉等位基因频率大于0.01的变异位点;所述已知的突变数据库包括千人基因组计划、dbSNP、ESP6500、ExAC或gnomAD;(2) 保留变异类型为移码突变、错义突变、非移码突变、剪接突变、起始密码子丢失突变、终止密码子获得突变或终止密码子丢失突变的变异位点;(3) 保留变异等位基因频率为0.25~0.75的杂合变异位点。

5. 根据权利要求1或4所述的方法,其特征在于,所述注释的区域包括:3'非翻译区、5'非翻译区和外显子区。

6. 根据权利要求1所述的方法,其特征在于,所述已知髓母细胞瘤易感基因包括:APC、BRCA2、PALB2、PTCH1、TP53和SUFU。

7. 根据权利要求1所述的方法,其特征在于,所述多个家系样本的家系数 $\geq 43$ 。

8. 根据权利要求1所述的方法,其特征在于,所述髓母细胞瘤新遗传易感基因包括:SSPO、SPECC1、RABEP1、MTUS1、MPRIIP和CSMD1。

9. 根据权利要求1所述的方法,其特征在于,所述血液包括外周血。

## 一种筛选髓母细胞瘤新遗传易感基因的方法

### 技术领域

[0001] 本发明涉及肿瘤易感基因技术领域,特别是涉及一种筛选髓母细胞瘤新遗传易感基因的方法。

### 背景技术

[0002] 癌症易感基因已被证明在指导癌症预防策略方面具有高度的临床相关性<sup>[1]</sup>。在髓母细胞瘤(MB)中,癌症易感综合征的重要性越来越得到认可<sup>[2]</sup>。易感基因的胚系突变约占MB诊断的5~6%。遗传缺陷会导致特定分子通路的失调,从而导致肿瘤发生<sup>[3]</sup>。MB已发现与以下癌症易感综合征有关,分别是:Gorlin综合征(*PTCH1*<sup>[6]</sup>和*SUFU*<sup>[7]</sup>基因突变),Li-Fraumeni综合征(*TP53*基因突变)<sup>[4]</sup>,Fanconi anaemia(*BRAC2*基因突变)<sup>[5]</sup>,家族性腺瘤性息肉病(*APC*基因突变)<sup>[6]</sup>。目前MB已经验证了的易感基因有六个分别是:*APC*,*BRCA2*,*PALB2*,*PTCH1*,*SUFU*和*TP53*,这些易感基因作用于对MB的不同亚型,其中,*APC*影响Wnt/ $\beta$ -catenin通路,*BRCA2*和*PALB2*影响DNA修复通路,*PTCH1*和*SUFU*影响Hedgehog通路,*TP53*影响细胞周期调控和DNA损伤修复通路。这些基因的突变或异常表达可能导致相应通路的失控,从而影响细胞增殖、分化、生长和凋亡等生物学进程,进而导致MB的发生,对MB遗传咨询和检测提供了指导<sup>[3]</sup>。但是目前髓母细胞瘤易感基因的胚系突变约占MB诊断率仅为5~6%,仍有大量的髓母患者无法在易感基因的分子机制上解释发病原因。

[0003] 参考文献:

[1] Mark A Glaire, Matthew Brown, David N Church, Ian Tomlinson. Cancer predisposition syndromes: lessons for truly precision medicine[J]. The Journal of Pathology,2017,241(2).

[0004] [2] Carta Roberto, et al. Cancer Predisposition Syndromes and Medulloblastoma in the Molecular Era[J]. Frontiers in oncology,2020,10.

[0005] [3] Waszak SM, Northcott PA, Buchhalter I, Robinson GW, Sutter C, Groebner S, et al. Spectrum and prevalence of genetic predisposition in medulloblastoma: a retrospective genetic study and prospective validation in a clinical trial cohort[J]. The Lancet Oncology,2018,19(6).

[0006] [4] Li FP, Fraumeni JF. Rhabdomyosarcoma in children: epidemiologic study and identification of a familial cancer syndrome[J]. Natl Cancer Inst 1969; 43: 1365-73.

[0007] [5] Offit K, Levrano O, Mullaney B, et al. Shared genetic susceptibility to breast cancer, brain tumors, and Fanconi anemia[J]. Natl Cancer Inst 2003; 95: 1548-51.

[0008] [6] Hamilton SR, Liu B, Parsons RE, et al. The molecular basis of Turcot's syndrome[J]. N Engl J Med 1995; 332: 839-47.

## 发明内容

[0009] 为了解决上述问题,本发明提供了一种筛选髓母细胞瘤新遗传易感基因的方法。本发明提供的方法可以在髓母细胞瘤群体的全基因组数据里准确地筛选检测出新的遗传易感基因,筛选结果假阳性低,对MB遗传咨询和检测提供了技术支持。

[0010] 为了实现上述目的,本发明提供如下技术方案:

本发明提供了一种筛选髓母细胞瘤新遗传易感基因的方法,包括以下步骤:

选择多个家系样本为待测样本;所述家系样本包括髓母细胞瘤患者的血液、髓母细胞瘤患者的肿瘤组织和髓母细胞瘤患者父母的血液;

将所述待测样本进行全基因组测序和变异位点检测,得到每个样本的VCF文件;按照家系关系将所述每个样本的VCF文件进行合并,得到家系VCF文件;

将所述家系VCF文件进行质控和注释,去除已知髓母细胞瘤易感基因发生突变的家系VCF文件,得到不含已知易感基因突变的家系VCF文件;

在所述不含已知易感基因突变的家系VCF文件中进行筛选,得到所述髓母细胞瘤新遗传易感基因;所述筛选的标准为:同时满足1)~3)的基因,具体为:

1) 出现2次以上杂合性缺失变异位点的基因;所述杂合性缺失变异位点为:变异位点在髓母细胞瘤患者的血液测序数据为杂合,肿瘤测序数据为野生型纯合或者突变型纯合,父亲和/或母亲的血液测序数据为杂合;

2) GDI为评级为Medium以下或无评级的基因;

3) 脑部平均表达量 $\geq 10$ TPM的基因。

[0011] 优选的,所述质控的标准包括:去掉Qual By Depth $< 2.0$ ,RMS Mapping Quality $< 40.0$ ,Fisher Strand $> 60.0$ ,Strand Odds Ratio $> 3.0$ ,Mapping Quality Rank Sum Test $< -12.5$ 和Read Pos Rank Sum Test $< -8.0$ 的变异位点。

[0012] 优选的,所述变异位点检测的变异位点包括点突变、插入突变和缺失突变。

[0013] 优选的,所述注释的标准包括:(1) 利用已知的突变数据库,过滤掉等位基因频率大于0.01的变异位点;所述已知的突变数据库包括千人基因组计划、dbSNP、ESP6500、ExAC或gnomAD;(2) 保留变异类型为移码突变、错义突变、非移码突变、剪接突变、起始密码子丢失突变、终止密码子获得突变或终止密码子丢失突变的变异位点;(3) 保留变异等位基因频率为0.25~0.75的杂合变异位点。

[0014] 优选的,所述注释的区域包括:3'非翻译区、5'非翻译区和外显子区。

[0015] 优选的,所述已知髓母细胞瘤易感基因包括:APC、BRCA2、PALB2、PTCH1、TP53和SUFU。

[0016] 优选的,所述多个家系样本的家系数 $\geq 43$ 。

[0017] 优选的,所述髓母细胞瘤新遗传易感基因包括:SSPO、SPECC1、RABEP1、MTUS1、MPRIP和CSMD1。

[0018] 优选的,所述血液包括外周血。

[0019] 有益效果:

本发明提供的方法筛选掉包含已知易感基因突变的家系后,留下的均为不包含已知易感基因突变的家系,在这些家系的突变数据里寻找候选易感基因,可以提高方法的可解释率(即髓母细胞瘤在分子机制上发病原因的解釋率);本发明通过筛选含有2次以上杂

合性缺失变异位点的基因,并过滤掉GDI中已经报道了的正常人群基因损害值较高的基因,可以降低筛选结果的假阳性,提高筛选结果的准确性。

### 具体实施方式

[0020] 本发明提供了一种筛选髓母细胞瘤新遗传易感基因的方法,包括以下步骤:

选择多个家系样本为待测样本;所述家系样本包括髓母细胞瘤患者的血液、髓母细胞瘤患者的肿瘤组织和髓母细胞瘤患者父母的血液;

将所述待测样本进行全基因组测序和变异位点检测,得到每个样本的VCF文件;按照家系关系将所述每个样本的VCF文件进行合并,得到家系VCF文件;

将所述家系VCF文件进行质控和注释,去除已知髓母细胞瘤易感基因发生突变的家系VCF文件,得到不含已知易感基因突变的家系VCF文件;

在所述不含已知易感基因突变的家系VCF文件中进行筛选,得到所述髓母细胞瘤新遗传易感基因;所述筛选的标准为:同时满足1)~3)的基因,具体为:

1) 出现2次以上杂合性缺失变异位点的基因;所述杂合性缺失变异位点为:变异位点在髓母细胞瘤患者的血液测序数据为杂合,肿瘤测序数据为野生型纯合或者突变型纯合,父亲和/或母亲的血液测序数据为杂合;

2) GDI为评级为Medium以下或无评级的基因;

3) 脑部平均表达量 $\geq 10$ TPM的基因。

[0021] 本发明选择多个家系样本为待测样本;所述家系样本包括髓母细胞瘤患者的血液、髓母细胞瘤患者的肿瘤组织和髓母细胞瘤患者父母的血液。在本发明中,所述多个家系样本的家系数优选为 $\geq 43$ ;所述血液优选包括外周血。

[0022] 得到待测样本后,本发明将所述待测样本进行全基因组测序和变异位点检测,得到每个样本的VCF文件。在本发明中,所述变异位点检测的变异位点包括点突变、插入突变和缺失突变。本发明对所述全基因组测序和变异位点检测的方法没有特殊要求,选择本领域技术人员所熟知的方法或委托基因检测公司进行检测即可。

[0023] 得到每个样本的VCF文件后,本发明按照家系关系将所述每个样本的VCF文件进行合并,得到家系VCF文件。在本发明中,所述合并所使用的工具优选包括GATK软件。

[0024] 得到家系VCF文件后,本发明将所述家系VCF文件进行质控和注释,去除已知髓母细胞瘤易感基因发生突变的家系VCF文件,得到不含已知易感基因突变的家系VCF文件。在本发明中,所述已知髓母细胞瘤易感基因优选包括:APC、BRCA2、PALB2、PTCH1、TP53和SUFU;所述质控的标准优选包括:去掉Qual By Depth(QD) $< 2.0$ ,RMS Mapping Quality(MQ) $< 40.0$ ,Fisher Strand(FS) $> 60.0$ ,Strand Odds Ratio(SOR) $> 3.0$ ,Mapping Quality Rank Sum Test(MQRankSum) $< -12.5$ 和Read Pos Rank Sum Test(ReadPosRankSum) $< -8.0$ 的变异位点。本发明通过去除已知髓母细胞瘤易感基因发生突变的家系,留下均为不包含已知易感基因突变的家系,在这些家系的突变数据里寻找候选易感基因,避免已知易感基因对新易感基因的干扰,可以提高本发明的准确性。

[0025] 本发明所述QD是变异位点的质量值(Quality)除以覆盖深度(Depth)得到的比值,这里的变异质量值就是VCF中QUAL的值,用来衡量变异的可靠程度,这里的覆盖深度是这个位点上所有含有变异碱基的样本的覆盖深度之和。

[0026] 本发明所述MQ是所有比对至该位点上的reads的比对质量值的均方根。

[0027] 本发明所述FS和SOR这两个指标是通过Fisher检验后得到的P值,描述的是测序或者比对时对于只含有变异的reads以及只含有参考序列碱基的reads是否存在着明显的正负链特异性(Strand bias,或者说是偏好性)。这个差异反应了测序过程中随机性不够,或者是比对算法在基因组的某些区域存在一定的选择偏好。

[0028] MQRankSum是一个变异质量评分指标,用于评估变异位点的质量。MQRankSum是Mapping Quality Rank Sum Test的缩写,它是一种非参数的Wilcoxon秩和检验,用来比较变异位点上参考碱基和变异碱基的测序质量。具体来说,MQRankSum检验会比较两组数据:一组是携带参考等位基因的读段的映射质量(mapping quality),另一组是携带变异等位基因的读段的映射质量。这个检验的目的是检测是否存在显著的映射质量差异,这可能表明变异的检测可能受到了测序或映射错误的影响。在GATK的变异过滤过程中,MQRankSum可以作为一个过滤标准,帮助本发明识别和排除那些可能由于技术问题而被错误识别的变异。

[0029] ReadPosRankSum也是一个变异检测的质量评分指标,用于评估变异位点的质量。ReadPosRankSum是Read Position Rank Sum Test的缩写,它是一种非参数的Wilcoxon秩和检验,用来比较变异位点上参考碱基和变异碱基的读段在其测序读段中的相对位置,可以帮助本发明识别和排除那些可能由于读段位置偏差而被错误识别的变异。

[0030] 在本发明中,所述注释的区域优选包括:3'非翻译区(3'UTR)、5'非翻译区(5'UTR)和外显子区(Exonic);所述注释的标准优选包括:(1)利用已知的突变数据库,过滤掉等位基因频率大于0.01的变异位点;所述已知的突变数据库包括千人基因组计划、dbSNP、ESP6500、ExAC或gnomAD;(2)保留变异类型为移码突变、错义突变、非移码突变、剪接突变、起始密码子丢失突变、终止密码子获得突变或终止密码子丢失突变的变异位点;(3)保留变异等位基因频率(alt/DP)为0.25~0.75的杂合变异位点。

[0031] 本发明的名词释义如下:

移码(Frame shift)突变:指插入或缺失的碱基数不是3的倍数,导致从变异点开始之后的阅读框改变,可能产生完全不同的蛋白质。

[0032] 错义(Missense)突变:指一个碱基的改变导致编码的氨基酸发生改变。中文名为“错义变异”。

[0033] 非移码(Nonframeshift)突变:指插入或缺失的碱基数是3的倍数,不会改变阅读框,但可能在蛋白质中增加或减少氨基酸。

[0034] 剪接(Splicing)突变:指变异发生在基因的剪接位点,可能影响RNA的剪接过程,导致异常的mRNA和蛋白质产物。

[0035] 起始密码子丢失(Startloss)突变:指变异导致起始密码子(通常是ATG)的丢失,可能导致蛋白质的合成无法开始。

[0036] 终止密码子获得(Stopgain)突变:指变异导致一个新的终止密码子的产生,导致蛋白质提前终止。

[0037] 终止密码子丢失(Stoploss)突变:指变异导致正常的终止密码子丢失,可能导致蛋白质延长。

[0038] alt:代表变异等位基因(Alternate Allele)的读数,即在该变异位点上检测到的

非参考等位基因的测序读段数。DP:代表该位点的总深度 (Depth of Coverage), 即在该变异位点上的所有测序读段数 (包括参考等位基因和变异等位基因的读段)。

[0039] alt/DP这个比例反映了变异等位基因在总测序深度中的比例,也就是变异等位基因的频率。在过滤变异数据时,设定一个阈值范围来选择那些等位基因频率在特定范围内的变异,以排除可能的假阳性或假阴性。

[0040] 如果alt/DP小于0.25,这可能意味着变异等位基因的频率太低,可能是由于测序错误或者污染导致的假阳性。

[0041] 如果alt/DP大于0.75,这可能意味着变异等位基因的频率太高,可能是由于参考基因组不准确或者样本中有多个拷贝的情况。

[0042] 因此,通过设定 $0.25 \leq \text{alt/DP} \leq 0.75$ 这个过滤条件,本发明可以选择那些等位基因频率适中、更可能是真实存在的变异。这样的过滤标准有助于提高变异检测的准确性。

[0043] 在所述不含已知易感基因突变的家系VCF文件中进行筛选,得到所述髓母细胞瘤新遗传易感基因;所述筛选的标准为:同时满足1)~3)的基因,具体为:

1) 出现2次以上杂合性缺失变异位点的基因;所述杂合性缺失变异位点为:变异位点在髓母细胞瘤患者的血液测序数据为杂合,肿瘤测序数据为野生型纯合或者突变型纯合,父亲和/或母亲的血液测序数据为杂合;

2) GDI为评级为Medium以下或无评级的基因;

3) 脑部平均表达量 $\geq 10$ TPM的基因。

[0044] 本发明所述出现2次以上杂合性缺失变异位点的基因是指在一个基因不同位置上发现了两次以上的所述杂合性缺失变异位点,该杂合性缺失变异位点可以是相同类型的变异位点,也可以是不同类型的变异位点。

[0045] 基因损伤指数 (GDI) 是健康人群中每个人类基因的累积突变损伤,基于健康个体的千人基因组计划数据库 (第3阶段) 基因变异。前人研究证明,在正常人群里经常高度受损的人类基因不太可能导致疾病。通过选择GDI为评级为Medium以下或无评级的基因,可以过滤掉不太可能致病的基因高频变异。

[0046] 本发明所述选脑部平均表达量是指脑部bulk-RNA的表达平均值。由于髓母细胞瘤发病于脑部,筛选在脑部高表达的基因,其筛选得到的髓母细胞瘤新遗传易感基因准确度更高。

[0047] 本发明筛选得到的所述髓母细胞瘤新遗传易感基因优选包括:SSPO、SPECC1、RABEP1、MTUS1、MPRIP和CSMD1。

[0048] 为了进一步说明本发明,下面结合实施例对本发明提供的一种筛选髓母细胞瘤新遗传易感基因的方法进行详细地描述,但不能将它们理解为对本发明保护范围的限定。

[0049] 实施例1

[0050] 1、病例纳入

[0051] 招募了髓母细胞瘤病例300例,家系样本43组,每组都包含患者肿瘤样本、血液样本、父母血液样本,所有肿瘤样本均经两名高级职称病理医师独立诊断确认。本项目涉及人类遗传资源采集和保藏,经首都医科大学附属北京天坛医院伦理委员会批准,所有入组患者均签署知情同意书,并遵照《中华人民共和国人类资源管理条例》报科技部国家人类遗传资源办公室审批。

[0052] 2、全基因组测序和变异检测

[0053] 取样:取了患者和父母的外周血和患者的肿瘤组织,提取DNA,DNA总量 $\geq 6 \mu\text{g}$ ,浓度: $\geq 50 \text{ ng}/\mu\text{l}$ ,纯度: $\text{OD}_{260/280} = 1.8\sim 2.0$ 。

[0054] 委托深圳华大生命科学研究院进行全基因组测序和测序数据变异检测,包括如下步骤:

全基因组测序:①将基因组DNA用超声波进行随机打断;②打断的DNA片段进行末端修复和接头的连接,构建DNA文库;③桥式PCR对文库DNA进行扩增;④对扩增后的DNA文库采用双端测序在MGISEQ-2000平台上进行测序和数据产出;

测序数据变异检测;①下机数据质控(QC):设置过滤阈值为Q30,除去低质量reads序列,去掉带有完整测序接头(adapter)的reads序列,去除N碱基的含量超过该条read长度比例的10% reads序列,丢弃长度小于100bp的reads序列,采用的工具有SAOPnuke、FastQC等;②比对:利用SOAP和BWA等软件进行测序得到的短reads片段比对到人类基因组(hg38)参考序列上。③变异检测:基于比对的结果,利用SOAPsnp、GATK等软件进行点突变、插入缺失突变检测。得到每个样本的VCF后,按照家系关系用GATK进行合并,得到合并后的家系VCF。

[0055] 3、变异的质控、注释和致病性分类

[0056] 变异数据VCF的质控:对全基因组变异数据做硬过率,人为设定一个或者若干个指标阈值(也可以叫数据特征值),然后把所有不满足阈值的变异位点舍弃掉。其中指标有5个,Qual By Depth(QD),Fisher Strand(FS),Strand Odds Ratio(SOR),RMS Mapping Quality(MQ),Mapping Quality Rank Sum Test(MQRankSum),设置的阈值为 $\text{QD} < 2.0$ ,  $\text{MQ} < 40.0$ ,  $\text{FS} > 60.0$ ,  $\text{SOR} > 3.0$ ,  $\text{MQRankSum} < -12.5$ ,  $\text{ReadPosRankSum} < -8.0$ 。

[0057] 质控后的变异信息注释:硬过率VCF后本发明得到了可信度较高的变异信息,使用kkgseq软件对质控后的VCF进行基因的注释,软件中使用以下一组参数执行序列变异的质量控制:(1)群体等位基因频率过滤,包括千人基因组计划,dbSNP,ESP6500,ExAC和gnomAD,过滤掉频率大于0.01的变异;(2)保留3UTR,5UTR,exonic区域变异类型为:frameshift,missense,nonframeshift,splicing,startloss,stopgain,stoploss的变异位点;(3)杂合变异位点满足 $0.25 \leq \text{alt}/\text{DP} \leq 0.75$ 。

[0058] 4、排除携带已知易感基因突变的家系

[0059] 根据前期广泛的文献调研,得知MB群体的易感基因有6个,具体如下:APC、BRCA2、PALB2、PTCH1、TP53和SUFU,鉴定过滤后的家系注释信息是否包含这些已知易感基因的突变,排除注释信息里包含已知易感基因的突变家系,在43个家系中找到了16个家系包含已知易感基因突变,有27个家系里没有找到已知6个易感基因的突变。

[0060] 5、无已知易感基因突变家系中杂合性缺失(Loss of heterogeneity,LOH)分析

[0061] 在没有已知易感基因突变的家系中,筛选变异位点里,患者血液数据杂合,肿瘤数据为野生型纯合或者突变型纯合,父母血液数据有任一方为杂合,满足这些条件的变异位点,即初步得到了MB易感基因的新候选位点,得到246个基因。

[0062] 6、人类基因损伤指数(GDI)评估

[0063] 利用add\_GDI.py(<https://lab.rockefeller.edu/casanova/GDI/programs>)筛选GDI为评级为Medium以下的和GDI中无评级的基因,在初步LOH分析得到的246个基因中,筛

选留下了102个候选易感基因。

[0064] 7、确定MB新的遗传易感基因列表

[0065] 通过核心家系的全基因组测序(WGS)分析,使用成人非癌症gnomAD对照队列进行家系遗传共分离分析,每组家系分别包含先证者的血液和肿瘤全基因组变异数据和其父母的全基因组变异数据,排除掉患者中前人报道过的突变易感基因后,筛选留下了102个候选易感基因;再从102个候选易感基因中筛选出家系群体里出现过两次以上的杂合性缺失(LOH)的基因,得到LOH高频基因列表(表1),经过GDI数据库过滤,候选基因筛选范围缩小至damage score值为NA或者medium以下的基因(表2),筛选脑部高表达(bulk-RNA的表达平均值 $\geq 10$ TPM)的基因,最后筛选出如下6个基因列表(表3)。

[0066] 表1为LOH高频基因列表

携带杂合性缺失的家系	家系总数	杂合性缺失频率	基因
7	27	0.26	<i>MUC3A</i>
4	27	0.15	<i>ZZEF1</i>
4	27	0.15	<i>PKHD1L1</i>
4	27	0.15	<i>HLA-DQA1</i>
3	27	0.11	<i>WDR81</i>
3	27	0.11	<i>TTN</i>
3	27	0.11	<i>LOC112268354</i>
2	27	0.07	<i>TRIM16</i>
2	27	0.07	<i>SSPO</i>
2	27	0.07	<i>SPECC1</i>
2	27	0.07	<i>SLC5A10</i>
2	27	0.07	<i>SLC3A1</i>
2	27	0.07	<i>RPH3AL</i>
2	27	0.07	<i>RABEP1</i>
2	27	0.07	<i>MYH13</i>
2	27	0.07	<i>MTUS1</i>
2	27	0.07	<i>MPRIP</i>
2	27	0.07	<i>MFSD4B</i>
2	27	0.07	<i>CSMD1</i>

[0067] 表2为GDI数据库过滤得到的结果

基因	人类基因损伤指数	人类基因损伤指数Phred	损伤预测
<i>MUC3A</i>	4720	13.877	High
<i>ZZEF1</i>	5318.261	15.074	High
<i>PKHD1L1</i>	12241.382	23.722	High
<i>HLA-DQA1</i>	16622.239	27.598	High
<i>WDR81</i>	3904.136	12.337	High
<i>TTN</i>	74772.866	42.913	High
<i>LOC112268354</i>	NA	NA	NA
<i>TRIM16</i>	533.078	4.711	Medium
<i>SSPO</i>	NA	NA	NA
<i>SPECC1</i>	1303.955	6.795	Medium
<i>SLC5A10</i>	259.295	3.461	Medium
<i>SLC3A1</i>	899.715	5.843	Medium
<i>RPH3AL</i>	298.799	3.687	Medium
<i>RABEP1</i>	184.904	2.952	Medium
<i>MYH13</i>	5547.032	15.4	High
<i>MTUS1</i>	1447.427	7.096	Medium
<i>MPRIP</i>	NA	NA	NA
<i>MFSD4B</i>	NA	NA	NA
<i>CSMD1</i>	1048.097	6.215	Medium

[0068] 表3 MB新遗传易感基因列表

基因名	细胞遗传学定位	基因组坐标 (GRCh38)
<i>SSPO</i>	7q36.1	7:149,776,042-149,833,965
<i>SPECCI</i>	17p11.2	17:20,009,359-20,319,026
<i>RABEP1</i>	17p13.2	17:5,282,284-5,386,340
<i>MTUS1</i>	8p22	8:17,643,802-17,801,521
<i>MPRIP</i>	17p11.2	17:17,042,457-17,192,643
<i>CSMD1</i>	8p23.2	8:2,935,361-4,994,914

[0069] 尽管上述实施例对本发明做出了详尽的描述,但它仅仅是本发明一部分实施例,而不是全部实施例,人们还可以根据本实施例在不经创造性前提下获得其他实施例,这些实施例都属于本发明保护范围。