



(12)发明专利申请

(10)申请公布号 CN 110257537 A

(43)申请公布日 2019.09.20

(21)申请号 201910596979.6

(22)申请日 2019.07.04

(71)申请人 北京京诺玛特科技有限公司

地址 101318 北京市顺义区临空经济核心区裕华路28号2幢3层301室(空港B区)

(72)发明人 赵金银 李宏志 刘琦 许立志

(74)专利代理机构 北京汉鼎理利专利代理事务所(特殊普通合伙) 11618

代理人 潘满根

(51) Int. Cl.

C12Q 1/689(2018.01)

C12Q 1/6806(2018.01)

C12N 15/10(2006.01)

C40B 50/06(2006.01)

C40B 40/06(2006.01)

权利要求书2页 说明书10页
序列表1页

(54)发明名称

用于梅毒病原体检测的液相杂交捕获文库及其构建方法

(57)摘要

本发明提供一种用于梅毒病原体检测的液相杂交捕获文库及其构建方法。所述液相杂交捕获探针文库包括梅毒基因组序列和人基因组参照序列；所述梅毒基因组序列每隔10-20bp截取50-70bp，得到探针序列；将所得探针序列与病原体数据库进行比对，选取其中特异比到梅毒上的序列作为液相杂交捕获文库的探针序列。本发明的梅毒病原体检测的液相杂交捕获文库对于目标捕获区域的合理设计，保证了各区域液相杂交捕获探针文库的均一性，可大幅提高检测灵敏度及准确性。

1. 一种用于梅毒病原体检测的液相杂交捕获文库,其特征在于,所述液相杂交捕获探针文库包括:梅毒基因组序列和人基因组参照序列;所述梅毒基因组序列每隔10-20bp截取50-70bp,得到探针序列;将所得探针序列与病原体数据库进行比对,选取其中特异比到梅毒上的序列作为液相杂交捕获文库的探针序列;所述基因组参照序列作为内参以保证液相杂交捕获效果,其选自人基因组保守序列。

2. 根据权利要求1所述的液相杂交捕获文库,其特征在于,所述梅毒基因组序列是GenBank号:CP003679.1的序列。

3. 根据权利要求1所述的液相杂交捕获文库,其特征在于,所述梅毒基因组序列每隔12bp截取65bp,得到探针序列;将所得探针序列与病原体数据库进行比对,选取其中特异比到梅毒上的序列作为液相杂交捕获文库的探针序列。

4. 根据权利要求1所述的液相杂交捕获文库,其特征在于,所述人基因组保守序列选取EFTUD2基因和/或SDHA基因部分序列作为人基因组参照序列;所述EFTUD2基因每隔10-20bp截取50-70bp,得到人基因组参照序列;所述SDHA基因每隔10-20bp截取50-70bp,得到人基因组参照序列。

5. 根据权利要求4所述的液相杂交捕获文库,其特征在于,选取EFTUD2基因和/或SDHA基因部分序列作为人基因组参照序列,所述EFTUD2基因每隔16bp截取65bp,得到人基因组参照序列;和/或所述SDHA基因每隔16bp截取65bp,得到人基因组参照序列。

6. 根据权利要求1-5任一所述的用于梅毒病原体检测的液相杂交捕获文库的构建方法,其特征在于,所述方法包括以下步骤:

步骤1:设计梅毒基因组序列和人基因组参照序列,所述梅毒基因组序列每隔10-20bp截取50-70bp,得到探针序列;将所得探针序列与病原体数据库进行比对,选取其中特异比到梅毒上的序列作为液相杂交捕获文库的探针序列;所述基因组参照序列作为内参以保证液相杂交捕获效果,其选自人基因组保守序列;

步骤2:将所述设计梅毒基因组序列和人基因组参照序列在其上下游分别加入引物池合成通用引物序列,以便后续PCR扩增制备探针所需;

步骤3:利用所述上下游通用序列引物进行PCR扩增制备得到带有生物素标记的液相杂交捕获探针文库。

7. 根据权利要求6所述的构建方法,其特征在于,所述梅毒基因组序列是GenBank号:CP003679.1的序列;和/或所述人基因组保守序列选取EFTUD2基因和/或SDHA基因部分序列作为人基因组参照序列;所述EFTUD2基因每隔10-20bp截取50-70bp,得到人基因组参照序列;所述SDHA基因每隔10-20bp截取50-70bp,得到人基因组参照序列。

8. 根据权利要求6所述的构建方法,其特征在于,所述引物池合成上游通用引物序列为SEQ No.1: TGTAATACGACTCACTATAGGGAGA;和所述引物池合成下游通用引物序列为SEQ No.2: TGTAATACGACTCACTATAGGGAGA;所述引物池合成的序列结构是: TGTAATACGACTCACTATAGGGAGANNNTGAGACGCGACCACGAA,其中其中加粗的NNNNN代表多种插入片段序列,即梅毒基因组序列和/或人基因组参照序列截取得到的探针序列。

9. 根据权利要求6所述的构建方法,其特征在于,通过PCR扩增制备得到带有生物素标记的液相杂交捕获探针文库步骤的PCR的反应体系中,dUTP与dTTP的比例为1:3时。

10. 根据权利要求6所述的构建方法,其特征在于,通过PCR扩增制备得到带有生物素标

记的液相杂交捕获探针文库步骤中PCR扩增后经过二次磁珠纯化回收得到液相杂交捕获探针文库。

用于梅毒病原体检测的液相杂交捕获文库及其构建方法

技术领域

[0001] 本发明属于分子生物学技术领域,涉及一种用于梅毒病原体检测的液相杂交捕获文库及其构建方法。

背景技术

[0002] 感染性疾病是目前引起人类疾病的主要原因之一。据预测,全球每年约4000万人感染,50万死亡。中国每年约1500万感染,20万死亡。对于感染性样本的检测,目前主要的检测方法还是免疫学/分子生物学和分离培养为主。临床感染患者表现类型多种多样,主要有血流感染、神经系统感染、呼吸道感染及消化道感染症候群,而引起人们感染的病原体多种多样,包括细菌、真菌、病毒、寄生虫、支原体、衣原体、立克次氏体、螺旋体等。很多病原体感染能够引起不同的临床表型,不同的病原体感染也能够引起同样的临床表现,这也决定了检测方法的多样性和复杂性。传统的病原微生物检测手段包括基于抗原抗体的免疫学方法和基于核酸检测的方法已经无法满足未知病原体快速检测的需要。因此,一种通用的快速方法对于急性感染患者来说尤为重要。而二代测序技术以能一次并行对几十万到几百万条DNA分子进行序列测定,通过生物信息学分析获得所测样本的基因序列信息。二代测序技术已经成为国内外新发突发病原微生物引起的疫情的首要确证手段,如我国2009年的蜱虫病,2009年开始的甲型H1N1流感和2013年H7N9流感。在世界范围内,还有2011年5月出现的德国大肠杆菌O104:H4疫情2011年7月出现的荷兰超级耐药肺炎克雷伯菌Oxa4疫情,以及2014年震惊世界的西非埃博拉病毒疫情。近年来,随着二代测序成本的降低,其已经走进了日常临床检验中,美国FDA和我国CFDA先后颁布了指导原则,解决目前临床急性感染和重症感染患者诊断难的问题。

[0003] 其中梅毒是一类典型的感染性疾病,梅毒是梅毒螺旋体引起的系统性疾病。一般将梅毒分为三期:一期梅毒(如感染部位溃疡或硬下疳)、二期梅毒(包括但不限于皮疹、皮肤黏膜病变及淋巴结病变)和三期梅毒(如心脏病变或树胶肿)。梅毒的发病率近年来逐年升高,是应该引起高度重视的乙类传染病。

[0004] 隐性梅毒,也称潜伏梅毒临床表现为无任何梅毒性的症状及体征,但其具有更高的传染风险。一期梅毒硬下疳和二期梅毒都具有自限性,且一般不留痕迹,如未能治疗或治疗不成功,则进入隐性梅毒阶段。

[0005] 而这其中神经性梅毒是较难诊断的一类,神经梅毒是梅毒螺旋体侵犯中枢神经系统所引起的慢性系统感染性疾病。病程可出现在梅毒的各个时期(包括早期梅毒),表现为颅神经功能障碍、脑膜炎、脑卒中、神经状态的急性病变、听力和视力下降等。

[0006] 因此,鉴于梅毒临床表现复杂,误诊率高,迫切需要一种能够快速、灵敏、便捷、准确检测梅毒的方法及试剂盒。

发明内容

[0007] 在一种实施方式中,本发明提供一种用于梅毒病原体检测的液相杂交捕获文库,

所述液相杂交捕获探针文库包括：梅毒基因组序列和人基因组参照序列；所述梅毒基因组序列每隔10-20bp截取50-70bp，得到探针序列；将所得探针序列与病原体数据库进行比对，选取其中特异比到梅毒上的序列作为液相杂交捕获文库的探针序列；所述基因组参照序列作为内参以保证液相杂交捕获效果，其选自人基因组保守序列。

[0008] 在一种实施方式中，所述梅毒基因组序列是GenBank号：CP003679.1的序列。

[0009] 在一种实施方式中，所述梅毒基因组序列每隔12bp截取65bp，得到探针序列；将所得探针序列与病原体数据库进行比对，选取其中特异比到梅毒上的序列作为液相杂交捕获文库的探针序列。

[0010] 在一种实施方式中，所述人基因组保守序列选取EFTUD2基因和/或SDHA基因部分序列作为人基因组参照序列；所述EFTUD2基因每隔10-20bp截取50-70bp，得到人基因组参照序列；所述SDHA基因每隔10-20bp截取50-70bp，得到人基因组参照序列。

[0011] 在一种实施方式中，选取EFTUD2基因和/或SDHA基因部分序列作为人基因组参照序列，所述EFTUD2基因每隔16bp截取65bp，得到人基因组参照序列；和/或所述SDHA基因每隔16bp截取65bp，得到人基因组参照序列。

[0012] 在一种实施方式中，本发明提供一种用于梅毒病原体检测的液相杂交捕获文库的构建方法，其特征在于，所述方法包括以下步骤：步骤1：设计梅毒基因组序列和人基因组参照序列，所述梅毒基因组序列每隔10-20bp截取50-70bp，得到探针序列；将所得探针序列与病原体数据库进行比对，选取其中特异比到梅毒上的序列作为液相杂交捕获文库的探针序列；所述基因组参照序列作为内参以保证液相杂交捕获效果，其选自人基因组保守序列；步骤2：将所述设计梅毒基因组序列和人基因组参照序列在其上下游分别加入引物池合成通用引物序列，以便后续PCR扩增制备探针所需；步骤3：利用所述上下游通用序列引物进行PCR扩增制备得到带有生物素标记的液相杂交捕获探针文库。

[0013] 在一种实施方式中，所述梅毒基因组序列是GenBank号：CP003679.1的序列；和/或所述人基因组保守序列选取EFTUD2基因和/或SDHA基因部分序列作为人基因组参照序列；所述EFTUD2基因每隔10-20bp截取50-70bp，得到人基因组参照序列；所述SDHA基因每隔10-20bp截取50-70bp，得到人基因组参照序列。

[0014] 在一种实施方式中，所述引物池合成上游通用引物序列为SEQ No.1：TGTAATACGACTCACTATAGGGAGA；和所述引物池合成下游通用引物序列为SEQ No.2：TGTAATACGACTCACTATAGGGAGA；所述引物池合成的序列结构是：TGTAATACGACTCACTATAGGGA GANNNTGAGACGCGACCACGAA，其中其中加粗的NNNN代表多种插入片段序列，即梅毒基因组序列和/或人基因组参照序列截取得到的探针序列。

[0015] 在一种实施方式中，通过PCR扩增制备得到带有生物素标记的液相杂交捕获探针文库步骤的PCR的反应体系中，dUTP与dTTP的比例为1:3时。

[0016] 在一种实施方式中，通过PCR扩增制备得到带有生物素标记的液相杂交捕获探针文库步骤中PCR扩增后经过二次磁珠纯化回收得到液相杂交捕获探针文库。

[0017] 本发明所构建的液相杂交捕获探针文库中序列合成技术高，合成序列准确率高，准确率能达到99.8%，保证了液相杂交捕获探针文库的准确性，进而保证较高的捕获效率表现。同时，对于目标捕获区域的合理设计保证了各区域液相杂交捕获探针文库的均一性，可大幅提高检测灵敏度及准确性。

[0018] 本发明提供了一种从人的血浆游离DNA样本中选择性富集梅毒病原体序列片段的方法,大大地提高了梅毒病原体检测灵敏度及准确性,从而提供了一种针对梅毒病原体检测的高效而低成本的检测试剂盒。

[0019] 本发明中,构建的液相杂交捕获探针文库可以同时杂交1-4个样品,样品可以通过本领域熟悉的二代建库方法先将待捕获样品分别进行测序文库构建,加上不同的双端index标记,将加上不同双端index的1-4个样品进行等量混合,然后用液相杂交捕获探针文库进行杂交捕获,最后将捕获的产物进行高通量测序,捕获的不同样品的目标区域后续分析时通过建库时加上的index标记进行区分。

具体实施方式

[0020] 为了使本领域技术领域人员更好地理解本申请中的技术方案,下面将结合实施例对本发明作进一步说明,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例,都应当属于本申请保护的范围。

[0021] 在本发明各实施方式中,为了使读者更好地理解本申请而提出了许多技术细节。但是,即使没有这些技术细节,也可以实现本申请各权利要求保护的技术方案。以下实施例的定量实验中,均是设置三次重复实验,结果取平均值。

[0022] 实施例1本发明的梅毒病原体检测液相杂交捕获探针文库构建

[0023] 具体而言,将本发明根据现有的或已获得的梅毒病原体基因组信息为基础,构建了针对梅毒病原体检测液相杂交捕获探针文库。

[0024] 用于液相杂交捕获的探针文库DNA序列要具有较高的特异性,不受其他同源核酸序列的影响。本发明按照以下方式获得梅毒病原体检测液相杂交捕获探针文库,具体描述如下:

[0025] 1. 探针设计:如上所述,梅毒病原体检测液相杂交捕获探针文库包含两部分:1. 梅毒基因组序列;2. 人基因组参照序列。其中梅毒基因组序列(GenBank号:CP003679.1),大小约1.14Mb,根据参考序列,每隔12bp截取65bp,共得到94792条探针序列,将得到的序列与病原体数据库进行比对,选取了其中92703条特异比到梅毒上的序列作为探针序列。同时选取人基因组保守序列作为内参以保证液相杂交捕获效果,选取EFTUD2基因、SDHA基因部分序列作为人基因组参照序列,EFTUD2基因(长约4.4k),每隔10-20bp截取50-70bp(约4x覆盖),共得到2762条探针序列;SDHA基因每隔10-20bp截取50-70bp,取其中535条序列。

[0026] 在设计好的探针片段在其上下游分别加入引物池(oligo pool)合成通用引物序列,以便后续PCR扩增制备探针所需,引物池合成通用引物序列所述引物池合成上游通用引物序列为SEQ No.1:TGTAATACGACTCACTATAGGGAGA;和所述引物池合成下游通用引物序列为SEQ No.2:TGTAATACGACTCACTATAGGGAGA;所述引物池合成的序列结构是:TGTAATACGACTCACTATAGGGAGANNNTGAGACGCGACCACGAA,其中加粗的NNNNN代表多种插入片段序列,即梅毒基因组序列和/或人基因组参照序列截取得到的探针序列。

[0027] 2. PCR制备液相杂交捕获探针文库:上述引物池合成成功后,将干粉状态的引物池用50ul洁净的DNase/RNase free ddH₂O溶解,充分震荡后离心。后利用上下游通用序列引物进行PCR扩增制备得到带有生物素标记的液相杂交捕获探针文库。

[0028] 具体的,对于上述技术方案,优选的情况下,所述PCR扩增体系为:

	5×GC buffer	10uL
	上下游引物 (10uM)	各 2uL
	dATP、dGTP、dCTP (10mM)	各 0.4uL
[0029]	dTTP (10mM)	0.3uL
	Biotin-11-dUTP (1mM)	1uL
	Thermo Phusion Taq	0.5uL
	模板 (引物池)	2uL (30ng)
	ddH ₂ O	31uL

[0030] PCR扩增条件如下:98℃、30s;98℃、10s,60℃、30s,72℃、30s,14个循环;72℃、7min;4℃、∞。

[0031] 上述操作涉及将生物素标记的Biotin-11-dUTP掺入合成的DNA产物中,通过优化PCR的反应体系,主要是优化dUTP与dTTP的比例,当dUTP:dTTP为1:3时,扩增带有Biotin-11-dUTP的PCR产物效果最佳。

[0032] PCR扩增后经过2×磁珠纯化回收得到液相杂交捕获探针文库。

[0033] 实施例2本发明构建了针对梅毒病原体检测液相杂交捕获的应用

[0034] 具体而言,本发明构建了针对梅毒病原体检测液相杂交捕获的检测试剂盒,试剂盒首先将提取到的血浆样本游离DNA进行样本文库构建,经过末端修复、加“A”及连接过程将DNA片段连上双端index接头,随后利用构建的液相杂交捕获探针文库与含有待捕获的样本文库DNA片段杂交,经过洗涤富集后对这些DNA片段进行测序,经过生物信息学数据分析后获得样本检测结果。该技术方案包括从血浆游离DNA到测序结果输出的全部实验流程,主要包括样本提取、文库构建、液相杂交捕获、测序及数据分析四方面内容。

[0035] 实验选取1例梅毒阳性患者的血浆游离DNA样本,对样本进行样本文库构建,实验组1利用上述梅毒病原体检测试剂盒进行液相杂交捕获后上机测序,实验组2不进行液相杂交捕获直接测序,得到测序结果对比。

[0036] 1. 血浆游离DNA样本提取及文库制备

[0037] 将采集好梅毒阳性血液样本的streack管经过3000g离心10min后分离血浆至EP管中,分离好的血浆样本管再经过14000rpm离心10min后取上清进行游离DNA提取。游离DNA提取采用天根游离DNA提取试剂盒TIANamp Micro DNA Kit,DP316。

[0038] 提取好的游离DNA样本分别进行不同index的样本文库构建,建库条件相同:

[0039] (1).利用酶反应将提取好的质量较高的cfDNA样本进行末端修复及加A(建库试剂来自商品化建库试剂盒KAPA Hyper Prep Kit Illumina platforms)。

	组分	用量
	End Repair & A-Tailing buffer	7 μL
[0040]	End Repair & A-Tailing Enzyme Mix	3 μL
	游离 DNA 样本	50 μL
	总计	60 μL

[0041] 反应条件:20℃30min,65℃30min,4℃ever。

[0042] (2).将上述经过末端修复、加A的cfDNA和双端adaptor接头进行连接反应。

组分	用量
Ligation buffer	30 μ L
DNA Ligase	10 μ L
[0043] 双端 adaptor	1 μ L
H2O	9 μ L
末端修复、加 A 产物	60 μ L
总计	110 μ L

[0044] 反应条件:20 $^{\circ}$ C15min。连接后产物经过0.8 \times 磁珠纯化后得到连接产物。

[0045] (3).将上一步的连接产物进行PCR扩增富集文库

组分	用量
2 \times KAPA HiFi HotStart ReadyMix	25 μ L
[0046] Oligo1 & 2, 5 μ M	5 μ L
连接产物	20 μ L
总计	50 μ L

[0047] 反应条件:98 $^{\circ}$ C45s,98 $^{\circ}$ C15s;60 $^{\circ}$ C30s;72 $^{\circ}$ C30s,10-13cycles,72 $^{\circ}$ C1min,4 $^{\circ}$ C ever。PCR后产物经过0.5-0.9 \times 磁珠筛选得到游离DNA样本文库。

[0048] 2.梅毒病原体序列的富集

[0049] 对实验组1的构建好的游离DNA样本进行液相杂交捕获,实验组2样本文库不做处理直接测序。

[0050] (1)游离DNA样本文库混合、封闭及干燥:将构建好的基因组DNA文库等量混合,总量1 μ g,随后将COT DNA、基因组DNA文库与封闭引物按照如下比例混合。其中COT DNA作为基因组中重复率较高的一部分DNA片段,在杂交时有助于提高杂交效率,封闭引物用来封闭文库中的测序接头。将上述混合好的样本在真空浓缩仪中60 $^{\circ}$ C蒸干,用于后续杂交。

组分	加入量
COT DNA	5 μ g
[0051] 游离 DNA 文库	1 μ g
样本文库封闭引物	2 μ mol
探针文库通用序列封闭引物	2 nmol

[0052] (2)重新溶解及变性:向上述干燥后的混合物中加入10.5 μ L的杂交缓冲液,充分涡旋溶解后于95 $^{\circ}$ C变性10min。

[0053] (3)杂交:将变性好的10.5 μ L混合物加入到4.5 μ L杂交捕获探针文库中,涡旋30s充分溶解混匀后全速离心30s。于55 $^{\circ}$ C杂交16h。

[0054] (4)生物亲和素磁珠捕获目标区域DNA文库片段:杂交反应后,将15 μ L样品转移至事先经过生物亲和素磁珠洗涤溶液洗涤重悬的100 μ L链霉亲和素标记磁珠中,混匀后于55 $^{\circ}$ C孵育45min,每隔15分钟吹打混匀一次,让捕获的片段与磁珠结合,特异吸附目标片段,将杂交到的片段抓取出来。

[0055] (5)洗涤非目标区域DNA片段:

- [0056] ①加入100 μ L 55 $^{\circ}$ C预热的洗涤溶液I,混匀后磁力悬浮弃上清,重复一次。
- [0057] ②加入200 μ L 55 $^{\circ}$ C预热的Stringent洗涤溶液,混匀后55 $^{\circ}$ C孵育5min,磁力悬浮弃上清,重复三次,共四次。
- [0058] ③加入200 μ L洗涤溶液I,涡旋2min,磁力悬浮弃上清。
- [0059] ④加入200 μ L洗涤溶液II,涡旋1min,磁力悬浮弃上清。
- [0060] ⑤加入200 μ L洗涤溶液III,涡旋30S,磁力悬浮弃上清。
- [0061] ⑥加入50 μ L PCR级别dd H₂O重悬磁珠。
- [0062] (6) 捕获后富集:洗脱后磁珠带着捕获的目标DNA片段,进入样本的LM-PCR富集。LM-PCR反应体系如下(一个捕获样本做两管PCR反应):

组分	用量
2 \times KAPA HiFi HotStart ReadyMix	25 μ L
[0063] Post-LM-PCR Oligos 1 & 2,5 μ M	5 μ L
带有捕获片段的磁珠溶液	20 μ L
总计	50 μ L

[0064] PCR反应条件为:98 $^{\circ}$ C预变性45s;98 $^{\circ}$ C变性15s,60 $^{\circ}$ C退火30s,72 $^{\circ}$ C延伸30s,共14个循环;72 $^{\circ}$ C延伸1min;4 $^{\circ}$ C保温。

[0065] 反应后,利用磁珠纯化PCR扩增产物,得到捕获后的样本文库,质检后用于测序分析。

[0066] 3.高通量测序与数据分析

[0067] 本发明采用本领域技术人员熟知的二代测序平台进行高通量测序,将捕获的序列在Illumina测序平台采用边合成边测序的方法进行序列测定(测序试剂均购自Illumina商业测序试剂盒),对得到的数据进行分析可知其上的标签序列,根据标签序列确定其对应的样本DNA来源。二代测序具有高通量的特点,为这一技术应用提供了快速高效的便利条件。本发明利用二代测序手段对上面方法获得的测序文库进行了测序并且获得了测序数据,有关分析结果如下:

[0068] 表1测序后数据分析结果

[0069]

实验组	RAW reads	Clean reads	Map reads	host	Map host Rate(%)	Map TP reads
1	35,259,381	11,979,136	11,761,176		98.18%	18,720
2	135, 357, 101	95,795,703	94,408,359		98.55%	42

[0070] 如表1所示,选取的梅毒阳性样本,实验组1利用上述梅毒病原体检测试剂盒进行液相杂交捕获后上机测序,实验组2不进行液相杂交捕获直接测序。测试数据分析结果显示,进行了液相杂交捕获实验的实验组1与直接建库测序的实验组2结果比较,实验组1检测到了18,720条梅毒reads,相比实验组2检测的数量(42)多了444倍。而测序数据产出量仅为后者的26%,因此利用本发明提供的梅毒病原体检测液相杂交捕获探针文库可大大提高梅

毒病原体检测的灵敏度与准确性,同时大大降低了检测成本,具有非常好地应用价值。

[0071] 实施例3本发明的不同探针长度对梅毒病原检测结果影响

[0072] 1. 探针设计:实验组1探针设计如上所述,梅毒病原体检测液相杂交捕获探针文库包含两部分:1. 梅毒基因组序列;2. 人基因组参照序列。其中梅毒基因组序列(GenBank号:CP003679.1),大小约1.14Mb,根据参考序列,每隔12bp截取65bp,共得到94792条探针序列,将得到的序列与病原体数据库进行比对,选取了其中92703条特异比到梅毒上的序列作为探针序列。同时选取人基因组保守序列作为内参以保证液相杂交捕获效果,选取EFTUD2基因、SDHA基因部分序列作为人基因组参照序列,EFTUD2基因(长约4.4k),每隔16bp截取65bp(约4x覆盖),共得到2762条探针序列;SDHA基因每隔16bp截取65bp,取其中535条序列。

[0073] 实验组2探针设计方法:根据梅毒基因组序列(GenBank号:CP003679.1),每隔15bp截取100bp,共得到63183条探针序列,经比对后选取62968条特异性比到梅毒上的序列作为探针序列;作为内参的人基因组保守基因EFTUD2和SDHA序列,每隔每隔15bp截取100bp,共得到2198条序列。

[0074] 在设计好的探针片段在其上下游分别加入引物池合成通用引物序列,以便后续PCR扩增制备探针所需,在设计好的探针片段在其上下游分别加入引物池(oligo pool)合成通用引物序列,以便后续PCR扩增制备探针所需,引物池合成通用引物序列所述引物池合成上游通用引物序列为SEQ No.1: TGTAATACGACTCACTATAGGGAGA;和所述引物池合成下游通用引物序列为SEQ No.2: TGTAATACGACTCACTATAGGGAGA;所述引物池合成的序列结构是: TGTAATACGACTCACTATAGGGAGANNNTGAGACGCGACCACGAA,其中加粗的NNNNN代表多种插入片段序列,即梅毒基因组序列和/或人基因组参照序列截取得到的探针序列。

[0075] 2. PCR制备液相杂交捕获探针文库:上述引物池合成成功后,将干粉状态的引物池用50ul洁净的DNase/RNase free ddH₂O溶解,充分震荡后离心。后利用上下游通用序列引物进行PCR扩增制备得到带有生物素标记的液相杂交捕获探针文库。

[0076] 具体的,对于上述技术方案,优选的情况下,所述PCR扩增体系为:

	5×GC buffer	10uL
	上下游引物(10uM)	各 2uL
	dATP、dGTP、dCTP(10mM)	各 0.4uL
[0077]	dTTP(10mM)	0.3uL
	Biotin-11-dUTP(1mM)	1uL
	Thermo Phusion Taq	0.5uL
	模板(引物池)	2uL(30ng)
	ddH ₂ O	31uL

[0078] PCR扩增条件如下:98℃、30s;98℃、10s,60℃、30s,72℃、30s,14个循环;72℃、7min;4℃、∞。

[0079] 上述操作涉及将生物素标记的Biotin-11-dUTP掺入合成的DNA产物中,通过优化PCR的反应体系,主要是优化dUTP与dTTP的比例,当dUTP:dTTP为1:3时,扩增带有Biotin-11-dUTP的PCR产物效果最佳。

[0080] PCR扩增后经过2×磁珠纯化回收得到液相杂交捕获探针文库。

[0081] 3. 梅毒阳性患者的血浆游离DNA样本检测:将提取到的血浆样本游离DNA进行样本

文库构建,经过末端修复、加“A”及连接过程将DNA片段连上双端index接头,随后利用构建的液相杂交捕获探针文库与含有待捕获的样本文库DNA片段杂交,经过洗涤富集后对这些DNA片段进行测序,经过生物信息学数据分析后获得样本检测结果。

[0082] (1) 血浆游离DNA样本提取及文库制备

[0083] 将采集好梅毒阳性血液样本的streck管经过3000g离心10min后分离血浆至EP管中,分离好的血浆样本管再经过14000rpm离心10min后取上清进行游离DNA提取。游离DNA提取采用天根游离DNA提取试剂盒TIANamp Micro DNA Kit,DP316。

[0084] 提取好的游离DNA样本分别进行不同index的样本文库构建,建库条件相同:

[0085] ①利用酶反应将提取好的质量较高的cfDNA样本进行末端修复及加A(建库试剂来自商品化建库试剂盒KAPA Hyper Prep Kit Illumina platforms)。

[0086]	组分	用量
	End Repair & A-Tailing buffer	7 μ L
	End Repair & A-Tailing Enzyme Mix	3 μ L
[0087]	游离 DNA 样本	50 μ L
	总计	60 μ L

[0088] 反应条件:20 $^{\circ}$ C30min,65 $^{\circ}$ C30min,4 $^{\circ}$ Cever。

[0089] ②将上述经过末端修复、加A的cfDNA和双端adaptor接头进行连接反应。

[0090]	组分	用量
	Ligation buffer	30 μ L
	DNA Ligase	10 μ L
	双端 adaptor	1 μ L
	H2O	9 μ L
	末端修复、加 A 产物	60 μ L
	总计	110 μ L

[0091] 反应条件:20 $^{\circ}$ C15min。连接后产物经过0.8 \times 磁珠纯化后得到连接产物。

[0092] ③将上一步的连接产物进行PCR扩增富集文库。

[0093]	组分	用量
	2 \times KAPA HiFi HotStart ReadyMix	25 μ L
	Oligo1 & 2, 5 μ M	5 μ L
	连接产物	20 μ L
	总计	50 μ L

[0094] 反应条件:98 $^{\circ}$ C45s,98 $^{\circ}$ C15s;60 $^{\circ}$ C30s;72 $^{\circ}$ C30s,10-13cycles,72 $^{\circ}$ C1min,4 $^{\circ}$ Cever。PCR后产物经过0.5-0.9 \times 磁珠筛选得到游离DNA样本文库。

[0095] (2) 梅毒病原体序列的富集

[0096] 将构建好的游离DNA样本分别与实验组1和实验组2制备得到的探针文库进行液相杂交捕获。

[0097] ①游离DNA样本文库混合、封闭及干燥:将构建好的基因组DNA文库等量混合,总量

1 μ g,随后将COT DNA、基因组DNA文库与封闭引物按照如下比例混合。其中COT DNA作为基因组中重复率较高的一部分DNA片段,在杂交时有助于提高杂交效率,封闭引物用来封闭文库中的测序接头。将上述混合好的样本在真空浓缩仪中60 $^{\circ}$ C蒸干,用于后续杂交。

组分	加入量
COT DNA	5 μ g
[0098] 游离 DNA 文库	1 μ g
样本文库封闭引物	2 μ mol
探针文库通用序列封闭引物	2 nmol

[0099] ②重新溶解及变性:向上述干燥后的混合物中加入10.5 μ L的杂交缓冲液,充分涡旋溶解后于95 $^{\circ}$ C变性10min。

[0100] ③杂交:将变性好的10.5uL混合物加入到4.5 μ L杂交捕获探针文库中,涡旋30s充分溶解混匀后全速离心30s。于55 $^{\circ}$ C杂交16h。

[0101] ④生物亲和素磁珠捕获目标区域DNA文库片段:杂交反应后,将15uL样品转移至事先经过生物亲和素磁珠洗涤溶液洗涤重悬的100uL链霉亲和素标记磁珠中,混匀后于55 $^{\circ}$ C孵育45min,每隔15分钟吹打混匀一次,让捕获的片段与磁珠结合,特异吸附目标片段,将杂交到的片段抓取出来。

[0102] ⑤洗涤非目标区域DNA片段:

[0103] i. 加入100 μ L 55 $^{\circ}$ C预热的洗涤溶液I,混匀后磁力悬浮弃上清,重复一次。

[0104] ii. 加入200 μ L 55 $^{\circ}$ C预热的Stringent洗涤溶液,混匀后55 $^{\circ}$ C孵育5min,磁力悬浮弃上清,重复三次,共四次。

[0105] iii. 加入200 μ L洗涤溶液I,涡旋2min,磁力悬浮弃上清。

[0106] iv. 加入200 μ L洗涤溶液II,涡旋1min,磁力悬浮弃上清。

[0107] v. 加入200 μ L洗涤溶液III,涡旋30S,磁力悬浮弃上清。

[0108] vi. 加入50 μ L PCR级别dd H₂O重悬磁珠。

[0109] ⑥捕获后富集:洗脱后磁珠带着捕获的目标DNA片段,进入样本的LM-PCR富集。LM-PCR反应体系如下(一个捕获样本做两管PCR反应):

组分	用量
2 \times KAPA HiFi HotStart ReadyMix	25 μ L
[0110] Post-LM-PCR Oligos 1 & 2,5 μ M	5 μ L
带有捕获片段的磁珠溶液	20 μ L
总计	50 μ L

[0111] PCR反应条件为:98 $^{\circ}$ C预变性45s;98 $^{\circ}$ C变性15s,60 $^{\circ}$ C退火30s,72 $^{\circ}$ C延伸30s,共14个循环;72 $^{\circ}$ C延伸1min;4 $^{\circ}$ C保温。

[0112] 反应后,利用磁珠纯化PCR扩增产物,得到捕获后的样本文库,质检后用于测序分析。

[0113] (3)高通量测序与数据分析

[0114] 本发明采用本领域技术人员熟知的二代测序平台进行高通量测序,将捕获的序列在Illumina测序平台采用边合成边测序的方法进行序列测定(测序试剂均购自Illumina商

业测序试剂盒),对得到的数据进行分析可知其上的标签序列,根据标签序列确定其对应的样本DNA来源。二代测序具有高通量的特点,为这一技术应用提供了快速高效的便利条件。本发明利用二代测序手段对上面方法获得的测序文库进行了测序并且获得了测序数据,有关分析结果如下:

[0115] 表2测序后数据分析结果

[0116]

实验组	RAW reads	Clean reads	Map reads	host	Map host Rate(%)	Map TP reads
1	35,259,381	11,979,136	11,761,176		98.18%	18,720

[0117]

2	35,766,325	12,076,113	11,961,390		99.05%	81
---	------------	------------	------------	--	--------	----

[0118] 如表2所示,对比了不同探针设计方案对梅毒病原体液相杂交捕获检测试验的影响,检测数据显示,实验组1探针液相杂交捕获结果显示检测到了18,720条梅毒reads,而在大致相同数据产出的情况下,实验组2探针液相杂交捕获检测结果检测到81条梅毒reads,实验组1相较实验组2检测到的梅毒reads多了230倍。由结果不难看出实验组1的探针设计方案在捕获效率上明显优于实验组2,大大提高了梅毒病原体检测灵敏度,因此本发明提供的液相杂交捕获探针文库可大大提高梅毒病原体检测的灵敏度与准确性,同时大大降低了检测成本,具有较好地应用价值。

[0119] 应该理解到披露的本发明不仅仅限于描述的特定的方法、方案和物质,因为这些均可变化。还应理解这里所用的术语仅仅是为了描述特定的实施方式方案的目的,而不是意欲限制本发明的范围,本发明的范围仅受限于所附的权利要求。

[0120] 本领域的技术人员还将认识到,或者能够确认使用不超过常规实验,在本文中所述的本发明的具体的实施方案的许多等价物。这些等价物也包含在所附的权利要求中。

序列表

<110> 北京京诺玛特科技有限公司

<120> 用于梅毒病原体检测的液相杂交捕获文库及其构建方法

<160> 2

<170> SIPOSequenceListing 1.0

<210> 1

<211> 25

<212> DNA

<213> 人工序列 (Artificial Sequence)

<400> 1

tgtaatacga ctcaactatag ggaga 25

<210> 2

<211> 25

<212> DNA

<213> 人工序列 (Artificial Sequence)

<400> 2

tgtaatacga ctcaactatag ggaga 25