

基因组文库和 cDNA 文库的构建及筛选

刘芝华 中国医学科学院肿瘤研究所 分子肿瘤学国家重点实验室

概念:

基因组文库是含有某种生物体全部基因的随机片段的重组 DNA 克隆群体;cDNA 文库是指含有所有重组 cDNA 的克隆群体。

基因组文库: 来源于基因组 DNA, 反映基因组的全部信息, 用于基因组物理图谱的构建, 基因组序列分析, 基因在染色体上的定位, 基因组中基因的结构和组织形式等。

cDNA 文库: 来源于细胞表达出的 RNA, 反映基因组表达的基因序列信息, 用于研究特定细胞中基因的表达状态和表达基因的功能等。

基因文库=基因组文库+cDNA 文库

载体的种类和特征:

	受体细胞	结构	插入片段	举例
质粒	E.coli	环状	很小, <8kb	pUC18/19 ,T-载体, pGEM3z
λ 噬菌体	E.coli	线状	9-24kb	EMBL, λgt 系列
丝状噬菌体及噬菌粒	E.coli	环状	<10kb	M13 系列
粘粒载体	E.coli	环状	35-45kb	
BAC	E.coli	环状	约 300kb	
PAC	E.coli	环状	100-800kb	PCYPAC
YAC	酵母细胞	线性染色体	100-2000kb	
MAC	哺乳动物细胞	环状	>1000kb	
病毒载体	动物细胞	环状		SV40 载体, 昆虫杆状病毒载体
穿梭载体	动物细胞和细	环状		pSVK3 质粒, PBV,Ti

	菌			质粒,
--	---	--	--	-----

· DNA 的制备：高分子量的外源 DNA 100-150kb

· 剪切：随机切割

· 载体：

λ 噬菌体：48.5kb, 插入型(最多可容纳 9kb)、取代型(可容纳 38-51kb, 最适 19-20kb), EMBL, Charon 粘性质粒(Cosmid): 4-6kb, 质粒+噬菌体的性质, 可容纳大片段的外源基因, 最多可容纳 46kb。

酵母人工染色体克隆体系 (YAC, Yeast Artificial Chromosome Cloning System)
插入大小 200—1000kb

· 载体 DNA 的制备：

载体 DNA 酶切反应-----载体臂的纯化-----载体脱磷处理 (常用 BamHI) (蔗糖密度梯度离心) (碱性磷酸酶)

· 连接和包装：外源片段与两臂之间的比例, 包装蛋白

· 感染宿主菌：选择合适的宿主菌, LE392, NM538, NM539, KW251 等

· 基因组文库的保存和扩增：完整性和代表性

测滴度, 要在 10⁶fu 以上; 保存, 氯仿 4℃, 7% DMSO -70℃

cDNA 文库的构建

· 分类：按表达类型分：表达型、非表达型

按载体类型分：质粒 cDNA 文库、噬菌体 cDNA 文库

· 关键：mRNA 的获得。完整、种类多、不能降解。电泳应在 0.5-8kb 之间。已知探针做 Northern 或 PCR。

· 反转录：AMV 鸟成髓细胞性白血病病毒

M-MLV 鼠白血病病毒

引物：oligo dT, Random primer

掺入同位素检测

· cDNA 第二链的合成

· 载体：lgt10, lgt11, lZAP 等。

克隆入 λ 噬菌体的效率是克隆入质粒中的 30 倍。

1 ZAP II: (i) 噬菌体高效; (ii) 质粒易操作; (iii) 蓝白斑; (iv) 插入大小 10kb; (v) 可表达融合蛋白, 可用 DNA 探针或抗体来筛选; (vi) 体内自我剪切

· 从少量 RNA 中建 cDNA 文库:

一般情况下, 只有 10% 的 mRNA 可变成可克隆的双链 cDNA, 而且体外包装和细菌转化都会再次降低效率, 很多人结合 PCR, 但有非特异性。

---结合 Solid-phase RNA Capture System, 可用 5ng mRNA, 这时 oligo dT 结合在一个磁珠上, 做第二链的模板;

---用 Universal Buffer, 从 cDNA 第一、第二链的合成, 到连 Linker 等全在一个系统内进行。

· cDNA 文库的标准化: 使高丰度的信息相对减少、低丰度的信息相对增多的过程。

减少反复重复, 但也会丢失一些信息。

----与基因组 DNA 杂交; ---应用二级动力学原理在溶液中使双链 DNA 再退火。

理论上, 在单细胞中, 去除高丰度者可使低丰度者聚集 3 倍或更少。因为细胞中 1/3 mRNA 在细胞中有 1-10 拷贝。

· 定向克隆的 cDNA 文库: 大量 EST 测序时会用到。

· 好的 cDNA 文库的标准:

---具有代表性, 含所有带 PolyA 尾的 RNA 群体且各类型出现频率也与之一致;

---大部分含有较长或全长的 cDNA 插入序列;

---少有基因组、线粒体 DNA、核糖体 RNA 的污染。

· cDNA 文库的筛选:

DNA 探针, 抗体, 标记蛋白

文库的扩增:

噬菌体文库加 SM, 粘粒和质粒文库 LB。

文库的保存:

噬菌体文库: 4°C (加氯仿) 可保存几年, -70°C (7% DMSO)

长期保存; 粘粒和质粒文库: 15%甘油, -70°C 冻存。

文库的筛选:

·选择文库

·所要筛选的克隆数:

--对基因组文库取决于克隆片段的大小和基因组的大小

$$N = \ln(1-P) / \ln[1-(I/G)]$$

其中 I: 克隆片段的随机大小, G: 靶基因组的大小, P: 分离到基因的概率。

--对 cDNA 文库取决于所要片段的表达丰度和基因组的大小

用核酸做探针还是用抗体做探针