

构建基因突变文库的方法

5 技术领域

本发明属于生物技术领域。具体地，本发明涉及构建基因突变文库的领域。

背景技术

10 基因突变体文库是大量 DNA 变体序列的组合，是基因合成、基因突变和定向进化研究相结合的产物。基因突变文库已经越来越多地应用于研究领域，如高通量药物靶点筛选，蛋白质工程定向进化，合成多样性抗体库用来筛选高亲和力和特异性的变异抗体等。基因突变体文库构建是利用分子定向进化模拟自然选择过程，改变原有蛋白的氨基酸序列，以期获得具有特定功能的突变蛋白。

15 分子定向进化是利用现代分子生物学方法创造大量的突变同源基因文库人为模拟自然进化机制，采用灵敏的定向筛选策略，创造出自然界不存在的或某些特性有显著改变的突变蛋白质或其他生物分子。分子定向进化已广泛应用于蛋白质的分子改造，被认为是改良蛋白质特性或调控序列的最高效方法。

基因突变文库可以改变蛋白质的特定位点或者 DNA 的调控区域，可以用来研究结构和功能之间的关系。根据不同设计目的，体外合成的文库可以更精确更随意的控制所研究序列的多样性。而且，合成的 DNA 文库可以显著提高获得核酸或蛋白质突变体的效率，有助于下游研究。

25 基因突变文库包括丙氨酸扫描文库、简并突变文库、三联体文库、定点饱和突变文库、随机突变文库等。

体外分子优化可非常有效地生成改良或新型的突变蛋白，识别调控序列，并可探测结构和功能的关键残基。用体外分子优化的方法构建合成文库对蛋白特性、调控和功能的系统性研究是非常有效的途径。

30 应用高通量半导体精准引物池合成构建突变文库理论所需的全部引物，能根据下游的不同表达系统做密码子优化，设置每个突变位置的不同突变

的分布率，从根本上杜绝非预期的或终止密码子的引入问题，文库只包含需要的突变体，为后续筛选节省时间和精力。

但是目前基因突变文库的构建方法步骤较复杂，耗时耗力。现有方法需要先将芯片合成的突变区域通过 PCR 进一步扩增，然后以需要突变的基因为模板，扩增突变区域前后的 2 个非突变区域，然后通过组装 PCR 将 3 个片段组装成一个片段，最后以组装产物为模板，使用基因首尾引物扩增全长片段。此外，每个步骤之后还必须进行纯化。

本发明描述了一种基因突变文库的构建方法，能通过较为简单的步骤完成基因突变文库的构建，大大的节约了基因突变文库的构建的成本。

此外，目前现有的构建基因突变文库的方法产生的基因突变文库还存在突变分布不符合预期的问题。

本发明的构建基因突变文库的方法还可以通过简单的优化反应条件而实现所获得的基因突变文库中的包含各种突变的序列的比例更接近所需要的比例。

发明内容

基因突变文库对于蛋白质工程和抗体药工程等非常重要，利用高通量半导体精准引物池合成能为后续的基因突变文库的构建以及筛选节省大量的时间和精力。但是目前基因突变文库的构建方法步骤较复杂。

本发明提供了一种新的构建基因突变文库的方法，其能通过较为简单的步骤完成基因突变文库的构建，解决了基因突变文库的构建步骤复杂的问题，同时节约了构建基因突变文库的成本。

第一方面，本发明提供了一种构建基因突变文库的方法，所述方法包括：

(1) 合成在一个或多个突变位点上包含突变核苷酸的寡核苷酸序列池；

(2) 对合成的包含突变核苷酸的寡核苷酸序列池中的寡核苷酸进行 PCR 扩增；

(3) 以包含已甲基化的位点且包含待突变的基因或其部分的质粒为模板，以步骤(2)中获得的扩增后的寡核苷酸序列池中的寡核苷酸为引物，进行 PCR 扩增；

(4) 用识别并切割所述已甲基化的位点的核酸内切酶切割步骤(3)的扩增后的体系中存在的模板质粒；和

(5) 向步骤(4)获得的体系中添加分别对应于待突变的基因或其部分的两端的正向引物和反向引物, 进行 PCR 扩增, 以得到包含多个含有突变核苷酸的基因或其部分的基因突变文库。

5 在一些实施方案中, 本发明的构建基因突变文库的方法的步骤(1)-步骤(5)之间不进行任何纯化步骤。

在一些实施方案中, 本发明的构建基因突变文库的方法进一步包括(6)对步骤(5)的扩增产物进行回收和纯化, 以得到最终产物。在一些实施方案中, 步骤(6)包括对步骤(5)的扩增产物进行琼脂糖凝胶电泳, 然后对其进行切胶回收并纯化, 优选使用 DNA 凝胶回收试剂盒进行回收和纯化。

10 在一些实施方案中, 本发明的构建基因突变文库的方法进一步包括(6)对步骤(5)的扩增产物进行回收和纯化, 以得到最终产物; 和(7)对步骤(6)中获得的最终产物进行测序, 以验证序列和/或检测基因突变分布。在一些实施方案中, 该测序为 Sanger 测序和/或高通量测序。

15 在一些实施方案中, 本发明的构建基因突变文库的方法的步骤(1)中, 合成的寡核苷酸序列池中的寡核苷酸序列的长度为 60-170 bp, 例如, 60、70、80、90、100、110、120、130、140、150、160 或 170 bp, 优选为 80-90 bp, 例如 80 或 90。

20 在一些实施方案中, 本发明的构建基因突变文库的方法的步骤(1)中, 在步骤(1)中合成的寡核苷酸序列池中的寡核苷酸在 1 至 10 个突变位点上包含突变核苷酸, 例如在 1、2、3、4、5、6、7、8、9 或 10 个突变位点上包含突变核苷酸。在一些实施方案中, 本发明的构建基因突变文库的方法的步骤(1)中, 在步骤(1)中合成的寡核苷酸序列池中的寡核苷酸在 2 至 10 个突变位点上包含突变核苷酸。在一些实施方案中, 该 2 至 10 个突变位点是相邻的或不相邻的。

25 在一些实施方案中, 本发明的构建基因突变文库的方法中, 其中所述含有突变核苷酸的基因编码的氨基酸序列与所述待突变的基因编码的氨基酸序列具有至少一个氨基酸差异。在一些实施方案中, 所述至少一个氨基酸差异选自下组: 取代、添加或缺失。优选地, 所述至少一个氨基酸差异是取代。在一些实施方案中, 所述取代是用与原氨基酸具有不同性质的氨基酸取代原氨基酸。30 优选地, 所述取代是用与原氨基酸具有不同性质的氨基酸取代原氨基酸, 其中所述性质选自下组: 酸碱性、极性、荷电性和侧链基团。按照氨

基酸的酸碱性，氨基酸分为酸性氨基酸(赖氨酸、精氨酸、组氨酸)和碱性氨基酸(天冬氨酸、谷氨酸)。按照侧链的极性，氨基酸分为非极性氨基酸(丙氨酸、缬氨酸、亮氨酸、异亮氨酸、脯氨酸、苯丙氨酸、色氨酸和蛋氨酸)和极性氨基酸，其中极性氨基酸分为极性不带电荷氨基酸(甘氨酸、丝氨酸、苏氨酸、半胱氨酸、酪氨酸、天冬氨酸、谷氨酰胺)、极性带正电荷的氨基酸(赖氨酸、精氨酸和组氨酸)、极性带负电荷的氨基酸(天冬氨酸、谷氨酸)。根据侧链基团的结构，氨基酸分为脂肪族氨基酸(丙氨酸、缬氨酸、亮氨酸、异亮氨酸、蛋氨酸、天冬氨酸、谷氨酸、赖氨酸、精氨酸、甘氨酸、丝氨酸、苏氨酸、天冬酰胺、半胱氨酸、谷氨酰胺)、芳香族氨基酸(苯丙氨酸、酪氨酸)、杂环族氨基酸(组氨酸、色氨酸)、杂环亚氨基酸(脯氨酸)。

在一些实施方案中，本发明的构建基因突变文库的方法的步骤(1)中，合成的寡核苷酸序列池中的不同突变寡核苷酸的突变核苷酸可以编码与待突变的基因的该位点的密码子编码的氨基酸不同的氨基酸的突变密码子，例如，待突变的基因中该待突变的位点为丙氨酸，则该合成的寡核苷酸序列池中的不同序列在该位点的核苷酸密码子可以是编码不同的常见氨基酸的密码子，例如编码甘氨酸、缬氨酸、亮氨酸、异亮氨酸、脯氨酸、苯丙氨酸、酪氨酸、色氨酸、丝氨酸、苏氨酸、半胱氨酸、蛋氨酸、天冬酰胺、谷氨酰胺、天冬氨酸、谷氨酸、赖氨酸、精氨酸、和组氨酸的密码子。作为另一个例子，特定突变的核苷酸密码子在突变基因文库中的不同序列中可以编码与待突变位点的原密码子编码的氨基酸具有不同性质的氨基酸。例如，待突变的基因中该待突变的位点为编码非极性氨基酸丙氨酸的密码子，则寡核苷酸序列池中的不同寡核苷酸序列中的该突变位点可以是编码极性氨基酸的密码子，例如编码色氨酸、丝氨酸、酪氨酸、半胱氨酸、蛋氨酸、天冬酰胺、谷氨酰胺或苏氨酸的密码子。

本发明的构建基因突变文库的方法的步骤(2)的目的是进一步扩大芯片合成的寡核苷酸的拷贝数。

在一些实施方案中，本发明的构建基因突变文库的方法的步骤(2)中，PCR 扩增使用分别对应于所述在一个或多个突变位点上包含突变核苷酸的寡核苷酸序列的两端的两个引物。

在本发明的构建基因突变文库的方法的步骤(2)中，尽量减少 PCR 扩增的循环数，以尽量避免影响突变点密码子编码的氨基酸的分布情况。PCR 循环

数越多，将指数性地扩大突变点密码子编码的氨基酸的频率差异。优选地，PCR 扩增的循环数为 10-30 个循环，可以为 10 至 28 个循环、10 至 26 个循环、10 至 28 个循环、10 至 25 个循环、10 至 22 个循环、10 至 20 个循环、10 至 18 个循环、10 至 16 个循环、12 至 30 个循环、12 至 28 个循环、12 至 26 个循环、12 至 24 个循环、12 至 22 个循环、12 至 20 个循环、12 至 18 个循环、12 至 16 个循环、12 至 14 个循环、14 至 30 个循环、14 至 28 个循环、14 至 26 个循环、14 至 24 个循环、14 至 22 个循环、14 至 20 个循环、14 至 18 个循环、14 至 16 个循环、16 至 30 个循环、16 至 28 个循环、16 至 26 个循环、16 至 24 个循环、16 至 22 个循环、16 至 20 个循环、16 至 18 个循环，更优选为 10 至 20 个循环，甚至更优选为 12 至 22 个循环、14 至 16 个循环，例如，10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30 个循环，最优选为 16 个循环。

在本发明的构建基因突变文库的方法的步骤(2)中，DNA 聚合酶可为任意 DNA 聚合酶，优选为高保真 DNA 聚合酶以减少扩增错误，更优选为高保真的 Phusion DNA 聚合酶。DNA 聚合酶还可为 PrimeSTAR HS DNA Polymerase 和 Q5® High-Fidelity DNA Polymerase 等高保真 DNA 聚合酶。

在本发明的构建基因突变文库的方法的步骤(3)中，以包含已甲基化的位点的待突变的基因或其部分的质粒为模板，以步骤(2)中获得的扩增后的寡核苷酸序列池中的寡核苷酸为引物，进行 PCR 扩增。该步骤的目的是为了获得在突变位点包含突变核苷酸的基因或其部分的序列。反应体系中没有其他引物，向前扩增的引物缺少对应的正向引物，向后扩增的引物缺少对应的反向引物。该质粒中含有需要被突变而未突变的全长基因或其部分的序列。步骤(2)中获得的扩增后的寡核苷酸序列池中的寡核苷酸序列为双链 DNA，在 PCR 的变性阶段该双链 DNA 解离为两条 DNA 单链，这两条 DNA 单链分别作为各自 PCR 反应的引物，在各自的 PCR 反应中以包含需要被突变的基因或其部分的质粒为模板分别向前后两端进行线性扩增。线性扩增是将 DNA 分子线性放大，最终得到的 DNA 分子数量远远小于指数扩增，引物需求量也远小于指数扩增。因此，在步骤(3)的线性扩增时需要严格控制引物的量和体系中模板的量，并且尽量减少 PCR 反应的循环数。

在一些实施方案中，在本发明的构建基因突变文库的方法的步骤(3)中，需要严格控制引物即步骤(2)中获得的扩增后的寡核苷酸序列池中的寡核苷

酸的量和体系中模板即包含已甲基化的位点的待突变的基因或其部分的质粒的量。发明人发现，当引物与模板的摩尔比为 20:1 时，扩增效果最好。在一些实施方案中，引物与模板的摩尔比为 15:1 至 100:1，优选为 15:1 至 90:1、15:1 至 80:1、15:1 至 70:1、15:1 至 60:1、15:1 至 50:1、15:1 至 40:1、15:1 至 30:1、15:1 至 28:1、15:1 至 26:1、15:1 至 24:1、15:1 至 22:1、15:1 至 20:1、16:1 至 100:1；优选为 16:1 至 90:1、16:1 至 80:1、16:1 至 70:1、16:1 至 60:1、16:1 至 50:1、16:1 至 40:1、16:1 至 30:1、16:1 至 28:1、16:1 至 26:1、16:1 至 24:1、16:1 至 22:1、16:1 至 20:1；优选为 17:1 至 90:1、17:1 至 80:1、17:1 至 70:1、17:1 至 60:1、17:1 至 50:1、17:1 至 40:1、17:1 至 30:1、17:1 至 28:1、17:1 至 26:1、17:1 至 24:1、17:1 至 22:1、17:1 至 20:1；优选为 18:1 至 90:1、18:1 至 80:1、18:1 至 70:1、18:1 至 60:1、18:1 至 50:1、18:1 至 40:1、18:1 至 30:1、18:1 至 28:1、18:1 至 26:1、18:1 至 24:1、18:1 至 22:1、18:1 至 20:1；优选为 19:1 至 90:1、19:1 至 80:1、19:1 至 70:1、19:1 至 60:1、19:1 至 50:1、19:1 至 40:1、19:1 至 30:1、19:1 至 28:1、19:1 至 26:1、19:1 至 24:1、19:1 至 22:1、19:1 至 20:1；优选为 20:1 至 90:1、20:1 至 80:1、20:1 至 70:1、20:1 至 60:1、20:1 至 50:1、20:1 至 40:1、20:1 至 30:1、20:1 至 28:1、20:1 至 26:1、20:1 至 24:1、20:1 至 22:1，最优选为 20:1。

在本发明的构建基因突变文库的方法的步骤(3)中，尽量减少 PCR 扩增的循环数，以尽量避免影响突变点核苷酸编码的氨基酸的分布情况。PCR 循环数越多，将指数性地扩大突变点编码的氨基酸的频率差异。优选地，PCR 扩增的循环数为 10-30 个循环，可以为 10 至 28 个循环、10 至 26 个循环、10 至 28 个循环、10 至 25 个循环、10 至 22 个循环、10 至 20 个循环、10 至 18 个循环、10 至 16 个循环、12 至 30 个循环、12 至 28 个循环、12 至 26 个循环、12 至 24 个循环、12 至 22 个循环、12 至 20 个循环、12 至 18 个循环、12 至 16 个循环、14 至 30 个循环、14 至 28 个循环、14 至 26 个循环、14 至 26 个循环、14 至 24 个循环、14 至 22 个循环、14 至 20 个循环、14 至 18 个循环、14 至 16 个循环、16 至 30 个循环、16 至 28 个循环、16 至 26 个循环、16 至 24 个循环、16 至 22 个循环、16 至 20 个循环、16 至 18 个循环，更优选为 10 至 20 个循环，甚至更优选为 12 至 22 个循环、14 至 16 个循环，例如，10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30 个循环，最优选为 15 个循环。

在本发明的构建基因突变文库的方法的步骤(3)中, 根据待扩增的基因或其部分的长度, 通过控制 PCR 反应的延伸时间, 使得以步骤(2)中获得的扩增后的寡核苷酸序列池中的寡核苷酸的两条链分别作为引物的两个 PCR 反应能够分别延伸至该基因或其部分的首尾两端。

5 在本发明的构建基因突变文库的方法的步骤(4)中, 用识别并切割所述已甲基化的位点的核酸内切酶切割步骤(3)的扩增后的体系中存在的模板质粒, 其目的是为了避免不包含突变的质粒作为下一步的模板而扩增得到不含突变的最终基因产物。由于提取自细菌的质粒都具有甲基化位点, 而步骤(3)中 PCR 扩增得到的产物不含甲基化位点, 因此使用识别并切割该甲基化位点的核酸内切酶能够选择性地将质粒切割成小片段, 从而不能作为下
10 一步 PCR 反应的模板。

在一些实施方案中, 在本发明的构建基因突变文库的方法中, 该质粒是从细菌中提取的。在一些实施方案中, 该质粒是从大肠杆菌中提取的。在一些实施方案中, 该质粒是从大肠杆菌 TOP10 中提取的。细菌中普遍存在甲基化酶, 以甲基化细菌自身的 DNA。例如, 大肠杆菌 TOP10 中存在
15 Dam 甲基化酶。Dam 甲基化酶识别并且甲基化 GATC 基序中的腺嘌呤。

在一些实施方案中, 在本发明的构建基因突变文库的方法的步骤(4)中, 识别并切割已甲基化的位点的核酸内切酶是本领域熟知的。例如, 识别并切割已甲基化的位点的核酸内切酶为 *DpnI*、*MspII* 或 *FspEI* 等。优选地, 该
20 识别并切割甲基化位点的核酸内切酶为 *DpnI*。限制性内切酶 *DpnI* 可有效识别并切断腺嘌呤甲基化的 G^mATC, 而不能切断非甲基化的 GATC 序列。提取自 Dam⁺细菌的质粒都含有 G^mATC, 可作为 *DpnI* 的底物。优选地, 可以被 *DpnI* 消化的质粒提取自大肠杆菌 TOP10 菌株。

在一些实施方案中, 在本发明的构建基因突变文库的方法中, 具有已
25 甲基化的位点的质粒为 pMB1、pBR322、pcDNA3.1 或 pUC57。在一些实施方案中, 具有已甲基化的位点的质粒为 pMB1。

在一些实施方案中, 在本发明的构建基因突变文库的方法的步骤(5)中, 向步骤(4)获得的体系中添加分别对应于待突变的基因或其部分的两端的正向引物和反向引物, 进行 PCR 扩增, 以得到包含多个含有突变核苷酸的基因或其部分的基因突变文库。其中该 PCR 反应的模板为通过碱基互补配对而连接在一起的两条具有不同方向的步骤(3)中获得的 PCR 扩增产物。
30

在本发明的构建基因突变文库的方法的步骤(5)中, 尽量减少 PCR 扩增的循环数, 以尽量避免影响突变点密码子编码的氨基酸的分布情况。PCR 循环数越多, 将指数性地扩大突变点氨基酸的频率差异。优选地, PCR 扩增的循环数为 10-30 个循环, 可以为 10 至 28 个循环、10 至 26 个循环、10 至 28 个循环、10 至 25 个循环、10 至 22 个循环、10 至 20 个循环、10 至 18 个循环、10 至 16 个循环、12 至 30 个循环、12 至 28 个循环、12 至 26 个循环、12 至 24 个循环、12 至 22 个循环、12 至 20 个循环、12 至 18 个循环、12 至 16 个循环、12 至 14 个循环、14 至 30 个循环、14 至 28 个循环、14 至 26 个循环、14 至 24 个循环、14 至 22 个循环、14 至 20 个循环、14 至 18 个循环、14 至 16 个循环、16 至 30 个循环、16 至 28 个循环、16 至 26 个循环、16 至 24 个循环、16 至 22 个循环、16 至 20 个循环、16 至 18 个循环, 更优选为 10 至 20 个循环, 甚至更优选为 12 至 22 个循环、14 至 16 个循环, 例如, 10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30 个循环, 最优选为 16 个循环。

在本发明的构建基因突变文库的方法的步骤(5)之后, 优选地, 该步骤还包括对扩增产物进行琼脂糖凝胶进行电泳、回收和/或纯化, 以得到最终产物。

进一步地, 本发明的构建基因突变文库的方法还包括步骤(6): 对步骤(5)中获得的最终产物进行测序, 以验证序列是否正确。优选地, 该测序为 Sanger 测序, 包括对最终产物进行平末端连接, 将其连接至线性化 Puc57-EV, 使用 M13F 引物对最终产物进行测序。优选地, 该测序为高通量测序或下一代测序(next generation sequencing technology, 简称 NGS), 以验证序列是否正确, 并且还可以检测每个突变位点的各种不同核苷酸突变分别占的比例, 以验证是否与所需要的比例一致。

第二方面, 本发明提供了一种基因突变文库, 其是使用本发明所述的构建基因突变文库的方法构建的。

第三方面, 本发明提供了一种用于根据上述的构建基因突变文库的方法的试剂盒, 其包含在一个或多个突变位点上包含突变核苷酸的寡核苷酸序列池、包含已甲基化的位点且包含待突变核苷酸的基因或其部分的质粒、和识别并切割所述已甲基化的位点的核酸内切酶。在一些实施方案中, 所述试剂盒还包含 DNA 聚合酶。

第四方面，本发明提供了一种解析蛋白质中的氨基酸突变与蛋白质的特性、调控和/或功能之间的关系的方法，其包括以下步骤：

(1) 使用根据本发明所述的方法构建编码所述蛋白质的基因的基因突变文库；

5 (2) 将所述基因突变文库中的突变基因编码的蛋白质的特征、调控和/或功能与未突变的蛋白质的特征、调控和/或功能进行比较；和

(3) 解析突变位置的氨基酸与蛋白质的特性、调控和/或功能的关系。

具体地，本发明涉及如下各项：

10 1. 一种构建基因突变文库的方法，所述方法包括：

(1) 合成在一个或多个突变位点上包含突变核苷酸的寡核苷酸序列池；

(2) 对合成的包含突变核苷酸的寡核苷酸序列池中的寡核苷酸进行 PCR 扩增；

15 (3) 以包含已甲基化的位点且包含待突变的基因或其部分的质粒为模板，以步骤(2)中获得的扩增后的寡核苷酸序列池中的寡核苷酸为引物，进行 PCR 扩增；

(4) 用识别并切割所述已甲基化的位点的核酸内切酶切割步骤(3)的扩增后的体系中存在的模板质粒；和

20 (5) 向步骤(4)获得的体系中添加分别对应于待突变核苷酸的基因或其部分的两端的正向引物和反向引物，进行 PCR 扩增，以得到包含多个含有突变的基因或其部分的基因突变文库。

2. 根据项 1 所述的方法，其中在步骤(1)-步骤(5)之间不进行纯化。

3. 根据项 1 所述的方法，其进一步包括：

(6) 对步骤(5)的扩增产物进行回收和纯化，以得到最终产物。

25 4. 根据项 3 所述的方法，其进一步包括：

(7) 对步骤(6)中获得的最终产物进行测序，以验证序列和/或检测基因突变分布。

5. 根据项 4 所述的方法，其中所述测序为 Sanger 测序和/或高通量测序。

30 6. 根据项 1-5 中任一项所述的方法，其中在步骤(1)中合成的寡核苷酸序列池中的寡核苷酸序列的长度为 60-170 bp，优选为 80-90 bp。

7. 根据项 1-6 中任一项所述的方法，其中在步骤(1)中合成的寡核苷酸序

列池在 1 至 10 个突变位点上包含突变核苷酸。

8. 根据项 1-7 中任一项所述的方法，其中所述突变位点是相邻或不相邻的。

9. 根据项 1-8 中任一项所述的方法，其中所述含有突变核苷酸的基因编码的氨基酸序列与所述待突变的基因编码的氨基酸序列具有至少一个氨基酸差异。

10. 根据项 9 所述的方法，其中所述至少一个氨基酸差异选自下组：取代、添加或缺失。

11. 根据项 10 所述的方法，其中所述至少一个氨基酸差异是取代。

12. 根据项 11 所述的方法，其中所述取代是用与原氨基酸具有不同性质的氨基酸取代原氨基酸。

13. 根据项 12 所述的方法，其中所述性质选自下组：酸碱性、极性、荷电性和侧链基团。

14. 根据项 1-13 中任一项所述的方法，其中步骤(2)中的 PCR 扩增使用分别对应于所述在一个或多个突变位点上包含突变核苷酸的寡核苷酸序列的两端的两个引物。

15. 根据项 1-14 中任一项所述的方法，其中步骤(2)中 PCR 扩增的循环数为 10-25 个循环，优选为 12-22 个循环、14-20 个循环、15-18 个循环或 15-16 个循环，最优选为 16 个循环。

16. 根据项 1-15 中任一项所述的方法，其中步骤(2)的 PCR 扩增使用的 DNA 聚合酶为高保真 DNA 聚合酶。

17. 根据项 1-16 中任一项所述的方法，其中步骤(3)中引物与模板的摩尔比为 15:1 至 50:1，优选为 16:1 至 40:1，17:1 至 30:1，18:1 至 25:1，或 19:1 至 22:1，更优选为 19:1 至 21:1，最优选为 20:1。

18. 根据项 1-17 中任一项所述的方法，其中步骤(3)中 PCR 扩增的循环数为 10-20 个循环，优选为 12-18 个循环、14-16 个循环，最优选为 15 个循环。

19. 根据项 1-18 中任一项所述的方法，其中步骤(3)的 PCR 扩增使用的 DNA 聚合酶为高保真 DNA 聚合酶，优选为高保真的 Phusion DNA 聚合酶。

20. 根据项 1-19 中任一项所述的方法，其中所述质粒提取自使质粒甲基化的细菌，优选提取自大肠杆菌，更优选提取自大肠杆菌 TOP10 菌株。

21. 根据项 1-20 中任一项所述的方法，其中步骤(4)中的核酸内切酶为

DpnI、*MspII* 或 *FspEI*。

22. 一种基因突变文库，其是使用根据项 1-21 中任一项所述的方法构建的。

23. 一种用于根据项 1-21 中任一项所述的构建基因突变文库的方法的试剂盒，其包含在一个或多个突变位点上包含突变核苷酸的寡核苷酸序列池、包含已甲基化的位点且包含待突变的基因或其部分的质粒、和识别并切割所述已甲基化的位点的核酸内切酶。

24. 根据项 23 所述的试剂盒，其中所述试剂盒还包含 DNA 聚合酶。

25. 一种解析蛋白质中的氨基酸突变与蛋白质的特性、调控和/或功能之间的关系的方法，其包括以下步骤：

10 (1) 使用根据项 1-21 中任一项所述的方法构建编码所述蛋白质的基因的基因突变文库；

(2) 将所述基因突变文库中的突变基因编码的蛋白质的特征、调控和/或功能与未突变的蛋白质的特征、调控和/或功能进行比较；和

(3) 解析突变位置的氨基酸与蛋白质的特性、调控和/或功能的关系。

15

附图说明

图 1 是本发明的基因突变文库构建方法的流程示意图。

图 2 是对合成的包含突变核苷酸的寡核苷酸序列池中的寡核苷酸进行 PCR 扩增获得的产物的电泳图。右侧泳道为 DL3000 标记物，左侧泳道为用 20 Chip-F 和 Chip-R 为引物，以合成的包含突变核苷酸的寡核苷酸序列池为模板的扩增产物。

图 3 是实施例 5 获得的最终产物的电泳图。左侧泳道为 DL3000 标记物，右侧泳道为用 Gene-R 和 Gene-F 为引物在实施例 3 中分别以合成的寡核苷酸的两条链为引物扩增获得的正义和反义 PCR 扩增产物单链 DNA 通过互补区 25 连接在一起作为模板扩增获得的产物经琼脂糖凝胶电泳、回收、纯化获得的最终产物。

图 4A-4C 是对最终产物的 Sanger 测序的结果。在目标突变位点，即在第 35 位氨基酸的密码子位置亦即第 103-105 位核苷酸，发生了突变。

图 5 是测试以合成的寡核苷酸为引物进行线性扩增步骤中模板与引物的 30 摩尔比为 1:254，PCR 反应的循环数为 25 的最终产物电泳图。左侧泳道为 DL3000 标记物，右侧泳道为用 Gene-R 和 Gene-F 为引物，以在实施例 7 中

分别以合成的寡核苷酸的两条链为引物扩增获得的正义和反义 PCR 扩增产物单链 DNA 通过互补区连接在一起作为模板，扩增获得的最终产物。

图 6 测试以合成的寡核苷酸为引物进行线性扩增步骤中模板与引物的摩尔比为 1:254，PCR 反应的循环数为 15。左侧泳道为 DL3000 标记物，右侧泳道为用 Gene-R 和 Gene-F 为引物，以在实施例 7 中分别以合成的寡核苷酸的两条链为引物扩增获得的正义和反义 PCR 扩增产物单链 DNA 通过互补区连接在一起作为模板，扩增获得的最终产物。

发明详述

如本说明书和所附权利要求书中使用的，单数形式“一个”，“一种”和“该/所述”包括复数提及物，除非上下文明确另有规定。如此，例如，提及“一个/种分子”任选包括两个/种或更多个/种此类分子的组合，诸如此类。

如本文中使用的，术语“约”指技术领域技术人员容易知道的相应数值的常规误差范围。本文中提及“约”某个数值或参数包括(并描述)涉及该数值或参数本身的实施方案。

理解的是，本文中描述的本发明的方面和实施方案包括“包含”，“由……组成”，和“基本上由……组成”的方面和实施方案。

术语“寡核苷酸”是指短链核苷酸的总称。在本文中，寡核苷酸可由 2-300 个核苷酸通过磷酸二酯键聚合而成。

术语“寡核苷酸序列池”是指由多种不同寡核苷酸序列组成的混合物。

术语“基因突变文库”是指由多种不同 DNA 变体序列组成的混合物。基因突变文库是基因合成、基因突变和定向进化研究相结合的产物。基因突变文库已经越来越多地应用于研究领域，如高通量药物靶点筛选，蛋白质工程定向进化，合成多样性抗体库用来筛选高亲和力和特异性的变异抗体等。

术语“PCR 扩增”是指聚合酶链式反应，为用于扩增特定的 DNA 片段的分子生物学技术，它可看作是生物体外的特殊 DNA 复制，PCR 的最大特点是能将微量的 DNA 大幅增加。具体而言，PCR 技术的基本原理类似于 DNA 的天然复制过程，其特异性依赖于与靶序列两端互补的寡核苷酸引物。PCR 由变性-退火-延伸三个基本反应步骤构成：①模板 DNA 的变性：模板 DNA 经加热至 90℃ 以上一定时间后，使模板 DNA 双链或经 PCR 扩增形成的双链 DNA 解离，使之成为单链，以便它与引物结合，为下轮反应作准备；

②模板 DNA 与引物的退火(复性): 模板 DNA 经加热变性成单链后, 温度降至较低温度, 引物与模板 DNA 单链的互补序列配对结合; ③引物的延伸: DNA 模板-引物结合物在约 70℃在 DNA 聚合酶(如 TaqDNA 聚合酶)的作用下, 以 dNTP 为反应原料, 靶序列为模板, 按碱基互补配对与半保留复制原理, 合成一条新的与模板 DNA 链互补的半保留复制链。重复循环变性-退火-延伸三过程就可获得更多的“半保留复制链”。

在本申请的构建基因突变文库的方法中的步骤(5)中, 向步骤(4)获得的体系中添加分别对应于待突变的基因或其部分的两端的正向引物和反向引物, 进行 PCR 扩增, 以得到包含多个含有突变核苷酸的基因或其部分的基因突变文库。在此步骤中, 首先进行融合 PCR, 即①模板 DNA 的变性: 模板 DNA 经加热至 90℃以上一定时间后, 使模板 DNA 保持原有的单链状态, 其中该模板 DNA 为步骤(3)中以含有待突变基因的质粒为模板以步骤(2)扩增的寡核苷酸池中的寡核苷酸为引物进行 PCR 扩增的产物, 其在对应于步骤(1)中合成的寡核苷酸池中的寡核苷酸序列位置处互补; ②退火: 模板 DNA 之间进行退火, 模板 DNA 单链的互补序列配对结合; ③延伸: DNA 聚合酶分别以互补的链为模板, 从 3'端开始延伸 DNA 链, 从而形成完整的 DNA 双链。之后的 PCR 反应与普通 PCR 相同, 以形成的完整的 DNA 双链为模板, 以添加的分别对应于待突变的基因或其部分的两端的正向引物和反向引物为引物, 经过变性/退火/延伸三个步骤扩增得到包含多个含有突变核苷酸的基因或其部分的基因突变文库。

术语“PCR 引物”是指一种合成的具有特定核苷酸序列的大分子, 其在聚合酶链式反应的退火(复性)阶段, 与模板 DNA 单链的互补序列配对结合, 并在延伸阶段在 DNA 聚合酶的作用下, 与反应体系中的核苷酸以共价键形式连接。

通常情况下, 引物对中的一条引物与靶区域一端的一条 DNA 模板链互补, 引物对中的另一条引物与靶区域另一端的另一条 DNA 模板链互补。其功能是作为核苷酸聚合作用的起始点, 核酸聚合酶可分别由这两条引物的 3'端开始合成新的核酸链。合成的 DNA 双链在下一循环的 DNA 变性阶段重新解链, 可以作为该循环的模板进行扩增。理论上, 这种扩增方式产生 $m \cdot 2^n$ 条 DNA 双链, 其中 m 为初始加入的模板链的数目, n 为循环数。因此, 使用这种引物对的 PCR 扩增也被称为指数扩增。

在本文中的引物对还可以是互补的两条寡核苷酸, 其中这两条寡核苷

酸在变性阶段解链，在退火阶段与其分别对应的模板互补配对，分别向前后两个方向延伸扩增。扩增后获得的 DNA 双链为体系中最初加入的模板链与由该引物起始延伸得到的产物链。在这种情况下，产物链不能作为下一轮扩增反应的模板。每一次 PCR 循环只能使用体系中最初加入的模板链作为模板。因此，理论上，这种扩增方式使得 n 次循环后每个方向最终分别产生 $m \times n$ 条 DNA 单链，其中 m 为初始加入的模板链的数目，n 为循环数。因此，使用这类引物对的 PCR 扩增为线性扩增。

术语“DNA 聚合酶”是以 DNA 为复制模板，将 DNA 由 5'端点开始复制制到 3'端的酶。DNA 聚合酶的主要活性是在具备模板、引物、dNTP 等的情况下催化 DNA 的合成。

术语“高保真 DNA 聚合酶”是指掺入错误率低且具有校正活性的 DNA 聚合酶，以保证目的 DNA 的忠实复制。

术语“待突变的基因”在本文中是指需要对其进行特定突变以构建基因突变文库的基因，该待突变的基因的序列是不包含突变的核苷酸的原始序列。

术语“质粒”是指细菌等生物中染色体 DNA(或拟核)以外的 DNA 分子，存在于细胞质中，具有自主复制能力，使其存在于子代细胞中，保持恒定的拷贝数，并表达所携带的遗传信息。一般情况下，从细菌中提取的质粒都具有已经甲基化的位点。

术语“DNA 甲基化”为 DNA 化学修饰的一种形式，能够在不改变 DNA 序列的前提下，改变遗传表现。广义上的 DNA 甲基化是指 DNA 序列上特定的碱基在 DNA 甲基转移酶(DNA methyltransferase, DNMT)的催化作用下，以 S-腺苷甲硫氨酸(S-adenosyl methionine, SAM)作为甲基供体，通过共价键结合的方式获得一个甲基基团的化学修饰过程。这种 DNA 甲基化修饰可以发生在胞嘧啶的 C-5 位、腺嘌呤的 N-6 位及鸟嘌呤的 N-7 位等位点。DNA 甲基化作为一种稳定的修饰状态，在 DNA 甲基转移酶的作用下，可随 DNA 的复制过程遗传给新生的子代 DNA，是一种重要的表观遗传机制。

常见的实验室 K12 的大肠杆菌菌株如 TOP10 包含 3 种甲基化酶识别不同的 DNA 序列：Dam 甲基化酶在 DNA 的 GATC 基序的 A 上添加一个甲基；Dcm 甲基化酶在 DNA 的 CCWGG 基序的第二个 C 上添加一个甲基；EcoKI 甲基化酶在 DNA 的 AACNNNNNNGTGC 基序或者 GCACNNNNNNGTT 基

序的 A 上添加甲基。

术语“限制性内切酶”是指可以识别并附着于特定的脱氧核苷酸序列，并对在每条链中特定部位的两个脱氧核糖核苷酸之间的磷酸二酯键进行切割的一类酶，简称“内切酶”。

- 5 术语“酶切位点”是指 DNA 链上一段特定的序列，限制性内切酶能够特异性地识别出这个序列并以特定的方式在此处将 DNA 序列切成两段。

- 术语“识别并切割已甲基化的位点的内切酶”是指能够识别并附着于特定的包含已甲基化的脱氧核苷酸的脱氧核苷酸序列，并对在每条链中特定部位的两个脱氧核糖核苷酸之间的磷酸二酯键进行切割的一类酶。例如，
10 识别并切割已甲基化的位点的内切酶可以是 *DpnI* 酶、*FspEI* 酶和 *MspII*。

DpnI 酶识别并切割提取自含有 *Dam* 甲基化酶的细菌的质粒中的甲基化的基序 G^mATC 。但是不切割未被甲基化的 PCR 产物，这是因为 PCR 体系中不含任何甲基化酶。其中基序“ G^mATC ”中的“ mA ”表示甲基化的腺嘌呤核苷酸。

- 15 *FspEI* 酶是修饰依赖性内切酶，其识别 C^mC 位点，并且在修饰的胞嘧啶 C 的 3' 端 N_{12}/N_{16} 处产生双链 DNA 断裂，即 $5' \dots C^mC(N) \dots 3'$ 和 $3' \dots GG(N) \dots 5'$ 。因此，*FspEI* 酶识别并且切割提取自含有 *Dcm* 甲基化酶的细菌的质粒中的甲基化的基序 C^mCWGG 。但是不切割未被甲基化的 PCR 产物，这是因为 PCR 体系中不含任何甲基化酶。其中，基序“ C^mC ”和“ C^mCWGG ”中的“ mC ”表示甲基化的胞嘧啶核苷酸。
20

MspII 酶是修饰依赖性内切酶，其识别 mCNNR 位点，并且在修饰的胞嘧啶的 3' 端 N_9/N_{13} 处产生双链 DNA 断裂，即 $5' \dots ^mCNNR(N) \dots 3'$ 和 $3' \dots GNNY(N) \dots 5'$ ，其中 $R = A$ 或 G ； $Y = C$ 或 T 。其中基序“ mCNNR ”中的“ mC ”表示甲基化的胞嘧啶核苷酸。

- Sanger 测序法是根据核苷酸在某一固定的点开始，随机在某一个特定的碱基处终止，并且在每个碱基后面进行荧光标记，产生以 A、T、C、G 结束的四组不同长度的一系列核苷酸，然后在尿素变性的 PAGE 胶上电泳进行检测，从而获得可见 DNA 碱基序列的一种方法。
25

- 高通量测序技术(High-throughput sequencing)又称“下一代”测序技术(“Next-generation” sequencing technology，缩写为 NGS)，是相对于传统的 Sanger 测序法而言的。NGS 以能一次并行对几十万到几百万条 DNA 分子进行序列测定和一般读长较短等为标志。目前高通量测序的主要平台代表有
30

罗氏公司(Roche)的 454 测序仪(Roch GS FLX sequencer)、Illumina 公司的 Solexa 基因组分析仪(Illumina Genome Analyzer)和 ABI 的 SOLiD 测序仪(ABI SOLiD sequencer)。其共同特点是(1)将目标 DNA 剪切为小片段；(2)单个小片段 DNA 分子结合到固相表面；(3)单分子独立扩增；(4)每次只复制一个
5 碱基(A,C,T,G)并检测信号；(5)高分辨率的成像系统。

术语“测序深度(Sequencing Depth)”是指测序得到的碱基总量(bp)与基因组大小的比值，它是评价测序量的指标之一。测序深度与基因组覆盖度之间是一个正相关的关系，测序带来的错误率或假阳性结果会随着测序深度的提升而下降。在本文中，测序深度应用于高通量测序基因突变文库所获得的测序结果。在本文中，测序深度特指在高通量测序结果中每种基因突
10 变体的测序深度。

术语“平均测序深度”在本文中是指在基因突变文库的测序结果中，平均每种基因突变体的测序深度。

术语“排名 10%测序深度”是指按照在基因突变文库测序结果中得到的所有基因突变体分别对应的测序深度从小到大进行排序，排名第 10%的
15 基因突变体所对应的测序深度。

术语“排名 90%测序深度”是指按照在基因突变文库测序结果中得到的所有基因突变体分别对应的测序深度从小到大进行排序，排名第 90%的基因突变体所对应的测序深度。

术语“偏移度”是指排名 90%测序深度/排名 10%测序深度，用于表征对基因突变文库进行高通量测序获得的所有基因突变体的测序深度是否均一，其表明在基因突变文库中，在某一突变位点包含不同突变的不同突变序列的拷贝数是否均一。1 是最优的偏移度，表示在基因突变文库中，在某一突变位点包含不同突变的不同突变序列的拷贝数是相同的。偏移度越接近 1，
20 在某一突变位点包含不同突变的不同突变序列的拷贝数越接近。

术语“基因突变分布”在本文中是指在最终得到的基因突变文库中所有序列在某一特定位置处的突变的总体分布。

30

序列说明

SEQ ID NO	序列说明
1	需要对其构建基因突变文库的原始基因所编码的蛋白质的氨基酸序列
2	需要对其构建基因突变文库的原始基因核苷酸序列
3	芯片合成的包含突变的寡核苷酸
4	芯片合成的包含突变的寡核苷酸
5	芯片合成的包含突变的寡核苷酸
6	芯片合成的包含突变的寡核苷酸
7	芯片合成的包含突变的寡核苷酸
8	芯片合成的包含突变的寡核苷酸
9	芯片合成的包含突变的寡核苷酸
10	芯片合成的包含突变的寡核苷酸
11	芯片合成的包含突变的寡核苷酸
12	芯片合成的包含突变的寡核苷酸
13	芯片合成的包含突变的寡核苷酸
14	芯片合成的包含突变的寡核苷酸
15	芯片合成的包含突变的寡核苷酸
16	芯片合成的包含突变的寡核苷酸
17	芯片合成的包含突变的寡核苷酸
18	芯片合成的包含突变的寡核苷酸
19	芯片合成的包含突变的寡核苷酸
20	芯片合成的包含突变的寡核苷酸
21	芯片合成的包含突变的寡核苷酸
22	用于 PCR 扩增合成的包含突变的寡核苷酸的正向引物
23	用于 PCR 扩增合成的包含突变的寡核苷酸的反向引物
24	用于 PCR 扩增包含突变的全长基因的正向引物
25	用于 PCR 扩增包含突变的全长基因的反向引物

具体实施例

实施例 1. 合成寡核苷酸

需要构建文库的氨基酸序列如下：

- 5 MKRAFIMVLDSFGIGATEDAERFGDVGADTLGHIAEACAKGEADNGRKG
 PLNLPNLTRLGLAKAHEGSTGFIPAGMDGNAEVIGAYAWAHMSSGKDS
 VSGHWEIAGVPVLFEGWGYFSDHENSFPQELLDKLVERANLPGYLGNCRS
 SGTVILDQLGEEHMKTGKWIFYTSAASVFQIACHEETFGLDKLYELCEIA
 REELTNGGYNIGRVIARPFIGDKAGNFQRTGNRRDLAVEPPAPT₁₀VLQKL
 V
 DEKHGQVVSVGKIADIYANCGITKKVKATGLDALFDATIKEMKEAGDNT

IVFTNFVDFDSSWGHRRDVAGYAAGLELFDRLPELMSLLRDDDILILTA
DHGCDPTWTGTDHTREHIPVLVYGPKVKPGSLGHRETADIGQTLAKYF
GTSDMEYGKAMF* (SEQ ID NO: 1), 共 407 个氨基酸, 其中第 35 个氨基酸为需要突变的氨基酸, 需要将丙氨酸(Ala)突变成另外 19 种常见氨基酸,
5 即, 甘氨酸(Gly)、缬氨酸(Val)、亮氨酸(Leu)、异亮氨酸(Ile)、苯丙氨酸(Phe)、脯氨酸(Pro)、色氨酸(Trp)、丝氨酸(Ser)、酪氨酸(Tyr)、半胱氨酸(Cys)、蛋氨酸(Met)、天冬酰胺(Asn)、谷氨酰胺(Gln)、苏氨酸(Thr)、天冬氨酸(Asp)、谷氨酸(Glu)、赖氨酸(Lys)、精氨酸(Arg)和组氨酸(His)。

需要突变的全长基因的核苷酸序列如下(共 1224 bp):

10 atgaaacgtgcattattatggtgctggactcattcggcatcgccgctacagaagatgcagaacgctttggtgacgtcgg
ggctgacaccctgggtcatatc**gcagaagccttgccaaaggcgaagctgataacggctgtaaaggcccgcctcaatc**
tgccaaatctgaccgtctggggctggcgaaagctcacgaaggttctaccggttcattccggcggaatggacggc
aacgctgaagttatcggcgctacgcatggcgccacgaaatgtcatccggtaaagatagcgtgtctggtcactggga
aattgccgggtgtccggttctgttgagtggggatattctccgatcacgaaaacagcttcccgaagagctgctggata
15 aactggtcgaacgcgctaacttgcgggttacctcggttaactgccgttctccgtacggctcattctggtacaaactggc
gaagagcacatgaaaaccggcaagtggattttctatacctccgctgcatccgtgtccagattgcctgccatgaagaaa
cttcggtctggataaactctacgaactgtgcgaaatcgccgtgaagagctgaccaacggcggtacaatatcggtc
gtgttatcgctcgtccgttatcggcgacaaagccggttaacttcacgctaccggttaaccgtcgtgacctgggtgtgag
ccgccagcaccgacctgctgcagaaactggtgatgaaaaacacggccaggtggtttctgtcggtaaaattgcgga
20 catctacgccaactgcggtatcaccaaaaaagtgaagcgactggcctggacgcgctgtttgacgccaccatcaaag
agatgaaagaagcgggtgataacaccatcgtcttcaccaactcgttgacttcgactcttctggggccaccgtcgga
cgtcgccggttatgcccggtctggaactgttcgaccgctctgccggagctgatgtctctgctgcgcgatgacga
catcctgacctcaccgctgaccacggttgcgatccgacctggaccggtactgaccacacgcgtgaacacattccggt
actggtatatggcccgaagtaaaaccgggctcactgggtcatcgtgaaaccttcgcggtatatcgccagactctggc
25 aaaatattttgttactctgatatggaatatggcaagccatgttctaa (SEQ ID NO: 2)

委托 Genscript 公司使用芯片合成包含突变的核苷酸的寡核苷酸序列池。每种寡核苷酸为对应于在第 35 位包含另外 19 种常见氨基酸之一的突变氨基酸的蛋白质的编码核苷酸序列。每种寡核苷酸的长度为 80 bp。其中每种氨基酸突变的比例为 1/19。序列如下:

30 agaagatgcagaacgctttggtgacgtcggggtgacaccctgggtcatatcGGCgaagccttgccaaaggcg
aagctg (SEQ ID NO: 3);

- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcGTTgaagcttgccaaaggcg
aagctg (SEQ ID NO: 4);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcCTGgaagcttgccaaaggcg
aagctg (SEQ ID NO: 5);
- 5 agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcATTgaagcttgccaaaggcg
aagctg (SEQ ID NO: 6);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcATGgaagcttgccaaaggcg
aagctg (SEQ ID NO: 7);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcTTTgaagcttgccaaaggcg
10 aagctg (SEQ ID NO: 8);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcTGGgaagcttgccaaaggcg
aagctg (SEQ ID NO: 9);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcCCGgaagcttgccaaaggcg
aagctg (SEQ ID NO: 10);
- 15 agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcTCAGaagcttgccaaaggcg
aagctg (SEQ ID NO: 11);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcACAGaagcttgccaaaggcg
aagctg (SEQ ID NO: 12);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcTGCgaagcttgccaaaggcg
20 aagctg (SEQ ID NO: 13);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcTATgaagcttgccaaaggcg
aagctg (SEQ ID NO: 14);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcAATgaagcttgccaaaggcg
aagctg (SEQ ID NO: 15);
- 25 agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcCAAGaagcttgccaaaggcg
aagctg (SEQ ID NO: 16);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcGATgaagcttgccaaaggcg
aagctg (SEQ ID NO: 17);
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcGAAGaagcttgccaaaggcg
30 aagctg (SEQ ID NO: 18)
- agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcAAAGaagcttgccaaaggcg

aagctg (SEQ ID NO: 19);

agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcAGAGaagcttgccaaaggcg

aagctg (SEQ ID NO: 20); 和

agaagatgcagaacgctttggtgacgtcggggctgacaccctgggtcatatcCATgaagcttgccaaaggcg

5 aagctg (SEQ ID NO: 21)。

实施例 2. 对芯片合成的寡核苷酸进行扩增

设计扩增引物，正向引物命名为 Chip-F，反向引物命名为 Chip-R。引物序列如下：Chip-F：agaagatgcagaacgctttggtgac (SEQ ID NO: 22)，反向引物序列如下：Chip-R：cagcttcgcctttggcacaagcttc (SEQ ID NO: 23)。配制 PCR 反应体系(参见表 1)，其中模板为芯片合成的寡核苷酸序列池，引物为上述 Chip-F/Chip-R、反应体系中还包含来自试剂盒 Phusion® High-Fidelity DNA Polymerase(购买自 New England Biolabs)的 Phusion DNA 聚合酶、dNTP 和 5xHF 缓冲液，添加水至总体积为 50 μ l，根据表 2 所示的 PCR 反应程序对芯片合成的寡核苷酸进行 PCR 扩增。扩增结果如图 2 所示。

表 1:

HF 缓冲液 (5x)	10 μ l
10 nM dNTP	1 μ l
模板(芯片合成的寡核苷酸)	100ng
正向引物: Chip-F	2 μ l
反向引物: Chip-R	2 μ l
BSA (20 mg/ml)	10 μ l
Phusion (2 U/ μ l)	0.5 μ l
H ₂ O	添加至 50 μ l

PCR 反应程序如表 2 所示:

表 2:

1	98°C 30 秒
2	98°C 10 秒
3	64°C 10 秒
4	72°C 20 秒
5	返回步骤 2, 16 个循环
6	72°C 5 分钟

20

实施例 3 以扩增后的芯片合成的寡核苷酸为引物进行线性扩增

以提取自大肠杆菌 TOP10 的包含待突变的全长基因和已甲基化的位点的质粒 PMB1 为模板，以实施例 2 中扩增后的芯片合成的寡核苷酸为引物进行 PCR 扩增。实施例 2 中扩增后的芯片合成的寡核苷酸为双链 DNA，两条链均能当引物，分别向前后两端扩增，两个扩增均为线性扩增。线性扩增引物需求量远小于指数扩增，因此，线性扩增时需要严格控制引物(即实施例 2 中扩增后的芯片合成的寡核苷酸)与体系中模板(即质粒)的比例。质粒模板和实施例 2 中扩增后的芯片合成的寡核苷酸(双链)的摩尔比为 1:20，PCR 反应的循环数为 15。其中 PCR 反应体系如表 3 所示：

表 3

HF 缓冲液 (5x)	10 μ l
10 nM dNTP	1 μ l
模板(PMB1)	1.5 ng
芯片合成的寡核苷酸(dsDNA)	0.473 ng
Phusion DNA 聚合酶(2 U/ μ l)	0.5 μ l
H ₂ O	添加至 50 μ l

PCR 反应程序如表 4 所示：

表 4

1	98 $^{\circ}$ C 30 秒
2	98 $^{\circ}$ C 10 秒
3	60 $^{\circ}$ C 30 秒
4	72 $^{\circ}$ C 15 秒
5	返回步骤 2，15 个循环
6	72 $^{\circ}$ C 5 分钟

实施例 4. *DpnI* 消化扩增体系中的质粒模板

如果后续扩增体系中含有 PMB1 质粒模板，在下一步扩增反应中会产生不包含突变的原始的全长基因，因此需要去除反应体系中的质粒。由于大肠杆菌 TOP10 菌株具有 Dam 甲基化酶，可以催化质粒中的 GATC 的腺嘌呤甲基化，形成 G^mATC。因此，从大肠杆菌 DH5 α 中提取的质粒 PMB1 具有甲基化修饰。*DpnI* 能识别和切割已甲基化的位点 G^mATC，加入 *DpnI* 将扩增体系中的质粒模板消化成小片段，而不能作为下一步 PCR 扩增的模板。配制酶切体系，加入 10 μ l 的 PCR 反应产物，*DpnI* (New England Biolabs, R0176L)，10xCutsamrt，添加水至总体积为 50 μ l。该反应在 *DpnI* 酶的最适温度 37 $^{\circ}$ C 下反应 2 小时。

酶切反应体系如表 5 所示：

表 5

实施例 3 中获得的 PCR 产物	10 μ l
Cutsamrt	5 μ l
<i>DpnI</i>	1 μ l
H ₂ O	添加至 50 μ l

实施例 5. PCR 扩增以获得最终产物

5 设计基因的首尾引物，在基因的头部位设置正向引物，命名为 Gene-F，在基因的尾部设置反向引物，命名为 Gene-R。引物序列如下，Gene-F：atgaaacgtgcatttattatgg (SEQ ID NO: 24)，Gene-R：ttagaacatggctttgccatat (SEQ ID NO: 25)。配制 PCR 反应体系，以实施例 4 中获得的核酸内切酶 *DpnI* 消化后的产物体系为模板，在反应体系中加入引物 Gene-F/Gene-R、Phusion
10 DNA 聚合酶、dNTP、5xHF 缓冲液，添加水至总体积为 50 μ l，以扩增基因的全长序列。在实施例 3 中分别以芯片合成的寡核苷酸的两条链为引物扩增获得的正义和反义 PCR 扩增产物(即两种单链 DNA，其在突变位点具有互补的核苷酸)会通过互补区连接在一起，作为该 PCR 扩增反应的模板。PCR 结束后，对产物进行琼脂糖凝胶电泳，将对应于最终产物大小的条带进行
15 切胶，使用 axygen DNA 凝胶回收试剂盒(购买自 AXYGEN 公司，货号：AP-GX-250)进行回收纯化，得到最终产物，如图 3 所示。

该 PCR 扩增的反应体系如表 6 所示：

表 6

HF 缓冲液 (5x)	10 μ l
10 nM dNTP	1 μ l
模板	2.5 μ l
正向引物 Gene-F	1 μ l
反向引物 Gene-R	1 μ l
Phusion (2 U/ μ l)	0.5 μ l
H ₂ O	添加至 50 μ l

该 PCR 扩增反应的反应程序如表 7 所示：

表 7

1	98℃ 30 秒
2	98℃ 10 秒
3	60℃ 30 秒
4	72℃ 20 秒

5	返回步骤 2, 15 个循环
6	72°C 5 分钟

实施例 6. 最终产物的鉴定

将实施例 5 中回收纯化的 DNA 进行 Sanger 测序, 测序结果如图 4A-4C 所示。图 4A 中显示, 在目标突变位点, 在第 35 位氨基酸的密码子位置(即第 103-105 位核苷酸)发生了突变。

将实施例 5 中回收纯化的 DNA 进行高通量测序, 分析获得的 DNA 序列中突变位点的突变核苷酸序列和各突变核苷酸序列的频率。高通量测序结果如表 8 所示。

表 8

预期的寡核苷酸种类	19
测得的寡核苷酸种类	19
最大的测序深度	284
平均测序深度	230.85
排名 10%测序深度	211
排名 90%测序深度	258
偏移度(90%/10%深度)	1.22

10

实施例 7. 以芯片合成的寡核苷酸为引物进行线性扩增步骤中模板与引物的不同比率的比较

测试以芯片合成的寡核苷酸为引物进行线性扩增步骤中模板与引物的不同比率。以提取自大肠杆菌 TOP10 的包含待突变的全长基因和已甲基化的位点的质粒 PMB1 为模板, 以实施例 2 中扩增后的芯片合成的寡核苷酸为引物进行 PCR 扩增。模板质粒和实施例 2 中扩增后的芯片合成的寡核苷酸(双链)的摩尔比为 1:254, PCR 反应的循环数为 25。其中 PCR 反应体系如表 8 所示, 反应程序如表 9 所示。然后重复实施例 4-6 相同的步骤, 最终得到的全长基因的电泳结果如图 5 所示, 条带较浅, 且弥散很强。

20

表 8

HF 缓冲液 (5x)	10 µl
10 nM dNTP	1 µl
模板(PMB1)	20 ng
芯片合成的寡核苷酸(dsDNA)	80 ng
Phusion DNA 聚合酶(2 U/µl)	0.5 µl
H ₂ O	添加至 50 µl

PCR 反应程序如表 9 所示：

表 9

1	98℃ 30 秒
2	98℃ 10 秒
3	60℃ 30 秒
4	72℃ 15 秒
5	返回步骤 2，25 个循环
6	72℃ 5 分钟

优化反应的循环数。以提取自大肠杆菌 TOP10 的包含待突变的全长基因和已甲基化的位点的质粒 PMB1 为模板，以实施例 2 中扩增后的芯片合成的寡核苷酸为引物进行 PCR 扩增。模板质粒和实施例 2 中扩增后的芯片合成的寡核苷酸(双链)的摩尔比为 1:254，减少 PCR 反应的循环数为 15。其中 PCR 反应体系如表 8 所示，反应程序如表 10。最终得到的全长基因的电泳结果如图 6 所示，条带较浅，且稍有弥散。

表 10

1	98℃ 30 秒
2	98℃ 10 秒
3	60℃ 30 秒
4	72℃ 15 秒
5	返回步骤 2，15 个循环
6	72℃ 5 分钟

10

实施例 8: 本发明的构建基因突变文库的方法与现有技术中构建基因突变文库的方法的比较

现有技术中用于构建基因突变文库的常规方法是：1. 先通过 PCR 对芯片合成的突变区域进行扩增，随后进行柱纯化，2. 以待突变的基因为模板扩增突变区域前后的非突变区域，突变区前边的区域和突变区后边的区域需分别扩增，各做一个 PCR 反应，之后进行琼脂糖凝胶电泳，通过切胶回收纯化 DNA 片段；3. 通过组装 PCR 将 3 个片段组装成一个片段，4. 以组装产物为模板(无需纯化)，加基因首尾引物扩增全长片段，之后进行琼脂糖凝胶电泳，通过切胶回收纯化 DNA 片段。

现有技术中用于构建基因突变文库的上述常规方法总耗时大约 6.5 小时，而本发明的用于构建基因突变文库的方法耗时大约 5.5 小时。由此可见，本发明的构建基因突变文库的方法步骤更少，更节省时间。

还需要说明的是，在可实施且不明显违背本发明的主旨的前提下，在本说明书中作为某一技术方案的构成部分所描述的任一技术特征或技术特征的组合同样也可以适用于其它技术方案；并且，在可实施且不明显违背本发明的主旨的前提下，作为不同技术方案的构成部分所描述的技术特征之间也可以以任意方式进行组合，来构成其它技术方案。本发明也包含在上述情况下通过组合而得到的技术方案，并且这些技术方案相当于记载在本说明书中。

以上通过具体实施方式和实施例对本发明进行了说明，但本领域技术人员应该理解的是，这些并非意图对本发明的范围进行限定，本发明的范围应由权利要求书确定。