

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2016 年 6 月 16 日 (16.06.2016)



(10) 国际公布号
WO 2016/090584 A1

(51) 国际专利分类号:
C12Q 1/68 (2006.01)

(21) 国际申请号: PCT/CN2014/093515

(22) 国际申请日: 2014 年 12 月 10 日 (10.12.2014)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 深圳华大基因研究院 (BGI SHENZHEN)
[CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(72) 发明人: 杨明 (YANG, Ming); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。赵鑫 (ZHAO, Xin); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。刘耿 (LIU, Geng); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。刘敬一 (LIU, Jingyi); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。侯勇 (HOU, Yong); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。吴迪 (WU, Kui); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。李波 (LI, Bo); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(74) 代理人: 北京清亦华知识产权代理事务所 (普通合伙) (TSINGYIHUA INTELLECTUAL PROPERTY LLC); 中国北京市海淀区清华园清华大学照澜院商业楼 301 室, Beijing 100084 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告 (条约第 21 条 (3))。

(54) Title: METHOD AND DEVICE FOR DETERMINING CONCENTRATION OF TUMOR NUCLEIC ACID

(54) 发明名称: 确定肿瘤核酸浓度的方法和装置

(57) Abstract: Disclosed is a method for determining the content of tumor nucleic acid. The tumor nucleic acid comes from a body fluid sample, the body fluid sample comprises free nucleic acid, and the free nucleic acid comprises tumor nucleic acid and non-tumor nucleic acid. The method comprises: measuring the sequence of at least one part of the free nucleic acid to obtain measurement data, the measurement data comprising multiple read segments; detecting mutation sites in the free nucleic acid on the basis of the measurement data; screening out the mutation sites with the mutation frequency obviously different from the mutation frequency of the same sites in a negative check sample to obtain tumor mutation sites; and determining the content of the tumor nucleic acid in the body fluid sample on the basis of the supported read segment quantity of at least one of the tumor mutation sites. Also disclosed are a device for determining the content of tumor nucleic acid, and a computer readable storage medium.

(57) 摘要: 本发明公开一种确定肿瘤核酸含量的方法, 肿瘤核酸来自体液样本, 体液样本包含游离核酸, 游离核酸包括肿瘤核酸和非肿瘤核酸, 所述方法包括: 对游离核酸的至少一部分进行序列测定, 获得测定数据, 所述测定数据包括多个读段; 基于所述测定数据, 检测所述游离核酸中的变异位点; 筛选出突变频率与阴性对照样本中相同位点的突变频率具有显著差异的变异位点, 获得肿瘤变异位点; 基于至少一个所述肿瘤变异位点的支持读段量, 确定所述体液样本中的肿瘤核酸含量。本发明还公开一种确定肿瘤核酸含量的装置和一种计算机可读存储介质。

WO 2016/090584 A1

确定肿瘤核酸浓度的方法和装置

技术领域

本发明涉及生物信息和生物医学领域，具体的，本发明涉及一种确定肿瘤核酸含量的方法、一种确定肿瘤核酸含量的装置和一种计算机可读存储介质。

5 背景技术

20 世纪 70 年代以来，我国癌症发病及死亡率一直呈上升趋势，至 90 年代的 20 年间，癌症死亡率上升 29.42%，年龄调整死亡率上升 11.56%。2000 年癌症发病人数约 180-200 万，死亡 140-150 万。恶性肿瘤发病率全国 35 岁至 39 岁年龄段为 87.07/10 万，40 岁至 44 岁年龄段几乎翻番，达到 154.53/10 万；50 岁以上人群发病占全部发病的 80%以上，60 岁
10 以上癌症发病率超过 1%，80 岁达到高峰。全国肿瘤死亡率为 180.54/10 万，每年因癌症死亡病例达 270 万例。我国居民因癌症死亡的几率是 13%[赫捷, 陈万青. (2012). 2012 中国肿瘤登记年报. 军事医学科学出版社:18-28., McBride, D. J. and A. K. Orpana, et al. (2010). "Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors." Genes Chromosomes Cancer 49(11): 1062-1069.]。癌症不仅严重影响劳动
15 人民的健康，而且成为医疗费用上涨的重要因素。我国每年耗费大量资金用于癌症病人的医疗费用，带来巨大的经济负担。

近年来，肺癌成为全球发病率和死亡率最高的癌种。在我国，肺癌为全国发病率和死亡率最高的恶性肿瘤，其发病率为 53.57/10 万，死亡率为 45.57/10 万，男性高于女性。其中 80%以上为非小细胞肺癌（NSCLC）、其余为小细胞肺癌（SCLC）。不同类型的肺癌
20 对治疗的反应不同，需要针对病理分期和组织学分型选择最佳治疗方案。如，EGFR 突变阳性肿瘤对厄洛替尼或吉非替尼有较好敏感性而对其他化疗药物敏感性较差。而临床上进行病理分期和组织学分型常采用纵隔镜、胸腔镜、PET/CT 扫描、MRI、组织穿刺活检等方法，不仅对患者造成创伤，而且操作繁琐、分辨率有限。

患者血液中，含有组织细胞释放的游离的 DNA（cell-free DNA，cfDNA），其中包含
25 肿瘤组织释放的肿瘤 DNA（circulating tumor DNA，ctDNA），癌症患者血浆游离 DNA 含有与肿瘤相关的多种突变基因。研究表明，ctDNA 的平均长度为 140-170bp，其基因突变情况能直接或者辅助临床反映肿瘤的进展情况。随着第二代测序技术的发展，基于基因组变异检测外周血样本的测序技术已经应用于产前诊断等临床层面。

Complete Genomics (CG)测序是一种新的二代测序技术，它采用高密度 DNA 纳米芯片
30 技术，在芯片上嵌入 DNA 纳米球，用非连续、非连锁联合探针锚定（cPAL）技术读取序列。HapMap 计划中一位 Caucasian 男性和一位 Yoruban 女性基因组利用 CG 技术进行重测序，检测出 94%该计划报道的 SNP 位点，其一致性达 99.15%。同时与 ILLUMINA 测序 SNP

分析结果相比，检测出其 96%SNP 位点，一致性高达 99.88%。CG 测序可避免测序中错误的积累，其准确性较合成法和连接法测序高，可达 99.999%。

发明内容

本发明旨在至少在一定程度上解决上述技术问题之一或至少提供一种商业选择。

5 依据本发明的一方面，本发明提供一种确定肿瘤核酸含量的方法，所述肿瘤核酸来自体液样本，所述体液样本包含游离核酸，所述游离核酸包括肿瘤核酸和非肿瘤核酸，所述方法包括：对所述游离核酸的至少一部分进行序列测定，获得测定数据，所述测定数据包括多个读段；基于所述测定数据，检测所述游离核酸中的变异位点；筛选出突变频率与阴性对照样本中相同位点的突变频率具有显著差异的变异位点，获得肿瘤变异位点；基于所有所述肿瘤变异位点的支持读段量，确定所述体液样本中的肿瘤核酸含量。所说的序列测定包括，对所述游离核酸的至少一部分进行测序文库构建，获得测序文库，以及对所述测序文库进行测序，获得测定数据。测序可选用的测序平台包括但不限于 Complete Genomics 公司 (CG) 的 CGA、Illumina 的 Hiseq2000/2500、Life Technologies 的 Ion Torrent 和 Roche 454 测序平台，依据所选测序平台进行相应的测序文库制备，接着将测序文库上机测序，
15 就能获得所说的测定数据。

在本发明的一个实施例中，利用 CG 的文库构建方法及测序平台来获得测定数据，CG 平台通过酶切构建多接头配对末端文库，利用特有的组合探针连接测序 (cPAL) 技术对所构建的环状文库进行测序，测读出接头两旁的碱基，因为其是利用酶切连接一个接头的两段来进行配对末端文库构建的，由于每一种酶都有一个首选的切割距离，而在实际酶切时
20 经常比首选距离多一个位置或少一个位置，这样使得读段中经常带有缺口 (gap)，缺口常为 +1nt 或者 -1nt，和/或，建库时倘若使用同一种酶多次酶切，每次的酶切位置易发生变化，酶切位置的变化也会使获得的读段带有缺口，例如在构建多接头单链环状文库时，利用 Alu 酶两次酶切来连接多个接头的不同部分，读测这些接头旁的碱基时，会产生带 +3/-3 的缺口的读段，文库构建可参考 US7897344，获得多接头单链环状文库。读段带负缺口显示为该
25 读段的两部分的末端有重叠，读段带正缺口显示为该读段的两部分的末端非连续。在本发明中，所说的缺口的大小还可以是 0。以 CG 平台目前的双接头 (two adaptors, 2-AD) 测序文库为例，2-AD 测序输出总长为 60bp，可分为两对读段对 (mate-paired reads)，每对读段对中的读段在 10bp 的位置都有小的 gap，在 20bp 位置有一个无效测序位点 N，一对读段对的两个读段之间在基因组上的距离一般小于 2000bp。来自多接头文库中的多个读段，一个
30 读段可以和任一其它读段组成一对读段对。在本发明的一个实施例中，对于血液样本中的游离核酸，通过参考改进 CG 的建库技术，利用一个接头的两部分分别连接于一段核酸片段 (插入片段) 的两端，单链分离，单链成环，获得 1-接头单链环状文库，该 1-接头单链

环状文库由所述核酸片段的一条链和连接所述一条链的两端的一个预定 DNA 序列构成, 滚环扩增形成 DNA 纳米球 (DNB), 利用 CG 测序 cPAL 技术对 DNB 进行测序, 关于 DNB 种植在芯片上以及 cPAL 技术可分别参考 US8278039B2 和 US8518640B2。所说的预定 DNA 序列为已知序列, 为所说的接头或者接头的一条链。对所述单链单接头环状文库进行测序, 获得测定数据, 测定数据包含多对读段对, 一对读段对包含两个读段, 两个读段分别来自所述单接头环状单链文库的接头两旁序列, 两个读段在文库上的距离即为插入片段的大小, 将一对读段对的两个读段分别称为左臂和右臂。构建所述单接头环状单链文库, 包括: (1) 对所述游离核酸进行末端磷酸化, 获得末端磷酸化产物; (2) 末端修复所述末端磷酸化产物, 获得末端修复产物; (3) 将第一序列和第二序列连接至所述末端修复产物的两端, 获得第一连接产物; (4) 利用第三序列对所述连接产物进行缺刻平移和扩增, 获得扩增产物, 所述第三序列为一对引物对, 所述引物对的至少一条引物带有生物素标记; (5) 利用所述生物素标记对所述扩增产物进行单链分离, 获得单链产物; (6) 利用第四序列环化所述单链产物, 获得所述测序文库; 其中, 所述第四序列能够连接所述第一序列的一端和所述第二序列的一端, 所述第一序列和/或所述第二序列的另一端为双脱氧核苷酸, 其中的步骤(1)和(2)无顺序限制。末端修复是为获得平末端核酸片段, 使得能够连接其它核苷酸或序列。末端磷酸化是为了减少样本核酸片段的互相连接, 使得核酸含量很低的样本也能够进行文库构建且满足文库上机量要求。根据 CG 建库方式不同, 每次测序能产生的数据量也不同, 例如采用所说的改进的 1-adapter 文库能产生 1080G 数据, 相当于 360 个人全基因组的数据量, 远大于 ILLUMINA、proton 等测序平台。CG 其更高通量的特点使每个基因组测序的费用低至 1000 美元, 并且还在不断降低[Drmanac, R. and A. B. Sparks, et al. (2010). "Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays." science 327(5961): 78-81.]. 利用 CG 测序对肿瘤血浆样本基因突变进行检测, 是一种经济、准确、对患者创伤小或无创的检测技术, 具有很高的推广价值, 将被广泛应用在临床疾病的筛查、治疗和监控中。单接头环状单链文库如图 1 所示, 单接头建库量较小, 适合 cfDNA 含量较少的情况, 此外还有建库时间短及建库成本低的优点。所说的第四序列能够连接第一序列和的第二序列形成所说的接头, 缺刻平移是为消除连接在末端修复产物两端的第一序列和/或第二序列的另一端的双脱氧核苷酸造成的缺刻 (nick), 利用至少一条引物带有生物素标记使扩增产物的至少一条链带有生物素标记, 使后续易于基于该生物素标记分离获得单链产物。在本发明的一个实施例中, 对构建的文库进行测序是利用组合探针锚定连接测序技术进行的, 例如利用 CG 测序平台进行。

在本发明的一个实施例中, 利用 CG 平台获得测定数据之后, 所述基于测定数据检测所述游离核酸中的变异位点, 包括: 将所述测定数据与参考序列进行比对, 获得比对结果;

消除所述比对结果中的读段的缺口，获得通用比对结果；基于所述通用比对结果，检测所述变异位点；其中，所述比对结果和/或通用比对结果包括多个所述读段对的比对结果，和/或，所述比对结果和/或通用比对结果包括多个所述左臂的比对结果和多个所述右臂的比对结果。比对可以利用已知比对软件进行，比如 SOAP、BWA 等，也可以利用 CG 平台的比对软件 TeraMap 进行。在本发明的一个实施例中，比对是利用 TeraMap 进行的，所得的比
5 对结果的格式为 TeraMap。在本发明的一个实施例中，所说的消除比对结果中读段的缺口是指，对带负缺口的读段去除掉其负缺口即去除掉重叠的碱基，对带正缺口的读段以 N 替代正缺口的大小，N 为 A、T、C 或 G，例如，对带负缺口比如为-2nt 的读段，基于缺口处该读段可分成两部分，两部分的末端有 2nt 重叠，比如该读段的两部分分别为
10 ATCGCTTAAG 和 AGTACGATTC，消除其负缺口即重叠的 AG，获得对应的读段为 ATCGCTTAAGTACGATTC。

在本发明的一个实施例中，所说的比对包括：将每对读段对的左臂和右臂分别与所述参考序列比对，获得一级左比对结果和一级右比对结果；分别以所述一级左比对结果和所述一级右比对结果的其中一个为参考，对另一个作比对，获得二级左比对结果和二级右比
15 对结果；基于所述二级左比对结果和所述二级右比对结果，获得多个所述读段对的比对结果，或者获得多个所述左臂的比对结果和多个所述右臂的比对结果。这样经过两次比对，可以获得准确的读段对比对结果，在本发明的一个实施例中，第一次比对是与参考序列作全局比对，以该次左臂/右臂比对结果为基准对右臂/左臂比对结果进行的第二次比对为局部比对，这样，能够将分别来自二级左比对结果和二级右比对结果中的比对到同一参考染色
20 体上且之间的距离符合预期的两个读段配对成一对读段对，获得读段对比对结果。

在本发明的一个实施例中，所说的比对包括，设置读段的缺口的大小以使每个左臂或者每个右臂与所述参考序列进行多次比对，获得多次比对结果进而获得最佳比对结果。例如，将所述每个左臂或者所述每个右臂的缺口分别设置为-3nt、-2nt、-1nt、0nt、1nt、2nt、3nt、4nt、5nt、6nt 和 7nt，获得对应的多个读段，分别将所述对应的多个读段与所述参考
25 序列比对，将最优比对的序列作为该左臂/右臂，这里对于比对结果的好坏可以基于所利用的比对软件对比对结果的默认评判。

在本发明的一个实施例中，在消除比对结果中的读段的缺口之前，提取所述比对结果中的唯一比对结果以替换所述比对结果，所述唯一比对结果包括唯一比对上所述参考序列的多个读段对，并且每一所述读段对比对到所述参考序列的相同染色体，每一所述读段对
30 中的两个读段在参考序列上的距离符合预期的其来自的位置之间的距离，例如读段对来自前述的单接头环状单链文库时，两个读段在参考序列上的距离符合插入片段的大小。

在本发明的一个实施例中，修正所述唯一比对结果，使所述唯一比对结果中的每一对读段对比对到所述参考序列的相同染色体的正链。这里所说的“正链”和“负链”是互补

的两条链，是相对的，称一条链为正链就可以称其互补链为负链，在本发明的一个实施例中，将与参考序列匹配的链称为正链，将另一条链称为负链。例如，对于分别比对上同一参考染色体的正负链的一对读段，将比对上负链的读段变成其反向互补链，这样以其反向互补链来替代该读段得以实现所说的修正。

5 在本发明的一个实施例中，对所述比对结果或所述唯一比对结果进行数据格式转换。在本发明的一个实施例中，希望通用比对结果的格式为 SAM 或 BAM，以利于后续的基于比对结果或唯一比对结果对数据进一步分析，SAM 或 BAM 是常见的二进制格式，BAM 是 SAM 的压缩格式。由于利用不同的比对软件，可能输出的比对结果或者唯一比对结果的格式不适用于现有的后续数据处理或者分析软件程序，例如前述的 TeraMap 格式的比对结果，
10 其输出的数据格式不满足现有的大部分变异检测软件 SOAPsnp、GATK 或 SOAPindel 等对输入数据格式的要求，转换数据格式使获得具有通用数据格式的通用比对结果，便于对比对数据进一步分析处理。

 所检测的变异位点包括 CNV 和 SNP。基于通用比对结果检测 SNP 可以利用目前已知的 SNP 检测方法和/或软件，比如利用 SOAP2、samtools、GATK 等。在本发明的一个实施例中，使用开源软件 samtools 对通用比对结果（bam 文件）进行预处理，包括排序、去除
15 由于建库 PCR 产生的重复 reads 等，之后使用 samtools mpileup 和/或 GATK 开源软件得到 SNP 结果，并用开源软件 bcftools 转换为 vcf 格式。相比其他工具 Samtools 操作简单，输出格式通用，在大数据处理过程中可以使用多线程提高效率。这里的基于通用比对结果的 SNP 检测方法，包括开发数据格式转换软件（TeraMap2Sam）获得通用比对结果，准确处
20 理掉测定数据中的 gap、将 CG 平台 TeraMap 的比对结果转化为通用的 SAM 格式，使后续能够直接使用 Samtools，GATK 等众多优秀的开源软件进行变异检测，使后续分析的选择更加广泛。利用本发明的 SNP 检测方法，包括使用 samtools 软件做 SNP 分析简单通用、速度快且可信度高。

 基于通用比对结果检测 CNV 可以利用目前已知的 CNV 检测方法，比如利用隐马可夫
25 模型、环状二元分割、等级分割或核平滑算法等。在本发明的一个实施例中，检测 CNV 包括：在参考序列上设置多个窗口，基于所述通用比对结果中到所述窗口的读段的量与对照样本的通用比对结果中匹配到相同窗口的读段的量的差异具有显著性，判定所述待测样本核酸存在所述 CNV，其中，所述窗口为所述参考序列的一部分。其中，窗口的大小可依据
30 预检测的 CNV 的大小来调整设置，对照样本的通用比对结果可以参照前述待测体液样本的通用比对结果的获得方式来获得，差异是否为显著性的判断可以利用统计检验比如 z-score(标准分数)计算 z 值来进行，当 z 值大于或小于某一预定阈值时则判定该窗口区域存在 CNV，比如正常对照为二倍体（CNV=2），当 z 值为正数时说明待测样本的该窗口的

CNV > 2, 为负数则说明待测样本的该窗口的 CNV < 2, 在本发明的一个实施例中, 设定预定阈值为 3, 即当 z 值得绝对值大于 3 时则确定该窗口发生 CNV。所说的匹配即为比对上。所说的读段的量可以为一个数目, 也可以为一个比值, 例如, 也可以基于待测样本的窗口的测序深度与对照样本相应窗口的测序深度的差异, 使用 z -score(标准分数)进行检验检测出拷贝数变异, 所说的窗口的测序深度=比对到该窗口的读段的量/该窗口的大小。在本发明的一个实施例中, 考虑到实际测序过程中读段 (reads) 中 GC 含量会对测序深度有一定的影响 [Alkan, Can, Jeffrey M Kidd, Tomas Marques-Bonet, Gozde Aksay, Francesca Antonacci, Fereydoun Hormozdiari, Jacob O Kitzman, et al. “Personalized Copy Number and Segmental Duplication Maps Using next-Generation Sequencing.” Nature Genetics 41, no. 10 (October 2009): 1061 – 67], 先进行 GC 含量校正, 消除 GC 含量对测序深度的影响。所说的 GC 含量校正, 可以利用多个对照样本的测序数据, 取多个窗口算窗口的 GC 含量和平均测序深度, 对 GC-测序深度的数据进行二维回归分析, 例如利用局部加权回归散点平滑法 (lowess 回归) 建立二者的关系, 根据回归得的关系对各个窗口的测序深度进行 GC 含量校正。所说的测序深度-GC 含量关系可以通过以下来建立: 获得多个对照样本核酸的测序数据, 所述测序数据由多个读段组成; 在所述参考序列上设置多个窗口, 将所述多个对照样本的测序数据分别与所述参考序列的窗口比对, 计算各个对照样本的测序数据中比对上每个窗口的读段的数目, 获得每个窗口的测序深度, 所述窗口为所述参考序列的一部分, 所述窗口的测序深度=各个对照样本的比对上所述窗口的读段的总数目/(对照样本个数*所述窗口的大小); 基于每个窗口的测序深度和该窗口的 GC 含量, 利用二维回归分析法建立所述测序深度和 GC 含量的关系。

在本发明的一个实施例中, CNV 检测包括: 在参考序列上设置多个窗口, 计算窗口的测序深度, 窗口的测序深度=所述通用比对结果中比对到所述窗口的读段的数量/所述窗口的大小; 利用测序深度和 GC 含量的关系校正所述窗口的测序深度, 获得窗口的校正测序深度; 基于所述窗口的校正测序深度与对照样本的相同窗口的校正测序深度的差异具有显著性, 判定所述体液样本核酸存在 CNV, 其中, 所述窗口为所述参考序列的一部分。较佳地, 前述的对照样本的个数不小于 30 个, 样本数目达到 30 个使样本数据呈现满足特定分布符合适于利用多数统计检验方法来检验, 例如, 利用 t 检验、 z 检验等来统计检验一般要求多个样本数据符合正态分布。所说的对照样本的相同窗口的校正测序深度是利用所述测序深度-GC 含量关系校正对照样本的相同窗口的测序深度获得的, 所述对照样本的相同窗口的测序深度=所述对照样本的测序数据中比对到所述窗口的读段的数目/所述窗口的大小。前述对照样本的测序数据、比对结果等可以通过参照前述本发明一方面的或者任一具体实施方式中的测序数据处理方法获得, 可以与待测样本的测序数据、比对结果同时获得,

也可以预先获得保存备用。利用前述的 CNV 检测方法开发的 CNV 检测程序基于标准分数方法做 CNV 分析，速度快，分辨率高。

在本发明的一个实施例中，在所说的体液样本中检测出来的变异位点，各变异位点的突变频率 = $\frac{\text{所述比对结果中支持该变异位点的读段的数目}}{\text{所述比对结果中比对到该变异位点的读段总数}}$ ，或者，各变异位点的突变频率

5 $= \frac{\text{所述通用比对结果中支持该变异位点的读段的数目}}{\text{所述通用比对结果中比对上该变异位点的读段总数}}$ 。所说的“支持”指比对上该位点的读段的对应位置

的碱基与该位点变异碱基一样，例如，变异位点 K 为 T/C，参考序列上的该位点碱基为 T，该样本核酸中的该位点发生突变变为 C，基本上所有比对上该位点的读段的相应位点不是为 T 碱基就是为 C 碱基，比对上的读段中相应位点为 C 的读段即为支持该变异位点 K 的读段。

10 在本发明的一个实施例中，筛选出突变频率与阴性对照样本中相同位点的突变频率具有显著差异的变异位点为肿瘤变异位点，该条件是发明人结合目前相关数据库信息、大量文献报道信息、检测统计大量阳性样本和大量阴性样本确定下来的，具有统计意义，较佳的，这里的阳性或者阴性对照样本不少于 5 个，对照样本的数据可以自己对照样本的核酸进行提取、序列测定来获得，也可以依照他人已公开或公开数据库中的样本测序数据，

15 多个对照样本数据使统计判定条件/结果具有统计意义，更加可信。筛选确定来自肿瘤的变异位点，和/或基于肿瘤变异位点估计样本中的肿瘤核酸含量，可以辅助用于临床的肿瘤诊断筛查，辅助用于了解所测样本个体癌变的可能性及病情发展情况等。需要说明的是，所说的突变频率具有显著差异指具有实质差异，例如对于待测样本中的变异位点 A，多个阳性样本（cancer 样本）中的 reads 支持比例都为 5/400（变异 5 条 reads，总 400 条 reads），即

20 阳性样本中该位点的平均变异频率 1.25%，而在多个阴性对照样本中的 reads 支持比例都为 1/200（变异 1 条 reads，总 200 条 reads），即阴性对照样本中的平均变异频率 0.5%，若待测样本中的该变异位点的变异频率更接近 1.25%，例如达到 0.9%，则达到所说的显著差异或者所说的实质差异。具有显著差异，也可以指统计学上的对数据差异性的评价——差异具有显著性，例如对待测样本中的变异位点 A 进行多次检测，获得该位点的多组比对结果

25 数据，从每组比对结果数据中都可获得一个突变频率，所说的突变频率=支持该变异位点的读段数/比对上该位点的总读段数，接着比较待测样本的变异位点 A 的突变频率与阴性对照样本中的该位点的突变频率的差异，例如可以利用 z 检验或 t 检验，差异具有显著性（ $p \leq 0.05$ ），即认为达到所说的具有显著差异。在本发明的一个实施例，对 8 个阳性样本进行检测，筛选出其共有的高频突变，当所测样本也包含该突变而且其突变频率不小于这 8 个阳性样本中的该位点的最小突变频率或 80%的最小突变频率时，则确定该突变来自肿瘤细胞/

30

组织。

在本发明的一个实施例中，依据所有的来自肿瘤的变异位点的读段支持量，计算所测体液样本中的来自肿瘤的游离核酸含量，所说的读段支持量，可以为支持该变异的读段的数目，也可以是支持该变异的读段所占的比例。所说的肿瘤核酸含量

5
$$= \frac{2 * \text{比对结果中支持所有肿瘤变异位点的读段的数目}}{\text{比对结果中支持所有变异位点的读段数目}}, \text{ 或者, 所述肿瘤核酸含量}$$

$$= \frac{2 * \text{通用比对结果中支持所有肿瘤变异位点的读段的数目}}{\text{通用比对结果中支持所有变异位点的读段数目}},$$
 这里的变异位点为 SNP。临床上，癌症患者的

10 ctDNA 主要来源于肿瘤细胞分泌、凋亡或坏死，其浓度平均可达 180ng/ml，是正常人的 20 倍以上，因此利用本发明这一方面的方法，检测外周血 ctDNA 的突变及其含量变化，可以应用于或者辅助应用于以下方面：肿瘤的早期诊断检测、遗传杂合性评估、反映肿瘤动态

变化、靶向治疗基因突变分析、肿瘤早期治疗反应评价、肿瘤微量残留检测和肿瘤耐药性

15 实时检测，相对于传统的检测定量方法，本发明的这一方法具有如下优势：1) 微创性，只需受检者提供外周血血浆样本和/或唾液样本；2) 高灵敏度，不受限制于病灶大小，3) 高

准确性和 4) 高通量，可一次检测多个样本的变异和/或定量肿瘤核酸含量。

依据本发明的另一方面，本发明提供一种确定肿瘤核酸含量的装置，所述肿瘤核酸来

15 自体液样本，所述体液样本包含游离核酸，所述游离核酸包括肿瘤核酸和非肿瘤核酸，所述装置用以实现上述本发明一方面的或者任一具体实施方式中的确定肿瘤核酸含量的方法的部分或全部步骤，所述装置包括：数据输入单元，用以接收所述游离核酸的测定数据，所述测定数据包括多个读段，所述测定数据是通过对所述游离核酸的至少一部分进行序列测定获得的；数据输出单元，用以输出数据，包括所述肿瘤核酸含量；存储单元，用于存储

20 数据，其中包括可执行的程序；处理器，与所述数据输入单元、数据输出单元及存储单元数据连接，用于执行所述可执行的程序，所述程序的执行包括本发明上一方面或者其任一具体实施方式中的确定肿瘤核酸含量的方法。上述对本发明一方面的方法的优点和技术特征的描述，同样适用本发明这一方面装置，在此不再赘述。

依据本发明的再一方面，提供一种计算机可读存储介质，用于存储供计算机执行的程

25 序，所述程序的执行包括完成本发明一方面或者其任一具体实施方式中的确定肿瘤核酸含量的方法。前述对本发明的确定肿瘤核酸含量的方法的优点和技术特征的描述也适用于该计算机可读存储介质，在此不再赘述。所称存储介质可以包括：只读存储器、随机存储器、磁盘或光盘等。

附图说明

30 本发明的上述和/或附加的方面和优点从结合下面附图对实施方式的描述中将变得明显和容易理解，其中：

图 1 是本发明的一个实施例中的单接头环状单链文库的结构示意图；

图 2 是本发明的一个实施例中的单接头环状单链文库构建和测序流程图；

图 3 是本发明的一个实施例中的 Teramap2Sam 软件的算法流程图；

图 4 是本发明的一个实施例中的 SNP 检测流程图；

图 5 是本发明的一个实施例中的八例肺癌患者的 cfDNA 中的 BRAF 基因 V600E 突变的突变频率示意图；

图 6 是本发明的一个实施例中的四个个体的外周血的肿瘤浓度在不同治疗阶段的变化示意图；

图 7 是本发明的一个实施例中的监测一个个体的肿瘤药物敏感以及耐药突变 EGFR 基因 L821X 变异和 T790M 变异的突变频率随着治疗进程实时变化的示意图。

10 具体实施方式

下面详细描述本发明的实施例，所述实施例的示例在附图中示出，其中，自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，仅用于解释本发明，而不能理解为对本发明的限制。需要说明的是在本文中所使用的术语“第一”、“第二”、“第三”、“第四”或者“一级”、“二级”等
15 仅为方便描述指代，而不能理解为指示或暗示相对重要性，也不能理解为之间有先后顺序关系。在本发明的描述中，除非另有说明，“多个”的含义是两个或两个以上。

以下实施例仅用于说明本发明的优选实施方式，实施例中未注明具体操作手段或条件的，可以按照本领域内的文献所描述的技术或条件（例如参考 J.萨姆布鲁克等著，黄培堂等译的《分子克隆实验指南》，第三版，科学出版社）或者按照产品说明书进
20 行。所用试剂、仪器或软件未注明生产厂商者或来源途径的，均为可以通过市购获得的常规产品、服务或是开源的。

实施例一

以下以肺癌患者的外周血血浆作为检测对象，样本来自西南医院，进行如下检测：

（一）文库建立及测序

25 建库及测序流程如图 2 所示，以下涉及的具体序列，都是从左到右为 5'端至 3'端，序列中的“//”中为末端修饰基团，“phos”表示磷酸化，“dd”表示双脱氧，“bio”表示生物素。

1、cfDNA 的提取(采用 SnoMag Circulating DNA Kit):

- 1) 取 200ul 血浆于 1.5mlEP 管，加入 600ul buffer LSB。
- 30 2) 加入 20ul NanoMag Circulating Beads 混匀，室温放置 10min，每 2-3min 混匀一次。
- 3) 将 EP 管置于磁力架上吸附 1min，弃上清。
- 4) 取下 EP 管加入 150uL Buffer WA，混匀。

5) 将 EP 管置于磁力架上吸附 1min, 弃上清。

6) 取下 EP 管加入 150uL 75%乙醇, 混匀。

7) 将 EP 管置于磁力架上吸附 1min, 弃上清。

8) 重复 6-7 一次。

5 9) 室温干燥磁珠 5min。

10) 加入 32ul elution buffer 混匀磁珠, 室温静置 5min。

11) 将 EP 管置于磁力架上吸附 1min, 转移上清至新的 1.5mlEP 管。

2、文库的构建:

1) rSAP 去磷酸化反应

cfDNA	30ul
10x NEBuffer 2	3.5ul
rSAP (1U/ul)	1.5ul
Total	35ul

10 反应条件:

37°C	15min
65°C	10min
4°C	0.1 °C/secs 降温 -hold

2) T4 DNA Polymerase 末端补平

rSAP DNA	35ul
Water	2.5ul
10x NEBuffer2	0.5ul
0.1M ATP	0.4ul
25 mM dNTP	0.4ul
BSA (10mg/ml)	0.2ul
T4 DNA polymerase (3U/ul)	1ul
Total	40ul

反应条件:

12°C	20min
4°C	hold

15 60ul AmpureXP beads 纯化以上反应产物, 22ulElution buffer 洗脱。

3) 第一序列和第二序列分别连接到末端补平的 DNA 片段的两端

End-Repaired DNA	20ul
Water	3.9ul
5 uM 第一序列	0.9ul
5 uM 第二序列	0.9ul
3x HB	13.3ul
T4 DNA Ligase (600U/ul)	1.0ul
Total	40ul

反应条件：

20℃	15min
4℃	hold

40ul AmpureXP

beads 纯化以上反应产物,

22ul Elution buffer 洗脱。

- 5 第一序列的两条链为：TTGGCCTCCGACT/3-ddT/ (SEQ ID NO: 1),
/5phos/AAGTCGGAGGCCAAGCGGTCTGT/ddC/ (SEQ ID NO: 2)。

第二序列的两条链分别为：/5Phos/GTCTCCAGTCGAAGCCCGACG/3ddC/ (SEQ ID NO: 3), GCTTCGACTGGAGA/3ddC/ (SEQ ID NO: 4)。

4) 缺刻平移 (Nick Translation)

5'+3'-Adap-B DNA (上步产物)	20ul
Water	7.74ul
10x taq buffer (LK)	3.2ul
0.1M ATP (LK)	0.4ul
25 mM dNTP (LK1)	0.16ul
20 uM 第三序列中的上游引物	0.5ul
10uM 第三序列中的下游引物	1ul
Total	32ul

- 10 第三序列中的上游引物/5-bio/TCCTAAGACCGCTTGGCCTCCGACT (SEQ ID NO: 5),
第三序列中的下游引物

5Phos/AGACAAGCTCxxxxxxxxxxGATCGGGCTTCGACTGGAGAC (SEQ ID NO: 6), 中间“x”处为可变的标签序列区域, 可以以 N 替代, N 为 A、T、C 或 G, 当没有其它样本文库一起混合上机, 只有一个样本文库上机, 不需要标签序列, 即第三序列可为

- 15 5Phos/AGACAAGCTCGATCGGGCTTCGACTGGAGAC (SEQ ID NO: 7), 在该示例中, 由于是肿瘤游离核酸样本, 混合核酸中的目标核酸 (ctDNA) 含量低, 若多个这样的样本文库混合上机获得混合数据, 需要拆分混合数据对应到各自样本, 会损失一部分数据, 且构建的是单接头环状文库读段相对短, 要准确检测变异需要深度测序获得相对大量的测定数据, 较佳的, 单个样本文库上机。

- 20 反应条件：

60℃	5min
37℃	0.1℃/secs-hold

向上步反应物中加入如下 8ul Nick Translation mix

Water	5.8ul
10x taq buffer (LK)	0.8ul
T4 DNA Ligase (600U/ul)	0.8ul

(LK3)	
Std Taq polymerase (5U/ul) (LK1)	0.6ul
Total	8ul

反应条件:

37°C	20min
4°C	hold

40ul AmpureXP beads 纯化以上反应产物, 37.4ul Elution buffer 洗脱。

5 5) PCR with Pfx

Nick Translated DNA	35.4ul
10x Pfx buffer	5.0ul
25 mM dNTP	0.8ul
50mM MgSO ₄	2.0ul
Pfx polymerase	0.8ul
20 uM 第三序列中的 上游引物	2.0ul
10uM 第三序列中的 上游引物	4ul
Total	50ul

反应条件:

Step 1	94°C	2min	
Step 2	94°C	15s	13cycles
	56°C	30s	
	72°C	30s	
Step 3	72°C	10min	
	4°C	∞	0.1°C/sec 降温

50ul AmpureXP beads 纯化以上反应产物, 22ul Elution buffer 洗脱。

6) Qubit 定量

利用 Qubit dsDNA HS assay kit 对 PCR 产物进行浓度测定。

10 7) 链分离 (Strand Separation)

a) 多个文库混合, 使 DNA 共约 160ng。样品补 1xTE 至总体积为 60ul。

b) 提前准备以下试剂: 4X BBB, Streptavidin Beads, 0.3M MOPS acid, 0.5% Tween20, 1X BBB/Tween Mix, 1X BWB/Tween Mix, 0.1M NaOH。其中 1X BWB/Tween Mix、0.1M NaOH, Streptavidin Beads 需现配现用。

15 c) 提前 15min 配置以下四种试剂

0.5% Tween20, 1X BBB/Tween Mix, 1X BWB/Tween Mix, 0.1M NaOH。

其中 0.5% Tween20 配置方法同前述, 其他三种配置方法如下:

d) 1X BBB/Tween Mix

1X BBB	30ul
--------	------

0.5% Tween20	0.3ul
Total	30.3ul

e) 1X BWB/Tween Mix

1X BWB	2000ul
0.5% Tween20	20ul
Total	2020ul

f) 0.1M NaOH

0.5M NaOH	15.6ul
Water	62.40ul
Total	78.0ul

g) Streptavidin Beads 洗涤方法如下:

- 每个样品取 30ul Streptavidin Beads: 加入 3-5 倍体积的 1XBBB, 混匀后置于磁力

5 架上静止吸附, 调整不粘管的方向, 使得 beads 在 1XBBB 洗液中前后游动, 弃上清液后, 重复上述操作一次,

- 取出不黏管加入 1 倍体积 (30ul) 1X BBB/Tween Mix 悬浮, 混匀后室温静置。

10 h) 向 60ulPCR 产物样品中加入 20ul 4XBBB 混匀, 然后转移到上步骤含有 30ul1X BBB/Tween Mix 溶解的 beads 的不粘管中混匀, 此 110ul 混合物室温下结合 15-20min, 中间轻轻弹匀一次。

i) 将上述不粘管磁力架放置 3-5min, 弃去上清液, 用 1ml 的 1X BWB/Tween Mix 洗涤 2 次, 方法同 Streptavidin Beads 的洗涤方法

j) 向上述 beads 中加入 26ul 0.1M NaOH, 吹打混匀后放置 10min, 再置于磁力架上 3-5min, 取上清到新的 1.5ml EP 管中。

15 k) 向上述 1.5mlEP 管中加入 13ul 0.3M MOPS, 混匀备用。

l) 此步骤产物可以冻存于 -20℃。

8) 环化 (Splint Circulation)

a) 向上一步得到的 39ul 的样品中加入 10ul 的 20uM 第四序列, 第四序列为 TCGAGCTTGTCTTCCTAAGACCGC (SEQ ID NO: 8);

20 b) 提前 5 分钟准备连接酶反应混合液, 配制如下:

Water	4.2ul
10x TA Buffer (LK1)	6ul
100mM ATP	0.6ul
600U/ul Ligase	0.2ul
Total	11ul

c) 将连接酶反应混合液震荡充分混匀, 离心后, 向已经加入引物反应混合液的 EP 管中加入连接酶反应混合液 11ul, 震荡 10s 混匀, 瞬时离心。

d) 置于 PCR 仪中 37℃ 孵育 1.5h。

e) 反应完成后, 取出 5ul 样品, 待 6%变性胶电泳检测, 剩余的约 55ul 体积, 进入下一步酶反应。

9) 酶切消化 (Exo I and III)

a) 提前 5 分钟左右准备引物反应混合液, 配制如下:

10x TA Buffer (LK1)	1ul
20U/ul Exo I	3ul
200/ul Exo III	1ul
Total	5ul

5 b) 将上述混合液震荡充分混匀, 离心后, 向上一步得到的 55ul 的样品中分别加入 5ul 的反应混合液;

c) 震荡 10s 混匀离心, 置于 PCR 仪中 37℃ 孵育 30min。

d) 酶切 30min 完成后, 向样品中加入 2.5ul 500mM EDTA 终止酶反应。

e) 上述样品用 PEG32 beads/tween20 纯化, 方法如下:

10 将上步骤样品 59ul 转移到 1.5ml 不粘管中, 加入 78ul 的 PEG32 beads/tween20 (PEG32 beads : tween20=100:1), 室温结合 15min, 期间吹打混匀一次;

f) 不粘管置于磁力架 3-5min 后弃去上清, 用 700ul 75%乙醇洗涤两次, 洗涤时将不粘管前后方向反转, 使得 beads 在乙醇中游动, 每次洗涤游动 2-3 次;

15 g) 室温下晾干后用 27ul TE/tween20 回溶 (TE:tween20=500:1), 溶解时间共计 15min, 中间混匀一次;

h) 上清转移到新的 1.5ml EP 管中, 将最终得到产物用 Qubit™ ssDNA Assay Kit 定量。Buffer 与染料比例为 199:1 混匀后 vortex 并离心混合备用, 取两份 190ul 稀释后染料工作液分别加入 10ul 的两种标准品 vortex 并离心混合备用, 取 198ul 稀释后染料工作液加入 2ul 样品, vortex 后并离心进行 Qubit 仪器定量。

20 i) 浓度标准化 (Normalization)

按照单链分子定量测定的浓度调整 DNB 制备使用的样本起始量统一调整为 35.3ng—53ng, 将对应体积样本 (<60ul) 转移至 Biorad PCR 板中, 使用 1XTE 补齐使总体积不超过 120ul。

25 终浓度为 5.625-7.5fmol/ul, 体积为 120ul, 则总量为 35.3ng-53ng, 1 adapter 测序中的 DNB 需要 120fmol, 7.5foml/ul, 16ul。故需要把文库稀释至 7.5fmol/ul。

10) CG 1-Adapter 测序

利用CG平台的标准化流程测序。DNA 纳米芯片是CG独创的一种高通量测序技术。该示例的对改进的单接头测序文库进行测序, 较其他测序方案成本更低、速度更快, 并集成质控确保其测序质量。

实施例二

对实施例一的下机数据进行处理。利用本发明的方法，基于 CG 平台测序技术，可以对超微量的 cfDNA 进行富集，文库建立、测序和数据分析工作。在该示例中，由于 CG 测序原理的特殊性，其测序的 reads 较短，且在特定位置存在重测序以及小 gap 的现象，难以不处理直接使用普通的比对软件对测序结果进行比对。针对 reads 的特殊结构，我们使用 CG 平台专有的 TeraMap 进行比对，其工作原理是：首先，它将读长中的两端（LeftArm，RightArm）分别做比对，其间 TeraMap 会尝试多种 gap 值来处理读长，以获得更多的比对结果；然后，将每一端的比对结果拿出来作为参考，对另一端做局部比对（例如 4-AD，局部比对的范围是 0~700bp）；如果两端可以良好比对到同一染色体，且 insert-size（插入片段）符合期望（例如 4-AD，一读段对的两读段的距离为 0~700bp），则只输出最佳比对结果，否则两端的多个比对结果全部输出。TeraMap 是 CG 测序平台的比对软件，它可将 CG 特有序列比对到参考基因组上，其输出格式由三部分组成，简要说明如下：第一行，是 reads 序列信息；第二行和第三行，是 reads 比对情况说明；第四行和第五行，是 reads 比对结果详细信息。

第一行：

列号	字段	类型	简介
1	QNAME	字符串	参考序列编号
2	POS	整型	比对到参考序列的位置
3	SEQ	字符串	比对片段的序列信息

第二行：

列号	字段	类型	简介
1	ARM	字符串	Reads 两段标识
2	resultType	字符串	比对结果类型
3	UnreportedMatch ed	布尔型	是否有未报告的匹配
4	UnreportedPerfect Matches	布尔型	是否有未报告的完美匹配
5	indexOverflow	布尔型	索引是否越界
6	matchCount	整型	比对到的位置数目统计

第四行：

列号	字段	类型	简介
----	----	----	----

1	chr	字符串	染色体号
2	POS	整型	比对到参考序列的位置
3	ARM/STRAND	字符串	Reads 两段标识及正负链信息
4	gaps	整型	Reads 中的 gap 位置及长度信息
5	editCount	整型	编辑数
6	hasConsistentMate	布尔型	是否为一致比对
7	foundByLocalSearch	布尔型	是否通过局部搜索找到
8	localSearchAfterOverflow	布尔型	是否在溢出后使用局部搜索
9	hasSameCostNearbyAlignments	布尔型	是否有同代价临近比对
10	mappedWithMaxAllowedCost	布尔型	是否用最大允许代价比对

因为 TeraMap 比对存在 gap 问题，使得无法进行下游分析，依据本发明的方法开发 Teramap2Sam 软件，将 TeraMap 比对结果中 gap 去除并转换为 SAM(sequence alignment/map format)。Teramap2Sam 软件的主要流程可分为三部分，算法流程图如图 3 所示。

第一步：提取唯一比对结果。根据 TeraMap 输出结果 matchCount 判定是否唯一比对，
5 同时要求插入片段长度满足要求以及两端 read 比对在用一条参考序列上。

第二步：去除 gap。根据 gaps 字段判定 reads 中的 gap 位置，并修正 read 序列。

第三步：计算 FLAG。根据双端 read 的比对方向，计算 SAM 文件中的 FLAG 参数，
获得比对情况。

SAM 是存储比对信息的一种较通用的格式，每一行是一个 reads 的比对结果，主要由
10 十一个字段组成，其后还可添加更多字段包含更多信息，比如 XT:A:U 就是表示此 reads 为 unique 比对。简要说明如下：

列号	字段	类型	简介
1	QNAME	字符串	比对片段的编号
2	FLAG	整型	标识符之和，用于表述比对情况
3	RNAME	字符串	参考序列的编号

4	POS	整型	比对上的位置
5	MAPQ	整型	比对质量值
6	CIGAR	字符串	简要比对信息表达式
7	RNEXT	字符串	下一片段比对上的参考序列的编号
8	PNEXT	整型	下一片段比对上的位置
9	TLEN	整型	比对片段长度
10	SEQ	字符串	序列片段的序列信息
11	QUAL	字符串	序列的质量信息

在实际使用中为了节约存储资源，主要使用其二进制压缩格式（BAM）。此外 CG 又针对其 reads 结构开发了 Assembly Software 将 reads 重新组装，组装完成后进行变异检测分析等后续工作。

由于 CG 单接头 reads 的特殊结构存在 reads 太短（最短 12bp）的缺点，在一些特殊数据的处理中 CG 原有的突变检测工具不再适用或者检测结果不佳。针对这种情况，我们首先开发工具将 TeraMap 的比对结果转化为通用的 SAM/BAM 格式，其中 SAM/BAM 是高通量测序中普遍使用的比对结果格式，所以我们采用这种通用的格式，然后再使用 BAM 数据检测 SNP 变异。SNP 检测可以利用已知开源软件，比如利用 SOAP2、samtools、GATK 等。在该示例中使用开源软件 samtools 对通用比对结果（SAM 格式的 bam 文件）进行预处理，包括排序、去除由于建库 PCR 产生的重复 reads 等，之后使用 samtools mpileup 和/或 GATK 开源软件得到 SNP 结果，并用开源软件 bcftools 转换为 vcf 格式。相比其他工具 Samtools 操作简单，输出格式通用，在大数据处理过程中可以使用多线程提高效率，速度快可信度高。

较之传统方法，我们使用的基于 CG 单接头测序的方法可以实现超微量建库测序，建库只需要 1-10ng 核酸，需要外周血量 2-5ml，并且 CG 的标准化流程简单快速，TeraMap 比对结果转换为 SAM 格式之后比闭源的 TeraMap 格式更加通用，可以使用 Samtools 等软件进行处理。该示例中利用 TeraMap 进行比对。测序完成后使用 CG 平台的集成工具 makeADF 得到原始 reads，然后用 TeraMap 进行比对，将测序得到的 reads 比对的参考序列上。得到的比对结果使用 TeraMap2Sam 转换为通用的 SAM 格式，接着利用 Samtools 等软件进行 SNP 检测，检测流程如图 4 所示，结果见表 1。

表 1

BAM(排序并去重)	198Gb/4h20min
SNP(vcf)	3.1Gb/2h38min
测序深度	85.62X
1X 深度以上覆盖度	91.66%
5X 深度以上覆盖度	89.97%
10X 深度以上覆盖度	87.85%
1X 深度以上 CDS 区域覆盖度	99.67%
5X 深度以上 CDS 区域覆盖度	98.37%
10X 深度以上 CDS 区域覆盖度	96.04%
总共检测到的肿瘤体细胞 SNP 数量	26533

目前已有的拷贝数变异检测方法有隐马可夫模型、环状二元分割、等级分割、核平滑算法等。我们基于格式为 SAM/BAM 的通用比对结果，根据总长达 1,000,000bp 的多个窗口的 reads 深度分布，使用 z-score(标准分数)得到拷贝数变异结果。

- 5 考虑到实际测序过程中 reads 中 GC 含量会对测序深度有一定的影响，我们对比对结果（BAM）进行 GC 含量校正，消除 GC 含量对深度的影响。具体的，取总长达 1,000,000bp 的多个窗口算窗口的 GC 含量和平均测序深度，对 GC-测序深度的数据进行 lowess 回归，根据回归曲线对 GC 含量进行校正。

- 标准分数（standard score）也叫 z 分数（z-score），是一个分数与平均数的差再除以标准差的过程。用公式表示为： $z=(x-\mu)/\sigma$ 。其中 x 为某一具体分数， μ 为平均数， σ 为标准差。Z 值的量代表着原始分数和母体平均值之间的距离，是以标准差为单位计算。在原始分数低于平均值时 Z 则为负数，反之则为正数。在该示例中，通过对 2000bp 窗口内 reads 计数（原始分数）和总体 reads 平均值（多个正常对照样本）之间的距离使用标准差进行度量，可以有效检测出拷贝数变异。Z 值为正数时反应为拷贝数大于 2（正常样本是 2 倍体），
10 比如重复，z 值为负数时反应拷贝数小于 2，比如缺失。将该实施例中的上述的 CNV 检测方法编写成程序，并将该程序命名为 calcu_zscore_query，将 z 绝对值大于 3 的区域判断为发生 CNV。

- 使用 z-score（标准分数）可以快速检测出 CNV，50 乘全基因组数据的 CNV 分析只需 4 小时，作为对比，CONTRA 软件[<http://sourceforge.net/projects/contra-cnv/>]需要 1 天以上
20 时间。

该示例中利用 TeraMap 进行比对。测序完成后使用 CG 平台的集成工具 makeADF 得到原始 reads，然后用 TeraMap 进行比对，将测序得到的 reads 比对的参考序列上。得到的比对结果使用 TeraMap2Sam 转换为通用的 SAM 格式，接着利用上述方法检测 CNV，表 2 展示检测结果。

表 2

获得 BAM 格式结果(排序并去重)	198Gb/4h20min
CNV(zscore)	660Mb/36min
测序深度	85.62X
1X 深度以上测序覆盖度	91.66%
5X 深度以上测序覆盖度	89.97%
10X 深度以上测序覆盖度	87.85%
1X 深度以上 CDS 区域覆盖度	99.67%
5X 深度以上 CDS 区域覆盖度	98.37%
10X 深度以上 CDS 区域覆盖度	96.04%
检测到肿瘤体细胞 CNV 数量(标准分数绝对值大于 3)	1606239 个

实施例三

图 5 是 BRAF 基因突变 V600E 在 8 例 (S1-S8) 肺癌患者 cfDNA 的突变频率情况, 样本获自西南医院, 从图 5 中的各例患者的该变异的突变频率可看出, BRAF 基因 V600E 突变是该癌阳性样本中共有的高频突变。该突变是临床上有意义的突变, 通过检测

BRAF 基因可筛选 EGFR 酪氨酸激酶抑制剂 (EGFR-TKIs) 受益的患者和 BRAF 基因突变靶向药物受益的患者, 该突变在甲状腺癌、结直肠癌以及黑色素瘤患者中, 可作为患者诊断的一项指标。

图 6 示意其中 4 个个体 (P1-P4) 的外周血的肿瘤浓度随着治疗进展 (T1—T3 三个治疗时间段) 的变化情况。外周血肿瘤核酸浓度的变化可用于评估或者辅助临床判断治疗效果。

图 7 反映其中一个个体的肿瘤药物敏感以及耐药突变 EGFR 基因 L821X 变异和 T790M 变异的突变频率随着治疗进程实时改变的监测情况。肿瘤变异的突变频率的改变可用于指导或者辅助指导临床选择用药。

权利要求书

1. 一种确定肿瘤核酸含量的方法，所述肿瘤核酸来自体液样本，所述体液样本包含游离核酸，所述游离核酸包括肿瘤核酸和非肿瘤核酸，所述方法包括：

5 对所述游离核酸的至少一部分进行序列测定，获得测定数据，所述测定数据包括多个读段；

基于所述测定数据，检测所述游离核酸中的变异位点；

筛选出突变频率与阴性对照样本中相同位点的突变频率具有显著差异的变异位点，获得肿瘤变异位点；

10 基于所有所述肿瘤变异位点的支持读段的量，确定所述体液样本中的肿瘤核酸含量。

2. 权利要求 1 的方法，其特征在于，所述对游离核酸的至少一部分进行序列测定包括，对所述游离核酸的至少一部分进行测序文库构建，获得测序文库，以及

对所述测序文库进行测序，获得测定数据，所述测序文库为单链环状 DNA 文库。

3. 权利要求 2 的方法，其特征在于，构建所述单链环状 DNA 文库，包括，

15 (1) 对所述游离核酸进行末端磷酸化，获得末端磷酸化产物；

(2) 末端修复所述末端磷酸化产物，获得末端修复产物；

(3) 将第一序列和第二序列连接至所述末端修复产物的两端，获得第一连接产物；

(4) 利用第三序列对所述连接产物进行缺刻平移和扩增，获得扩增产物，所述第三序列为一对引物对，所述引物对的至少一条引物带有生物素标记；

20 (5) 利用所述生物素标记对所述扩增产物进行单链分离，获得单链产物；

(6) 利用第四序列环化所述单链产物，获得所述测序文库；其中，

所述第四序列能够连接所述第一序列的一端和所述第二序列的一端，所述第一序列和/或所述第二序列的另一端为双脱氧核苷酸。

4. 权利要求 2 的方法，其特征在于，构建所述单链环状 DNA 文库，包括，

25 (a) 对所述游离核酸进行末端修复，获得末端修复产物；

(b) 末端磷酸化所述末端修复产物，获得末端磷酸化产物；

(c) 将第一序列和第二序列连接至所述末端磷酸化产物的两端，获得第一连接产物；

(d) 利用第三序列对所述连接产物进行缺刻平移和扩增，获得扩增产物，所述第三序列为一对引物对，所述引物对的至少一条引物带有生物素标记；

30 (e) 利用所述生物素标记对所述扩增产物进行单链分离，获得单链产物；

(f) 利用第四序列环化所述单链产物，获得所述测序文库；其中，

所述第四序列能够连接所述第一序列的一端和所述第二序列的一端，所述第一序列和/或所述第二序列的另一端为双脱氧核苷酸。

5. 权利要求 2 的方法，其特征在于，所述测序利用组合探针锚定连接测序技术进行。

6. 权利要求 2-5 任一方法，其特征在于，所述测定数据包括多对读段对，每对所述读段对中的两个读段分别来自同一单链环状 DNA 文库的两个位置，读段带有缺口，将一对读段对的两个读段分别称为左臂和右臂。

7. 权利要求 6 的方法，其特征在于，所述基于测定数据检测所述游离核酸中的变异位点，包括，

将所述测定数据与参考序列进行比对，获得比对结果；

消除所述比对结果中的每个读段的缺口，获得通用比对结果；

基于所述通用比对结果，检测所述变异位点；其中，

所述比对结果和/或通用比对结果包括多个所述读段对的比对结果，和/或，

所述比对结果和/或通用比对结果包括多个所述左臂的比对结果和多个所述右臂的比对结果。

8. 权利要求 7 的方法，其特征在于，所述比对包括，

将每对读段对的左臂和右臂分别与所述参考序列比对，获得一级左比对结果和一级右比对结果，

分别以所述一级左比对结果和所述一级右比对结果的其中一个为参考，对另一个作比对，获得二级左比对结果和二级右比对结果，

基于所述二级左比对结果和所述二级右比对结果，获得多个所述读段对的比对结果，或者获得多个所述左臂的比对结果和多个所述右臂的比对结果。

9. 权利要求 7 的方法，其特征在于，所述比对包括，

设置缺口的大小以使每个左臂或者每个右臂与所述参考序列进行多次比对，包括将每个所述左臂或者所述每个右臂的缺口分别设置为-3nt、-2nt、-1nt、0nt、1nt、2nt、3nt、4nt、5nt、6nt 和 7nt，获得对应的多个读段，分别将所述对应的多个读段与所述参考序列比对。

10. 权利要求 7-9 任一方法，其特征在于，所述比对结果的格式为 TeraMap。

11. 权利要求 7-9 任一方法，其特征在于，在消除所述比对结果中的每个读段的缺口之前，提取所述比对结果中的唯一比对结果以替换所述比对结果，所述唯一比对结果包括唯一比对上所述参考序列的多个读段对，并且每一所述读段对比对到所述参考序列的相同染色体，每一所述读段对中的两个读段在所述参考序列上的距离符合插入片段的大小。

12. 权利要求 11 的方法，其特征在于，修正所述唯一比对结果使所述唯一比对结果中的每一对读段对比对到所述参考序列的相同染色体的正链。

13. 权利要求 7-12 任一方法，其特征在于，对所述比对结果或所述唯一比对结果的数据格式进行转换，以获得所述通用比对结果。

14. 权利要求 7-13 任一方法，其特征在于，消除所述比对结果或者所述唯一比对结果中的每个读段的缺口包括，

若所述读段包含正缺口，以 N 填补所述正缺口的大小，

若所述读段包含负缺口，去除所述负缺口，其中，

N 为 A、T、C 或 G。

15. 权利要求 7-14 任一方法，其特征在于，所述通用比对结果的格式为 SAM 或 BAM。

5 16. 权利要求 15 的方法，其特征在于，所述变异位点包括 CNV 和 SNP，检测所述 CNV 包括，

在所述参考序列上设置多个窗口，

基于所述通用比对结果中匹配到所述窗口的读段的量与对照样本的通用比对结果中匹配到相同窗口的读段的量具有显著差异，判定所述体液样本游离核酸中存在 CNV，其中，

10 所述窗口为所述参考序列的一部分。

17. 权利要求 15 的方法，其特征在于，所述变异位点包括 CNV 和 SNP，检测所述 CNV 包括，

在所述参考序列上设置多个窗口，计算窗口的测序深度，窗口的测序深度=所述通用比对结果中比对到所述窗口的读段的数量/所述窗口的大小；

15 利用测序深度-GC 含量关系校正所述窗口的测序深度，获得窗口的校正测序深度；

基于所述窗口的校正测序深度与对照样本的相同窗口的校正测序深度具有显著差异，判定所述体液样本存在所述 CNV，其中，

所述窗口为所述参考序列的一部分。

18. 权利要求 16 或 17 的方法，其特征在于，所述对照样本的个数不小于 30 个。

20 19. 权利要求 17 的方法，其特征在于，所述对照样本的相同窗口的校正测序深度是利用所述测序深度-GC 含量关系校正对照样本的相同窗口的测序深度获得的，所述对照样本的相同窗口的测序深度=所述对照样本的测序数据中比对到所述窗口的读段的数目/所述窗口的大小。

25 20. 权利要求 17-19 任一方法，其特征在于，建立所述测序深度-GC 含量关系包括，获得多个对照样本核酸的测序数据，所述测序数据由多个读段组成；

在所述参考序列上设置多个窗口，将所述多个对照样本的测序数据分别与所述参考序列的窗口比对，计算各个对照样本的测序数据中比对上每个窗口的读段的数目，获得每个窗口的测序深度，所述窗口为所述参考序列的一部分，所述窗口的测序深度=各个对照样本的比对上所述窗口的读段的总数目/（对照样本个数*所述窗口的大小）；

30 基于多组窗口的测序深度和该窗口的 GC 含量数值，利用二维回归分析法建立所述测序深度-GC 含量关系。

21. 权利要求 20 的方法，其特征在于，所述二维回归分析法为局部加权回归散点平滑法。

22. 权利要求 7-21 任一方法，其特征在于，变异位点的突变频率

= $\frac{\text{所述比对结果中支持该变异位点的读段的数目}}{\text{所述比对结果中比对到该变异位点的读段总数}}$ ，或者，变异位点的突变频率
 = $\frac{\text{所述通用比对结果中支持该变异位点的读段的数目}}{\text{所述通用比对结果中比对上该变异位点的读段总数}}$ 。

23. 权利要求 7-22 任一方法，其特征在于，所述肿瘤核酸含量

= $\frac{2 * \text{比对结果中支持所有肿瘤变异位点的读段数目}}{\text{比对结果中支持所有变异位点的读段数目}}$ ，或者，所述肿瘤核酸含量
 5 = $\frac{2 * \text{通用比对结果中支持所有肿瘤变异位点的读段数目}}{\text{通用比对结果中支持所有变异位点的读段数目}}$ 。

24. 一种确定肿瘤核酸含量的装置，所述肿瘤核酸来自体液样本，所述体液样本包含游离核酸，所述游离核酸包括肿瘤核酸和非肿瘤核酸，其特征在于，包括，

数据输入单元，用以接收所述游离核酸的测定数据，所述测定数据包括多个读段，所述测定数据是通过对所述游离核酸的至少一部分进行序列测定获得的；

10 数据输出单元，用以输出数据，包括所述肿瘤核酸含量；

存储单元，用于存储数据，其中包括可执行的程序；

处理器，与所述数据输入单元、数据输出单元及存储单元数据连接，用于执行所述可执行的程序，所述程序的执行包括完成权利要求 1-23 任一方法。

25. 一种计算机可读存储介质，其特征在于，用于存储供计算机执行的程序，所述程
 15 序的执行包括完成权利要求 1-23 任一方法。

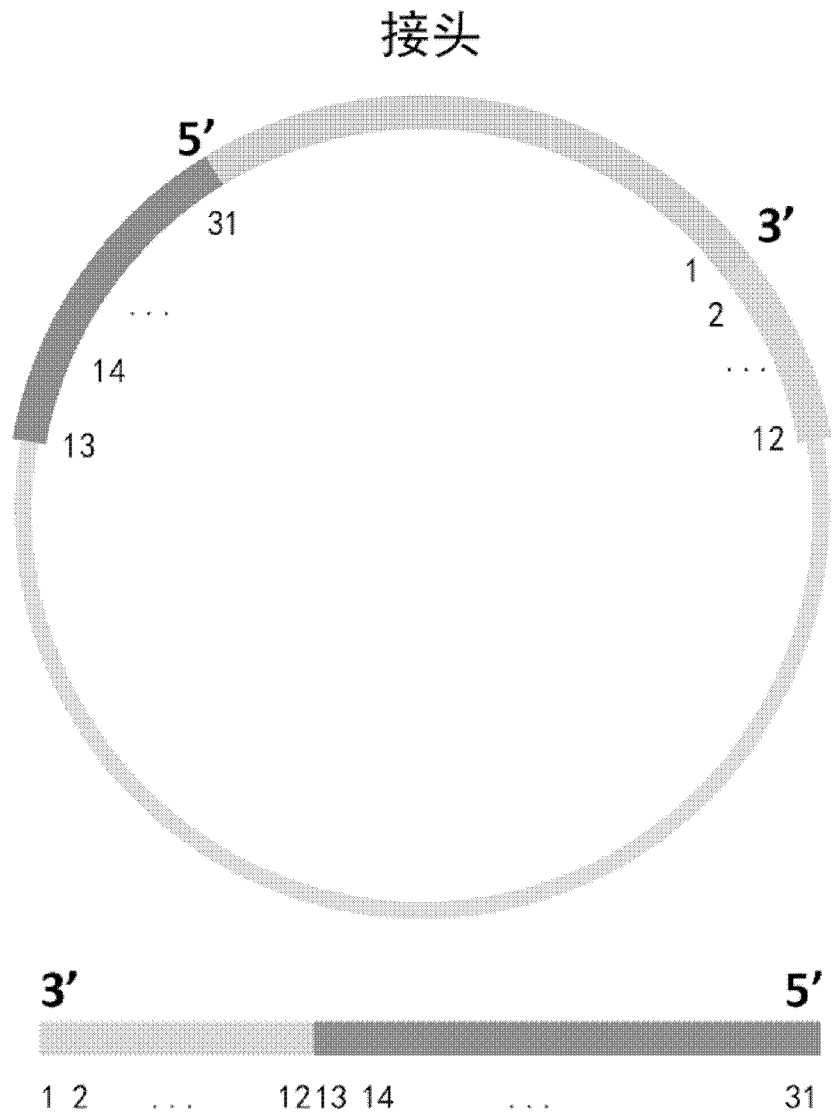


图 1

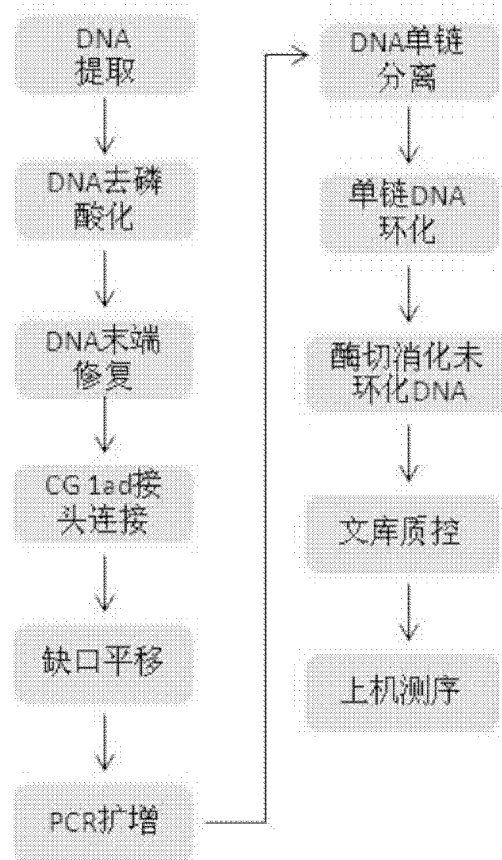


图 2

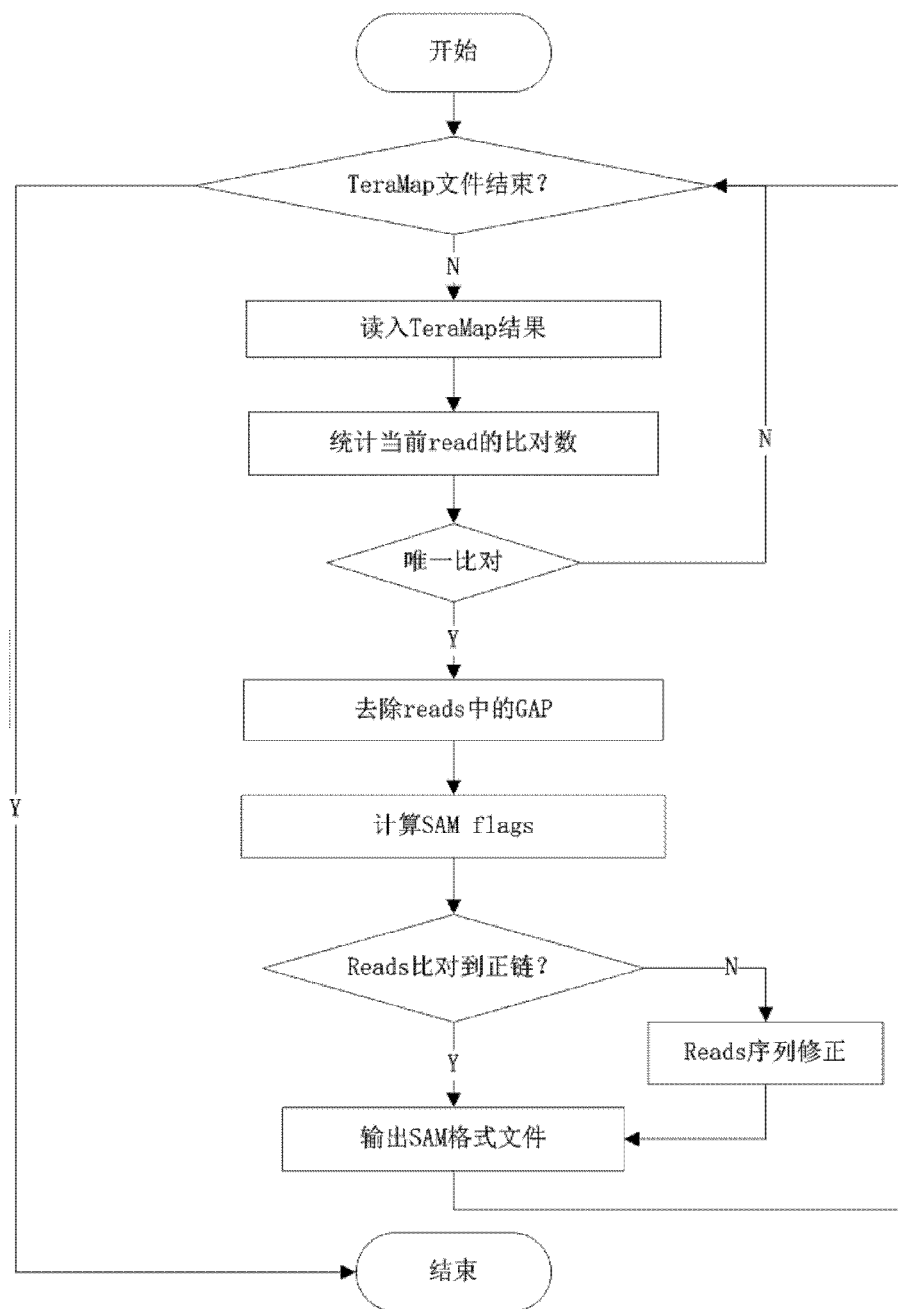


图 3

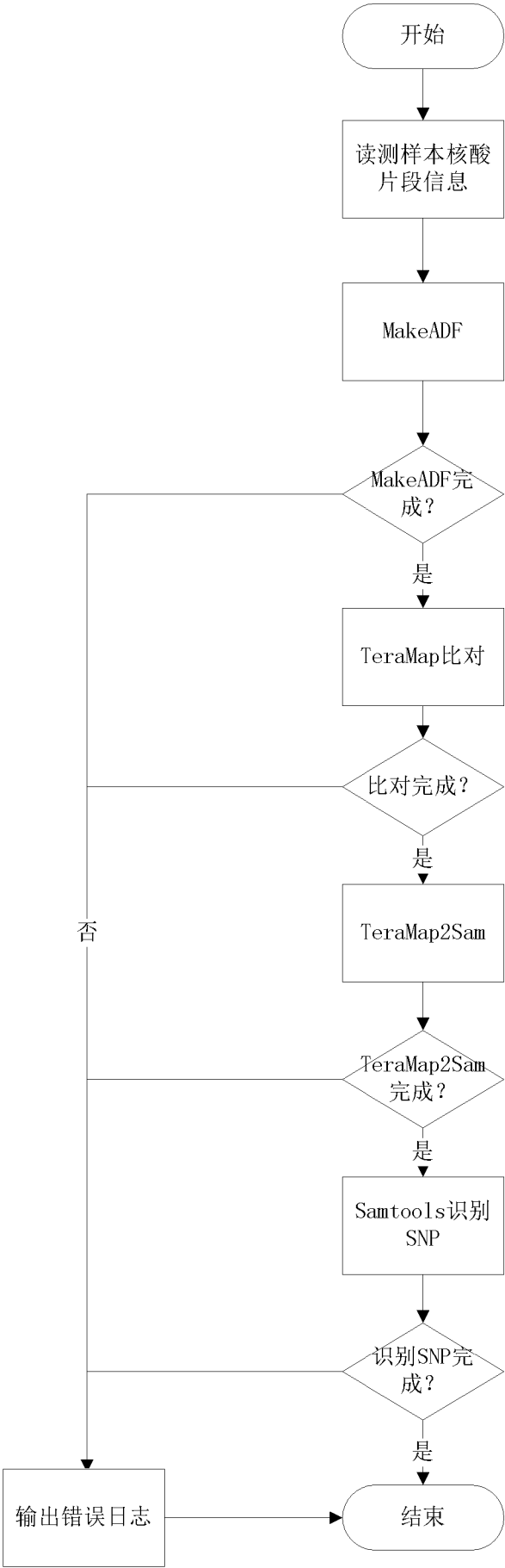


图 4

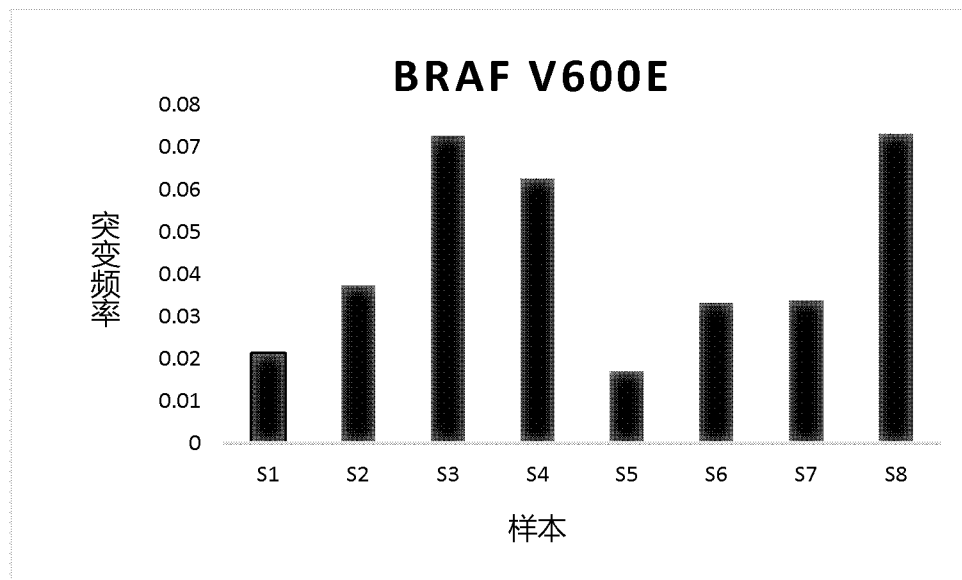


图 5

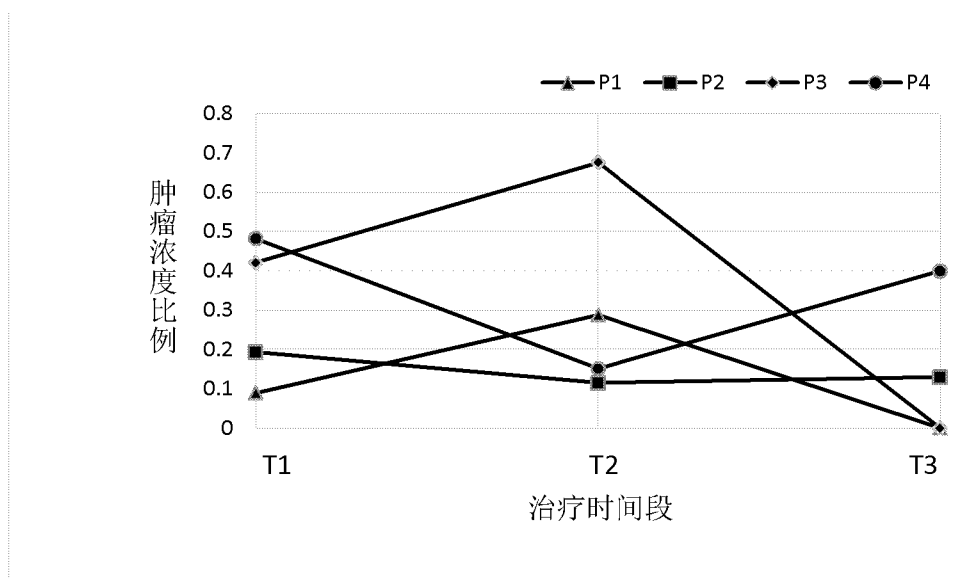


图 6

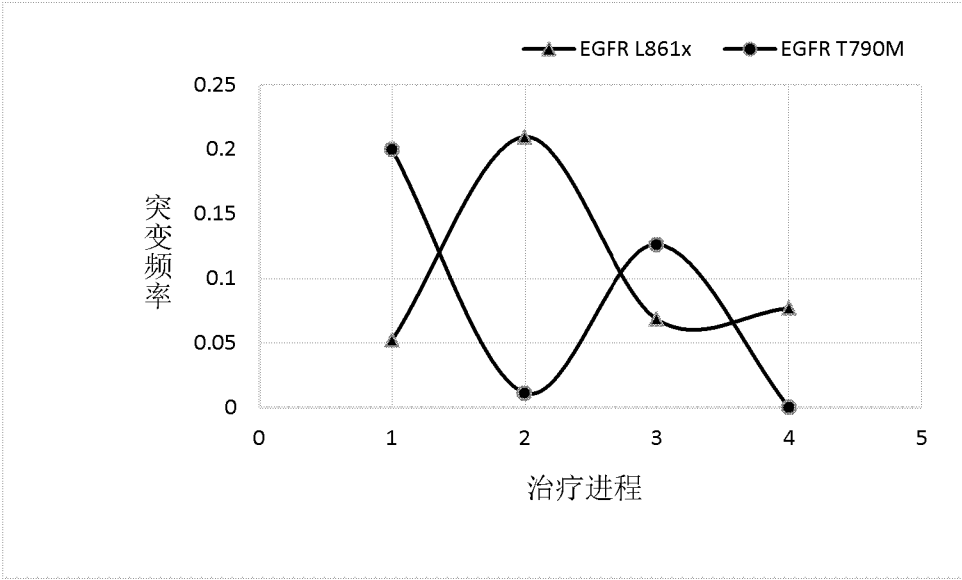


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2014/093515

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:
- a. (means)
- ☐ on paper
- ☒ in electronic form
- b. (time)
- ☒ in the international application as filed
- ☐ together with the international application in electronic form
- ☐ subsequently to this Authority for the purposes of search
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/093515

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI, SIPOABS, CPRSABS, CPEA, TWABS, CNABS, JPABS, MOABS, HKABS, CNMED, TWMED, KRABS, AUABS, DEABS, RUABS, CNKI, ISI web of Knowledge: free, cell free, cell-free, cfDNA, cfNA, circulating, nucleic acid, tumor, tumour, neoplasm, neoplasma, cancer, carcinoma, ctDNA, sequenc+, varia+, mutat+, frequen+, read

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2014039556 A1 (GUARDANT HEALTH INC et al.) 13 March 2014 (13.03.2014) see the abstract, claims 1 and 2, and description, paragraphs [0322]-[0331]	1, 24, 25
Y	WO 2014039556 A1 (GUARDANT HEALTH INC et al.) 13 March 2014 (13.03.2014) see the abstract, claims 1 and 2, and description, paragraphs [0322]-[0331]	2-23
X	WO 2014151117 A1 (UNIV LELAND STANFORD JUNIOR) 25 September 2014 (25.09.2014) see the abstract, and description, paragraphs [0019] and [0035]	1
Y	WO 2014151117 A1 (UNIV LELAND STANFORD JUNIOR) 25 September 2014 (25.09.2014) see the abstract, and description, paragraphs [0019] and [0035]	2-23
Y	CN 101932729 A (COMPLETE GENOMICS INC.) 29 December 2010 (29.12.2010) see the whole document	2-23

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 09 September 2015	Date of mailing of the international search report 21 September 2015
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer XING, Ying Telephone No. (86-10) 62411044

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2014/093515

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO 2014039556 A1	13 March 2014	CA 2883901 A1	13 March 2014
		KR 20150067161 A	17 June 2015
		CN 104781421 A	15 July 2015
		GB 201509071 D0	08 July 2015
		EP 2893040 A1	15 July 2015
WO 2014151117 A1	25 September 2014	US 2014296081 A1	02 October 2014
CN 101932729 A	29 December 2010	SG 193776 A1	30 October 2013
		AU 2008335362 A1	18 June 2009
		JP 5362738 B2	11 December 2013
		EP 2610351 B1	08 July 2015
		EP 2227563 B1	06 June 2012
		EP 2565279 B1	03 December 2014
		CN 103290106 A	11 September 2013
		CN 101932729 B	27 March 2013
		CN 103290106 B	29 July 2015
		AU 2008335362 B2	25 July 2013
		EP 2227563 A2	15 September 2010
		EP 2565279 A1	06 March 2013
		EP 2610351 A1	03 July 2013
		CA 2707901 A1	18 June 2009
		JP 2011520420 A	21 July 2011

国际检索报告

国际申请号

PCT/CN2014/093515

A. 主题的分类

C12Q 1/68(2006.01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

C12Q

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

DWPI, SIPOABS, CPRSABS, CPEA, TWABS, CNABS, JPABS, MOABS, HKABS, CNMED, TWMED, KRABS, AUABS, DEABS, RUABS, CNKI, ISI web of Knowledge: 游离, 无细胞, 循环, 核酸, 肿瘤, 癌, 测序, 变异, 突变, 频率, 读数, free, cell free, cell-free, cfDNA, cfNA, circulating, nucleic acid, tumor, tumour, neoplasm, neoplasma, cancer, carcinoma, ctDNA, sequenc+, varia+, mutat+, frequen+, read

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	WO 2014039556 A1 (GUARDANT HEALTH INC等) 2014年 3月 13日 (2014 - 03 - 13) 参见摘要, 权利要求1-2, 说明书第[0322]-[0331]段	1, 24, 25
Y	WO 2014039556 A1 (GUARDANT HEALTH INC等) 2014年 3月 13日 (2014 - 03 - 13) 参见摘要, 权利要求1-2, 说明书第[0322]-[0331]段	2-23
X	WO 2014151117 A1 (UNIV LELAND STANFORD JUNIOR) 2014年 9月 25日 (2014 - 09 - 25) 参见摘要, 说明书第[0019]、[0035]段	1
Y	WO 2014151117 A1 (UNIV LELAND STANFORD JUNIOR) 2014年 9月 25日 (2014 - 09 - 25) 参见摘要, 说明书第[0019]、[0035]段	2-23
Y	CN 101932729 A (考利达基因组股份有限公司) 2010年 12月 29日 (2010 - 12 - 29) 参见全文	2-23

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2015年 9月 9日

国际检索报告邮寄日期

2015年 9月 21日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
北京市海淀区蓟门桥西土城路6号
100088 中国

传真号 (86-10)62019451

受权官员

幸颖

电话号码 (86-10)62411044

第I栏 核苷酸和/或氨基酸序列(续第1页第1.c项)

1. 关于国际申请中所公开的任何对要求保护的发明必要的核苷酸和/或氨基酸序列，国际检索是在下列基础上进行的：

a. (提交提供)

☐ 纸件形式

☒ 电子形式

b. (提交时间)

☒ 含在申请提交时的国际申请中

☐ 以电子形式与国际申请一起提交

☐ 为检索之用随后提交本单位

2. ☐ 另外，在提交/提供了多个版本或副本的序列表的情况下，提供了关于随后提交的或附加的副本中的信息与申请时提交的申请中的信息相同或未超出申请时提交的申请中的信息范围（如适用）的所需声明。

3. 补充意见：

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2014/093515

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
WO	2014039556	A1	2014年 3月 13日	CA	2883901	A1	2014年 3月 13日
				KR	20150067161	A	2015年 6月 17日
				CN	104781421	A	2015年 7月 15日
				GB	201509071	D0	2015年 7月 8日
				EP	2893040	A1	2015年 7月 15日
WO	2014151117	A1	2014年 9月 25日	US	2014296081	A1	2014年 10月 2日
CN	101932729	A	2010年 12月 29日	SG	193776	A1	2013年 10月 30日
				AU	2008335362	A1	2009年 6月 18日
				JP	5362738	B2	2013年 12月 11日
				EP	2610351	B1	2015年 7月 8日
				EP	2227563	B1	2012年 6月 6日
				EP	2565279	B1	2014年 12月 3日
				CN	103290106	A	2013年 9月 11日
				CN	101932729	B	2013年 3月 27日
				CN	103290106	B	2015年 7月 29日
				AU	2008335362	B2	2013年 7月 25日
				EP	2227563	A2	2010年 9月 15日
				EP	2565279	A1	2013年 3月 6日
				EP	2610351	A1	2013年 7月 3日
				CA	2707901	A1	2009年 6月 18日
				JP	2011520420	A	2011年 7月 21日

表 PCT/ISA/210 (同族专利附件) (2009年7月)