

(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 2025/067173 A1

(43) 国际公布日
2025 年 4 月 3 日 (03.04.2025)

- (51) 国际专利分类号:
G16B 30/00 (2019.01) *G06F 16/35* (2025.01)
G16B 40/20 (2019.01) *G06F 18/24* (2023.01)
G06F 16/33 (2025.01) *G06F 40/289* (2020.01)
- (21) 国际申请号: PCT/CN2024/120794
- (22) 国际申请日: 2024 年 9 月 24 日 (24.09.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202311261520.3 2023年9月27日 (27.09.2023) CN
- (71) 申请人: 北京立康生命科技有限公司 (BEIJING LIKANG LIFE SCIENCES AND TECH CO., LTD.) [CN/CN]; 中国北京市大兴区北京

经济技术开发区经海四路 22 号院四区 5 号楼 3 层 311 100023 (CN)。

- (72) 发明人: 陈立 (CHEN, Li); 中国北京市大兴区北京经济技术开发区经海四路 22 号院四区 5 号楼 3 层 311 100023 (CN)。贾明明 (JIA, Mingming); 中国北京市大兴区北京经济技术开发区经海四路 22 号院四区 5 号楼 3 层 311 100023 (CN)。李情 (LI, Qing); 中国北京市大兴区北京经济技术开发区经海四路 22 号院四区 5 号楼 3 层 311 100023 (CN)。刘三阳 (LIU, Sanyang); 中国北京市大兴区北京经济技术开发区经海四路 22 号院四区 5 号楼 3 层 311 100023 (CN)。姚明月 (YAO, Mingyue); 中国北京市大兴区北京经济技术开发区经海四路 22 号院四区 5 号楼 3 层 311 100023 (CN)。

(54) Title: METHOD FOR IDENTIFYING PEPTIDE FRAGMENT BOUND TO CLASS MHC-I/HLA-I AND RECOGNIZED BY TCR, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 一种用于鉴定与MHC-I/HLA-I类结合及TCR识别肽段的方法、设备及存储介质

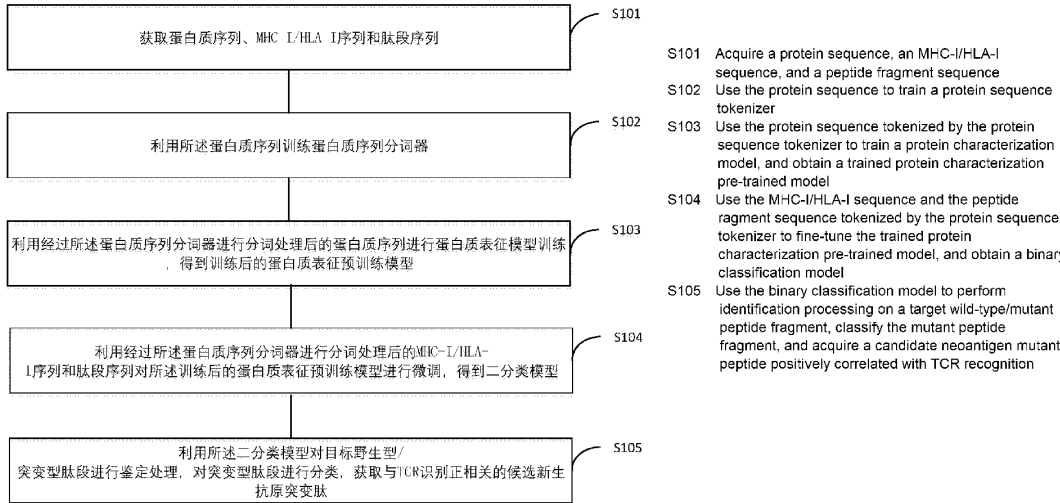


图 1

(57) Abstract: A method for identifying a peptide fragment bound to a class MHC-I/HLA-I and recognized by TCR, a device, and a storage medium. The method comprises: acquiring a protein sequence, an MHC-I/HLA-I sequence, and a peptide fragment sequence; using the protein sequence to train a protein sequence tokenizer; using the protein sequence tokenized by the protein sequence tokenizer to train a protein characterization model, and obtaining a trained protein characterization pre-trained model; using the MHC-I/HLA-I sequence and the peptide fragment sequence tokenized by the protein sequence tokenizer to fine-tune the trained protein characterization

[续页]

(74) 代理人:北京谨礼知识产权代理事务所(特殊普通合伙)(**KOI IP OFFICE**); 中国北京市昌平区沙河镇育荣园区1号楼264室 102206 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))。
- 包括说明书序列表部分(细则5.2(a))。

pre-trained model, and obtaining a binary classification model; using the binary classification model to perform identification processing on a target wild-type/mutant peptide fragment, classifying the mutant peptide fragment, and acquiring a candidate neoantigen mutant peptide positively correlated with TCR recognition.

(57) 摘要: 一种用于鉴定与MHC-I/HLA-I类结合及TCR识别肽段的方法、设备及存储介质, 其方法包括: 获取蛋白质序列、MHC-I/HLA-I序列和肽段序列; 利用蛋白质序列训练蛋白质序列分词器; 利用经过蛋白质序列分词器进行分词处理后的蛋白质序列进行蛋白质表征模型训练, 得到训练后的蛋白质表征预训练模型; 利用经过蛋白质序列分词器进行分词处理后的MHC-I/HLA-I序列和肽段序列对训练后的蛋白质表征预训练模型进行微调, 得到二分类模型; 利用二分类模型对目标野生型/突变型肽段进行鉴定处理, 对突变型肽段进行分类, 获取与TCR识别正相关的候选新生抗原突变肽。