

(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
2013 年 7 月 4 日 (04.07.2013)

W I P O | P C T

(10) 国际公布号
W O 2013/097744 A 1

(51) 国际分类号：

C12N 15/11 (2006.01) C12Q 1/68 (2006.01)
C12N 15/13 (2006.01) C40B 40/08 (2006.01)
C12N 15/10 (2006.01) C40B 50/06 (2006.01)
C12N 15/12 (2006.01) C12M 1/34 (2006.01)

(21) 国际申请号：PCT/CN20 12/087668

(22) 国际申请日：2012 年 12 月 27 日 (27.12.2012)

(25) 申请语言：中文

(26) 公布语言：中文

(30) 优先权：

201110444740.0 2011 年 12 月 27 日 (27.12.2011) CN
201210011443.1 2012 年 1 月 13 日 (13.01.2012) CN
201210017173.5 2012 年 1 月 19 日 (19.01.2012) CN
201210055230.9 2012 年 3 月 5 日 (05.03.2012) CN

(71) 申请人：深圳华大基因科技有限公司 (BGI SHEN - ZHEN CO., LIMITED) [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼 11F-3, Guangdong 518083 (CN)。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：刘晓 (LIU, Xia o); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。洪雪玉 (HONG, Xue y u); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。曾晓静 (ZENG, Xiaojing); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。张瑜芳 (ZHANG, Ruifang); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。苏政 (SU, Zheng); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。张俊杰 (ZHANG, Junjie); 中国广东省深圳市盐田区北山工业区综合

楼, Guangdong 518083 (CN)。郑春录 (ZHENG, Chunting); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。王俊 (WANG, Jun); 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 PROP - ERTY 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, ML, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, 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, 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))。

(54) Title: PRIMER COMPOSITION FOR USE IN AMPLIFYING CDR3 CODING SEQUENCE AND USE OF THE PRIMER COMPOSITION

(54) 发明名称：用于扩增 CDR3 编码序列的引物组合物及其用途

(57) Abstract: Provided in the present invention are a primer composition for use in amplifying CDR3 coding sequence and a use of the primer composition. The primer composition comprises: a first primer set and a second primer set. The first primer set consists of at least one type of V-region primer; each type of the at least one type of V-region primer contains a sequence complementary to at least one V-gene fragment. The second primer set consists of at least one type of J-region primer; each type of the at least one type of J-region primer contains a sequence complementary to at least one J-gene fragment. The source of the CDR3 coding sequence is selected from at least one among immunoglobulin heavy chain, immunoglobulin light chain, T cell receptor α chain, and T cell receptor β chain.

(57) 摘要：本发明提供了用于扩增 CDR3 编码序列的引物组合物及其用途。引物组合物包括：第一引物组和第二引物组，该第一引物组由至少一种 V 区引物组成，该至少一种 V 区引物的每一种均包含与至少一个 V 基因片段互补的序列；第二引物组由至少一种 J 区引物组成，该至少一种 J 区引物的每一种均包含与至少一个 J 基因片段互补的序列，其中，该 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。



WO 2013/097744 A1

用于扩增 CDR3 编码序列的引物组合物及其用途

优先权信息

本申请请求 2011 年 12 月 27 日向中国国家知识产权局提交的、专利申请号为 201110444740.0 的专利申请, 2012 年 1 月 13 日向中国国家知识产权局提交的、专利申请号为 201210011443.1 的专利申请, 2012 年 1 月 19 日向中国国家知识产权局提交的、专利申请号为 201210017173.5 的专利申请, 以及 2012 年 3 月 5 日向中国国家知识产权局提交的、专利申请号为 201210055230.9 的专利申请的优先权和权益, 并且通过参照将其全文并入此处。

技术领域

本发明涉及生物技术领域。具体地, 本发明涉及用于扩增 CDR3 编码序列的引物组合物及其用途。更具体地, 本发明涉及用于扩增 CDR3 编码序列的引物组合物、富集 CDR3 编码序列的方法、构建 CDR3 编码序列的测序文库的方法、确定 CDR3 编码序列的序列信息的方法、确定个体免疫状态的方法、确定个体免疫状态的系统以及试剂盒。

背景技术

免疫系统是机体抵御病原菌入侵与免疫调节的重要系统, 对其进行研究有很重要的意义。免疫球蛋白、T 细胞受体 (TCR) 和 HLA (人类白细胞抗原) 是人类基因组中最活跃的免疫大分子, 决定和反映了人和环境的相互作用。免疫大分子的多样性使得机体能识别无数的外来物质和清除体内产生的部分代谢物。免疫大分子多样性产生的机制主要有基因重排、体细胞突变、非模板核苷酸的插入与缺失等。

免疫球蛋白重链重排指 γ 、D、J 三种基因片段重排形成多种 CDR3 编码序列, 首先是 DJ 基因片段重排, 重排后的基因片段再与 V 基因进行组合, 得到编码 CDR3 的基因; 在免疫球蛋白轻链中, 先进行 κ 链基因重排, 若重排失败, 则 λ 链基因开始重排, 由重排、非模板核苷酸的插入与缺失等机制产生多种多样特异性的基因片段。由此, 一个个体内可能产生 10^{11} 个特异性的分子。

T 细胞受体的多样性主要是通过人外周血 T 细胞主要表达的 T 细胞受体的 α 链和 β 链的重排引起的, 而 T 细胞受体 α 链和 β 链重排引起的 T 细胞受体 α 链和 β 链 CDR3 编码序列的多样性能较好地代表 T 细胞受体的多样性甚至个人的免疫状态。

因此, 对免疫球蛋白重链和轻链、T 细胞受体 α 链和 β 链的 CDR3 编码序列进行研究, 意义重大。然而, 目前对于上述 CDR3 编码序列的研究仍有待改进。

发明内容

本发明旨在至少解决现有技术中存在的技术问题之一。为此, 本发明的一个目的在于提出一种能够有效地对免疫球蛋白重链和轻链、T 细胞受体 α 链和 β 链的 CDR3 编码序列进行富集的手段。

根据本发明的一个方面, 本发明提供了一种用于扩增 CDR3 编码序列的引物组合物。根据本发明的实施例, 该引物组合物包括第一引物组, 该第一引物组由至少一种 V 区引物组成, 该至少一种 V 区引物的每一种均包含与至少一个 V 基因片段互补的序列; 以及第二引物组, 该第二引物组由至少一种 J 区引物组成, 该至少一种 J 区引物的每一种均包含与至少一个 J 基因片段互补的序列, 其中, 该 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。利用该引物组合物, 能够有效地对免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的 CDR3 编码序列进行富集, 从而为对上述 CDR3 进行深入研究提供了便利的工具。

根据本发明的另一方面, 本发明还提供了一种富集 CDR3 编码序列的方法。根据本发明的实施例, 该方法包括以下步骤: 提供核酸样本, 所述核酸样本中包含编码 CDR3 的核酸序列; 以及利用前面所述的本发明的引物组合物, 以所述核酸样本作为模板, 进行 PCR 扩增, 以便获得富集所述 CDR3 编码序列的扩增产物, 其中, 所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。由此, 能够有效地富集免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的 CDR3 编码序列。

根据本发明的又一方面, 本发明提供了一种构建 CDR3 编码序列的测序文库的方法。根据本发明的实施例, 该方法包括以下步骤: 根据前面所述的方法, 获得富集所述 CDR3 编码序列的扩增产物; 以及针对所述扩增产物, 构建 DNA 测序文库, 所述 DNA 测序文库构成所述 CDR3 编码序列的测序文库, 其中, 所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。由此, 可以在对免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β

链的 CDR3 编码序列进行富集的基础上,构建可以用于测序的测序文库。

根据本发明的再一方面,本发明提出了一种确定 CDR3 编码序列的序列信息的方法。根据本发明的实施例,该方法包括以下步骤:根据前面所述的方法,构建 CDR3 编码序列的测序文库;以及对所述 CDR3 编码序列的测序文库进行测序,以便确定所述 CDR3 编码序列的序列信息,其中,所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。由此,能够有效确定免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的 CDR3 编码序列的序列信息。

根据本发明的又一方面,本发明提供了一种确定个体免疫状态的方法。根据本发明的实施例,该方法包括以下步骤:根据前面所述的方法,对所述个体的 CDR3 编码序列进行测序,以便获得由多个测序数据构成的测序结果;以及基于所述测序结果,确定所述个体的免疫状态,其中,所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。通过该方法,能够有效地获得个体的免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β 链的 CDR3 编码序列的序列信息,从而可以有效地确定个体免疫状态。

根据本发明的另一方面,本发明提出了一种确定个体免疫状态的系统。根据本发明的实施例,该系统包括:CDR3 编码序列富集装置,所述 CDR3 编码序列富集装置中设置有前面所述的引物组合物,以便对所述个体的核酸样本富集 CDR3 编码序列,其中,所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种;文库构建装置,所述文库构建装置与所述 CDR3 编码序列富集装置相连,以便针对所述经过富集的 CDR3 编码序列构建 CDR3 编码序列的测序文库;测序装置,所述测序装置与所述文库构建装置相连,用于对所述 CDR3 编码序列的测序文库进行测序,以便获得由多个测序数据构成的测序结果;以及分析装置,所述分析装置与所述测序装置相连,用于基于测序结果,确定所述个体的免疫状态。由此,利用该系统,能够有效地实施前述确定个体免疫状态的方法,从而能够有效地确定个体的免疫状态。

根据本发明的又一方面,本发明还提供了一种试剂盒。根据本发明的实施例,该试剂盒中设置有前面所述的引物组合物。由此,该试剂盒能够用于检测免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的 V-J 重排,或者用于检测免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的 CDR3 的编码序列。

本发明的附加方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本发明的实践了解到。

附图说明

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例的描述中将变得明显和容易理解,其中:

图 1 是免疫球蛋白重链中由 V_H 、 D_H 和 J_H 构成的 CDR3 的结构示意图;

图 2 是根据本发明一个实施例的对 CDR3 编码序列进行富集和测序的流程示意图;

图 3 是根据本发明一个实施例的用于确定个体的免疫状态的系统结构示意图;

图 4 是本发明实施例 1 中的多重 PCR 产物琼脂糖凝胶电泳结果;

图 5 是本发明实施例 1 中,对免疫球蛋白重链 CDR3 编码序列测序后,所得到的 V-J 基因片段配对分布及丰度的结果示意图;

图 6 是本发明实施例 2 中,引物扩增免疫球蛋白轻链中 V-J 重排构成的 CDR3 序列的示意图;

图 7 是本发明实施例 2 中的多重 PCR 产物琼脂糖凝胶电泳结果;

图 8 是本发明实施例 2 中,对免疫球蛋白轻链 CDR3 序列测序后,所得到的 V_K-J_K 基因片段配对分布及丰度的结果示意图;

图 9 是本发明实施例 2 中,对免疫球蛋白轻链 CDR3 序列测序后,所得到的 $V_\lambda-J_\lambda$ 基因片段配对分布及丰度的结果示意图;

图 10 是本发明实施例 3 中的多重 PCR 产物琼脂糖凝胶电泳结果;

图 11 是本发明实施例 3 中,对 T 细胞受体 α 链 CDR3 编码序列进行测序后,所得到的 V-J 基因片段配对分布及丰度的结果示意图;

图 12 是本发明实施例 4 中的多重 PCR 产物琼脂糖凝胶电泳结果;以及

图 13 是本发明实施例 4 中,对 T 细胞受体 β 链 CDR3 编码序列进行测序后,所得到的 V-J 基因片段配对分布及丰度的结果示意图。

发明详细描述

下面详细描述本发明的实施例。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，术语“第一”、“第二”仅用于描述目的，而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此，限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。进一步地，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。

用于扩增 CDR3 编码序列的引物组合物

根据本发明的一个方面，本发明提出了一种用于扩增 CDR3 编码序列的引物组合物。根据本发明的实施例，该引物组合物包括第一引物组和第二引物组。其中，第一引物组由至少一种 V 区引物组成，第二引物组由至少一种 J 区引物组成，该 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。

在本文中所使用的术语“V 区引物”指的是这样的一种引物，其能够特异性地识别免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β 链家族中的 V 基因片段，从而可以引导进行聚合酶链式反应，由此第一引物组中所包含的 V 区引物的每一种均包含与至少一个 V 基因片段互补的序列。类似的，在本文中所使用的术语“J 区引物”指的是这样的一种引物，其能够特异性地识别免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β 链家族中的 J 基因片段，从而可以引导进行聚合酶链式反应，由此第一引物组中所包含的 J 区引物的每一种均包含与至少一个 J 基因片段互补的序列。进而，在 V 区引物和 J 区引物的引导下，可以通过扩增反应例如 PCR 扩增，从含有编码 CDR3 的核酸序列的核酸样本中特异性地扩增包含 V 基因片段和 J 基因片段的编码序列。由于免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链，是由 V、D、J 基因片段或着 V、J 基因片段重排产物所编码的，因而通过特异性识别 V 和 J 基因片段的引物，即第一引物组和第二引物组，能够有效地从包含上述 CDR3 编码序列的核酸样本中扩增获得 CDR3 编码序列的扩增产物。进而，通过采用根据本发明实施例的引物组合物，能够高特异性地从样本中扩增富集获得 CDR3 的编码序列，且不会存在其他的干扰序列。从而能够实现免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链以及 T 细胞受体 β 链的 CDR3 编码序列的有效富集，进而为针对 CDR3 进行深入研究提供了便利的工具。

下面针对不同来源的 CDR3 编码序列，例如来源于免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β 链的 CDR3 编码序列，分别详细描述本发明的用于扩增 CDR3 编码序列的引物组合物。

1、CDR3 编码序列来源于免疫球蛋白重链

当 CDR3 编码序列来源于免疫球蛋白重链时，本发明的用于扩增 CDR3 编码序列的引物组合物的可以具有以下特征：

根据本发明的实施例，V 区引物和 J 区引物在 PCR 扩增过程中的作用，并不受特别限制。根据本发明的一个具体示例，V 区引物可以作为正义链引物，J 区引物可以作为反义链引物。发明人发现，通过如此设置 V 区引物和 J 区引物，能够进一步提高将引物组合物用于富集 CDR3 编码序列时的富集效率。根据本发明的实施例，上述引物组合物能够适用的免疫球蛋白类型，不受特别限制，本领域技术人员可以根据研究需要，选择适当的免疫球蛋白作为研究对象。根据本发明的一个实施例，所述免疫球蛋白为人类免疫球蛋白。由此，可以有效地将引物组合物用于对人类免疫球蛋白重链 CDR3 编码序列进行富集，从而可以有效地用于对人体免疫状况进行研究。

另外，根据本发明的实施例，可以通过对 V 区引物和 J 区引物的序列进行选择，实现一条引物可以识别多种 V 基因片段，从而可以进一步提高扩增的效率，减少引物的数目，降低成本。根据本发明的一个实施例，第一引物组的至少一种引物包含与多个 V 基因片段的保守区互补的序列。由此，可以在减少引物的数量的同时，提高对免疫球蛋白重链 CDR3 编码序列进行富集的效率，发明人还发现，这样操作能够提高各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。类似的，根据本发明的一个实施例，第二引物组的至少一种包含与多个 J 基因片段的保守区互补的序列。由此，可以在减少引物的数量的同时，提高对免疫球蛋白重链 CDR3 编码序列进行富集的效率，发明人还发现，这样操作能够提高各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。另外，根据 J 基因片段的序列特点，第二引物组可以仅由一条引物构成。因而根据本

发明的一个实施例，第二引物组由一种简并引物构成，该引物包含与多个 J 基因片段的保守区互补的序列。由此，可以在减少引物的数量的同时，提高对免疫球蛋白重链 CDR3 编码序列进行富集的效率，发明人还发现，这样操作能够提高各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。

- 5 具体地，根据本发明的实施例，针对人类免疫球蛋白重链 CDR3 的序列特征，本发明提出了一组 V 区引物 (SEQ ID NO: 1-20) 和一种 J 区引物 (SEQ ID NO: 21)，其序列和名称分别总结如下：

引物序号	序列	SEQ ID NO:
V _H 1.1	AGAGTCACCATAACCGCGGACA	1
V _H 1.2	GTCACCAGTACCAGGGACACG	2
V _H 1.3	AGTCACCATTACCAGGGACACA	3
V _H 1.4	AGAGTCACCATGACCAGGAACA	4
V _H 1.5	AGTCACCATGACCACAGACACA	5
V _H 1.6	AGAGTCACCATGACCGAGGACA	6
V _H 1.7	AGTCACCATTACYAGGGACAGG	7
V _H 1.8	GTCACCATGACCAGGGACACG	8
V _H 1.9	AGTCACCATTACCAGGGACATG	9
V _H 1.10	GAGTCACGATTACCRCGGACG	10
V _H 1.11	GAGTCACGATTACCGCGGACA	11
V _H 2	ACCAGGCTCACCATYWCCAAGG	12
V _H 3.1	GGCCGRTTCACCATCTCMAG	13
V _H 3.2	GGCMGATTACCATCTCCAG	14
V _H 3.3	GGCCGATTCGCCATCTCCAG	15
V _H 4.1	AGTCGAGTCACCATGTCAGTAG	16
V _H 4.2	AGTCGAGTYACCATATCMGTAG	17
V _H 5	GGCCASGTCACCATCTCAGC	18
V _H 6	AGTCGAATAACCATCAACCCAG	19
V _H 7	GGACGGTTTGTCTTCTCCTTGG	20
J _H	CTGAGGAGACGGTGACCAKKG	21

R=A/G, Y=C/T, M=A/C, K=G/T, S=C/G, W=A/T

上表中所列出的序列，可以至少识别一种 V 区或 J 区亚家族的编码序列，其所识别的代表性亚家族总结如下：

引物序号	亚家族
V _H 1.1	IGHV1-F-F2
V _H 1.2	IGHV1-2-F2
V _H 1.3	IGHV1-3-F2
V _H 1.4	IGHV1-8-F2
V _H 1.5	IGHV1-18-F2

V _H 1.6	IGHV1-24-F2
V _H 1.7	IGHV1-45-F2
V _H 1.8	IGHV1-46/2-F2
V _H 1.9	IGHV1-58-F2
V _H 1.10	IGHV1-69-F2.1
V _H 1.11	IGHV1-69-F2.2
V _H 2	IGHV2-70, 2-26, 2-5
V _H 3.1	IGHV3-23*2, 3-73, 3-15, 3-49, 3-72
V _H 3.2	IGHV3-64, 3-66, 3-D, IGH3-30, 3-30-3, 3-NL1, 3-33, 3-53, 3-43, 3-48, 3-20, 3-74, 3-7, 3-9, 3-21, 3-11, 3-13, 3-23
V _H 3.3	IGHV3-30*9
V _H 4.1	IGHV4-28, 4-4, 4-34*10, IGH4-59*02/*04
V _H 4.2	IGHV4-4, 4-30-2, 4-30-4, 4-34, 4-39, 4-59, 4-61, 4-b, 6-1, 4-31
V _H 5	IGH5-51, 5-a
V _H 6	IGH6-1
V _H 7	IGHV7-4-1
J _H	IGHJ1, 2,3,4,5,6

需要说明的是，上表中所列出的 V 区或 J 区亚家族并不是穷尽的，而仅仅是目前有报道的所有亚家族。事实上，发明人惊奇地发现，采用上述具体的引物序列，能够全面覆盖人类免疫球蛋白 CDR3 的全部亚家族，包括发现新的 CDR3 亚家族，从而能够全面的富集人类免疫球蛋白重链 CDR3 编码序列，另外，发明人还发现通过采用上述引物序列，能够同时在一个 PCR 反应体系中进行多重引物 PCR（有时也称为“多重 PCR”），可以有效地对样本中所包含的 CDR3 编码序列进行扩增，并且能够保证各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。根据本发明的实施例，多重引物 PCR 的退火温度可以为 60 摄氏度。发明人惊奇地发现，当退火温度为 60 摄氏度时，多重引物 PCR 扩增 CDR3 的效率得到显著提高。

由此，现有技术相比，本发明含有可扩增 IMGT 数据库免疫球蛋白 CDR3 序列所有 V 区和 J 区的特异性引物，可以最全面的富集人类免疫球蛋白重链 CDR3 的编码序列，扩增出来的产物能够区分免疫球蛋白重链的各个亚家族，且同一模板不会被两组引物特异性结合扩增。本发明中的引物能更好的呈现机体免疫大分子的集合与分布，对找到与疾病相关的信息或者免疫系统变化的信息具有更可靠的作用。

2、CDR3 编码序列来源于免疫球蛋白轻链

当 CDR3 编码序列来源于免疫球蛋白轻链时，本发明的用于扩增 CDR3 编码序列的引物组合物的可以具有以下特征：

由于免疫球蛋白轻链有 λ 、 κ 两种类型，存在 V _{λ} -J _{λ} 重排和 V _{κ} -J _{κ} 重排两种方式，具体地，在本发明实施例中，通过特异性识别 V _{λ} 和 J _{λ} 基因片段引物，即第一引物组中的 SEQ ID NO: 22-36 和第二引物组中的 SEQ ID NO: 41-43，能够选择性地有效地从包含 CDR3 编码序列的核酸样本中扩增获得 λ 型轻链 CDR3 序列的扩增产物； κ 型轻链 CDR3 扩增产物的选择性获得类似，利用第一引物组中的 SEQ ID NO: 37-40 和第二引物组中的 SEQ ID NO: 44 和 SEQ ID NO: 45 引物组来进行；类似的，混合使用第一引物组 SEQ ID NO: 22-40 和第二引物组的 SEQ ID NO: 41-45，多重 PCR 可一并获得 λ 、 κ 两种类型轻链的 CDR3 序列；在轻链中，CDR3 是 V-J 重排的产物，因而通过采用根据本发明实施例的引物组合物，能够高特异性地从样本中扩增富集获得 CDR3 序列，而不会存在其他的干扰序列。从而能够实现轻链 CDR3 序列的有效富集，进而为针对 CDR3 进行深入研究提供了便利的工具。

根据本发明的实施例，V 区引物和 J 区引物在 PCR 扩增过程中的作用，并不受特别限制。根据本发明的一个具体示例，V 区引物可以作为正义链引物，J 区引物可以作为反义链引物。发明人发现，通过如此设置 V 区引物和 J 区引物，能够进一步提高将引物组合物用于富集免疫球蛋白轻链 CDR3 序列时的富

集效率。根据本发明的实施例，上述引物组合物能够适用的免疫球蛋白类型，不受特别限制，本领域技术人员可以根据研究需要，选择适当的免疫球蛋白作为研究对象。根据本发明的一个实施例，所述免疫球蛋白为人类免疫球蛋白。由此，可以有效地将引物组合物用于对人类免疫球蛋白轻链 CDR3 序列进行富集，从而可以有效地用于对人体免疫状况进行研究。

- 5 另外，根据本发明的实施例，可以通过对 V 区引物和 J 区引物的序列进行选择，实现一条引物可以识别多种 V 基因片段，从而可以进一步提高扩增的效率，减少引物的数目，降低成本。根据本发明的一个实施例，第一引物组的至少一种引物包含与多个 V 基因片段的保守区互补的序列。由此，可以在减少引物的数量的同时，提高对免疫球蛋白轻链 CDR3 序列进行富集的效率，发明人还发现，这样操作能够提高各 CDR3 序列扩增的均一性，从而能够真实地反映 CDR3 序列在宿主中的分布比例。类似的，根据
- 10 本发明的一个实施例，反向引物组的至少一种包含与多个 J 基因片段的保守区互补的序列。由此，可以在减少引物的数量的同时，提高对免疫球蛋白轻链 CDR3 序列进行富集的效率，发明人还发现，这样操作能够提高各 CDR3 序列扩增的均一性，从而能够真实地反映 CDR3 序列在宿主中的分布比例。

具体地，根据本发明的实施例，针对人类免疫球蛋白轻链 CDR3 的序列特征，本发明提出了一组 V 区引物 (SEQ IDNO: 22-40) 和一组 J 区引物 (SEQ IDNO: 41-45)，其序列和名称分别总结如下：

引物序号	引物序列	SEQ ID NO:
V _λ 1	CCAAGTCTGGC RC STCAGCC	22
V _λ 2.1	CCTCAGGGGTCCCTGATCGCT	23
V _λ 2.2	CGGCCCTCAGGGGTTTCTAATC	24
V _λ 3.1	GGATCCCTGAGM G ATTCTCTGG	25
V _λ 3.2	ATTCTCYGGCTCCAGCTCAG	26
V _λ 3.3	GGCCCTCAGGGATCCCAGAC	27
V _λ 3.4	AACGATTCTCTGGGTCCACCTC	28
V _λ 4.1	ATCGCTTCATGGGCTCCAGTTC	29
V _λ 4.2	ATCGCTTCTCAGGCTCCAGCTC	30
V _λ 5	CCCAGCCGCTTCTCTGGAT	31
V _λ 6	TGGGGTCCCTGATCGGTTCTCT	32
V _λ 7	GGGCAAAGCTGCCCTGAC M CT	33
V _λ 8	TGGCTCCATCCTTGGAACA	34
V _λ 9	GCCTGAATCGGTACCTGACCATC	35
V _λ 10	TCTGCATCCAGGTCAGGAAACAC	36
V _κ 1.1	CAGATTTCACTCTCACH A TCA	37
V _κ 1.2	CAGATTTTACTTTCA C CATCAG	38
V _κ 2	GGACAGATTTYACACTGAAAATCAG	39
V _κ 5	GATTCAGTGGCAGCGGGTATGGA	40
J _λ 1/2/3	CTAGGACGGT S ASCTTGGT	41
J _λ 6	CGAGGACGGTCACCTTGGT	42
J _λ 7	CGAGG R CGGTCAGCTGGGT	43
J _κ 1.2	CTTACGTTTGATCTCCACCTTGGT	44
J _κ 5	CTTACGTTTAATCTCCAGTCGTGT	45

15 R=A/G, Y=C/T, M=A/C, H=A/C/T, S=C/G, W=A/T

上表中所列出的序列，可以至少识别一种 V 区或 J 区亚家族的序列，其所识别的代表性亚家族总结如下：

引物 序号	亚家族
V _λ 1	IGLV1-36, 1-44,1-47, 1-40,1-51
V _λ 2.1	IGLV2-8,2-11,2-18, 2-23,
V _λ 2.2	IGLV2-14
V _λ 3.1	IGLV3-1,3-9,3-12,3-21
V _λ 3.2	IGLV3-10, 3-16, 3-25, 3-27
V _λ 3.3	IGLV3-19
V _λ 3.4	IGLV3-22
V _λ 4.1	IGLV4-3
V _λ 4.2	IGLV4-60,4-69
V _λ 5	IGLV5-45,5-39,5-37,5-52
V _λ 6	IGLV6-57
V _λ 7	IGLV7-46,7-43
V _λ 8	IGLV8-61
V _λ 9	IGLV9-49
V _λ 10	IGLV10-54
V _κ 1.1	IGKV1-17,1D-17,1-9,1-5,1-6,1-13,1D-13,1D-43,1-12,1D-12,1-16,1D-16,1-8,1D-8,1-27,1-39,1D-39, 1-NL ,3-15,3D-15,3-NL1,3-NL2,3-NL3,3-NL4,3-NL5,3-11,3D-11,3-20,3D-20,3D-7,4-1
V _κ 1.2	IGKV1-33,1D-33
V _κ 2	IGKV2-40,2D-40,2-28,2D-28,2-30,2D-30,2-24,2D-26,2-29,2D-29,
V _κ 5	IGKV5-2
J _λ 1/2/3	IGLJ 1,J2,J3
J _λ 6	IGLJ6
J _λ 7	IGLJ7
J _κ 1.2	IGKJ1-4
J _κ 5	IGKJ5

需要说明的是，上表中所列出的 V 区或 J 区亚家族并不是穷尽的，而仅仅是目前有报道的所有亚家族。事实上，发明人惊奇地发现，采用上述具体的引物序列，能够全面覆盖人类免疫球蛋白轻链 CDR3 的全部亚家族，包括发现新的 CDR3 亚家族，从而能够全面的富集人类免疫球蛋白轻链 CDR3 序列，另外，发明人还发现通过采用上述引物序列，能够进行多重 PCR，可以有效地对样本中所包含的 CDR3 序列进行扩增，并且能够保证各 CDR3 序列扩增的均一性，从而能够真实地反映 CDR3 序列在宿主中的分布比例。根据本发明的实施例，多重引物 PCR 的退火温度可以为 60 摄氏度。发明人惊奇地发现，当退火温度为 60 摄氏度时，多重 PCR 扩增 CDR3 的效率得到显著提高。

由此，现有技术相比，本发明含有可扩增 IMGT 数据库免疫球蛋白 CDR3 序列所有 V 区和 J 区的特异性引物，可以最全面的富集人类免疫球蛋白轻链 CDR3 序列，扩增出来的产物能够区分免疫球蛋白轻链的各个亚家族，且同一模板不会被两条引物特异性结合扩增。本发明中的引物能更好的呈现机体免疫大分子的集合与分布，对找到与疾病相关的信息或者免疫系统变化的信息具有更可靠的作用。

3、CDR3 编码序列来源于 T 细胞受体 α 链

当 CDR3 编码序列来源于 T 细胞受体 α 链时，本发明的用于扩增 CDR3 编码序列的引物组合物的可以具有以下特征：

根据本发明的实施例，V 区引物和 J 区引物在 PCR 扩增过程中的作用，并不受特别限制。根据本发明的一个具体示例，V 区引物可以作为正义链引物，J 区引物可以作为反义链引物。发明人发现，通过如此设置 V 区引物和 J 区引物，能够进一步提高将引物组合物用于富集 T 细胞受体 α 链 CDR3 编码序列时的富集效率。根据本发明的实施例，上述引物组合物能够适用的 T 细胞受体类型，不受特别限制，本领域技术人员可以根据研究需要，选择适当的 T 细胞受体作为研究对象。根据本发明的一个实施例，所述 T 细胞受体为人类 T 细胞受体。由此，可以有效地将引物组合物用于对人类 T 细胞受体 α 链 CDR3 编码序列进行富集，从而可以有效地用于对人体免疫状况进行研究。

另外，根据本发明的实施例，可以通过对 V 区引物和 J 区引物的序列进行选择，实现一条引物可以识别多种 V 基因片段，从而可以进一步提高扩增的效率，减少引物的数目，降低成本。根据本发明的一个实施例，第一引物组的至少一种引物包含与多个 V 基因片段的保守区互补的序列。由此，可以在减少引物的数量的同时，提高对 T 细胞受体 α 链 CDR3 编码序列进行富集的效率，发明人还发现，这样操作能够提高各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。类似的，根据本发明的一个实施例，第二引物组的至少一种包含与多个 J 基因片段的保守区互补的序列。由此，可以在减少引物的数量的同时，提高对 T 细胞受体 α 链 CDR3 编码序列进行富集的效率，发明人还发现，这样操作能够提高各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。

具体地，根据本发明的实施例，针对人类 T 细胞受体 α 链 CDR3 的序列特征，本发明提供了一组 V 区引物 (SEQ ID NO: 46-84) 和一组 J 区引物 (SEQ ID NO: 85-134)，其序列和名称分别总结如下：

引物类型	引物序号	引物序列 (SEQ ID NO:)
V 区引物	TRAV1- 1-F	GTTATGGTTACCTCCTTCTACAGGA(46)
	TRAV1-2-F	AGGGTACAGTTACCTCCTTTTGAAG(47)
	TRAV2-F	CTTCATCGCTGCTCATCCTCCAG(48)
	TRAV3-F	GAAACCATCTGCCCTTGTGAGCG(49)
	TRAV4-F	CCTGTTTATCCCTGCCGACAGAA(50)
	TRAV5-F2	TGAAACAAGACCAAAGACTCACTG(51)
	TRAV6-F	GGTCACCTTTGATACCACCCTTAA(52)
	TRAV7-F2	GCTACATTACTGAAGAATGGAAGCA(53)
	TRAV8- 1/8-3-F	TAATCTGAGGAAACCCTCTGTGCA(54)
	TRAV8-2/8-4-F	CCTGACGAAACCCKCAGCCCAT(55)
	TRAV8-6-F	AGGAAACCCTCAGTCCATATAAGC(56)
	TRAV9- 1/9-2-F	AAGGTTTTGAAGCCAYGTACCGTA(57)
	TRAV10-F	CAAAGCTCTCTGCACATCACAGC(58)
	TRAV12-1/12-2-F	ATTTCCTGCTCATCAGAGACTCC(59)
	TRAV12-3-F	ATCTCCTTGTTTCATCAGAGACTCAC(60)
	TRAV13-1-F	GACAGCCAAACATTTCTCCCTGC(61)
	TRAV13-2-F	GCAAATTGCAGCTACTCAACCTGG(62)

	TRAV14-F	TCCGCCAACCTTGTCATCTCCG(63)
	TRAV16-F	AGACATCTTTCCACCTGAAGAAACC(64)
	TRAV17-F	CAGTTCCTTGTTGATCACGGCTTC(65)
	TRAV18-F3	GTCCATCAAGAGTGACAGTTCCTT(66)
	TRAV19-F	TCACCATCACAGCCTCACAAGTC(67)
	TRAV20-F	GCACATCACAGCCCCTAAACCTG(68)
	TRAV21-F	CTTTATACATTGCAGCTTCTCAGCC(69)
	TRAV22-F	TACATTTCTCTTCCCAGACCACA(70)
	TRAV23-F	TCATTGCATATCATGGATTCCCAGC(71)
	TRAV24-F	ACAGCTATTTGTACATCAAAGGATC(72)
	TRAV25-F2	CTGACATTTTCAGTTTGGAGAAGC(73)
	TRAV26-1/26-2-F	TCATCACWGAAGACAGAAAGTCCAG(74)
	TRAV27-F	AGGACAGTTCTCTCCACATCACTG(75)
	TRAV29-F	CAAAAGTGCCAAGCACCTCTCTCT(76)
	TRAV30-F	GCAAAGCTCCCTGTACCTTACGG(77)
	TRAV34-F	GAAAAGATAACTGCCAAGTTGGATG(78)
	TRAV35-F	GCTTCCTGAATATCTCAGCATCCAT(79)
	TRAV36-F	CAGCATCCTGAACATCACAGCCA(80)
	TRAV38-1/38-2-F	AGTCTCAAGATCTCAGACTCACAG(81)
	TRAV39-F	TCAGCACCTCCACATCACAGC(82)
	TRAV40-F	CCCCCATTGTGAAATATTCAGTCC(83)
	TRAV41-F3	GAAGCATGGAAGATTAATTGCCAC(84)
	TRAJ16-R	TAAGATCCACCTTTAACATGGTTCC(85)
	TRAJ7-R	TTGGTATGACCACCACTTGGTTC(86)
	TRAJ29-R	TTGCAATCACAGAAAGTCTTGTGCC(87)
	TRAJ56-R	CTGGTCTAACACTCAGAGTTATTC(88)
	TRAJ20-R	TTGCTCTTACAGTTACTGTGGTTCC(89)
	TRAJ46-R	TGGGCCTAACTGCTAAACGAGTC(90)
	TRAJ33-R	CTGGCTTTATAATTAGCTTGGTCC(91)
	TRAJ8-R	TTGGACTGACCAGAAGTCGGGTG(92)
	TRAJ9-R	GCTTTAACAAATAGTCTTGTTCCCTGC(93)
	TRAJ38-R2	GACTCACTCGGATTTACTGCCAG(94)
	TRAJ4-R2	CATACTCACATGGGTGTACAGCCA(95)

J 区引物	TRAJ12-R	AGGCCTGACCAGCAGTCTGGT(96)
	TRAJ54-R	TTGGGTTGATAGTCAGCCTGGTTC(97)
	TRAJ45-R	AGGGCTGGATGATTAGATGAGTC(98)
	TRAJ4 1-R	TGTGACCAACAGCGAGGTGCC(99)
	TRAJ22-R	CAGGTAAAACAGTCAATTGTGTCC(100)
	TRAJ36-R	AGGGAATAACGGTGAGTCTCGTTC(101)
	TRAJ6-R	ACGGATGAACAATAAGGCTGGTTC(102)
	TRAJ26-R	GGCAGCACGGAC AATCTGGTTC(103)
	TRAJ5-R	TTGGTTGCACTTGGAGTCTTGTTTC(104)
	TRAJ34-R	TTGGAAAGACTTGTAACTGGTCC(105)
	TRAJ1 1-R	CTGGAGAGACTAGAAGCATAGTCC(106)
	TRAJ14-R	CAGGTTTTACTGATAATCTTGTTCC(107)
	TRAJ3 1-R	GGGCTTCACCACCAGCTGAGTT(108)
	TRAJ39-R	GGGTTTGACCATTAACCTTGTTTC(109)
	TRAJ1 8-R	AGGCCAGACAGTCAACTGAGTTC(110)
	TRAJ30-R	TGGGGAGAATATGAAGTCGTGTC(111)
	TRAJ48-R	TGGGTATGATGGTGAGTCTTGTTTC(112)
	TRAJ10-R	TGAGTTCCACTTTTAGCTGAGTGC(113)
	TRAJ3-R2	TACTTACTTGGCCGATGCTGAG(114)
	TRAJ44-R	GAGCGTGACCTGAAGTCTTGTTTC(115)
	TRAJ53-R	TTGGATTCACGGTTAAGAGAGTTC(116)
	TRAJ17-R	TTGGTTTAACTAGCACCTGGTTC(117)
	TRAJ40-R	TGCTAAAACCTTCAGCCTGGTGC(118)
	TRAJ57-R	ATGGGTTTACTGTCAGTTTCGTTTC(119)
	TRAJ43-R	TTGGTTTTACTGTCAGTCTGGTCC(120)
	TRAJ37-R	CTGGTTTTACTTGTAAGTTGTCCC(121)
	TRAJ42-R	TTGGTTTAAAC AGAGAGTTTAGTGC C(122)
	TRAJ2 1-R	TTGGTTTTACATTGAGTTTGGTCCC(123)
	TRAJ28-R	TTGGTATGACCGAGAGTTTGGTCC(124)
	TRAJ27-R	TTGGCTTCAC AGTGAGCGTAGTC(125)
	TRAJ52-R	TTGGATGGACAGTCAAGATGGTCC(126)
	TRAJ49-R	TTGGAATGACCGTCAAACCTTGTTCC(127)
	TRAJ50-R	TTGGAATGACTGATAAGCTTGTTCC(128)

	TRAJ15-R	TGGAACCTCACTGATAGGTGGGTTC(129)
	TRAJ13-R	GGGATGACTTGGAGCTTTGTTCC(130)
	TRAJ24-R2	GAATGGGGCTTACCTGGGGTGAC(131)
	TRAJ23-R	GGGTTTCACAGATAGCTCCGTTC(132)
	TRAJ47-R	GGACTTGACTCTCAGAATGGTTCC(133)
	TRAJ32-R	TGGCTGGACAGCAAGCAGAGTG(134)

根据本发明的实施例，本发明的 V 区引物 (SEQ ID NO: 46-84) 和 J 区引物 (SEQ ID NO: 85-134) 能够区分各亚家族，且能够更直观地呈现基因重排后 V-J 配对的情况和 V 基因的使用偏好性，此外，引物合成容易，错配率低、特异性高，且所有引物的退火温度接近，从而能够降低扩增的偏向性。

发明人惊奇地发现，采用上述具体的引物序列，能够全面覆盖人类 T 细胞受体 CDR3 的全部亚家族，包括发现新的 CDR3 亚家族，从而能够全面的富集人类 T 细胞受体 CDR3 编码序列，另外，发明人还发现通过采用上述引物序列，能够同时在一个 PCR 反应体系中进行多重引物 PCR (有时也称为“多重 PCR”)，能够有效地对样本中所包含的 CDR3 编码序列进行扩增，并且能够保证各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。根据本发明的实施例，多重引物 PCR 的退火温度可以为 60 摄氏度。发明人惊奇地发现，当退火温度为 60 摄氏度时，多重引物 PCR 扩增 CDR3 的效率得到显著提高。

由此，与现有技术相比，本发明含有可扩增 IMGT 数据库 T 细胞受体 α 链 CDR3 序列所有 V 区基因和 J 区基因的特异性引物，可以最全面的富集人类 T 细胞受体 α 链 CDR3 的编码序列，扩增出来的产物能够区分 T 细胞受体 α 链的各个亚家族，且同一模板不会被两组引物特异性结合扩增。本发明中的引物能更好的呈现机体免疫大分子的集合与分布，对找到与疾病相关的信息或者免疫系统变化的信息具有更可靠的作用。

4、CDR3 编码序列来源于 T 细胞受体 β 链

当 CDR3 编码序列来源于 T 细胞受体 β 链时，本发明的用于扩增 CDR3 编码序列的引物组合物的可以具有以下特征：

根据本发明的实施例，V 区引物和 J 区引物在 PCR 扩增过程中的作用，并不受特别限制。根据本发明的一个具体示例，V 区引物可以作为正义链引物，J 区引物可以作为反义链引物。发明人发现，通过如此设置 V 区引物和 J 区引物，能够进一步提高将引物组合物用于富集 T 细胞受体 β 链 CDR3 编码序列时的富集效率。根据本发明的实施例，上述引物组合物能够适用的 T 细胞受体类型，不受特别限制，本领域技术人员可以根据研究需要，选择适当的 T 细胞受体作为研究对象。根据本发明的一个实施例，所述 T 细胞受体为人类 T 细胞受体。由此，可以有效地将引物组合物用于对人类 T 细胞受体 β 链 CDR3 编码序列进行富集，从而可以有效地用于对人体免疫状况进行研究。

另外，根据本发明的实施例，可以通过对 V 区引物和 J 区引物的序列进行选择，实现一条引物可以识别多种 V 基因片段，从而可以进一步提高扩增的效率，减少引物的数目，降低成本。根据本发明的一个实施例，第一引物组的至少一种引物包含与多个 V 基因片段的保守区互补的序列。由此，可以在减少引物的数量的同时，提高对 T 细胞受体 β 链 CDR3 编码序列进行富集的效率，发明人还发现，这样操作能够提高各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。类似的，根据本发明的一个实施例，第二引物组的至少一种包含与多个 J 基因片段的保守区互补的序列。由此，可以在减少引物的数量的同时，提高对 T 细胞受体 β 链 CDR3 编码序列进行富集的效率，发明人还发现，这样操作能够提高各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。

具体地，根据本发明的实施例，针对人类 T 细胞受体 β 链 CDR3 的序列特征，本发明提供了一组 (SEQ ID NO: 135-164) 和一组 J 区引物 (SEQ ID NO: 165-177)，其序列和名称分别总结如下：

引物类型	引物序号	引物序列 (SEQ ID NO:)
------	------	--------------------

V 区引物	TRBV2F	ATTCACTCTGAAGATCCGGTCCAC(135)
	TRBV9F	CCTGACTTGCACTCTGAACCTA(136)
	TRBV14F	GGAAGGACGTATCTACTGAAGG(137)
	TRBV15F	TTCTTGACATCCGGCTCACGAGG(138)
	TRBV19F	TCCTTTCTCTCACTGTGACATCGG(139)
	TRBV3-1-F4	AAACAGTTCCAAATCGMTTCTCAC(140)
	TRBV4-1/2/3-F4	CAAGTCGCTTCTCACTGTAATG(141)
	TRBV5-1-F4	GCCAAGTTCTTAACCTCGCTCT(142)
	TRBV5-4/5/6/8-F4	TCAGGTCGCCAGTTCCCTAAVYAT(143)
	TRBV6-1/2/3/5/8-F4	CAATGGCTACAAATGTCCTCYAGAT(144)
	TRBV6-4-F4	TGATGGTTATAGTGTCTCCAGAG(145)
	TRBV6-9-F4	CGATGGCTACAAATGTATCCAGAT(146)
	TRBV6-6-F4	GATGGCTACAAAGTCTCCAGAT(147)
	TRBV7-2/4/6/7/8-F4	GGATCCGCTCTCCACTCTGAMGAT(148)
	TRBV7-3-F4	GGATCCGCTCTCTACTCTGAAGAT(149)
	TRBV7-9-F4	GGATCTTTCTCCACCTTGAGAT(150)
	TRBV10-1-F4	CCTCACTCTGGAGTCTGTGCTGC(151)
	TRBV10-2/3-F4	CCTCACTCTGGAGTCTMGCTACC(152)
	TRBV11-1/2/3-F4	GCAGAGAGGCTCAAAAGGAGTAGACT(153)
	TRBV12-3/4-F4	ATCGATTCTAGCTAAGATGCCT(154)
	TRBV12-5-F4	ATCGATTCTCAGCAGAGATGCCT(155)
	TRBV13-F4	TCGATTCTCAGCTCAACAGTTCT(156)
	TRBV18-F4	TAGATGAGTCAAGGAATGCCAAAG(157)
	TRBV20-1-F4	AACCATGCAAGCCTGACCTT(158)
	TRBV24-1-F2	CTCCCTGTCCCTAGAGTCTGCCAT(159)
	TRBV25-1F	GCCCTCACATACCTCTCAGTACCTC(160)
	TRBV27/28-F4	GGAGATGTTCTCTGARGGGTACA(161)
	TRBV29-1-F4	AACTCTGACTGTGAGCAACATGAG(162)
	TRBV16-F2	CTGTAGCCTTGAGATCCAGGCTACGA(163)
	TRBV30-F5	CAGATCAGCTCTGAGGTGCCCA(164)
	TRBJ1.1R	CTTACCTACAACTGTGAGTCTGGTG(165)
	TRBJ1.2R	CTTACCTACAAAGGTTAACCTGGTC(166)

J 区引物	TRBJ1.3R	CTTACCTACAACAGTGAGCCAACTT(167)
	TRBJ1.4R	CATACCCAAGACAGAGAGCTGGGTTC(168)
	TRBJ1.5R	CTTACCTAGGATGGAGAGTCGAGTC(169)
	TRBJ1.6R	CATACCTGTCACAGTGAGCCTG(170)
	TRBJ2.1R	CCTTCTTACCTAGCACGGTGA(171)
	TRBJ2.2R	CTTACCCAGTACGGTCAGCCT(172)
	TRBJ2.3R	CCGCTTACCGAGCACTGTCAG(173)
	TRBJ2.4R	CCAGCTTACCCAGCACTGAGA(174)
	TRBJ2.5R	CGAGCACCAGGAGCCGCGT(175)
	TRBJ2.6R	CTCGCCCAGCACGGTCAGCCT(176)
	TRBJ2.7R	CTTACCTGTGACCGTGAGCCTG(177)

注：R=A/G，Y=C/T，M=A/C。

根据本发明的实施例，本发明的 V 区引物和 J 区引物能够区分各亚家族，且能够更直观地呈现基因重排后 V-J 配对的情况和 V 基因的使用偏好性，此外，引物合成容易，错配率低、特异性高，且所有引物的退火温度接近，从而能够降低扩增的偏向性。

5 发明人惊奇地发现，采用上述具体的引物序列，能够全面覆盖人类 T 细胞受体 CDR3 的全部亚家族，包括发现新的 CDR3 亚家族，从而能够全面的富集人类 T 细胞受体 CDR3 编码序列，另外，发明人还发现通过采用上述引物序列，能够同时在一个 PCR 反应体系中进行多重引物 PCR(有时也称为"多重 PCR")，能够有效地对样本中所包含的 CDR3 编码序列进行扩增，并且能够保证各 CDR3 编码序列扩增的均一性，从而能够真实地反映 CDR3 编码序列在宿主中的分布比例。根据本发明的实施例，多重引物 PCR 的退火
10 温度可以为 60 摄氏度。发明人惊奇地发现，当退火温度为 60 摄氏度时，多重引物 PCR 扩增 CDR3 的效率得到显著提高。

由此，与现有技术相比，本发明含有可扩增 IMGT 数据库 T 细胞受体 β 链 CDR3 序列所有 V 区基因和 J 区基因的特异性引物，可以最全面的富集人类 T 细胞受体 β 链 CDR3 的编码序列，扩增出来的产物能够区分 T 细胞受体 β 链的各个亚家族，且同一模板不会被两组引物特异性结合扩增。本发明中的引物
15 能更好的呈现机体免疫大分子的集合与分布，对找到与疾病相关的信息或者免疫系统变化的信息具有更可靠的作用。

富集 CDR3 编码序列的方法

在本发明的又一方面，本发明还提出了一种富集 CDR3 编码序列的方法。根据本发明的实施例，该方法包括以下步骤：首先，提供核酸样本，该核酸样本中包含编码 CDR3 的核酸序列；接下来，利用前面所述的引物组合物，以所提供的核酸样本作为模板，进行 PCR 扩增，如前所述，基于引物组合物的特点，可以通过 PCR 扩增获得扩增产物，该扩增产物中富集了 CDR3 编码序列。其中，所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。利用该方法，能够有效地对免疫球蛋白重链 CDR3 编码序列进行扩增，从而有效地实现免疫球蛋白 CDR3 编码
20 序列的富集。

25

根据本发明的实施例，核酸样本的来源不受特别限制。本领域技术人员可以根据研究需要，选择可以获得核酸样本的来源。例如为了研究某一组织的特有免疫状态，可以从该组织或其附近提取免疫细胞作为核酸样本的来源。根据本发明一个实施例，可以采用从人外周血能够分离含有上述核酸样本的单个核细胞，并且通过分离核酸来获得上述核酸样本。由此，提供核酸样本的步骤可以进一步包括：首先，
5 从人外周血分离外周血单个核细胞；接下来，从外周血单个核细胞提取核酸样本。由此，可以有效地获得含有编码 CDR3 的核酸序列的核酸样本，从而，可以进一步提高富集 CDR3 编码序列的效率。本领域技术人员可以通过任何常规的手段从外周血中提取外周血单个核细胞。根据本发明的一个实施例，可以通过密度梯度离心分离所述外周血单个核细胞。并且可以采用常规的手段从所分离的外周血单个核细胞中提取基因组 DNA 和总 RNA 作为用于扩增的核酸样本。由此，可以方便快捷且低成本地获得核酸样本。
10 本领域技术人员能够理解的是，当采用总 RNA 时作为核酸样本进行扩增时，根据试验需要，可以首先通过逆转录将总 RNA 转换为 cDNA。

根据本发明的实施例，上述 PCR 的类型并不受特别限制，即可以依次进行多次 PCR 反应，也可以在一个 PCR 体系中完成多轮 PCR 扩增。根据本发明的一个实施例，PCR 扩增为多重引物 PCR 扩增。由此，可以同时在一个反应体系中完成对目标序列，即多种样本的 CDR3 编码序列的扩增，并且能够保证各样本的 CDR3 编码序列扩增的均一性，从而能够真实地反映各样本的 CDR3 编码序列的真实相对比例。在进行 PCR 扩增之后，可以进一步包括通过选 自琼脂糖凝胶电泳、磁珠纯化和纯化柱纯化的至少一种分离纯化所得到的扩增产物。由此，可以提高扩增产物的纯度，从而提高富集免疫球蛋白重链 CDR3 编码序列的效率。

此外，需要说明的是，当 CDR3 编码序列来源于免疫球蛋白重链时，本发明的富集 CDR3 编码序列的方法中 PCR 扩增所采用的引物组合物，其第一引物组中的 V 区引物可以具有 SEQ ID NO: 1-20 所示的核苷酸序列，其第二引物组中的 J 区引物具有如 SEQ ID NO: 21 所示的核苷酸序列，获得的扩增产物的长度为 140-280bp；当 CDR3 编码序列来源于免疫球蛋白轻链时，本发明的富集 CDR3 编码序列的方法中 PCR 扩增所采用的引物组合物，其第一引物组中的 V 区引物可以具有 SEQ ID NO: 22-40 所示的核苷酸序列，其第二引物组中的 J 区引物可以具有 SEQ ID NO: 41-45 所示的核苷酸序列，获得的扩增产物的长度为 110-250bp；当 CDR3 编码序列来源于 T 细胞受体 α 链时，本发明的富集 CDR3 编码序列的方法中 PCR 扩增所采用的引物组合物，其第一引物组中的 V 区引物可以具有 SEQ ID NO: 46-84 所示的核苷酸序列，其第二引物组中的 J 区引物可以具有如 SEQ ID NO: 85-134 所示的核苷酸序列，获得的扩增产物的长度为 100-200bp；当 CDR3 编码序列来源于 T 细胞受体 β 链时，本发明的富集 CDR3 编码序列的方法中 PCR 扩增所采用的引物组合物，其第一引物组中的 V 区引物可以具有 SEQ ID NO: 135-164 所示的核苷酸序列，其第二引物组中的 J 区引物可以具有如 SEQ ID NO: 165-177 所示的核苷酸序列，获得的扩增产物的长度为 100-200bp。由此，可以进一步提高扩增产物中 CDR3 编码序列的纯度，从而能够提高富集 CDR3 编码序列的效率。

构建 CDR3 编码序列的测序文库及确定其序列信息的方法

在本发明的又一方面，本发明提出了一种构建 CDR3 编码序列的测序文库的方法。根据本发明的实施例，该方法可以包括以下步骤：首先，根据前面所述的方法，获得富集所述 CDR3 编码序列的扩增产物。接下来，针对所得到的扩增产物，构建 DNA 测序文库，该 DNA 测序文库可以作为 CDR3 编码序列的测序文库。其中，所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。由此，可以在对免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β 链的 CDR3 编码序列进行富集的基础上，构建可以用于测序的测序文库。
40

根据本发明的实施例，针对扩增产物构建 DNA 测序文库的方法并不受特别限制。根据本发明的一个实施例，针对扩增产物，构建 DNA 测序文库可以进一步包括：

首先，对扩增产物进行末端修复，以便获得经过末端修复的扩增产物。根据本发明的一个实施例，所述末端修复是利用 Klenow 片段、T4 DNA 聚合酶和 T4 多核苷酸激酶进行的，所述 Klenow 片段具有 5' $\rightarrow$ 3' 聚合酶活性和 3' $\rightarrow$ 5' 外切酶活性，但缺少 5' $\rightarrow$ 3' 外切酶活性。由此，可以进一步提高末端修复的效率，从而可以进一步提高构建测序文库的效率。
45

接下来,对经过末端修复的扩增产物进行 3'端添加碱基 A,以便获得 3'末端添加碱基 A 的扩增产物。根据本发明的一个实施例,利用 Klenow (3'-5' exo-) 对所述经过末端修复的扩增产物进行 3'端添加碱基 A。由此,可以进一步提高在扩增产物的 3'末端添加碱基 A 的效率,从而可以进一步提高构建测序文库的效率。

接着,将所得到的 3'末端添加碱基 A 的扩增产物与接头相连,以便获得连接产物;对所得到的连接产物进行 PCR 扩增,以便获得第二扩增产物。根据本发明的一个实施例,将所述具有粘性末端 A 的扩增产物与接头相连是利用 T4 DNA 连接酶进行的。由此,可以进一步提高扩增产物与接头连接的效率,进而可以进一步提高构建测序文库的效率。

最后,将所得到的第二扩增产物进行纯化回收,以便获得回收产物,所得到的回收产物构成所述测序文库。根据本发明的实施例,对第二扩增产物进行纯化回收的方法并不受特别限制,根据本发明的具体实例,可以通过选自琼脂糖凝胶电泳、磁珠纯化和纯化柱纯化的至少一种分离纯化所述第二扩增产物。由此,可以进一步提高构建测序文库的效率。

由此,可以有效地构建测序文库,从而便于后续的测序及进一步分析。

根据本发明的再一方面,本发明提出了一种确定 CDR3 编码序列的序列信息的方法。根据本发明的实施例,该方法可以包括以下步骤:

首先,根据前面所述的方法,构建 CDR3 编码序列的测序文库。其中,该 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。

接下来,对该 CDR3 编码序列的测序文库进行测序,以便确定该 CDR3 编码序列的序列信息。根据本发明的实施例,利用选自 HiSeq2000、SOLiD、454、和单分子测序装置的至少一种进行所述测序。由此,能够高通量高精度地对所得到的免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β 链的 CDR3 编码序列,从而进一步提高了确定 CDR3 编码序列的序列信息的方法的效率。

确定个体免疫状态的方法和系统、试剂盒

免疫组库作为多样性的免疫细胞在一个个体内某一时刻的总和,它反应了个体遗传因素、抗原接触史和个体时刻的免疫调控。免疫组库能用于疾病相关研究,对疾病机理进行探讨,可作为寻找生物标记物的一个有效手段,免疫组库的研究结果可以促进对更多疾病的早期诊断,治疗甚至预防。目前已有相关研究表明 IgH、TCR 与免疫系统疾病的发生有一定关系,某一种克隆的增多或减少直接影响疾病的发生和进展,例如有研究表明特定 IgH 克隆的重排是非霍奇金淋巴瘤的生物标志物;特异性的 B 细胞单克隆增多是慢性淋巴细胞性白血病的首个生物标志物。由此,本发明再一方面提出了一种确定个体免疫状态的方法。

根据本发明的实施例,本发明的确定个体免疫状态的方法可以包括以下步骤:首先,根据前面所述的方法,对所述个体的免疫球蛋白重链 CDR3 编码序列进行测序,以便获得由多个测序数据构成的测序结果;然后,基于所得到的测序结果,确定该个体的免疫状态。其中,所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。通过该方法,能够有效地获得个体的免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β 链的 CDR3 编码序列的序列信息,从而可以有效地确定个体免疫状态。在本文中所使用的术语"免疫状态"应作广义理解,其是指任何可以通过免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β 链的 CDR3 编码序列的序列信息反映出的免疫信息。

根据本发明的一个实施例,基于所得到的测序结果,确定个体的免疫状态可以进一步包括:将所得到的测序结果与对照序列进行比对,以便确定所述个体中所包含的 CDR3 的亚家族类型,以及各亚家族类型的相对比例。由此,可以有效地判断个体免疫系统的组成和分布情况,从而有效地确定个体的免疫状态。另外,根据本发明的实施例,可以对个体进行多次监控,判断 CDR3 的亚家族类型,以及各亚家族类型的相对比例随时间的变化。为此,根据本发明的一个实施例,在多个不同的时间点,从相同的个体提取样品,并分别根据前面所述的方法对样品,获得多个测序结果;以及将所述多个测序结果进行比对,以便确定所述个体中 CDR3 亚家族类型以及相对比例的变化。由此,可以基于不同时间点的样品的测序结果的比对,有效地确定个体中 CDR3 亚家族类型以及相对比例的变化,从而更加有效地判断个体

的免疫状态。由此，可以在不同时间对同一个体或多个体进行采样，分析例如疾病前后或某种特定事件、时期前后个体免疫组库的变化，了解个体对特定事件、在特定时期的免疫系统变化。例如，能够从单一克隆水平知道当前样本的细致变化，从而寻找与疾病发生发展史相关的信息。

本发明的又一方面提出了一种确定个体免疫状态的系统。根据本发明的实施例，参考图3，该确定个体免疫状态的系统1000包括：CDR3编码序列富集装置100、文库构建装置200、测序装置300以及分析装置400。其中，CDR3编码序列富集装置100中设置有前面所述的引物组合物，以便对个体的核酸样本富集CDR3编码序列，其中所述CDR3编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T细胞受体 α 链和T细胞受体 β 链的至少一种。文库构建装置200与CDR3编码序列富集装置100相连，以便针对经过富集的CDR3编码序列构建CDR3编码序列的测序文库。根据本发明的实施例，关于针对扩增产物，构建测序文库的方法和流程，本领域技术人员可以根据不同的测序技术进行适当选择，关于流程的细节，可以参见测序仪器的厂商例如Illumina公司所提供的规程，例如参见Illumina公司Multiplexing Sample Preparation Guide (Part# 100536 1; Feb 2010)或Paired-End SamplePrep Guide (Part# 1005063; Feb 2010)，通过参照将其并入本文。在本文中所使用的术语“相连”应作广义理解，既可以是直接相连，也可以是间接相连，只要能够实现上述功能上的衔接即可。测序装置300与文库构建装置200相连，用于对CDR3编码序列测序文库进行测序，以便获得由多个测序数据构成的测序结果。分析装置400与测序装置300相连，用于基于测序结果，确定所述个体的免疫状态。由此，利用该系统，能够有效地实施前述确定个体免疫状态的方法，从而有效地确定个体的免疫状态。根据本发明的一个实施例，分析装置400可以进一步包括比对单元，比对单元中存储有对照序列，用于将测序结果与对照序列进行比对，以便确定个体中所包含的CDR3的亚家族类型，以及各亚家族类型的相对比例。根据本发明的实施例，可以采用在分析装置300中预存有相关的序列信息，也可以采用分析装置300与远程数据库（图中未显示）相连，进行联网操作。由此，可以通过将测序结果与对照序列例如已知的免疫基因组数据库IMGT进行比对，确定CDR3的亚家族类型分布以及各亚家族类型的分布，从而进一步提高确定个体免疫状态的效率。

根据本发明的又一方面，本发明还提出了一种试剂盒。根据本发明的实施例，该试剂盒中设置有前面所述的引物组合物。利用该试剂盒可以有效地用于检测免疫球蛋白重链、免疫球蛋白轻链、T细胞受体 α 链和T细胞受体 β 链的CDR3的编码序列，从而可以有效地确定个体的免疫状态。

根据本发明的实施例，该试剂盒可以包含根据前面所述的引物组合物。由此，该试剂盒可以用于检测免疫球蛋白重链、免疫球蛋白轻链、T细胞受体 α 链和T细胞受体 β 链的V-J重排。根据本发明的另一个实施例，本发明的试剂盒可以用于检测免疫球蛋白重链、免疫球蛋白轻链、T细胞受体 α 链和T细胞受体 β 链的CDR3的编码序列。根据本发明的实施例，在该试剂盒中，第一引物组和第二引物组可以设置在不同的容器中，也可以设置在相同的容器中以组合物的方式存在。需要说明的是，前面描述的关于本发明的用于扩增CDR3编码序列的引物组合物的优点，同样适用于该试剂盒，在此不再赘述。

下面参考具体实施例，对本发明进行说明，需要说明的是，这些实施例仅仅是说明性的，而不能理解为对本发明的限制。

若未特别指明，实施例中所采用的技术手段为本领域技术人员所熟知的常规手段，可以参照《分子克隆实验指南》第三版或者相关产品进行，所采用的试剂和产品也均为可商业获得的。未详细描述的各种过程和方法是本领域中公开的常规方法，所用试剂的来源、商品名以及有必要列出其组成成分者，均在首次出现时标明，其后所用相同试剂如无特殊说明，均以首次标明的内容相同。

一般方法

参考图2，本发明的富集CDR3编码序列、构建测序文库和测序的方法，一般可以包括以下步骤：

S100：从外周血分离单个核细胞（PBMC）

抽取正常人外周血，具体地，用含抗凝剂的无菌采血管进行采血，然后利用Ficoll-Paque PLUS或Percoll淋巴细胞分离液进行密度梯度离心分离PBMC。

S200：提取核酸样本

即提取基因组DNA和总RNA，具体地，利用蛋白酶K消化或者酚氯仿抽提的方法提取基因组DNA；利用Trizol法提取总RNA。

S300: 利用引物组合物扩增

以上述核酸样本为模板,采用本发明的引物组合物进行多重 PCR 扩增,以便获得富集 CDR3 编码序列的扩增产物。

S400: 回收&纯化扩增产物

5 然后,利用琼脂糖凝胶电泳以及 MiniElute PCR 纯化试剂盒 (Qiagen) 回收纯化扩增产物。其中,当以总 RNA 为模板时,需要先将 RNA 逆转录为 cDNA。

S500: 构建测序文库

5.1、末端修复及 3'末端添加碱基 A

10 将回收纯化的扩增产物通过 T4 DNA 聚合酶、Klenow 片段和 T4 多核苷酸激酶等酶的作用以 dNTP 为作用底物进行末端修复,获得经过末端修复的扩增产物。然后利用 Klenow 片段(3'-5'exo-) 聚合酶及 dATP 在经过末端修复的扩增产物的 3'末端添加碱基 A,以便获得 3'末端添加碱基 A 的扩增产物。然后,利用 MiniElute PCR 纯化试剂盒 (Qiagen) 回收纯化 3'末端添加碱基 A 的扩增产物。

5.2、接头连接

15 将所得到的 3'末端添加碱基 A 的扩增产物在 T4 DNA 连接酶的作用下与接头进行连接,以便获得连接产物。然后,利用琼脂糖凝胶电泳以及 MiniElute PCR 纯化试剂盒 (Qiagen) 回收纯化连接产物。

5.3、PCR 扩增

以连接产物为模板,加入通用 PCR 引物和测序引物,用 Phusion 酶进行 PCR 扩增,以便获得第二扩增产物。然后,利用琼脂糖凝胶电泳回收纯化第二扩增产物,以便获得回收产物,该回收产物构成 DNA 测序文库。

20 S600: 测序

将上述获得的 DNA 测序文库经安捷伦 2100 检测和 Q-PCR 定量后,利用 Hiseq2000 测序平台进行测序,以便获得由多个测序数据构成的测序结果,然后,以 IMGT 上免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链或 T 细胞受体 β 链的 CDR3 的序列作为参考序列,对测序结果进行分析。

25 实施例 1

参照上述一般方法,按照以下步骤,对免疫球蛋白重链的 CDR3 编码序列进行富集、构建测序文库、测序及分析:

1.分离人外周血单个核细胞

1) 取健康人新鲜外周血 10mL, 注入含有抗凝剂的无菌试管中摇匀,并加入等量的无菌 PBS 摇匀。
30 2) 取适量体积 Percoll 分层液 (外周血和分层液的体积比为 2:1~3:1) 于试管中,将稀释的血液慢慢加到分层液液面上,形成清晰界面,置于水平离心机中离心,2000 r/m, 20 分钟。

3) 离心后,可见从离心管底部到液面分为四层,分别为红细胞和粒细胞层、分层液层、单个核细胞层、血浆层。用吸管插入液面下,吸取单个核细胞层,置于另一个无菌试管中。

4) 加入 4 倍体积的无菌 PBS 充分混匀,离心,1000 r/m, 10 分钟;重复洗涤一次。

35 5) 用 PBS 稀释,进行细胞计数,提取的单个核细胞保存于 4℃ 备用。整个分离过程常温下进行。

2. DNA 提取

DNA 提取试剂盒提取 PBMC DNA。(Qiagen)

1) 将 2(μ L) QIAGEN 蛋白酶加入 200 μ L 样品中,混匀。

2) 将 200 μ L 缓冲液 A1 加入样品中,充分混匀,56℃ 孵育 10 分钟。

40 3) 加入 200 μ L 无水乙醇,充分混匀,将混合物转移到吸附柱上,离心,8000r/m, 1 分钟,弃废液。

4) 加入 50(μ L) 缓冲液 AW1,离心,8000r/m, 1 分钟,弃废液。

5) 加入 500mL 缓冲液 AW2,离心,8000 r/m, 1 分钟,弃废液。

6) 全速 (14000 r/m) 空甩 1 分钟。

7) 加入 200 μ L 缓冲液 AE,室温静置 1 分钟,离心,8000 r/m, 1 分钟。DNA 保存于-20℃ 备用。

45 3. 多重 PCR 扩增 CDR3

将上一步得到的 DNA 作为模板 DNA 进行多重 PCR 扩增。按下表在 200 μ L 的 PCR 管中配制 4 个 PCR

反应体系，分别为 LX500，LX800，ZXJ500，ZXJ800，4 个平行实验。

组分	体积 (μL)
QIAGEN Mutiplex PCR master mix	25
V _H 混合引物 (SEQ ID NO: 1-20)	1
J _H 引物 (SEQ ID NO: 21)	1
Q 溶液	5
无 RNA 酶水	13
模板 DNA	5
总体积	50

PCR 反应条件为：

95 °C	15min	} 30 个循环
94 °C	30s	
60 °C	90s	
72 °C	30s	
72 °C	5min	
15 °C	∞	

5 PCR 产物用 2% 琼脂糖凝胶电泳，电压为 4V/cm，电泳时间为 2h。然后切胶，取 140-280 bp 片段，用 QIAquick Gel 纯化试剂盒 (Qiagen) 进行纯化，最后将样品溶于 30μL 洗脱缓冲液。结果如图 4 所示，有多个扩增条带，在正式建库之前做过摸索实验，Sanger 测序结果表明 140-280bp 条带才是目的产物即重链 CDR3 编码序列，建库时回收 140-280bp 条带。

4. PCR 产物末端修复

将上一步得到的 DNA 分别按下表在 1.5mL 的离心管中配制 4 个末端修复反应体系

组分	体积 (μL)
DNA	30
T4 DNA 连接酶缓冲液	10
dNTP 混合物	4
T4 DNA 酶	5
Klenow 酶	1
T4 PNK	5
超纯水	45
总体积	100

10 然后将管子放到调至 20 °C 的 Thermomixer(Eppendorf) 上反应 30 min，反应结束后，用 QIAquick PCR 纯化试剂盒 (Qiagen) 进行纯化，最后将样品溶于 32μL 洗脱缓冲液。

5. DNA 片段 3' 末端添加碱基 A

将上一步所得的 DNA 分别按下表在 200μL 的 PCR 管中配制 4 个 3' 末端添加碱基 A 反应体系

组分	体积 (μL)
DNA	32

Klenow 緩冲液	5
dATP	10
Klenow (3'-5' exo-)	3
总体积	50

然后将管子放到调至 37°C 的 Thermomixer(Eppendorf) 上反应 30 min。反应结束后，用 MiniElute PCR 纯化试剂盒 (Qiagen) 进行纯化，最后将样品溶于 10 μ L 洗脱緩冲液。

6. 连接接头

5 将上一步得到的 DNA 分别按下表在 1.5mL 的离心管中配制 4 个末端修复反应体系

组分	体积 (μ L)
DNA	10
DNA 连接酶缓冲液	25
接头	10
DNA 连接酶	5
总体积	50

其中接头序列为：

接头 1: 5',GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG (SEQ ID NO: 178);

接头 2: 5'TACTACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 179)。

10 然后将管子放到调至 20°C 的 Thermomixer(Eppendorf) 上反应 15 min。反应结束后，用 2%琼脂糖凝胶电泳分离，电压为 120V，电泳时间为 120 分钟。电泳结束后切下目的条带，用 QIAquick PCR 纯化试剂盒 (Qiagen) 进行纯化，最后将样品溶于 30 μ L 洗脱緩冲液。

7. PCR 扩增目的产物，并加入测序引物

将上一步得到的 DNA 分别按以下体系配制 4 个 PCR 反应体系

组分	体积 (μ L)
接头连接的 DNA	1
P1 公用引物	1
Index N	1
超纯水	22
Phusion DNA 酶	25
总体积	50

P1 公用引物：

15 AATGATACGGCGACCACCGAGATCTACTACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 180);

Index N:

CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT (SEQ ID NO: 181)，其中碱基 N 为 A、T、C、G 四个碱基的任意组合作为区别标识。

20 PCR 反应条件为：

98°C	30s	} 12 个循环
98°C	10s	
65°C	30s	
72°C	30s	
72°C	5min	
4°C	∞	

反应结束后，用 QIAquick PCR 纯化试剂盒 (Qiagen) 进行纯化，最后将样品溶于 50 μ L 洗脱緩冲液。

8. 文库检测

Agilent 2100 Bioanalyzer analysis system 检测文库插入片段大小及含量 ;Q-PCR 精确定量文库的浓度。

9. 测序与数据分析

文库检测合格后将按照双末端 151 个碱基的读长在 Hiseq2000 测序仪上进行序列分析。

5 测序所得的原始数据的基本分析过程主要包括以下步骤：首先进行数据处理，通过接头或 PCR 引物上的序列标签区分不同样本的文库数据，对测序所得的原始数据进行去污染、去接头和去低质量过滤；然后将 reads 与 IMGT 数据库的参考序列进行 Y、D、J 基因比对，分析。

过滤之后的测序原始数据进行比对分析的结果，见下表 1，我们平均得到的总序列有 4.146 兆条，特异性的克隆平均有 2.27 兆条。

10

表 1

样品及平均值	总序列数	含 V、D、J 基因的序列数	含 ORF（开放阅读框）的序列	特异性克隆数	特异性氨基酸数
LX500	43,306,679	25,728,202	18,013,679	2,394,334	2,155,319
LX800	40,165,373	22,188,037	18,034,254	2,420,085	2,172,834
ZXJ500	43,416,993	24,344,437	17,090,162	2,180,659	1,944,255
ZXJ800	38,970,166	21,483,081	15,577,095	2,091,716	1,870,261
平均值	41,464,803	23,435,939	17,178,198	2,271,698	2,035,667

文库 LX500 测序后获得的各种 V-J 基因配对的序列的数量示于表 2 中。

表 2

<div><div>J 区</div><div>V-J 序列数</div><div>V 区</div></div>	IGHJ1	IGHJ2	IGHJ3	IGHJ4	IGHJ5	IGHJ6
IGHV1-18	437	538	4317	15399	6045	11615
IGHV1-24	143	219	1239	3301	1279	2531
IGHV1-2	1112	1465	11478	41431	15374	29736
IGHV1-3	446	661	3609	15468	5299	9318
IGHV1-45	4	53	81	321	144	192
IGHV1-46	222	291	1996	7114	2816	6266
IGHV1-58	209	261	1483	6091	2086	4711
IGHV1-69	887	1426	9335	32912	12728	28278
IGHV1-8	780	855	5337	22054	7827	16730
IGHV1-f	78	76	642	2258	807	1488
IGHV2-26	271	589	2210	8205	2946	4743
IGHV2-5	1225	1359	8911	32807	13119	7190
IGHV2-70	331	857	5384	17717	5831	9345
IGHV3-11	583	966	5070	19519	5233	11607
IGHV3-13	128	769	2478	6544	1873	6906
IGHV3-15	206	250	1238	7024	1812	3463

IGHV3-20	86	280	1568	3902	1099	2503
IGHV3-2 1	629	1124	5739	22727	5974	13876
IGHV3-23	4048	4958	22856	105585	27584	43089
IGHV3-30-3	279	379	1820	8628	198 1	3796
IGHV3-30	3053	5355	27359	1182 10	29122	68466
IGHV3-33	9 13	1545	7584	33983	8 111	18606
IGHV3-43	15	17	139	502	142	262
IGHV3-48	1176	20 13	10459	4 1097	10972	2478 1
IGHV3-49	77	116	733	3305	868	2 118
IGHV3-53	938	140 1	8393	30037	8680	19023
IGHV3-64	266	447	22 14	10022	2480	5449
IGHV3-66	135 1	2707	12582	49294	13083	2742 1
IGHV3-72	4 1	37	295	1226	308	577
IGHV3-73	48	80	300	1230	28 1	770
IGHV3-74	440	1104	4068	144 17	4482	8542
IGHV3-7	502	88 1	4756	18 178	4742	10843
IGHV3-9	123	116	683	3042	765	1253
IGHV3-d	62	63	352	143 1	440	794
IGHV3-NL1	122	128	765	3905	778	1702
IGHV4-28	2 16	734	3 178	7763	4504	5715
IGHV4-30-2	1089	2557	8825	28693	13927	12948
IGHV4-30-4	30	89	4 18	1280	111	1363
IGHV4-3 1	360	1146	483 1	11602	66 14	9944
IGHV4-34	24 17	4052	17750	62345	26275	39808
IGHV4-39	1933	426 1	159 10	53278	24095	28 184
IGHV4-4	33 1	939	4037	9958	5469	8225
IGHV4-59	170 1	399 1	16976	48053	228 18	29273
IGHV4-6 1	797	23 14	1015 1	25 134	12674	17368
IGHV4-b	183	394	1752	4962	2334	2684
IGHV5-5 1	1120	2040	1114 1	33538	12097	17690
IGHV5-a	335	565	309 1	9606	3522	4 134
IGHV6-1	564	1288	4896	20 188	6636	12746
IGHV7-4-1	494	893	5 175	22487	8672	139 10

图 5 是基于表 2 中数据，作出的 4 个平行文库的 V-J 基因配对分布图，图 5 显示了所有的 V 基因亚家族与 J 基因亚家族重排后相互配对和各种 VJ 配对 CDR3 的丰度，图右侧坐标是 J 基因全部亚家族的分类，图中横坐标是 V 基因的每一种亚家族分类，两坐标相交点是 V-J 基因配对产生的一种 CDR3 序列，左侧坐标则是每种 V-J 配对的丰度。从表 2 及图 5 可以得出结论，使用本发明提供的引物能够全面地富集人类免疫球蛋白重链 CDR3 的编码序列，扩增出来的产物能够区分免疫球蛋白重链的各个亚家族，并且通过上述本发明的方法能够确定各种重链 CDR3 编码序列的相对比例，从而可以进一步基于这些编码序列的相对比例或该比例的变化，确定个体的免疫状态。

实施例 2

参照上述一般方法，按照以下步骤，对免疫球蛋白轻链的 CDR3 编码序列进行富集、构建测序文库、测序及分析：

1. 分离人外周血单个核细胞

1) 取健康人新鲜外周血 10mL, 注入含有抗凝剂的无菌试管中摇匀, 并加入等量的无菌 PBS 摇匀, 样品名 ZXJ。

2) 取适量体积 Percoll 分层液 (外周血和分层液的体积比为 2:1~3:1) 于试管中, 将稀释的血液慢慢加到分层液液面上, 形成清晰界面, 置于水平离心机中离心, 2000 r/m, 20 分钟。

3) 离心后, 可见从离心管底部到液面分为四层, 分别为红细胞和粒细胞层、分层液层、单个核细胞层、血浆层。用吸管插入液面下, 吸取单个核细胞层, 置于另一个无菌试管中。

4) 加入 4 倍体积的无菌 PBS 充分混匀, 离心, 1000 r/m, 10 分钟; 重复洗涤一次。

5) 用 PBS 稀释, 进行细胞计数, 提取的单个核细胞保存于 4℃ 备用。整个分离过程常温下进行。

2. DNA 提取

DNA 提取试剂盒提取 PBMC DNA。(Qiagen)

1) 将 20 μ L QIAGEN 蛋白酶加入 200 μ L 样品中, 混匀。

2) 将 20 μ L 缓冲液 AL 加入样品中, 充分混匀, 56℃ 孵育 10 分钟。

3) 加入 200 μ L 无水乙醇, 充分混匀, 将混合物转移到吸附柱上, 离心, 8000r/m, 1 分钟, 弃废液。

4) 加入 500 μ L 缓冲液 AW1, 离心, 8000r/m, 1 分钟, 弃废液。

5) 加入 500 μ L 缓冲液 AW2, 离心, 8000 r/m, 1 分钟, 弃废液。

6) 全速 (14000 r/m) 空甩 1 分钟。

7) 加入 20 μ L 缓冲液 AE, 室温静置 1 分钟, 离心, 8000 r/m, 1 分钟。DNA 保存于 -20℃ 备用。

3. 多重 PCR 扩增 CDR3

将上一步得到的 DNA 作为模板 DNA 进行多重 PCR 扩增。按下表在 200 μ L 的 PCR 管中配制 PCR 反应体系, 为 ZXJ。

组分	体积 (μ L)
QIAGEN Mutiplex PCR master mix (一种市售的含聚合酶、缓冲体系和 dNTP 的混合试剂)	25
V $_{\lambda}$ 、V $_{\kappa}$ 混合引物 (SEQ ID NO: 22-40, 10 μ M)	1
J $_{\lambda}$ 、J $_{\kappa}$ 混合引物 (SEQ ID NO: 41-45, 10 μ M)	1
Q 溶液	5
无 RNA 酶水	13
模板 DNA	5
总体积	50

PCR 反应条件为：

95℃	15min	
94℃	30s	} 30 个循环
60℃	90s	
72℃	30s	
72℃	5min	
15℃	∞	

其中, 图 6 是引物扩增免疫球蛋白轻链中 V-J 重排构成的 CDR3 序列的示意图。

PCR 产物用 2% 琼脂糖凝胶电泳, 电压为 4V/cm, 电泳时间为 2h。然后切胶, 取 110-250 bp 片段, 用 QIAquick Gel 纯化试剂盒 (Qiagen) 进行纯化, 最后将样品溶于 30 μ L 洗脱缓冲液。结果如图 7 所示, 有多个扩增条带, 在正式建库之前做过摸索实验, Sanger 测序结果表明 110-250bp 条带才是目的产物即轻链 CDR3 序列, 建库时回收 110-250bp 条带。

4. PCR 产物末端修复

将上一步得到的 DNA 分别按下表在 1.5mL 的离心管中配制 4 个末端修复反应体系

组分	体积 (μL)
DNA	30
T4 DNA 连接酶缓冲液	10
dNTP 混合物	4
T4 DNA 酶	5
Klenow 酶	1
T4 PNK	5
超纯水	45
总体积	100

然后将管子放到调至 20℃ 的 Thermomixer(Eppendorf) 上反应 30 min, 反应结束后, 用 QIAquick PCR 纯化试剂盒 (Qiagen) 进行纯化, 最后将样品溶于 32μL 洗脱缓冲液。

5. DNA 片段 3' 末端添加碱基 A

5 将上一步所得的 DNA 分别按下表在 200μL 的 PCR 管中配制 4 个 3' 末端添加碱基 A 反应体系

组分	体积 (μL)
DNA	32
Klenow 缓冲液	5
d ATP	10
Klenow 片段(3'-5' exo-)	3
总体积	50

然后将管子放到调至 37℃ 的 Thermomixer(Eppendorf) 上反应 30 min。反应结束后, 用 MiniElute PCR 纯化试剂盒 (Qiagen) 进行纯化, 最后将样品溶于 10μL 洗脱缓冲液。

6. 连接接头

将上一步得到的 DNA 分别按下表在 1.5mL 的离心管中配制 4 个末端修复反应体系

组分	体积 (μL)
DNA	10
DNA 连接酶缓冲液	25
接头	10
DNA 连接酶	5
总体积	50

10 其中接头序列为:

接头 1: 5'GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG (SEQ ID NO: 178);

接头 2: 5'TACTACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 179)。

然后将管子放到调至 20℃ 的 Thermomixer(Eppendorf) 上反应 15 min。反应结束后, 用 2% 琼脂糖凝胶电泳分离, 电压为 120V, 电泳时间为 120 分钟。电泳结束后切下目的条带, 用 QIAquick PCR 纯化试剂盒 (Qiagen) 进行纯化, 最后将样品溶于 30μL 洗脱缓冲液。

15

7. PCR 扩增目的产物, 并加入测序引物

将上一步得到的 DNA 分别按以下体系配制 4 个 PCR 反应体系

组分	体积 (μL)
接头连接的 DNA	1
P1 公用引物	1

index 5 物	1
超 纯 水	22
Phusion DNA 酶	25
总体积	50

PI 公用引物：

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 180);

index 引物：

5 CAAGCAGAAGACGGCATAACGAGATNNNNNNNGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC T (SEQ ID NO: 181)，其中碱基 N 为 八、T、C、G 四个碱基的任意组合作为区别标识。

PCR 反应条件为：

98℃	30s	} 12 个循环
98℃	10s	
65℃	30s	
72℃	30s	
72℃	5min	
4℃	∞	

反应结束后，用 QIAquick PCR 纯化试剂盒 (Qiagen) 进行纯化，最后将样品溶于 50μL 洗脱缓冲液。

8. 文库检测

10 Agilent 2100 Bioanalyzer analysis system 检测文库插入片段大小及含量 ;Q-PCR 精确定量文库的浓度。

9. 测序与数据分析

文库检测合格后将按照双末端 151 个碱基的读长在 Hiseq2000 测序仪上进行序列分析。

15 测序所得的原始数据的基本分析过程主要包括以下步骤：首先进行数据处理，通过接头或 PCR 引物上的序列标签区分不同样本的文库数据，对测序所得的原始数据进行去污染、去接头和去低质量过滤；然后将 reads 与 IMGT 数据库的参考序列进行 Y、J 基因比对，分析。

过滤之后的测序原始数据进行比对分析的结果，见表 3，我们平均得到的总序列有 0.4 兆条，特异性的克隆平均有 0.16 兆条。

表 3

样品	总序列数	含 V、J 基因的序列数	含 ORF (开放阅读框) 的序列	特异性克隆数	特异性氨基酸数
ZXJ	71938668	48588975	24007284	288579	131822

20 文库 ZXJ 测序后获得的各种 V-J 基因配对的序列的数量示于表 4 和表 5 中 ,表 4 为 κ 型 IGL V-J 配对的序列的数量，表 5 为 λ 型 IGL V-J 配对的序列的数量。

表 4

<div>JK 基因</div> <div>Vκ-Jκ</div> <div>序列数</div> <div>Vκ 基因</div>	IGKJ1	IGKJ2	IGKJ3	IGKJ4	IGKJ5
IGKV1-12	143	157	48	657	412
IGKV1-13	136	136	41	612	400
IGKV1-16	503	486	133	1585	893
IGKV1-17	164	108	25	517	221
IGKV1-27	253	123	71	738	466
IGKV1-33	91	188	63	631	356

IGKV1-39	484	691	136	1969	1072
IGKV1-5	711	752	64	1225	479
IGKV1-6	77	41	14	163	60
IGKV1-8	133	103	25	630	236
IGKV1-9	296	287	74	1555	780
IGKV1D-12	396	418	83	1879	967
IGKV1D-13	73	50	23	348	204
IGKV1D-16	490	459	137	1563	893
IGKV1D-17	229	153	38	794	325
IGKV1D-33	88	182	61	619	353
IGKV1D-39	474	681	134	1942	1052
IGKV1D-43	18	8	3	56	49
IGKV1D-8	78	104	10	285	121
IGKV1-NL1	18	9	3	56	46
IGKV2-24	38	45	4	261	96
IGKV2-28	291	340	107	1492	855
IGKV2-29	470	576	137	2173	1080
IGKV2-30	247	350	67	803	353
IGKV2-40	27	89	3	157	91
IGKV2D-26	152	272	97	1787	1009
IGKV2D-28	309	329	104	1520	897
IGKV2D-29	236	230	62	823	553
IGKV2D-30	117	165	33	394	173
IGKV2D-40	10	48	3	71	35
IGKV3-11	364	381	149	1917	1280
IGKV3-15	284	340	70	723	383
IGKV3-20	575	921	187	2565	1730
IGKV3D-11	148	115	60	944	552
IGKV3D-15	263	341	66	722	396
IGKV3D-20	564	887	196	2574	1731
IGKV3D-7	60	61	11	198	50
IGKV3-NL1	2	3	1	11	2
IGKV3-NL2	1	0	0	36	15
IGKV3-NL3	0	0	0	11	4
IGKV3-NL4	23	28	3	146	77
IGKV3-NL5	41	36	8	340	186
IGKV4-1	88	89	8	361	135
IGKV5-2	191	417	23	632	171

表 5

J_h 基因 V_h-J_h 序列数	IGLJ1	IGLJ2	IGLJ3	IGLJ6	IGLJ7
IGLV10-54	432	523	2681	101	121
IGLV1-36	294	382	1473	108	173
IGLV1-40	1029	1224	4037	163	212
IGLV1-44	416	551	2244	117	101
IGLV1-47	344	594	2332	69	106

IGLV1-5 1	198	294	1187	55	8 1
IGLV2-1 1	1835	2252	9958	603	463
IGLV2-14	5869	5300	17601	659	477
IGLV2-1 8	1035	1095	4879	498	128
IGLV2-23	2145	2273	9 111	447	333
IGLV2-8	1213	1046	3 12 1	103	193
IGLV3-10	182	57 1	1898	147	2 15
IGLV3-12	217	337	119 1	35	23
IGLV3-16	220	337	1050	66	200
IGLV3-19	777	925	3070	149	192
IGLV3-1	953	1707	4049	199	2 17
IGLV3-2 1	1452	2022	6594	223	346
IGLV3-22	33	84	347	14	49
IGLV3-25	532	1340	4094	93	6 11
IGLV3-27	114	257	1185	7 1	105
IGLV3-9	82	228	768	26	37
IGLV4-3	7	3	32	0	0
IGLV4-60	169	486	1936	75	226
IGLV4-69	540	1243	6 134	322	350
IGLV5-37	163	2 13	980	25	3 1
IGLV5-39	67	110	5 11	14	18
IGLV5-45	392	522	2209	50	69
IGLV5-52	14 1	184	908	5	15
IGLV6-57	373	768	396 1	166	87
IGLV7-43	79	236	1095	28	46
IGLV7-46	92	290	1192	29	138
IGLV8-6 1	117	174	1006	29	40
IGLV9-49	139	224	1250	45	46

图 8 是基于表 4 数据作出的文库 ZXJ V_K-J_K 基因配对分布图，图 9 是基于表 5 数据作出的 v_λ-J_λ 基因配对分布图，图 8-9 显示了所有的 V 基因亚家族与 J 基因亚家族重排和各种 CDR3 的丰度，图中，纵坐标是 J 基因全部亚家族的分类，横坐标是 V 基因的每一种亚家族分类，两坐标相交点是 V-J 基因配对产生的一种 CDR3 序列，相交点每一方格颜色深浅代表每种 V-J 配对的丰度，图右侧对应有丰度颜色指示带。从表 4-5 及图 8-9 可以得出结论，使用本发明提供的引物能够全面地富集人类免疫球蛋白轻链 CDR3 序列，扩增出来的产物能够区分免疫球蛋白轻链的各个亚家族，并且通过上述本发明的方法能够确定各种轻链 CDR3 序列的相对比例，从而可以进一步基于这些序列的相对比例或该比例的变化，确定个体的免疫状态。

5

10 实施例 3

参照上述一般方法，按照以下步骤，对 T 细胞受体 α 链的 CDR3 编码序列进行富集、构建测序文库、测序及分析：

1. 分离人外周血单个核细胞

15

1) 取健康人新鲜外周血 5ml 于 15ml 的离心管中，并添加 5ml PBS，混匀，然后将其靠管壁缓慢加入

到装有 6ml 的单核细胞分离液的 50ml 离心管中，1 800rpm 离心 15min。

2) 吸取单核细胞层于新的 15ml 离心管中，加入三倍体积的 PBS，轻轻混匀，1 800rpm 离心 10min，弃上清，获得沉淀。

3) 用 1ml 的 PBS 将上述沉淀重悬，然后转移至新的 1.5ml EP 管中，1 800rpm 离心 10min，弃上清，获得沉淀。

20

4) 用 100μl 的 PBS 将上述步骤获得的沉淀重悬，则单核细胞提取完成。

2. 核酸样本提取

利用 DNA 提取试剂盒 (Qiagen) 提取人外周血单个核细胞基因组 DNA，具体地，包括：

1) 将 20μl QIAGEN 蛋白酶加入 200 μl 单核细胞样品中，混匀。

- 2) 将 200 μl 缓冲液 AL 加入样品中，充分混匀，56℃ 孵育 10 分钟。
- 3) 加入 200 μl 无水乙醇，充分混匀，将混合物转移到吸附柱上，8000r/m 离心 1 分钟，弃废液。
- 4) 然后加入 500 μl 缓冲液 AW1，8000r/m 离心 1 分钟，弃废液。
- 5) 再加入 500ml 缓冲液 AW2，8000r/m 离心 1 分钟，弃废液。
- 5 6) 然后全速 (14000 r/m) 空甩 1 分钟。
- 7) 再加入 200 μl 缓冲液 AE，室温静置 1 分钟，8000r/m 离心 1 分钟，以便提取获得 DNA，保存于 -20℃，备用。

3. 多重 PCR 扩增

以上一步得到的 DNA 作为模板，采用具有 SEQ ID NO: 46-84 所示的核苷酸序列的 V 区引物组成的第一引物组和具有 SEQ ID NO: 85-134 所示的核苷酸序列的 J 区引物组成的第二引物组，进行多重 PCR 扩增。

具体地，首先，将浓度均为 100 μM 的 39 种 V 区引物各取 1 μl 混合，然后加 1 l μl 的水稀释混匀，以便获得第一引物组，则获得的第一引物组中各 V 区引物的浓度均为 2 μM。然后，同样将浓度均为 100 μM 的 50 种 J 区引物各取 1 μl 混合，以便获得第二引物组，则获得的第二引物组中各 J 区引物的浓度均为 2 μM。

然后，按下表的配比，配制多重 PCR 扩增的反应体系：

组分	体积 (μl)
2 × QIAGEN Mutiplex PCR master mix	25
第一引物组 (SEQ ID NO: 46-84)	5
第二引物组 (SEQ ID NO: 85-134)	5
5 × Q 溶液	5
无 RNA 酶水	9
DNA	1
总体积	50

然后，将配制好的反应体系按以下反应条件进行 PCR 扩增：

95℃	15min	} 30 个循环
94℃	30s	
60℃	90s	
72℃	30s	
72℃	10min	
12℃	∞	

由此，获得扩增产物。

然后，将扩增产物进行 2% 琼脂糖凝胶电泳，并切取 10-200bp 的胶 (如图 10 所示)，并用 QIAquick Gel 纯化试剂盒 (Qiagen) 对其进行纯化回收，然后将回收产物溶于 34 μl 的洗脱缓冲液，备用。

4. 末端修复

将上一步得到的扩增产物，按下表中的配比于 1.5ml 离心管中配制末端修复反应体系：

扩增产物	32 μl
10 × 多聚核苷酸激酶缓冲液	10 μl
dNTP 溶液 (每种 10mM)	3 μl
T4 DNA 聚合酶	5 μl
双蒸水	44 μl

Klenow 片段	1 μl
T4 多聚核苷酸激酶	5 μl
总体积	100 μl

然后将离心管放置于调至 20℃ 的 Thermomixer(Eppendorf) 上反应 30 min, 以便获得经过末端修复的扩增产物, 然后利用 QIAquick PCR 纯化试剂盒 (Qiagen) 将经过末端修复的扩增产物进行纯化, 最后将其溶于 34μl 洗脱缓冲液中, 备用。

5. 3'末端添加碱基 A

5 将经过末端修复的扩增产物, 按下表中的配比于 1.5ml 离心管中配制 3'末端添加碱基 A 的反应体系:

经过末端修复的扩增产物	32 μl
10 × blue 缓冲液	5 μl
dATP(稀释为 1mM, GE 公司)	10 μl
Klenow (3'-5' exo-)	3 μl
总体积	50 μl

然后将离心管放置于调至 37℃ 的 Thermomixer(Eppendorf) 上反应 30 min, 以便获得 3 末端添加碱基 A 的的扩增产物, 然后利用 MiniElute PCR 纯化试剂盒 (Qiagen) 将 3'末端添加碱基 A 的的扩增产物进行纯化, 最后将其溶于 20μl 洗脱缓冲液中, 备用。

10 6. 连接接头

将 3'末端添加碱基 A 的的扩增产物, 按下表中的配比于 1.5ml 离心管中配制接头连接的反应体系:

3'末端添加碱基 A 的的扩增产物	18 μl
2 × Rapid 连接缓冲液	25 μl
PEI-接头 (接头 1 和接头 2, 50μM)	4 μl
T4 DNA 连接酶(Rapid, L603-HC-L)	3 μl
总体积	50 μl

其中 PEI-接头的序列为:

接头 1: TACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 179);

15 接头 2: 5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTC A C (SEQ ID NO: 182)。

然后将离心管放置于调至 20℃ 的 Thermomixer(Eppendorf) 上反应 15min, 以便获得连接产物, 然后利用 MiniElute PCR 纯化试剂盒 (Qiagen) 将连接产物进行纯化, 最后将其溶于 28μl 洗脱缓冲液中, 备用。

7. PCR 扩增

将上一步得到的连接产物, 按下表中的配比配制 PCR 反应体系:

连接产物	12.5 μl
P1 公用引物	1 μl
标签 7 引物 (Primer index 7)	1 μl
双蒸水	10.5 μl
2 × phusion master mix	25 μl
总体积	50 μl

PI 公用引物：

AATG[^] ATACGGCGACCA[^]CCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 180);

标签 7 引物 (Primer index7)：

5 CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
T (SEQ ID NO: 183)。

然后，将配制好的反应体系于 PCR 仪上进行 PCR 反应，以便获得第二扩增产物，其中 PCR 反应条件为：

98℃	1min	} 10 个循环
98℃	20s	
65℃	30s	
72℃	30s	
72℃	5min	
12℃	∞	

10 然后，利用 QIAquick PCR 纯化试剂盒 (Qiagen) 纯化回收 200-300bp 的第二扩增产物，以便获得回收产物，该回收产物构成 DNA 测序文库。然后，将获得的测序文库样品溶于 20μl 洗脱缓冲液。

8. 文库检测

利用 Agilent 2100 Bioanalyzer analysis system 检测获得的测序文库中插入片段大小及含量；利用 Q-PCR 精确定量该文库的浓度。

9. 测序与数据分析

15 文库检测合格后将按照双末端 151 个碱基的读长在 Hiseq2000 测序仪上进行测序及序列分析。

具体地，将检测合格的文库按照 Q-PCR 浓度在 Hiseq-2000 测序平台上采用边合成边测序的方法进行序列测定。为了将来源于不同样本制备的 DNA 文库在测序后区分开来，将 6bp 或 8bp 的标签序列通过接头或者 PCR 引物引入到片段的一侧，以便将不同文库直接混合后上机测序。测序时，先用 SP1 引物对样品 DNA 的一端进行测序，再用 SP2 引物对样品 DNA 的另一端进行测序。

20 然后，对获得的测序结果进行数据分析，其中分析数据时参考的序列为 IMGT 上 TCRA 的参考序列。具体地，首先对测序所得的原始数据进行基本分析，其主要包括以下步骤：对进行原始数据进行数据处理，通过接头或 PCR 引物上的序列标签区分不同样本的文库数据，对测序所得的原始数据进行去污染、去接头和去低质量过滤，以便确定 reads；然后，将 reads 与 IMGT 数据库的参考序列进行 V、J 基因比对、分析，结果见下表 6 以及图 11:

25

表 6: reads 分析比对结果

样品	总序列数	含 V、J 基因的序列数	含 ORF (开放阅读框) 的序列	特异性克隆数	特异性氨基酸数
ZRF TRA DNA	64875032	58614802	26575870	26575870	1645276

30 此外，图 11 显示了所有的 V 基因亚家族与 J 基因亚家族重排后相互配对和各种 VJ 配对分布以及 CDR3 的丰度。如图 11 所示，左侧坐标是 J 基因全部亚家族的分类，横坐标是 V 基因的每一种亚家族分类，两坐标相交点是 V-J 基因配对产生的一种 CDR3 序列，右侧颜色带指示每一方格的颜色深浅，代表每种 V-J 配对的丰度。从图 11 中可以看出，使用本发明提供的引物组合物能够全面地富集人类 T 细胞受体 CDR3 的序列，扩增出来的产物能够区分 T 细胞受体 α 链的各个亚家族。

实施例 4

参照上述一般方法，按照以下步骤，对 T 细胞受体 β 链的 CDR3 编码序列进行富集、构建测序文库、测序及分析：

1. 分离人外周血单个核细胞

1) 取健康人新鲜外周血 5mL 于 15ml 的离心管中，并添加 5ml PBS，混匀，然后将其靠管壁缓慢加入

2) 吸取单核细胞层于新的 15ml 离心管中，加入三倍体积的 PBS，轻轻混匀，800rpm 离心 10min，弃上清，获得沉淀。

3) 用 1ml 的 PBS 将上述沉淀重悬，然后转移至新的 1.5ml EP 管中，800rpm 离心 10min，弃上清，获得沉淀。

4) 用 100 μ L 的 PBS 将上述步骤获得的沉淀重悬，则单核细胞提取完成。

2. 核酸样本提取

2.1 DNA 提取

利用 DNA 提取试剂盒 (Qiagen) 提取人外周血单个核细胞基因组 DNA，具体地，包括：

1) 将 2 μ L QIAGEN 蛋白酶加入 200 μ L 单核细胞样品中，混匀。

2) 将 20 μ L 缓冲液 AL 加入样品中，充分混匀，56 $^{\circ}$ C 孵育 10 分钟。

3) 加入 200 μ L 无水乙醇，充分混匀，将混合物转移到吸附柱上，8000r/m 离心 1 分钟，弃废液。

4) 然后加入 50 μ L 缓冲液 AW1，8000r/m 离心 1 分钟，弃废液。

5) 再加入 500 μ L 缓冲液 AW2，8000r/m 离心 1 分钟，弃废液。

6) 然后全速 (14000 r/m) 空甩 1 分钟。

7) 再加入 200 μ L 缓冲液 AE，室温静置 1 分钟，8000r/m 离心 1 分钟，以便提取获得 DNA，保存于 -20 $^{\circ}$ C，备用。

2.2 总 RNA 提取及逆转录

也可以以总 RNA 为核酸样本。首先，利用 Trizol 法提取总 RNA。然后按照以下步骤将获得的总 RNA 逆转录为 cDNA：

(1) 按下表中的配比配置混合物：

总 RNA	10 μ L
C 区引物 (2 μ M)	1 μ L
总体积	11 μ L

其中，C 区引物 (TRBC-R1) 为：

CTCAAACACAGCGACCTC (SEQ ID NO: 184)。

然后，将获得的混合物于 PCR 仪上 65 $^{\circ}$ C 变性 5min 后，立即置于冰上，并向混合物中继续添加下表中的试剂：

5 $\times$ 1st stand 缓冲液	4 μ L
d NTP 混合物 (10mm)	2 μ L
0.1M DTT	2 μ L
无 RNA 酶水(40U/ μ L)	1 μ L
总体积	20 μ L

(2) 混匀后于室温下放置 2min，然后加入 1 μ L superscript II (20U/ μ L)。

(3) 混匀后，于 PCR 仪上按下列条件进行反应：

42 $^{\circ}$ C 50min

72 $^{\circ}$ C 15min

以便提取获得 cDNA, 保存备用。

然后, 分别以上一步得到的 DNA 和 cDNA 进行后续的步骤, 其中, 以 DNA 为例, 具体步骤如下所述:

3. 多重 PCR 扩增

5 以上一步得到的 DNA 作为模板, 采用具有 SEQ ID NO: 135-164 所示的核苷酸序列的 V 区引物组成的第一引物组和具有 SEQ ID NO: 165-177 所示的核苷酸序列的 J 区引物组成的第二引物组, 进行多重 PCR 扩增。

10 具体地, 首先, 将浓度均为 100μM 的 30 种 V 区引物各取 1μΣ 混合, 然后加 20μL 的水稀释混匀, 以便获得第一引物组, 则获得的第一引物组中各 V 区引物的浓度均为 2μM。然后, 同样将浓度均为 100μM 的 13 种 J 区引物各取 1μL 混合, 然后加 37μL 的水稀释混匀, 以便获得第二引物组, 则获得的第二引物组中各 J 区引物的浓度均为 2μM。

然后, 按下表的配比, 配制多重 PCR 扩增的反应体系:

组分	体积 (μL)
2 × QIAGEN Mutiplex PCR master mix	25
第一引物组 (SEQ ID NO: 135-164)	5
第二引物组 (SEQ ID NO: 165-177)	5
5 × Q 溶液	5
无 RNA 酶水	9
DNA	1
总体积	50

然后, 将配制好的反应体系按以下反应条件进行 PCR 扩增:

95 °C	15min	} 30 个循环
94 °C	30s	
60 °C	90s	
72 °C	30s	
72 °C	10min	
12 °C	∞	

由此, 获得扩增产物。

15 然后, 将扩增产物进行 2% 琼脂糖凝胶电泳, 其中电压为 100V, 电泳时间为 2 小时 20 分钟。然后切取 100-200 bp 的胶, 并用 QIAquick Gel 纯化试剂盒 (Qiagen) 对其进行纯化回收, 然后将回收产物溶于 30μL 的洗脱缓冲液, 备用。其中, 电泳结果见图 12。如图 12 所示, 其中 Mark 为 50bp DNA Ladder。

4. 末端修复

将上一步得到的扩增产物, 按下表中的配比于 1.5mL 离心管中配制末端修复反应体系:

扩增产物	32μl
10 × 多聚核苷酸激酶缓冲液	10 μΣ
dNTP 溶液 (每种 10mM)	3μΣ
T4 DNA 聚合酶	5 μL
双蒸水	44μΣ
Klenow 片段	1 μΣ
T4 多聚核苷酸激酶	5 μL

总体积	100 μ L
-----	-------------

然后将离心管放置于调至 20℃ 的 Thermomixer(Eppendorf) 上反应 30 min, 以便获得经过末端修复的扩增产物, 然后利用 QIAquick PCR 纯化试剂盒 (Qiagen) 将经过末端修复的扩增产物进行纯化, 最后将其溶于 34 μ L 洗脱缓冲液中, 备用。

5. 3'末端添加碱基 A

5 将经过末端修复的扩增产物按下表中的配比于 1.5mL 离心管中配制 3'末端添加碱基 A 的反应体系:

经过末端修复的扩增产物	32 μ L
10 $\times$ blue 缓冲液	5 μ L
dATP(稀释为 1mM, GE 公司)	10 μ L
Klenow (3'-5' exo-)	3 μ L
总体积	50 μ L

然后将离心管放置于调至 37℃ 的 Thermomixer(Eppendorf) 上反应 30 min, 以便获得 3'末端添加碱基 A 的扩增产物, 然后利用 MiniElute PCR 纯化试剂盒 (Qiagen) 将 3'末端添加碱基 A 的扩增产物进行纯化, 最后将其溶于 20 μ L 洗脱缓冲液中, 备用。

10 6. 连接接头

将 3'末端添加碱基 A 的扩增产物, 按下表中的配比于 1.5mL 离心管中配制接头连接的反应体系:

3'末端添加碱基 A 的扩增产物	18 μ L
2 $\times$ Rapid 连接缓冲液	25 μ L
PEI-接头 (接头 1 和接头 2, 50 μ M)	4 μ L
T4 DNA 连接酶(Rapid, L603-HC-L)	3 μ L
总体积	50 μ L

其中 PEI-接头的序列为:

接头 1: TACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO: 179);

15 接头 2: 5-Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTC A C (SEQ ID NO: 182)。

然后将离心管放置于调至 20℃ 的 Thermomixer(Eppendorf) 上反应 15min, 以便获得连接产物, 然后利用 MiniElute PCR 纯化试剂盒 (Qiagen) 将连接产物进行纯化, 最后将其溶于 32 μ L 洗脱缓冲液中, 备用。

7. PCR 扩增

将上一步得到的连接产物, 按下表中的配比配制 PCR 反应体系:

连接产物	12.5 μ L
P1 公用引物	1 μ L
标签 5 引物 (Primer index 5)	1 μ L
双蒸水	10.5 μ L
2 $\times$ phusion master mix	25 μ L
总体积	50 μ L

20 P1 公用引物:

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID

NO: 180);
标签 5 引物 (Primer index 5) :
CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
(SEQ ID NO: 185)。

5 然后，将配制好的反应体系于 PCR 仪上进行 PCR 反应，以便获得第二扩增产物，其中 PCR 反应条件为：

98℃	1min	} 10 个循环
98℃	20s	
65℃	30s	
72℃	30s	
72℃	5min	
12℃	∞	

然后，利用 QIAquick PCR 纯化试剂盒 (Qiagen) 对第二扩增产物进行纯化回收，以便获得回收产物，该回收产物构成 DNA 测序文库。然后，将获得的测序文库样品溶于 20μL 洗脱缓冲液。

8. 文库检测

10 利用 Agilent 2100 Bioanalyzer analysis system 检测获得的测序文库中插入片段大小及含量；利用 Q-PCR 精确定量该文库的浓度。

9. 测序与数据分析

文库检测合格后将按照双末端 151 个碱基的读长在 Hiseq2000 测序仪上进行测序及序列分析。

15 具体地，将检测合格的文库按照 Q-PCR 浓度在 Hiseq-2000 测序平台上采用边合成边测序的方法进行序列测定。为了将来源于不同样本制备的 DNA 文库在测序后区分开来，将 6bp 或 8bp 的标签序列通过接头或者 PCR 引物引入到片段的一侧，以便将不同文库直接混合后上机测序。测序时，先用 SP1 引物对样品 DNA 的一端进行测序，再用 SP2 引物对样品 DNA 的另一端进行测序。

20 然后，对获得的测序结果进行数据分析，其中分析数据时参考的序列为 IMGT 上 TCRB 的参考序列。具体地，首先对测序所得的原始数据进行基本分析，其主要包括以下步骤：对进行原始数据进行数据处理，通过接头或 PCR 引物上的序列标签区分不同样本的文库数据，对测序所得的原始数据进行去污染、去接头和去低质量过滤，以便确定 reads；然后，将 reads 与 IMGT 数据库的参考序列进行 V、D、J 基因比对、分析，结果见下表 7、表 8 以及图 13：

表 7: TCR V-J 分布图原始数据

	TRBJ1-1	TRBJ1-2	TRBJ1-3	TRBJ1-4	TRBJ1-5	TRBJ1-6	TRBJ2-1	TRBJ2-2	TRBJ2-3	TRBJ2-4	TRBJ2-5	TRBJ2-6	TRBJ2-7
TRBV10-1	25	66	5	14	37	22	54	20	172	1	26	7	53
TRBV10-2	2707	3048	347	433	1420	1396	1210	583	1083	160	1183	200	3375
TRBV10-3	29314	31838	4666	4396	13280	12213	12138	6549	12004	1509	11617	3190	36972
TRBV11-1	2266	2771	373	262	720	1346	1626	881	1453	251	1809	233	6988
TRBV11-2	12920	21158	3982	2467	5169	11250	9181	5429	8592	1668	11430	1545	43979
TRBV11-3	3469	4871	1070	468	1663	3039	3729	2653	3643	485	3309	400	11081
TRBV12-3	2731	3924	576	375	1598	1656	944	566	717	86	1403	119	3605
TRBV12-4	2810	3990	609	401	1656	1691	983	569	790	107	1388	104	3658
TRBV12-5	538	701	95	78	316	243	236	121	155	16	282	20	653
TRBV13	2182	3891	697	379	896	1858	1533	820	1550	207	1976	203	11896
TRBV14	6030	8454	1281	961	1345	3872	3625	1746	2757	626	3521	479	12361
TRBV15	13926	18762	3532	3712	13805	9367	9965	4936	7618	1705	11237	2386	31029
TRBV16	696	668	100	108	361	223	225	79	158	30	144	37	779
TRBV18	2641	4585	1063	817	1811	1775	1785	1198	1750	275	2113	333	10487
TRBV19	64973	65229	9764	10655	53226	28275	37465	19839	24160	5408	20406	7038	70569
TRBV20-1	93436	122945	17221	19009	67842	65166	86276	25155	77527	14419	86126	7282	244399
TRBV24-1	2911	7168	643	703	3527	1301	5270	3264	4672	461	4324	466	6454
TRBV25-1	61	211	21	7	56	115	67	33	61	5	38	11	240
TRBV27	5635	14875	1510	1713	5120	6085	4101	2901	3839	634	3420	647	21292
TRBV28	7421	18005	2555	2489	9659	6562	4356	5789	4361	812	4615	922	29196
TRBV29-1	30933	26858	2263	2816	4142	2866	7937	4369	6047	764	10419	1697	29438
TRBV2	58329	73266	9774	12848	37342	32237	32796	17659	30163	5572	33690	6738	95090
TRBV30	1716	4015	804	551	1866	2202	1236	845	1123	207	1072	171	5980
TRBV3-1	762	1595	268	181	373	717	299	224	335	47	574	43	1777
TRBV4-1	2437	4935	819	650	1827	2229	1396	715	1202	225	1654	176	7487

TRBV4-2	1038	1707	249	247	511	786	566	266	545	70	659	77	2455
TRBV4-3	2175	3769	644	491	919	1761	1074	565	985	138	1229	140	4498
TRBV5-1	38347	46266	7094	6556	9518	20751	21866	13354	20816	3976	28185	4857	10217
TRBV5-4	3067	6351	607	424	678	2097	1546	967	1779	236	2573	357	8389
TRBV5-5	4401	7497	981	1000	2639	3628	2402	1097	2594	325	3776	454	11300
TRBV5-6	5326	8532	1148	1151	2837	3703	3102	1516	2941	540	4524	579	11764
TRBV5-8	724	1603	216	124	348	623	380	257	423	70	633	68	2184
TRBV6-1	1026	2905	570	313	1379	2154	1624	811	1661	600	1708	166	6951
TRBV6-2	311	707	128	57	482	535	864	433	792	221	796	54	3656
TRBV6-3	283	712	133	56	459	514	848	397	800	205	793	58	3524
TRBV6-4	3065	6682	755	651	2744	3243	3241	1996	3717	463	1817	386	6733
TRBV6-5	16581	36376	5095	3594	12469	16830	9283	6757	10044	1894	10078	1958	41897
TRBV6-6	11175	24571	3869	2742	11815	12364	5633	4245	6240	1218	6760	1189	31368
TRBV6-8	10	41	8	3	16	26	27	9	24	5	20	0	81
TRBV6-9	211	699	75	57	293	339	124	84	149	23	148	21	806
TRBV7-2	51355	90271	15992	10047	32032	46916	44532	22319	43077	7409	60186	8811	166829
TRBV7-3	6318	11662	1866	1083	3233	5823	10290	4646	9407	1236	7979	1225	22531
TRBV7-4	586	1159	190	127	516	600	535	201	490	77	557	84	2064
TRBV7-6	3964	6384	1775	823	2070	4061	6552	2964	5424	1397	5982	837	17670
TRBV7-7	412	485	159	68	258	416	1496	614	1085	257	1665	113	2301
TRBV7-8	24612	44453	8562	5221	16932	22841	23977	11485	22189	3904	28500	3903	84675
TRBV7-9	72813	106751	18406	16135	29798	47101	55358	27528	48277	7719	65827	9929	178471
TRBV9	44647	64535	16244	9026	16391	51079	45227	24539	39786	5611	34401	5884	11787

WO 2013/097744

PCT/CN2012/087668

表 8: 表 7 的结果总结

样品	总序列数	含 V、D、J 基因的序列数	含 ORF（开放阅读框）的序列	特异性克隆数	特异性氨基酸数
ZRF TRB DNA	43913570	39573122	28802345	5940470	3766846

此外，图 13 显示了所有的 V 基因亚家族与 J 基因亚家族重排后相互配对和各种 VJ 配对分布以及 CDR3 的丰度。如图 13 所示，左侧坐标是 J 基因全部亚家族的分类，横坐标是 V 基因的每一种亚家族分类，两坐标相交点是 V-J 基因配对产生的一种 CDR3 序列，右侧颜色带指示每一方格的颜色深浅，代表每种 V-J 配对的丰度。从图 13 中可以看出，使用本发明提供的引物组合物能够全面地富集人类 T 细胞受体 β 链 CDR3 的序列，扩增出来的产物能够区分 T 细胞受体 β 链的各个亚家族。

工业实用性

本发明的用于扩增 CDR3 编码序列的引物组合物、富集 CDR3 编码序列的方法、构建 CDR3 编码序列的测序文库的方法、确定 CDR3 编码序列的序列信息的方法、确定个体免疫状态的方法、确定个体免疫状态的系统以及试剂盒，能够有效地应用于对个体的免疫球蛋白重链和轻链、T 细胞受体 α 链和 β 链的 CDR3 编码序列进行富集和测序，并且基于测序结果，能够准确、高效地确定个体的免疫状态。

尽管本发明的具体实施方式已经得到详细的描述，本领域技术人员将会理解。根据已经公开的所有教导，可以对那些细节进行各种修改和替换，这些改变均在本发明的保护范围之内。本发明的全部范围由所附权利要求及其任何等同物给出。

在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示意性实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本发明的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不一定指的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任何的一个或多个实施例或示例中以合适的方式结合。

权 利 要 求 书

1、一种用于扩增 CDR3 编码序列的引物组合物，其特征在于，包括：

第一引物组，所述第一引物组由至少一种 V 区引物组成，所述至少一种 V 区引物的每一种均包含与至少一个 V 基因片段互补的序列；以及

第二引物组，所述第二引物组由至少一种 J 区引物组成，所述至少一种 J 区引物的每一种均包含与至少一个 J 基因片段互补的序列，

其中，所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。

2、根据权利要求 1 所述的引物组合物，其特征在于，所述 CDR3 编码序列来源于免疫球蛋白重链。

3、根据权利要求 2 所述的引物组合物，其特征在于，所述 V 区引物是正义链引物，所述 J 区引物是反义链引物。

4、根据权利要求 2 所述的引物组合物，其特征在于，所述免疫球蛋白为人类免疫球蛋白。

5、根据权利要求 2 所述的引物组合物，其特征在于，所述第一引物组的至少一种引物包含与多个 V 基因片段的保守区互补的序列。

6、根据权利要求 5 所述的引物组合物，其特征在于，所述 V 区引物具有选自 SEQ ID NO: 1-20 所示的核苷酸序列。

7、根据权利要求 2 所述的引物组合物，其特征在于，所述第二引物组的至少一种包含与多个 J 基因片段的保守区互补的序列。

8、根据权利要求 2 所述的引物组合物，其特征在于，所述第二引物组由一种引物构成，该引物包含与多个 J 基因片段的保守区互补的序列。

9、根据权利要求 8 所述的引物组合物，其特征在于，所述 J 区引物具有如 SEQ ID NO: 21 所示的核苷酸序列。

10、根据权利要求 1 所述的引物组合物，其特征在于，所述 CDR3 编码序列来源于免疫球蛋白轻链。

11、根据权利要求 10 所述的引物组合物，其特征在于，所述 V 区引物是正义链引物，所述 J 区引物是反义链引物。

12、根据权利要求 10 所述的引物组合物，其特征在于，所述免疫球蛋白为人类免疫球蛋白。

13、根据权利要求 10 所述的引物组合物，其特征在于，所述第一引物组的至少一种引物包含与多个 V 基因片段的保守区互补的序列。

14、根据权利要求 13 所述的引物组合物，其特征在于，所述 V 区引物具有选自 SEQ ID NO: 22-40 所示的核苷酸序列。

15、根据权利要求 10 所述的引物组合物，其特征在于，所述第二引物组的至少一种包含与多个 J 基因片段的保守区互补的序列。

16、根据权利要求 15 所述的引物组合物，其特征在于，所述 J 区引物具有选自 SEQ ID NO: 41-45 所示的核苷酸序列。

17、根据权利要求 1 所述的引物组合物，其特征在于，所述 CDR3 编码序列来源于 T 细胞受体 α 链。

18、根据权利要求 17 所述的引物组合物，其特征在于，所述 V 区引物是正义链引物，所述 J 区引物是反义链引物。

19、根据权利要求 17 所述的引物组合物，其特征在于，所述 T 细胞受体为人类 T 细胞受体。

20、根据权利要求 17 所述的引物组合物，其特征在于，所述第一引物组的至少一种引物包含与多个 V 基因片段的保守区互补的序列。

21、根据权利要求 20 所述的引物组合物，其特征在于，所述 V 区引物具有选自 SEQ ID NO: 46-84 所示的核苷酸序列。

22、根据权利要求 17 所述的引物组合物，其特征在于，所述第二引物组的至少一种引物包含与多个 J 基因片段的保守区互补的序列。

23、根据权利要求 22 所述的引物组合物，其特征在于，所述 J 区引物具有选自 SEQ ID NO: 85-134 所示的核苷酸序列。

24、根据权利要求 1 所述的引物组合物，其特征在于，所述 CDR3 编码序列来源于 T 细胞受体 β 链。

25、根据权利要求 24 所述的引物组合物，其特征在于，所述 V 区引物是正义链引物，所述 J 区引物是反义链引物。

26、根据权利要求 24 所述的引物组合物，其特征在于，所述 T 细胞受体为人类 T 细胞受体。

27、根据权利要求 24 所述的引物组合物，其特征在于，所述第一引物组的至少一种引物包含与多个 V 基因片段的保守区互补的序列。

28、根据权利要求 27 所述的引物组合物，其特征在于，所述 V 区引物具有选自 SEQ ID NO: 135-164 所示的核苷酸序列。

29、根据权利要求 24 所述的引物组合物，其特征在于，所述第二引物组的至少一种引物包含与多个 J 基因片段的保守区互补的序列。

30、根据权利要求 29 所述的引物组合物，其特征在于，所述 J 区引物具有选自 SEQ ID NO: 165-177 所示的核苷酸序列。

31、一种富集 CDR3 编码序列的方法，其特征在于，包括以下步骤：

提供核酸样本，所述核酸样本中包含编码 CDR3 的核酸序列；以及

利用权利要求 1-30 任一项所述的引物组合物，以所述核酸样本作为模板，进行 PCR 扩增，以便获得富集所述 CDR3 编码序列的扩增产物，

其中，所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。

32、根据权利要求 31 所述的方法，其特征在于，进一步包括：

从人外周血分离外周血单个核细胞；以及

从所述外周血单个核细胞提取所述核酸样本。

33、根据权利要求 32 所述的方法，其特征在于，所述外周血来源于正常人。

34、根据权利要求 32 所述的方法，其特征在于，通过密度梯度离心分离所述外周血单个核细胞。

35、根据权利要求 31 所述的方法，其特征在于，所述 PCR 扩增为多重引物 PCR 扩增。

36、根据权利要求 35 所述的方法，其特征在于，所述多重引物 PCR 扩增的退火温度为 60 摄氏度。

37、根据权利要求 31 所述的方法，其特征在于，进一步包括通过选自琼脂糖凝胶电泳、磁珠纯化和纯化柱纯化的至少一种分离纯化所述扩增产物。

38、根据权利要求 31 所述的方法，其特征在于，所述 CDR3 编码序列来源于免疫球蛋白重链。

39、根据权利要求 38 所述的方法，其特征在于，所述 PCR 扩增采用权利要求 2-9 任一项所述的引物组合物进行。

40、根据权利要求 39 所述的方法，其特征在于，所述扩增产物的长度为 140-280bp。

41、根据权利要求 31 所述的方法，其特征在于，所述 CDR3 编码序列来源于免疫球蛋白轻链。

42、根据权利要求 41 所述的方法，其特征在于，所述 PCR 扩增采用权利要求 10-16 任一项所述的引物组合物进行。

43、根据权利要求 42 所述的方法，其特征在于，所述扩增产物的长度为 110-250bp。

44、根据权利要求 31 所述的方法，其特征在于，所述 CDR3 编码序列来源于 T 细胞受体 α 链。

45、根据权利要求 44 所述的方法，其特征在于，所述 PCR 扩增采用权利要求 17-23 任一项所述的引物组合物进行。

46、根据权利要求 45 所述的方法，其特征在于，所述扩增产物的长度为 100-200bp。

47、根据权利要求 31 所述的方法，其特征在于，所述 CDR3 编码序列来源于 T 细胞受体 β 链。

48、根据权利要求 47 所述的方法，其特征在于，所述 PCR 扩增采用权利要求 24-30 任一项所述的引物组合物进行。

49、根据权利要求 48 所述的方法，其特征在于，所述扩增产物的长度为 100-200bp。

50、一种构建 CDR3 编码序列的测序文库的方法，其特征在于，包括以下步骤：

根据权利要求 31-49 任一项所述的方法，获得富集所述 CDR3 编码序列的扩增产物；以及
针对所述扩增产物，构建 DNA 测序文库，所述 DNA 测序文库构成所述 CDR3 编码序列的测序文库，
其中，所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。

5 51、根据权利要求 50 所述的方法，其特征在于，针对所述扩增产物，构建 DNA 测序文库进一步包括：

对所述扩增产物进行末端修复，以便获得经过末端修复的扩增产物；

对所述经过末端修复的扩增产物进行 3' 端添加碱基 A，以便获得 3' 末端添加碱基 A 的扩增产物；

将所述 3' 末端添加碱基 A 的扩增产物与接头相连，以便获得连接产物；

10 对所述连接产物进行 PCR 扩增，以便获得第二扩增产物；以及

将所述第二扩增产物进行纯化回收，以便获得回收产物，所述回收产物构成所述 DNA 测序文库。

52、根据权利要求 51 所述的方法，其特征在于，所述末端修复是利用 Klenow 片段、T4 DNA 聚合酶和 T4 多核苷酸激酶进行的，所述 Klenow 片段具有 5' $\rightarrow$ 3' 聚合酶活性和 3' $\rightarrow$ 5' 外切酶活性，但缺少 5' $\rightarrow$ 3' 外切酶活性。

15 53、根据权利要求 51 所述的方法，其特征在于，利用 Klenow (3'-5' exo-) 对所述经过末端修复的扩增产物进行 3' 端添加碱基 A。

54、根据权利要求 51 所述的方法，其特征在于，将所述具有粘性末端 A 的扩增产物与接头相连是利用 T4 DNA 连接酶进行的。

20 55、根据权利要求 51 所述的方法，其特征在于，通过选自琼脂糖凝胶电泳、磁珠纯化和纯化柱纯化的至少一种分离纯化所述第二扩增产物。

56、一种确定 CDR3 编码序列信息的方法，其特征在于，包括以下步骤：

根据权利要求 50-55 任一项所述的方法，构建 CDR3 编码序列的测序文库；以及

对所述 CDR3 编码序列的测序文库进行测序，以便确定所述 CDR3 编码序列的序列信息，

25 其中，所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。

57、根据权利要求 56 所述的方法，其特征在于，利用选自 HiSeq2000、SOLiD、454 和单分子测序装置的至少一种进行所述测序。

58、一种确定个体免疫状态的方法，其特征在于，包括以下步骤：

30 根据权利要求 56 或 57 所述的方法，对所述个体的 CDR3 编码序列进行测序，以便获得由多个测序数据构成的测序结果；以及

基于所述测序结果，确定所述个体的免疫状态，

其中，所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种。

35 59、根据权利要求 58 所述的方法，其特征在于，基于所述测序结果，确定所述个体的免疫状态进一步包括：

将所述测序结果与对照序列进行比对，以便确定所述个体中所包含的 CDR3 的亚家族类型，以及各亚家族类型的相对比例。

60、根据权利要求 59 所述的方法，其特征在于，在多个不同的时间点，从相同的个体提取样品，并分别根据权利要求 56 或 57 所述的方法，获得多个测序结果；以及

40 将所述多个测序结果进行比对，以便确定所述个体中 CDR3 的亚家族类型以及相对比例的变化。

61、一种确定个体免疫状态的系统，其特征在于，包括：

CDR3 编码序列富集装置，所述 CDR3 编码序列富集装置中设置有权利要求 1-30 任一项所述的引物组合物，以便对所述个体的核酸样本富集 CDR3 编码序列，其中，所述 CDR3 编码序列来源于选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种；

45 文库构建装置，所述文库构建装置与所述 CDR3 编码序列富集装置相连，以便针对所述经过富集的 CDR3 编码序列构建 CDR3 编码序列的测序文库；

测序装置，所述测序装置与所述文库构建装置相连，用于对所述 CDR3 编码序列的测序文库进行测序，以便获得由多个测序数据构成的测序结果；以及

分析装置，所述分析装置与所述测序装置相连，用于基于所述测序结果，确定所述个体的免疫状态。

5 62、根据权利要求 61 所述的系统，其特征在于，所述分析装置进一步包括比对单元，所述比对单元中存储有对照序列，用于将所述测序结果与所述对照序列进行比对，以便确定所述个体中所包含的 CDR3 的亚家族类型，以及各亚家族类型的相对比例。

63、一种试剂盒，其特征在于，所述试剂盒中设置有权利要求 1-30 任一项所述的引物组合物。

64、根据权利要求 63 所述的试剂盒，其特征在于，所述试剂盒用于检测选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种的 V-J 重排。

10 65、根据权利要求 63 所述的试剂盒，其特征在于，所述试剂盒用于检测选自免疫球蛋白重链、免疫球蛋白轻链、T 细胞受体 α 链和 T 细胞受体 β 链的至少一种的 CDR3 的编码序列。

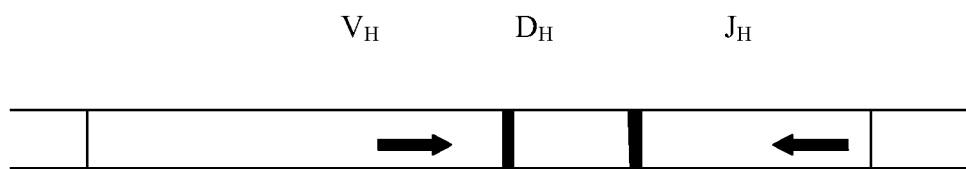


图1

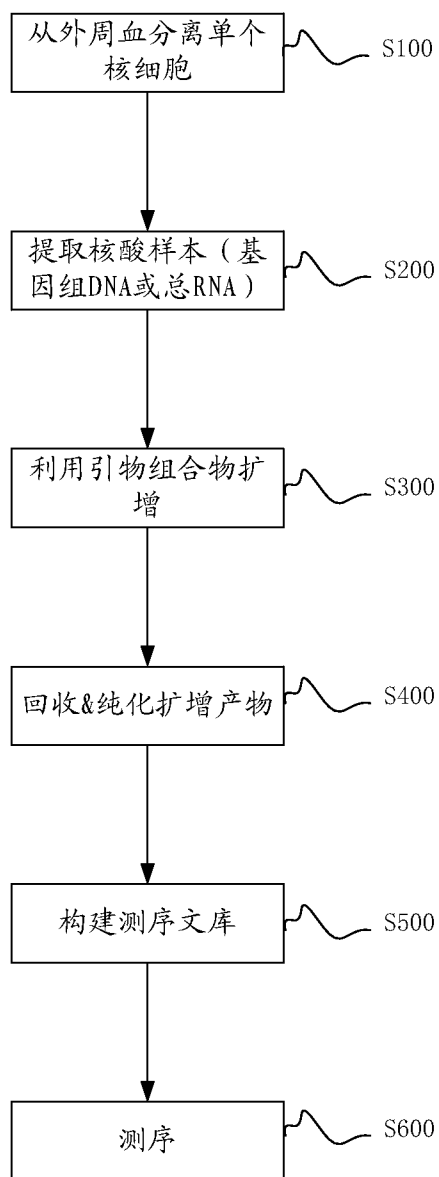


图2

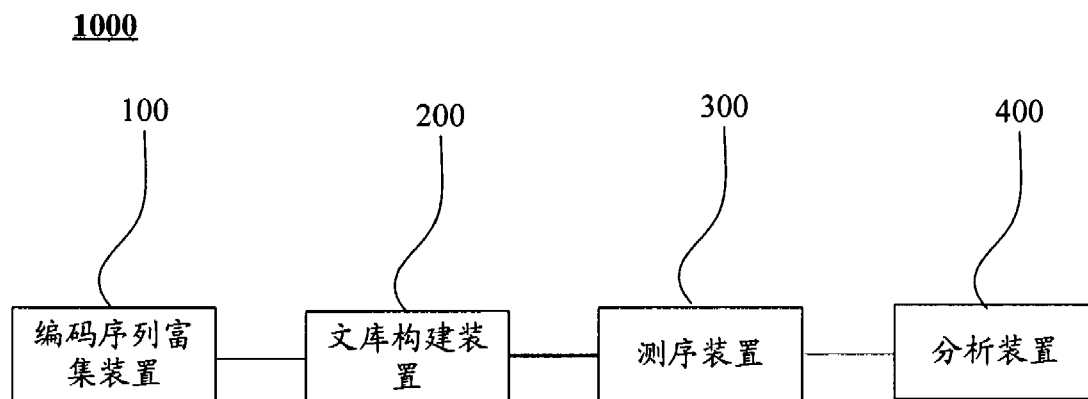


图 3

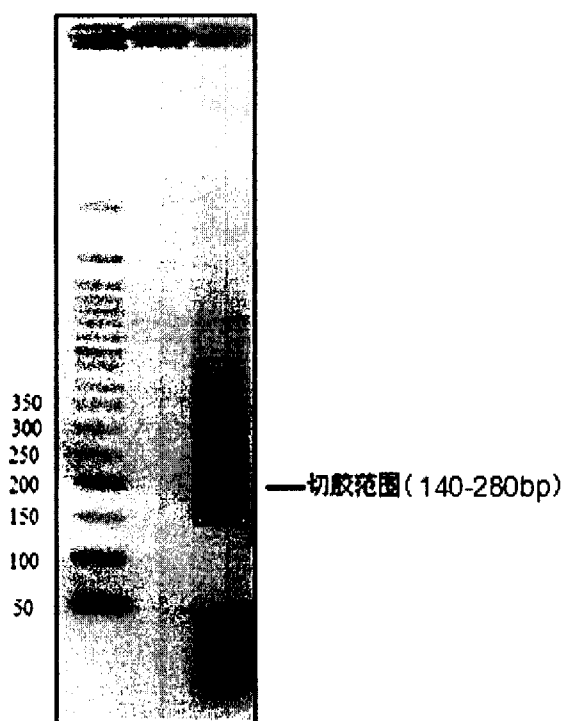


图 4

V-J基因配对分布及丰度

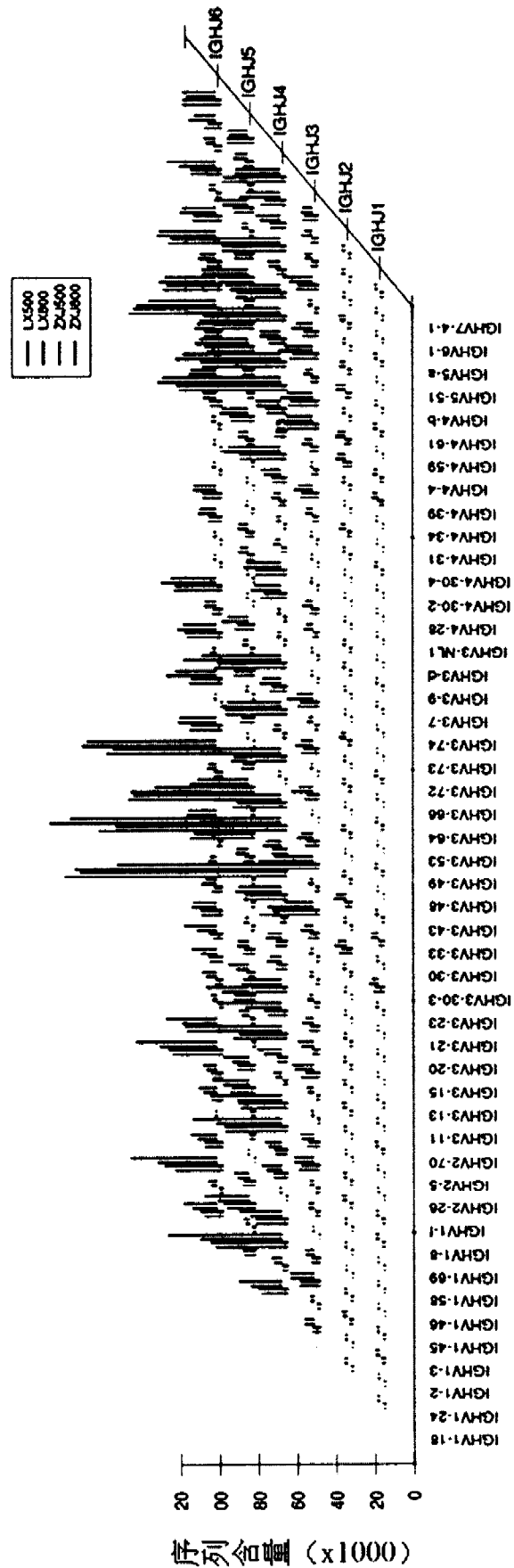


图 5

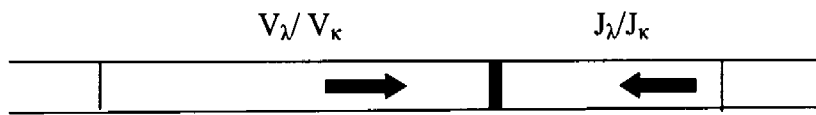


图6



图 7

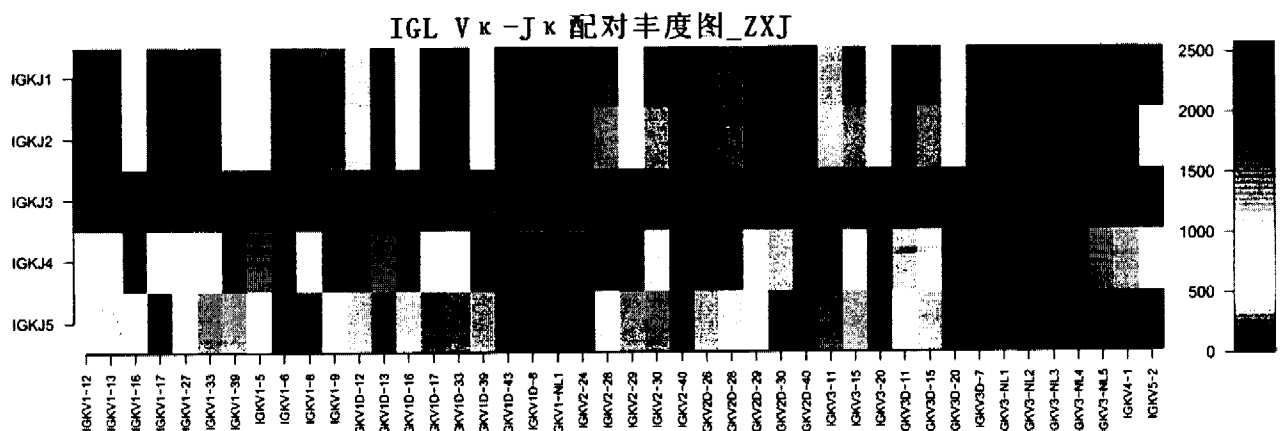


图 8

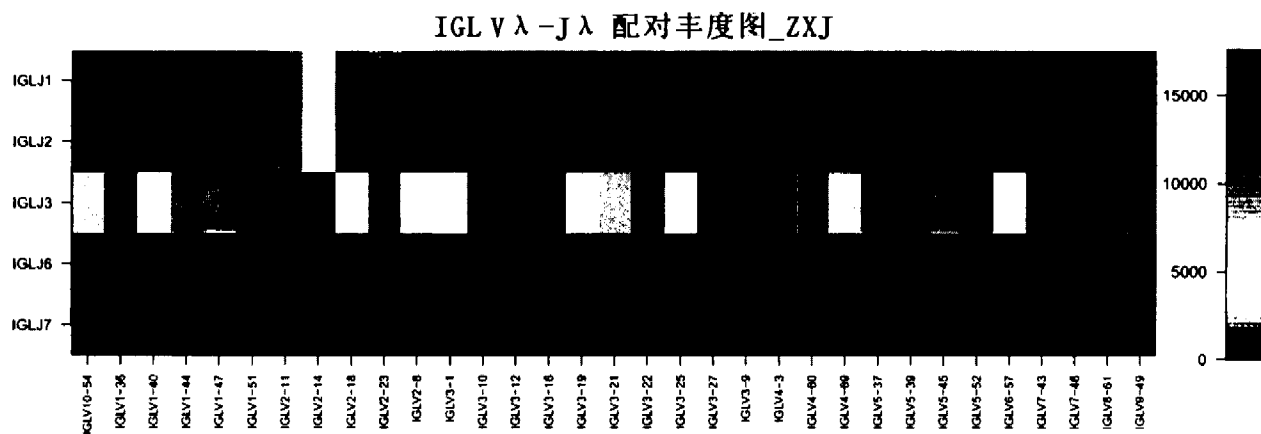


图 9

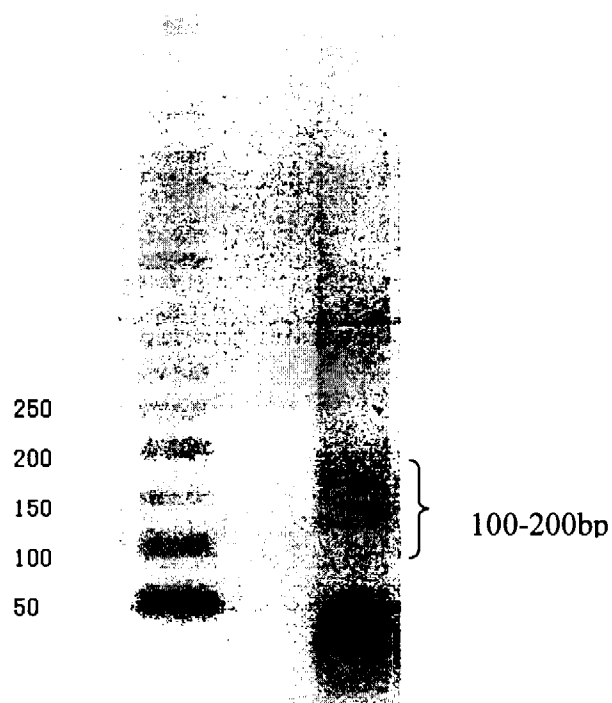


图 10

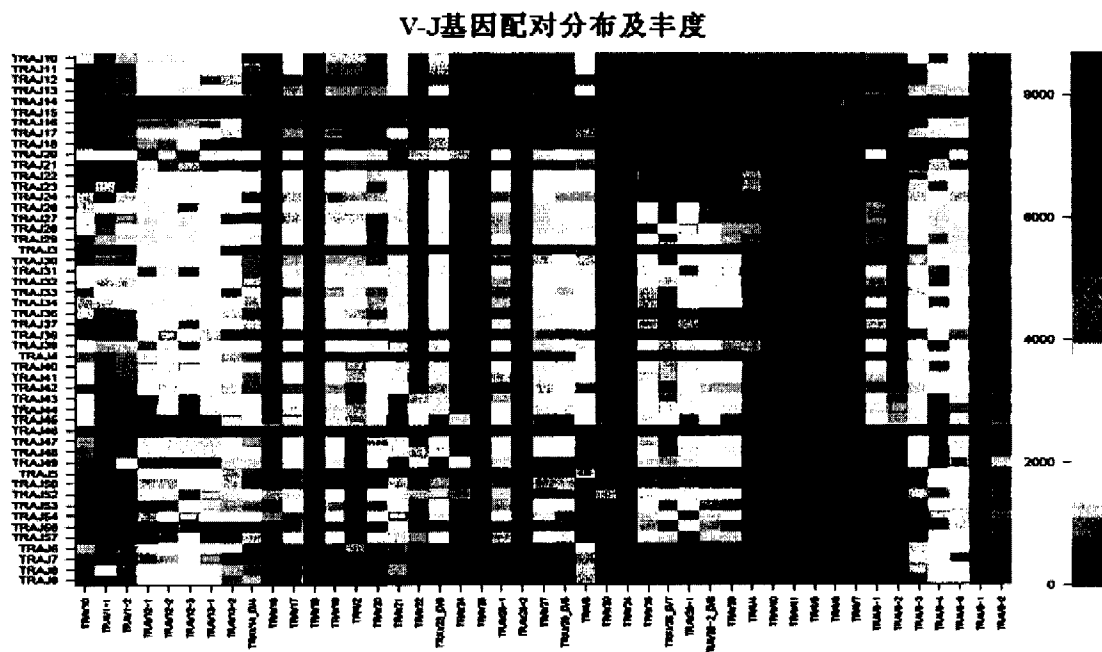


图 11

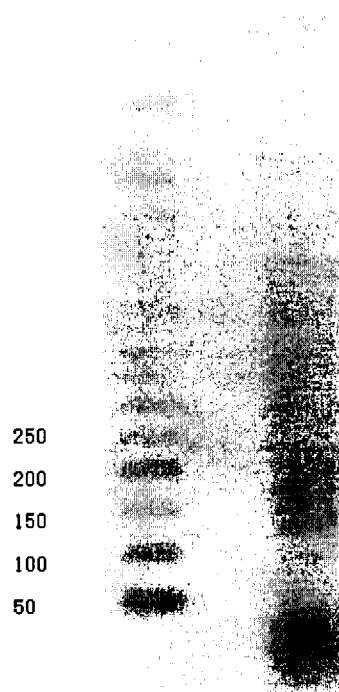


图 12

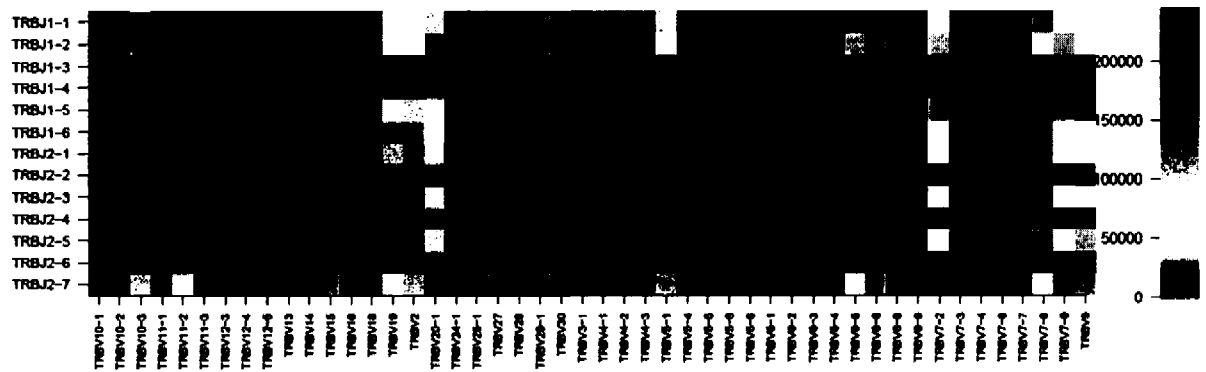


图 13

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/087668

A. CLASSIFICATION OF SUBJECT MATTER

See the extra sheet

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)

IPC: C12N; C12Q; C40B; C12M

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, CNKI, ISI web of Knowledge, EMBASE, GenBank: immunoglobulin, Ig, heavy chain, light chain, T cell receptor, TCR, complementary determining region, CDR3, primer, library, coding sequence, CDS, amplif+, sequence search on SEQ ID NO: 1-177

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2010/151416 A1 (FRED HUTCHINSON CANCER RESEARCH CENTER [US]) 29 December 2010 (29.12.2010) see claims 1-45, description, paragraphs [0002], [0004] -[0006], [0008]-[0010], [0014]-[0019], [0021], [0024]-[0030], [0034] -[0040], [0044]	1-50, 56-65
Y		51-55

☒ Further documents are listed in the continuation of Box C.
 ☒ See patent family annex.

* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "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) "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		"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 "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 "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 "&" document member of the same patent family
Date of the actual completion of the international search 21 March 2013 (21.03.2013)	Date of mailing of the international search report 04 April 2013 (04.04.2013)	
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

International application No.

PCT/CN2012/087668

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category [*]	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Teer J K , et al. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. GENOME RESEARCH. 2010, vol. 20, pages 1420-1431. see the abstract, page 1429, right column, paragraph [0007] to page 1430, right column, paragraph [0006], fig. 1C	51-55

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2012/087668

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO 2010151416 A I	29.12.2010	US 2010330571 A I	30.12.2010
		CA 2765949 A I	29.12.2010
		AU 2010263172 A I	19.01.2012
		EP 2446052 A 1	02.05.2012
		KR 20120044941 A	08.05.2012
		CN 102459643 A	16.05.2012
		INDELNP 2011 10153 E	19.10.2012

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2012/087668

A. CLASSIFICATION OF SUBJECT MATTER

C12N 15/11 (2006.01) i

C12N 15/13 (2006.01) i

C12N 15/10 (2006.01) i

C12N 15/12 (2006.01) i

C12Q 1/68 (2006.01) i

C40B 40/08 (2006.01) i

C40B 50/06 (2006.01) i

C12M 1/34 (2006.01) i

A. 主题的分类

见附加页

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)

IPC: C12N, C12Q, C40B, C12M

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, CNKI, ISI web of Knowledge, EMBASE, GenBank: 免疫球蛋白, 重链, 轻链, T 细胞受体, α 链, β 链, 互补决定区, 引物, 文库, 编码序列, 扩增, immunoglobulin, Ig, heavy chain, light chain, T cell receptor, TCR, complementary determining region, CDR3, primer, library, coding sequence, CDS, amplif+, 关于 SEQ ID NO: 1-177 的序列检索

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 2010/151416 A1 (FRED HUTCHINSON CANCER RESEARCH CENTER [US]) 29.12 月 2010 (29.12.2010) 参见权利要求 1-45, 说明书第 [0002]、[0004]-[0006]、[0008]-[0010]、[0014]-[0019]、[0021]、[0024]-[0030]、[0034]-[0040]、[0044] 段	1-50, 56-65
Y		51-55



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 3/4 之前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

21.3 月 2013 (21.03.2013)

国际检索报告邮寄日期

04.4 月 2013 (04.04.2013)

ISA/CN 的名称和邮寄地址:

中华人民共和国国家知识产权局

中国北京市海淀区蓟门桥西土城路 6 号 100088

传真号: (86-10)62019451

受权官员

幸颖

电话号码: (86-10) 62411044

c (续). 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
Y	Teer J K, 等. Systematic comparison of three genomic enrichment methods for massively parallel DNA sequencing. GENOME RESEARCH. 2010 年. 第 20 卷, 第 1420-1431 页 参见摘要, 第 1429 页右栏第 7 段至第 1430 页右栏第 6 段, 图 1C	51-55

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2012/087668

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
W O 2010151416 A 1	29.12.2010	U S 2010330571 A I	30.12.2010
		CA 2765949 A I	29.12.2010
		A U 2010263172 A I	19.01.2012
		EP 2446052 A I	02.05.2012
		K R 20120044941 A	08.05.2012
		CN 102459643 A	16.05.2012
		IN DELNP201 110153 E	19.10.2012

A. 主题的分类

C12N 15/11 (2006.01) i
 C12N 15/13 (2006.01) i
 C12N 15/10 (2006.01) i
 C12N 15/12 (2006.01) i
 C12Q 1/68 (2006.01) i
 C40B 40/08 (2006.01) i
 C40B 50/06 (2006.01) i
 C12M 1/34 (2006.01) i