

(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



(10) 国际公布号
WO 2020/227952 A1

(43) 国际公布日
2020 年 11 月 19 日 (19.11.2020) **WIPO | PCT**

(51) 国际专利分类号:
G16B 20/20 (2019.01) *G16B 50/00* (2019.01)
G16B 40/00 (2019.01)

(21) 国际申请号: PCT/CN2019/086972

(22) 国际申请日: 2019 年 5 月 15 日 (15.05.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: 深圳华大基因股份有限公司 (**BGI GENOMICS CO., LTD.**) [CN/CN]; 中国广东省深圳市盐田区洪安三街21号华大综合园7栋7层-14层, Guangdong 518083 (CN)。

(72) 发明人: 刘斯洋 (**LIU, Siyang**); 中国广东省深圳市盐田区洪安三街21号华大综合园7栋7层-14层, Guangdong 518083 (CN)。 黄树嘉 (**HUANG,**

Shujia); 中国广东省深圳市盐田区洪安三街21号华大综合园7栋7层-14层, Guangdong 518083 (CN)。 金鑫 (**JIN, Xin**); 中国广东省深圳市盐田区洪安三街21号华大综合园7栋7层-14层, Guangdong 518083 (CN)。

(74) 代理人: 北京品源专利代理有限公司 (**BEYOND ATTORNEYS AT LAW**); 中国北京市海淀区莲花池东路39号西金大厦6层, Beijing 100036 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,

(54) **Title:** BASE MUTATION DETECTION METHOD AND APPARATUS BASED ON SEQUENCING DATA, AND STORAGE MEDIUM

(54) 发明名称: 基于测序数据的碱基突变检测方法、装置及存储介质

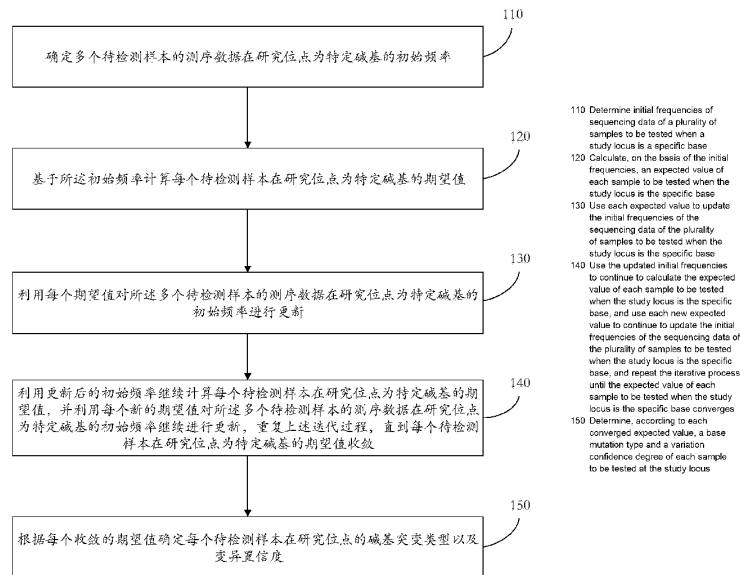


图 1

(57) **Abstract:** Disclosed are a base mutation detection method and apparatus based on sequencing data, and a storage medium. The method comprises: determining initial frequencies of sequencing data of a plurality of samples to be tested when a study locus is a specific base; calculating, on the basis of the initial frequencies, an expected value of each sample to be tested when the study locus is the specific base; using each expected value to update the initial frequencies of the sequencing data of the plurality of samples to be tested when the study locus is the specific base; using the updated initial frequencies to continue to calculate the expected value of

[见续页]

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))。

each sample to be tested when the study locus is the specific base until the expected value of each sample to be tested when the study locus is the specific base converges; and determining, according to each converged expected value, a base mutation type and a variation confidence degree of each sample to be tested at the study locus.

(57) 摘要: 本文公布一种基于测序数据的碱基突变检测方法、装置及存储介质, 所述方法包括确定多个待检测样本的测序数据在研究位点为特定碱基的初始频率; 基于所述初始频率计算每个待检测样本在研究位点为特定碱基的期望值; 利用每个期望值对所述多个待检测样本的测序数据在研究位点为特定碱基的初始频率进行更新; 利用更新后的初始频率继续计算每个待检测样本在研究位点为特定碱基的期望值, 直到每个待检测样本在研究位点为特定碱基的期望值收敛; 根据每个收敛的期望值确定每个待检测样本在研究位点的碱基突变类型以及变异置信度。