

[19]中华人民共和国专利局

[51]Int.Cl⁶



[12] 发明专利申请公开说明书

C12N 15/90
C12N 15/85 C12N 15/12
C12N 15/62

[21] 申请号 96195654.2

[43]公开日 1998 年 8 月 26 日

[11] 公开号 CN 1191574A

[22]申请日 96.6.7

[30]优先权

[32]95.6.7 [33]US[31]08 / 487,989

[32]96.5.6 [33]US[31]08 / 643,554

[86]国际申请 PCT / US96 / 10248 96.6.7

[87]国际公布 WO96 / 40965 英 96.12.19

[85]进入国家阶段日期 98.1.19

[71]申请人 凯斯韦斯顿瑞瑟弗大学

地址 美国俄亥俄州

[72]发明人 G · B · 冯伯克里恩 J · J · 哈里格顿

H · F · 威拉德

[74]专利代理机构 中国国际贸易促进委员会专利商标
事务所
代理人 李 瑛

权利要求书 8 页 说明书 51 页 附图页数 12 页

[54]发明名称 合成的哺乳动物染色体和其构建方法

[57]摘要

本发明涉及基因治疗、基因表达和这些用途的载体领域。具体地说,本发明涉及用于产生 DNA 的结构完整的大重复单元(尤其是对 α 卫星 DNA 的稳定克隆有用的大重复单位)的方法,以及开发和利用在哺乳动物(尤其是人类)中进行基因表达和基因治疗的合成或人工染色体的方法。本发明使得可以控制性地构建由纯化的 DNA 的分离的区段构成的稳定的合成或人工染色体。功能性最小区段优选地包括着丝粒 DNA、端粒 DNA 和基因组 DNA。所述人工染色体发挥天然存在的染色体的必需染色体功能,以使所述染色体作为基因治疗的有效载体起作用。

权 利 要 求 书

1. 一种人工哺乳动物染色体, 该染色体实质上包含着丝粒、端粒和基因组 DNA。

2. 一种人工哺乳动物染色体, 该染色体实质上包含着丝粒 DNA、端粒 DNA 和基因组 DNA, 其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

3. 一种人工哺乳动物染色体, 该染色体实质上包含着丝粒 DNA、端粒 DNA 和基因组 DNA, 其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段, 所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列, 并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

4. 一种人工哺乳动物染色体, 该染色体是通过用纯化的 DNA 转染哺乳动物细胞的方法产生的, 所说 DNA 实质上包含端粒 DNA、着丝粒 DNA 和基因组 DNA, 其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

5. 一种人工哺乳动物染色体, 该染色体是通过用纯化的 DNA 转染哺乳动物细胞的方法产生的, 所说 DNA 实质上包含端粒 DNA、着丝粒 DNA 和基因组 DNA, 其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段, 所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列, 并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

6. 一种人工哺乳动物染色体, 该染色体是通过用纯化的裸 DNA 转染哺乳动物细胞的方法产生的, 所说 DNA 实质上包含端粒 DNA、着丝粒 DNA 和基因组 DNA, 其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

7. 一种人工哺乳动物染色体, 该染色体是通过用纯化的裸 DNA 转染哺乳动物细胞的方法产生的, 所说 DNA 实质上包含端粒 DNA、着丝粒 DNA 和基因组 DNA, 其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切

片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

8. 一种人工哺乳动物染色体，该染色体是通过用纯化的凝聚 DNA 转染哺乳动物细胞的方法产生的，所说 DNA 实质上包含端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

9. 一种人工哺乳动物染色体，该染色体是通过用纯化的凝聚 DNA 转染哺乳动物细胞的方法产生的，所说 DNA 实质上包含端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

10. 一种人工哺乳动物染色体，该染色体是通过将纯化的包被 DNA 转染进哺乳动物细胞的方法产生的，所说 DNA 实质上包含端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

11. 一种人工哺乳动物染色体，该染色体是通过将纯化的包被 DNA 转染进哺乳动物细胞的方法产生的，所说 DNA 实质上包含端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

12. 权利要求 4-11 之任一的人工哺乳动物染色体，其中所说的着丝粒 DNA、所说的端粒 DNA 和所说的基因组 DNA 不是相互连接的。

13. 权利要求 4-11 之任一的人工哺乳动物染色体，其中一种或多种所说的着丝粒 DNA、所说的端粒 DNA 和所说的基因组 DNA 不是相互连接的。

14. 一种包含权利要求 1-11 之任一的人工哺乳动物染色体的组合体。

15. 权利要求 1-11 之任一的人工哺乳动物染色体，其中所说的着丝粒 DNA 包括 α -卫星 DNA。

16. 一种包含权利要求1-11之任一的人工哺乳动物染色体的哺乳动物细胞。

17. 权利要求1-11之任一的人工哺乳动物染色体，其中所说的染色体还包含异源DNA，当所说的染色体引入到哺乳动物细胞中时，该异源DNA从所说的染色体表达，或者引起基因产物的表达。

18. 纯化的DNA，该DNA实质上包含端粒DNA、着丝粒DNA和基因组DNA，其中所说的基因组DNA是选自由限制酶消化片段和机械剪切片段组成的组的次基因组DNA片段。

19. 纯化的DNA，该DNA实质上包含端粒DNA、着丝粒DNA和基因组DNA，其中所说的基因组DNA是选自由限制酶消化片段和机械剪切片段组成的组的次基因组DNA片段，所说的着丝粒DNA包含在有丝分裂期间与CENP-E相关联的DNA序列，并且所说的端粒DNA包含串联重复的序列TTAGGG。

20. 纯化的裸DNA，该DNA实质上包含端粒DNA、着丝粒DNA和基因组DNA，其中所说的基因组DNA是选自由限制酶消化片段和机械剪切片段组成的组的次基因组DNA片段。

21. 纯化的裸DNA，该DNA实质上包含端粒DNA、着丝粒DNA和基因组DNA，其中所说的基因组DNA是选自由限制酶消化片段和机械剪切片段组成的组的次基因组DNA片段，所说的着丝粒DNA包含在有丝分裂期间与CENP-E相关联的DNA序列，并且所说的端粒DNA包含串联重复的序列TTAGGG。

22. 纯化的凝聚DNA，该DNA实质上包含端粒DNA、着丝粒DNA和基因组DNA，其中所说的基因组DNA是选自由限制酶消化片段和机械剪切片段组成的组的次基因组DNA片段。

23. 纯化的凝聚DNA，该DNA实质上包含端粒DNA、着丝粒DNA和基因组DNA，其中所说的基因组DNA是选自由限制酶消化片段和机械剪切片段组成的组的次基因组DNA片段，所说的着丝粒DNA包含在有丝分裂期间与CENP-E相关联的DNA序列，并且所说的端粒DNA包含串联重复的序列TTAGGG。

24. 纯化的包被DNA，该DNA实质上包含端粒DNA、着丝粒DNA和基因组DNA，其中所说的基因组DNA是选自由限制酶消化片段和机械剪切片

段组成的组的次基因组 DNA 片段。

25. 纯化的包被 DNA，该 DNA 实质上包含端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

26. 由体外组合端粒 DNA、着丝粒 DNA 和基因组 DNA 制备的纯化的 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

27. 由体外组合端粒 DNA、着丝粒 DNA 和基因组 DNA 制备的纯化的 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

28. 由体外组合端粒 DNA、着丝粒 DNA 和基因组 DNA 制备的纯化的裸 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

29. 由体外组合端粒 DNA、着丝粒 DNA 和基因组 DNA 制备的纯化的裸 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

30. 由体外组合端粒 DNA、着丝粒 DNA 和基因组 DNA 制备的纯化的凝聚 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

31. 由体外组合端粒 DNA、着丝粒 DNA 和基因组 DNA 制备的纯化的凝聚 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

32. 由体外组合端粒 DNA、着丝粒 DNA 和基因组 DNA 制备的纯化的包被 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段

组成的组的次基因组 DNA 片段。

33. 由体外组合端粒 DNA、着丝粒 DNA 和基因组 DNA 制备的纯化的包被 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

34. 包含权利要求 18-33 之任一的 DNA 的组合体。

35. 包含权利要求 18-33 之任一的纯化的 DNA 的哺乳动物细胞。

36. 权利要求 18-33 之任一的纯化的 DNA，其中所说的着丝粒 DNA、所说的端粒 DNA 和所说的基因组 DNA 不是相互连接的。

37. 权利要求 18-33 之任一的纯化的 DNA，其中一种或多种所说的着丝粒 DNA、所说的端粒 DNA 和所说的基因组 DNA 是相互连接的。

38. 权利要求 18-33 之任一的纯化的 DNA，其中所说的着丝粒 DNA 包括 α -卫星 DNA。

39. 权利要求 18-33 之任一的纯化的 DNA，其中所说的 DNA 还包含异源 DNA，当所说的 DNA 引入到哺乳动物细胞中时，该异源 DNA 从所说的染色体表达，或者引起基因产物的表达。

40. 包含权利要求 18-33 之任一的 DNA 的载体。

41. 包含权利要求 40 的载体的细胞。

42. 包含权利要求 40 的载体的组合体。

43. 一种克隆重复串联族 DNA 的方法，该方法包括

(a) 制备第一 DNA 单位，以使所说 DNA 单位的相反的末端包含互补的但非同裂酶的限制位点；

(b) 将所说的 DNA 单位连接进载体；

(c) 在所说的限制位点之一使所说的载体线性化；

(d) 连接如步骤(a)制备的第二 DNA 单位，使其与所说的第一单位串联，以形成定向的重复族；

(e) 将所说的族转化进细菌宿主细胞；

(f) 选择包含所说的族的稳定的克隆；和

(g) 重复步骤(c)-(f)，直到达到所需的族的大小。

44. 权利要求 43 的方法，其中所说的 DNA 是 α 卫星 DNA。
45. 权利要求 44 的方法，其中所说的 α 卫星 DNA 长度大于 100 kb。
46. 权利要求 45 的方法，其中所说的 α 卫星 DNA 长度大于 140 kb。
47. 权利要求 44 的方法，其中所说的 α 卫星 DNA 是人类 α 卫星 DNA。
48. 一种包含由定向的重复的 DNA 族组成的序列的载体，所说的族包含重复 DNA 单位，其中每个 DNA 单位的相反的末端包含互补的但非同裂酶的限制位点。
49. 权利要求 48 的载体，其中所说的 DNA 是 α 卫星 DNA。
50. 权利要求 49 的载体，其中所说的 α 卫星 DNA 的所说的族长度大于 100 kb。
51. 权利要求 50 的载体，其中所说的 α 卫星 DNA 的所说的族长度大于 140 kb。
52. 权利要求 47 的载体，其中所说的 α 卫星 DNA 是人类 α 卫星 DNA。
53. 一种用权利要求 47 -52 之任一的载体稳定转化的宿主细胞。
54. 权利要求 53 的宿主细胞，其中所说的宿主细胞是原核细胞。
55. 权利要求 54 的宿主细胞，其中所说的原核细胞是大肠杆菌。
56. 一种制造人工哺乳动物染色体的方法，该方法包括把权利要求 18-33 之任一纯化的 DNA 引入到哺乳动物细胞中。
57. 一种制造人工哺乳动物染色体的方法，该方法包括将权利要求 34 的组合体引入到哺乳动物细胞中。
58. 一种制备纯化的 DNA 组合体的方法，该方法包括体外组合纯化的端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。
59. 一种制备纯化的 DNA 组合体的方法，该方法包括体外组合纯化的端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。
60. 一种制备纯化的裸 DNA 组合体的方法，该方法包括体外组合纯化的端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由

限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

61. 一种制备纯化的裸 DNA 组合体的方法，该方法包括体外组合纯化的端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

62. 一种制备纯化的凝聚 DNA 组合体的方法，该方法包括体外组合纯化的端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

63. 一种制备纯化的凝聚 DNA 组合体的方法，该方法包括体外组合纯化的端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

64. 一种制备纯化的包被 DNA 组合体的方法，该方法包括体外组合纯化的端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段。

65. 一种制备纯化的包被 DNA 组合体的方法，该方法包括体外组合纯化的端粒 DNA、着丝粒 DNA 和基因组 DNA，其中所说的基因组 DNA 是选自由限制酶消化片段和机械剪切片段组成的组的次基因组 DNA 片段，所说的着丝粒 DNA 包含在有丝分裂期间与 CENP-E 相关联的 DNA 序列，并且所说的端粒 DNA 包含串联重复的序列 TTAGGG。

66. 权利要求 56-65 之任一的方法，其中所说的着丝粒 DNA、所说的端粒 DNA 和所说的基因组 DNA 不是相互连接的。

67. 权利要求 56-65 之任一的方法，其中所说的一种或多种所说的着丝粒 DNA、所说的端粒 DNA 和所说的基因组 DNA 是相互连接的。

68. 一种在哺乳动物细胞中表达基因的方法，该方法包括培养包含权利要求 1-11 之任一的人工染色体的哺乳动物细胞，其中所说的染色体含有所说的基因或者含有使得所说的基因可以表达的 DNA 序列。

69. 一种在哺乳动物细胞中表达异源基因的方法，该方法包括培养包

含权利要求 18-33 之任一的 DNA 的哺乳动物细胞，其中所说的 DNA 含有所说的基因或者含有使得所说的基因可以表达的 DNA 序列。

70. 权利要求 68 的方法，其中所说的基因表达对包含所说细胞的哺乳动物提供治疗作用。

71. 权利要求 69 的方法，其中所说的基因表达对包含所说细胞的哺乳动物提供治疗作用。

说明书

合成的哺乳动物染色体和其构建方法

有关联邦政府资助下进行的研究和开发的发明权利的声明

在本发明的开发期间完成的部分工作利用了美国政府的资金。美国政府对本项发明具有一定的权利。

发明领域

本发明涉及基因表达和基因治疗领域，并涉及有关这些用途的新载体。具体地说，本发明涉及开发和利用在基因表达和基因治疗中(尤其是在人类中)作为载体的合成或人工染色体的方法。本发明使得可以从分离的纯化的DNA控制性地构建稳定的合成或人工染色体。就这一DNA而言，功能性染色体在细胞中形成并作为染色体外因子被保持。当治疗DNA包含的染色体中时，所述人工染色体发挥天然存在的染色体的必需染色体功能，以使所述染色体作为基因治疗的有效载体起作用。

发明背景

为治疗遗传性或者获得性疾病对细胞进行遗传操作称为基因治疗。迄今为止，此领域中的大多数临床研究集中于病毒的基因治疗载体的应用。基于这些研究的结果，能明显看到现行的病毒基因治疗载体有严重临床限制性。这些包括免疫原性，细胞病变，不稳定的基因表达及对治疗基因片段大小的限制。因此，近来人们更注重非病毒基因治疗载体的应用。

尤其是，对活细胞进行多种遗传操作时，合成的哺乳动物染色体将会是有用的载体。其优越性在于高度的有丝分裂稳定性，一致和可调的基因表达，高的克隆容量，及非免疫原性。

1983年，人工染色体首次构建于啤酒糖酵母(Murray等，自然，305: 189-193(1983)，1989年，构建于*S. pombe*(Hahnenberger等，美国科学院学报，86: 577-581(1989)。然而由于很多原因，是否能在哺乳动物细胞中构建相似的载体并不清楚。

第一，十多亿年前，多细胞有机体(即哺乳动物细胞的祖先)从酵母中分化出来。尽管在活的有机体之间存在着相似之处，但总体上，相似性与他们进化上的趋异程度呈反相关。很明显，酵母，一种单细胞有机体，在生物学上与复杂的多细胞脊椎动物有截然的不同。

第二，酵母染色体小于哺乳动物染色体几个级别长度。啤酒糖酵母和 *S. pombe* 染色体长度分别是 0.2—2Mb 及 3.5—5.5Mb，而哺乳动物的染色体大约在 50 Mb—250 Mb。由于在长度上的明显区别，能否将类似酵母人工染色体的构建体构建成功及转染进哺乳动物仍不清楚。

第三，酵母染色体压缩程度少于哺乳动物的染色体。这意味着，哺乳动物染色体的高级结构依赖于染色质复杂的相互作用。哺乳动物染色体的这种复杂结构(包括 DNA 结构和高级染色质结构)使人们怀疑人工染色体是否能在哺乳动物细胞中构建。

第四，酵母着丝粒比哺乳动物的着丝粒更简单。例如，啤酒糖酵母的着丝粒由 125bp 组成。*S. pombe* 的着丝粒包括 2 到 3 个拷贝的 14kb 序列元件及一段反向重复序列，中间为一段 7Kb 的核心区。相比之下，人类着丝粒由几百个 Kb 到几个 Mb 的高度重复的 α 卫星 DNA 组成。而且，在哺乳动物的着丝粒中，未证明有象在 *S. pombe* 中发现的中央核心区及反转重复序列。这样，不同于酵母着丝粒，哺乳动物的着丝粒非常大且高度重复。

第五，酵母着丝粒比哺乳动物的着丝粒有更少的纺锤体吸附(Bloom, 细胞, 73: 621-624(1993))。例如，啤酒糖酵母有一个微管吸附到着丝粒上。*S. pombe* 的每个着丝粒有 2—4 个微管。在人类中每个染色体的着丝粒的几十个微管吸附(Bloom, 细胞, 73: 621-624(1993))。这进一步说明相比于酵母着丝粒，哺乳动物着丝粒的复杂性。

总而言之，这些区别是重要的，因此在酵母中成功的结果并不能合理地预测能转化入哺乳动物。

正常的哺乳动物染色体由连续的线性 DNA 链组成，长度约在 50 至 250 Mb。为使这些遗传单位能忠实复制且在细胞分裂时分开，至少必须包括三种类型的功能元件：端粒，复制起点和着丝粒。

哺乳动物中的端粒是由重复序列(TTAGGG)组成，对染色体末端的复制和稳定性来说是必要的。复制的起点在细胞周期的 S 期对染色体 DNA 的有效

及可控复制是必要的。

虽然哺乳动物的复制起点还没有在序列水平上弄清,但我们相信它们在哺乳动物的 DNA 上丰富存在。最后,着丝粒在有丝分裂期间对染色单体分开进入子细胞是必要的,这保证了每一个子细胞获得一个单体,且只有一个的所有染色体的拷贝。正如复制起点一样。着丝粒也未能在序列水平上搞清。 α 卫星 DNA 可能是一个重要的着丝粒组成成分(Haaf 等, 细胞, 70: 681-696(1992); Larin 等, Hum. Mol. Genet. 3: 689-695(1994); Willard, Trends in Genet. 6: 410-415(1990)。但某些有丝分裂稳定的异常染色体衍生物缺乏 α 卫星 DNA(Callen 等, Am. J. MedGenet. 43: 709-715(1992); Crolla 等, J. Med. Genet 29: 699-703(1992); Voullaire 等, J. Hum. Genet. 52: 1153-1163(1993); Blennow 等, Am. J. Hum. Genet. 54: 877-853(1994); Ohashi 等, Am. J. Hum. Genet. 55: 1202-1208(1994))。因此,对哺乳动物的着丝粒的组成仍然了解很少。

虽然一些人声称在哺乳动物的细胞中构建成"人工"染色体,但没有任何人曾经构建成一个仅仅包含外源 DNA 的人工染色体。在以前的研究中,研究人员要么修饰现有的染色体使其变小("削短"法),要么将外源 DNA 插入到现有的染色体中,这种方法将使染色体断裂形成包含内源序列的片段("断裂"方法)。在本发明中,外源 DNA 序列可引入人类细胞,能形成稳定的合成染色体,但并未整合入内源染色体中。

在削短法中,利用了三种特殊的设计:(1)通过非法重组,端粒引导截短(Barnett, M. A. 等,核酸研究, 21: 27-36(1993); Farr. C. J. 等, ENBO J. 14: 5444-54(1995)), (2)通过同源重组, α 卫星 DNA 引导端粒插入/截短(Brown, K. E. 等, Hum Mol. Genet. 3: 1227-37(1994)), (3)双着丝粒染色体形成/断裂(Hadlaczky, G., 哺乳动物人工染色体,美国专利 5, 288, 625(1994))。

Barnett 等(核酸研究, 21: 27-36(1993)), Farr 等。(EMBO J. 14: 5444-54(1995))和 Brown 等(Hum Mol. Genet. 3: 1227-37(1994))描述了将内源染色体片断化的方法,即将端粒 DNA 和一种可选择标记同时转染进哺乳动物细胞。在上述实验中可构建一个小于初始染色体的截短染色体。但形成的截短染色体包含大量内源染色体序列,也包含内源着丝粒。

这样，这些截短染色体并不是从头产生。

Hadlaczky (哺乳动物人工染色体，美国专利 5,288,625(1994))描述了一种细胞系，这种细胞系由于双着丝粒染色体断裂现象可用于增殖一条染色体。但除了可选择的标记外所有序列来自于初始的全功能双着丝粒染色体。因此，这种所谓的"人工"染色体也不是从头产生的。

在"断裂"方法中，Haaf 等(细胞: 681-696(1992))和Praulovszky 等(美国科学院学报, 88: 11042-11046(1991))报道了通过把转染 DNA 插入内源染色体产生染色体片段的方法。转染以后，插入的 DNA 序列开始扩增(拷贝数增加)，在某些克隆中，一部分内源染色体断裂形成游离于染色体之外的片断。上述两个报道中，可在内源染色体及染色体外片段上广泛发现插入的转染 DNA。

在 Haaf 等的实验中(细胞, 70: 681-696(1992))，他们将人类 α 卫星 DNA 和新霉素抗性基因共转染进非洲绿色猴子细胞中。每一次转染不包括任何其它外源 DNA。在所有的转染克隆中，发现 DNA 被整合进内源染色体。在一个含有染色体外片段的克隆中，转染的 α 卫星 DNA 整合后被广泛地扩增。基于 Southern 印迹杂交和原位荧光杂交，作者得出结论，非洲绿色猴子序列与转染的 α 卫星 DNA 共扩增，且被扩散到转染的 α 卫星 DNA 中。在对包含扩增 α 卫星 DNA 的染色体的进一步分析中发现，被转染染色体的"数量，大小，转染位置(端粒，间质，或者是着丝粒)在 3-31 号细胞系之间是不同的。这提示了转染序列的不稳定性。最后，分析包含扩增 α 卫星 DNA 的染色体的有丝分裂行为，发现有高频率的后期桥形成，提示染色体具有双着丝粒(或者是多着丝粒)。这样，已发现的高度结构不稳定性及后期桥结构的高频发生与染色体片断形成来自于插入/扩增/断裂的想法是一致的。最后，值得注意的是，在某些包含未扩增的 α 卫星 DNA 的克隆中，未发现染色体外片段，进一步说明此方法中扩增对染色体片断化过程是重要的。

Prazovszky 等(Pvoc. Natl. Acad. Sci. USA 88: 11042-11046(1991))通过插入一段非着丝粒区人类 DNA 到内源染色体上产生染色体片段(后来 McGill 等和 Cooper 等在人类染色体 9qter 图谱上显示(Hum. Mol. Genet. 1: 749-751(1992))(Hum. Mol. Genet. 1: 753-754(1992))。同 Haaf 实验一样，插入的转染 DNA 广泛扩增，且散布于小鼠基因组序列中。作者认为转

染 DNA 的插入/扩增导致双着丝粒染色体的形成，然后断裂产生染色体片段。对染色体片段的分析毫无疑问的说明染色体片段来源于含有插入扩增 DNA 的小鼠染色体。

Haafer 等和 Praulovszky 等所做的实验有许多重要的相似之处。第一，都表明转染 DNA 可插入内源染色体。第二，插入以后，转染 DNA 能广泛扩增。第三，通过扩增的 DNA 序列，内源 DNA (受体细胞中未被转染的染色体序列) 可散布分布。第四，含有扩增转染序列的内源染色体能被 CREST 抗血清所着色。第五，含有扩增的转染序列的内源染色体在有丝分裂期间行为相似，即形成双着丝粒染色体。最后，含有扩增的转染序列的内源染色体都表现出结构不稳定。因此，这些大量重要的相似之处，及被 Pranovszky 等证明的染色体断裂都显示了在两种实验中染色体插入/扩增/断裂机制。

Larin 等 (Hum. Mol. Genet. 3: 689-95 (1994)) 获得进一步的证据，证明若只有转染和插入 α 卫星 DNA 而缺乏扩增则不足以产生染色体外片断。在这些实验中， α 卫星 DNA 与一个可选择的标记连接，并转染进入人类细胞。在所有抗药克隆中， α 卫星 DNA 插入内源染色体。当这些插入形成着丝粒样结构时 (例如：主缢痕，CREST 抗血清着色，后期染色体延迟)，没有一个克隆可形成染色体外片段。由于这些实验没能发现一种克隆，其染色体包含 α 卫星 DNA，但无内源着丝粒，因此这并不是一种可靠的方法能证明是否着粒样结构的形成可促进染色体分离。

由于是从已存的染色体构建成“削短”染色体，或插入 DNA 到已存的染色体形成“断裂”染色体，对从转染的裸露的 DNA 从头合成染色体，这些文献并未提供任何指导。

而且，由于多种原因，当作为基因治疗载体时，构建的染色体及构建方法有严重的限制性。第一，构建的方法仅能应用于构建培养细胞的染色体。由于病人的细胞断裂现象非常少见，及/或产生的染色体结构不可预测，这些方法不适用于直接应用于病人。另外，断裂方法中扩增序列的不稳定性也不适用于病人，因为有因基因组重排导致细胞转化及癌变的风险。

因此，如果有带有限定结构的一个预先构建的染色体作为载体能被直接引入病人的细胞，将会非常理想。尤其是这种载体可不依赖于插入内源染

色体或者是插入后的扩增，而且在细胞中构建体的结构与转染前的结构本质上相同。

第二，削短染色体和染色体片段包含不可限定的内源序列，而且并不能为鉴别功能重要的序列提供指导。

如果可构建由限定序列组成的载体，如果有一种构建限定的人工染色体的方法，通过这种方法能迅速鉴别出功能重要的序列，那将非常理想。

第三，通过现有的方法，不能将削短法或断裂法产生的染色体完全纯化。因此，当将削短染色体导入哺乳动物细胞时，很难不导进其它的哺乳动物染色体。

第四，因为还不能将削短染色体和染色体片段作为裸 DNA 分离，从而可重新导入细胞中，因此直到现在，一个外源 DNA 导入细胞后能否从头生成一条功能染色体仍不清楚（首先不插入宿主染色体中）。

因此，通过导入纯化的 DNA，能提供从头合成的人工哺乳动物染色体，将非常理想。

最后，在削短染色体和染色体片段上增加新的 DNA 序列（例如：治疗基因）十分困难。

因此，如果能在体外构建载体，载体上插入新的 DNA 序列的位置是确定而有效的，那将非常理想。

Sun 等（自然，遗传学，8: 3341 (1994)）报道设计了一种以病毒为基础的载体应用于人类细胞中。这种载体叫作“人类合成的游离染色体”。然而，载体依赖于 EBNA-1 的存在，这是一种有毒和有免疫原性的病毒蛋白质。而且载体依赖于病毒复制起点，而不是天然的哺乳动物染色体的复制起点。同时，这条“染色体”不包含功能着丝粒或者端粒 DNA，有丝分裂时不能形成动粒，因此这样的载体不能以一种可控的方式进行分离。最后，不同于内源染色体，这些载体在细胞中拷贝数增加，每个细胞约 50—100 个拷贝。基于定义哺乳动物的染色体的标准，这个载体叫做“人类人工染色体”不适当，因为它通过不相关的机制具有不同的性质和功能。

因此 仍然需要完全的合成或者人工染色体，他们来自于能在体外操作的 DNA，同时，转染进细胞后有功能染色体结构，且以可控方式指导基因表达。

在人类人工微染色体和基因治疗载体的发展构建中，克隆大的，高度重复 DNA 的能力是一个重要的步骤。此外，微生物中重复 DNA 的稳定克隆对制作哺乳动物染色体的高分辨率物理图谱非常重要。

有许多克隆系统能促进在微生物中克隆，扩增外源 DNA。已成功地应用质粒，噬菌体和酵母人工染色体(YACs)克隆许多哺乳动物的 DNA 序列。然而，某些类型的重复 DNA 似乎在這些载体中结构不稳定(Schalkwyk 等, Curr. Opin. Biotechnol. 6(1): 3793(1995); Brutlag, D. 等 细胞, 10: 509-519(1977))。这导致基因组物理图谱上的空缺，同时也不能应用这些载体来扩增高度重复的哺乳动物着丝粒 DNA。

已构建细菌人工染色体(BACs)可在大肠杆菌中克隆大的 DNA 片段(O Conner 等, 科学, 244(4910): 307-12(1989); Shizuya 等, 美国科学院学报, 89(18): 8794-7(1992); Hosoda 等, 核酸研究, 18(13): 3863-9(1990))。虽然这个系统看起来能稳定地扩增至少 300 kb 以上的哺乳动物 DNA，几乎没有什么独立的哺乳动物 DNA 片段曾被分析(Shizuya 等, 美国科学院学报, 89(18): 8794-7(1992))。同时在 BAC 载体中检测的结构稳定的片段，广泛的特征就是与某种类型的序列相关。因此，这些片段是否包含重复 DNA 元件还不清楚。尤其是，依据限制性位点和 Southern 分析，这些片段不包含 α 卫星 DNA 是明确的。

许多哺乳动物的 DNA 序列看起来似乎在酵母人工染色体(YAC)载体中结构稳定，然而相似长度的某些重复元件却不稳定(Neil 等, 核酸研究, 18(6): 1421-8(1990))。对 YAC 系统来源的 DNA 性质的了解表明大的系列重复单位在遗传上是不稳定的，即使在相似长度的非重复 DNA 能稳定的条件下。因此，在 BAC 载体上，象 α 卫星 DNA 中发现的大(>100bp)系列重复序列的结构稳定性不能用合理的必然性来预测。此外，即使某系列的 α 卫星族在 BAC 载体中结构上稳定，但可以推定，那些有足够长度及序列组成以至能促进着丝粒功能的系列能否在这个载体中稳定扩增仍不清楚。

与所引用的方法不同，本发明的几个实施方案描述了有确定结构及组成的预先构建的染色体载体，其可直接导入病人细胞。因为本发明中所描述的载体不依赖于插入内源染色体或者尔后的扩增，因此细胞中构建体的结构与转染前本质上是相同的。

与所引用的方法不同，本发明描述的载体由确定的序列组成。而且，构建这些合成的染色体的方法使得可以迅速鉴别其他功能重要序列。

与所引用的方法不同，本发明是发明人首次证明可通过将纯化 DNA 导入哺乳动物细胞能重新构建人工哺乳动物染色体。

与所引用的方法不同，因为本发明中所描述的载体在体外合成，插入新的 DNA 序列是简单而有效的。

发明概要

本发明的目的是提供单一和杂交合成族 DNA (特别是 α 卫星 DNA) 的方法。本发明的另一个目的是提供克隆、繁殖和稳定重组产生重复 DNA (特别是天然存在的或合成的 α 卫星族) 的方法。

因此，本发明涉及稳定克隆大重复 DNA 序列的方法、包含以上族的载体和用这些载体转化的宿主。

本发明发明人已开发了产生大量长度多达 736 kb 的纯化的完整 α 卫星族的方法。通过这些族与端粒 DNA 和人类基因组 DNA 序列一同转染进人类细胞，产生了在缺少选择的情况下显示出高度有丝分裂稳定性的若干全合成人染色体。

与以前试图产生人工哺乳动物染色体的方法不同，这种方法不依赖于现有的内源染色体的修饰。此外，在内源染色体内它不产生多重整合事件。这些染色体在染色体外形成和保持，因此避免了整合到内源染色体。

合成染色体形成相对高的频率和与染色体形成相关联的其它基因组重排的缺乏使得本发明发明人制备的合成染色体可以在异源基因表达和基因治疗中用作有效的载体。

因此，本发明基于发明人发现，借助于单独分离的纯化的 DNA，合成或人工染色体在细胞被从头产生(从纯化的 DNA)，并作为染色体外因子被保持。这种染色体保持天然哺乳动物染色体的必需功能之处在于，它作为非整合的构建体被稳定地保持在单个的哺乳动物细胞(没有选择性压力)中，就象天然存在的染色体一样被遗传。对于线性的染色体，这表示着丝粒、端粒和复制起点功能。

因此，本发明涉及一种合成或人工哺乳动物染色体。该染色体从分离的纯化的 DNA 产生。分离的纯化的 DNA 被转染进哺乳动物细胞。在没有整

合进内源染色体的情况下，其形成功能性染色体。这种染色体不是来源于原位内源天然存在的染色体。起始材料是分离的纯化的着丝粒 DNA 和允许染色体形成而不整合的 DNA。对线性染色体，包括端粒 DNA。在一个优选的实施方案中，允许染色体形成而不整合的 DNA 是基因组 DNA (来源于有机体天然存在的基因组)。

这样人工哺乳动物线性染色体最好实质上包含着丝粒、端粒和基因组 DNA。在一个实施方案中，人工染色体是环形染色体。在这种情况下，不存在端粒 DNA，因为对复制染色体末端其不是必需的。

所说的基因组 DNA 是次基因组 DNA 片段，其是限制酶消化片段、通过对基因组 DNA 的机械剪切产生的片段或体外合成的片段。该基因组 DNA 的起始材料(即转染的)可以是异源片段的混合物(例如限制消化物)，或可以是克隆的片段(同源的)。

着丝粒 DNA 包括指导或支持动粒形成从而能够使得进行适当的染色体分离的 DNA。在活性功能性着丝粒上的着丝粒 DNA 在有丝分裂期间与 CENP-E 相关联，如免疫荧光或免疫电子显微镜法所说明的。"相关联"意指着丝粒 DNA 和 CENP-E 由荧光原位杂交 (FISH) 免疫荧光共-定位。

端粒 DNA 包括提供端粒功能的串联重复的 TTAGGG。即线性 DNA 分子的复制末端。端粒 DNA 作为一个可有可无的在线性染色体是所需的时所使用的组分被包含。本文通过将"端粒的 / "端粒"放入括号中表示。

在转染之前，DNA 可以是裸露的、用一种或多种 DNA-凝聚剂凝聚的或用一种或多种 DNA-结合蛋白包被的。

本发明也涉及一种人工哺乳动物染色体，其是通过向哺乳动物细胞引入以上分离的纯化的 DNA 片段产生的。在一个优选的实施方案中，所述方法使用实质上包含着丝粒、端粒和基因组 DNA 的 DNA。

各种片段可以单独地转染或者在转染之前连接一种或多种。这样，着丝粒(端粒)和基因组 DNA 被单独地(unligated)引入，或者一种或多种分离的纯化的 DNA 相互连接。

本发明也涉及包含人工哺乳动物染色体的哺乳动物细胞和包含人工哺乳动物染色体的组合体。

本发明也涉及以上描述的分离的纯化的 DNA，当引入到哺乳动物细胞

中时，其形成人工哺乳动物染色体。在优选的实施方案中，分离的纯化的 DNA 实质上包含着丝粒、端粒和基因组 DNA。

本发明也涉及包含纯化的 DNA 的哺乳动物细胞和包含纯化的 DNA 的组合物。

本发明也涉及包含纯化的 DNA 的载体。

本发明也涉及包含所述载体的哺乳动物细胞和包含所述载体的组合物。

本发明也涉及以上描述的分离的纯化的 DNA，其是通过组合一种或多种以上所描述的 DNA 的方法产生的。在优选的实施方案中，所述 DNA 包括：(1) 着丝粒 DNA，(2) 端粒 DNA，(3) 基因组 DNA。DNA 可以是未连接的或一种或多种可以是相互连接的。

本发明也涉及通过把以上所述的纯化的 DNA 引入哺乳动物细胞制造人工哺乳动物染色体的方法。

本发明也涉及使 DNA 能够形成人工染色体的方法，该方法包括体外组合以上所描述的 DNA。

本发明也涉及通过把纯化的 DNA 引入哺乳动物细胞并使染色体复制，在哺乳动物细胞中繁殖人工染色体的方法。

在一个优选的实施方案中，本发明也涉及通过从人工哺乳动物染色体表达异源基因在哺乳动物细胞中表达所述基因的方法。

这样，本发明也涉及通过在人工染色体上包括进所需基因以表达兴趣基因来提供所需基因产物的方法。在优选的实施方案中，本发明提供了一种基因治疗的方法，其通过在人工哺乳动物染色体上包括进异源治疗性 DNA，以便对包含所述染色体的哺乳动物产生治疗效果。

在本发明的一个优选的实施方案中，着丝粒 DNA 是 α -卫星 DNA。

在本发明的一个优选的实施方案中，人工哺乳动物染色体完全来源于人类 DNA 序列，并且在人类细胞中是功能性的。

附图简要描述

图 1. 本发明的方法的一个示意性图。数字"1-16"代表 α 卫星 DNA (约 171 bp) 的单体单位的 1-16 个拷贝，在线性簇中串联排列。"X"代表在携带

所说族的载体的骨架中在扩展至所需大小期间的所需限制酶位点。

图 2. 百分重组体(在 50 代之后)和族大小(kb)的相互关系的图解。

图 3. 用于产生大的头-对-尾串联的 α 卫星DNA族的方法。pVJ104-Y α 16 以 BamHI 和 SjiI 线性化, 经脉冲电场凝胶电泳(PFGE)纯化。同样地, pBac-Y α 16 以 BamHI 和 BglII 线性化, 经 PFGE 纯化 α 卫星族。A) 在连接酶, BamHI 和 BglII 存在下将纯化的族一起温育。因为 BamHI 和 BglII 是互补/非同裂酶突出端, 形成 BamHI/BglII 接合的连接事件(头-对-尾连接正是这种情况)将消除两种位点。这样, 头-对-尾接合将阻止被 BamHI 和 BglII 切割。相反, 头-对-头, 或尾-对-尾的连接事件将分别再产生 BamHI 或 BglII 位点。因为存在 BamHI 和 BglII, 这些连接产物将被切开产生其组成单体(或头-对-尾多聚体)。通过控制连接酶的量, 温育时间和 DNA 的浓度, 必要时可以变化头-对-尾产物的长度。B) 连接后, 由 PFGE 分析产物。第 1 道, 分子量标准(NEBL Midrange II 标记); 第 2 道, 在 BamHI 和 BglII 存在 4 小时下连接的 Y α 16 (BamHI/BglII 片段); 第 3 道, 在 BamHI/BglII 存在下连接 12 小时的 Y α 16 (BamHI/BglII 片段), 第四道, 在 BamHI 和 BglII 存在下模拟连接的 Y α 16 (BamHI/BglII 片段); 第 5 道, 无限制酶连接 12 小时的 VK75 (BssHII 片段); 第 6 道, 在 BssHII 存在下连接 12 小时的 VK75 (BssHII 片段); 第七道, 模拟连接的 VK75 (BssHII 片段)。连接产物的分子量显示在左边。注: 虽然这些样品运行在相同的凝胶上, 但在第 4 和 5 道之间去除了若干不相干的道。

图 4. 用以制造合成染色体的策略。

图 5. 克隆 22-7 和 22-13 的合成染色体的荧光原位杂交(FISH)分析。收获细胞, 滴加到载玻片上, 如实验方法(参见本文实施例)中所述杂交到 Y α 卫星 DNA 上。用得克萨斯红抗生物素蛋白检测生物素化的探针, 并用两层生物素化的抗生物素蛋白和得克萨斯红抗生物素蛋白扩增。A) 克隆 22-7 的中期扩散的 DAPI 图象。B) 除了 α 卫星探针用三立方滤器可视化外, 其它与 A) 同。C) 克隆 22-13 的中期扩散的 DAPI 图象。D) 除了 α 卫星探针用三立方滤器可视化外, 其它与 C) 同。在每种情况之下, 合成染色体用一白箭头表明。

图 6. 克隆 22-6 和 23-1 的合成染色体的 FISH 分析。收获细胞, 滴加

到载玻片上，如实验方法(参见本文实施例)中所述杂交到 Y α 卫星 DNA(克隆 22-6)或 17 α 卫星 DNA(克隆 23-1)上。用得克萨斯红抗生物素蛋白检测生物素化的探针，并用两层生物素化的抗生物素蛋白和得克萨斯红抗生物素蛋白扩增。A) 克隆 22-6 的中期扩散的 DAPI 图象。B) 除了 α 卫星探针用三立方滤器可视化外，其它与 A) 同。C) 克隆 23-1 的中期扩散的 DAPI 图象。D) 除了 α 卫星探针用三立方滤器可视化外，其它与 C) 同。在每种情况之下，合成染色体用一白箭头表明。在 D) 中，黄箭头表明在整合位点的 C 组染色体的位置。

图 7. 克隆 22-11 和 17-15 的合成染色体的 FISH 分析。收获细胞，滴加到载玻片上，如实验方法(参见本文实施例)中所述杂交到 Y α 卫星 DNA(克隆 22-11)或 17 α 卫星 DNA(克隆 17-15)上。用得克萨斯红抗生物素蛋白检测生物素化的探针，并用两层生物素化的抗生物素蛋白和得克萨斯红抗生物素蛋白扩增。A) 克隆 22-11 的中期扩散的 DAPI 图象。B) 除了 α 卫星探针用三立方滤器可视化外，其它与 A) 同。C) 克隆 17-15 的中期扩散的 DAPI 图象。D) 除了 α 卫星探针用三立方滤器可视化外，其它与 C) 同。在每种情况之下，合成染色体用一白箭头表明。

图 8. 在包含合成染色体的克隆中存在的转染 α 卫星 DNA 的量的测定。A) 收获总基因组 DNA，消化，如实验方法中的描述电泳。第 1 道，HT1080；第 2 道，克隆 22-6；第 3 道，克隆 22-7；第 4 道，克隆 22-11；第 5 道，克隆 22-13；第 6 道，克隆 23-1。B) 显示了各克隆合成 Y α 卫星 DNA 估计的量。注：克隆 23-1 用 17 α 卫星 DNA 转染，因此不包含合成的 Y α 卫星 DNA。

图 9. 在有丝分裂期间 CENP-E 与合成染色体相关联。如实验方法中的描述，在从含有合成染色体的克隆收获的中期染色体上进行免疫荧光分析。A) 克隆 22-11 的 DAPI-染色的染色体。B) 除了抗 CENP-E 抗体的位置用三立方滤器可视化外，其它与 A) 同。C) 克隆 23-1 的 DAPI-染色的染色体。D) 除了抗 CENP-E 抗体的位置用三立方滤器可视化外，其它与 C) 同。在每种情况之下，合成染色体用一白箭头表明。

图 10. 在没有选择生长 70 天后克隆 22-11 的 X-Gal 平板染色。如本文实验方法中的描述收获细胞并染色。A) HT1080；B) 克隆 22-11。在克

隆 22-11 中存在蓝细胞而在 HT1080 中不存在表明在这些细胞中仍然表达 β -geo。

优选实施方案的详尽描述

本发明已发现功能性哺乳动物染色体可以从引入哺乳动物细胞的纯化的 DNA 构建。对各种用途而言利用这些染色体具有若干优点。

首先，因为它们被形成和自我复制，通过插入到宿主基因组中它们将不引起插入突变。

第二，由于大的转移载体(在兆碱基范围内)，具有容纳包括其所有调节元件的大基因的全部内容的能力。其本身可以包括几兆碱基的 DNA。

第三，因为一些遗传疾病是一个以上基因缺陷的结果，由于可以容纳多于一个基因的哺乳动物人工染色体。

第四，染色体是稳定的，这样可以在许多细胞分裂上提供治疗益处。

第五，染色体是非致免疫的。

因此，本发明提供了一种在微生物中产生结构上完整的高度重复的 DNA 的方法，其被用于构建人工染色体。限定长度、组成、方向和定相的族是可能的。"适当的定相"指在族中任何高级的重复的精确长度和方向不被族构建改变(从天然存在的序列改变)，在重复单位的接合时也不存在非重复 DNA。例如，对于 alphoid 序列，高级重复单位的长度和方向完全相同于天然存在的高级重复单位，在高级重复单位的接合时不存在非-alphoid 序列，除了产生限制位点的修饰碱基外。

因此，提供了一种用于克隆重复串联族 DNA 的方法，其中制备第一 DNA 单位，使得该 DNA 单位的相对的端包含互补的但非同裂酶的限制位点。将这种 DNA 连接进载体，该载体在限制位点之一线性化。如步骤 a) 制备第二 DNA 单位，然后与第一单位串联连接，以形成定向重复族。将这一族转化进宿主细胞，尤其是细菌宿主细胞，选择包含族的和稳定的克隆。以载体线性化开始，重复这些步骤，直到达到所需的族大小。

本发明的定向克隆方案在图 1 中说明，其中说明了克隆的 α 卫星 DNA 高级的重复。如图所示，本发明的方法利用"建造"方法，其中较短的单位(优选地是高级的重复)加合在引起产生较长的串联族重复单位。所述单位以导

致限定的方向的方式相互加合，其由两个不同的限制位点(每个在各重复单位的各端)建立。优选地，基于串联族的重复单位(尤其高级的重复单位)包含互补但非同裂酶的限制位点。这样的末端可以利用聚合酶链式反应之类的方法设计到所说单位中。在本发明的方法中，互补末端旨在包括互补的突出末端和平末端两者。

在一个优选的实施方案中，DNA 族是 alphoid DNA。如实施例 1 所示，聚合酶链式反应可以用来扩大一个单一 2.7 kb DNA alphoid 单位(实际长度 2.712 kb)，以便从人类染色体 17 在高级的 α 卫星重复的相对的末端产生互补限制位点(BamHI 和 Bgl II)。可以用聚合酶链式反应-介导的定点诱变修饰高级重复的末端，以便在各重复的相对的末端产生互补的限制位点。然后将修饰的高级重复克隆进小-F 克隆载体 pBAC108L(Shizuya, H. 等, 美国科学院学报, 89: 8794-8797(1992), 本文一并参考)中。将这些互补限制位点用于连接非互补侧翼限制位点，以直接克隆来源于单一修饰的高级重复的合成族 α 卫星 DNA(图 3)。合成的族可以从任何高级的 alphoid 重复(包括来源于从染色体 17、Y 染色体或其它染色体的这样的 alphoid DNA)产生，此外，可以制备由两种染色体高级重复组成的杂交族。在一个优选的实施方案中，所说 DNA 是人类 DNA。

长达 200-215 kb 的族在本发明的载体与宿主中是稳定的。在一个实施方案中，构建了一个 87 kb-215kb 长的族。在一个优选的实施方案中，构建了一个至少 100 kb 长的族。在一个非常优选的实施方案中，构建了一个至少 140 kb，尤其是至少 174 kb 长的族；174 kb 的族超过最小已知的观察到的功能性 α 卫星族的长度。

在产生这样的位点中有用的互补但非同裂酶限制酶的例子包括：SalI 和 XhoI；MunI 和 EcoRI；AflIII 和 NcoI/StyI(同裂酶：两者之一都是 either 逐-非同裂酶配偶体的配偶体)；NheI 和 XbaI 和 StyI/AvrII(同裂酶)和 SpeI(任何组合)；ClaI 和 BstBI 和 AccI(任何组合)；MluI/AflIII(同裂酶)和 BssHII 和 AscI；和 NotI 和 EagI。BclI 是 BamHI 与 BglIII 两者的互补/非同裂酶。

然后，用例如(1)BamHI 和 SfiI 或(2)BglIII 和 SfiI 消化扩增的 DNA。在用能够分离这样的 DNA 的物理的方法(例如，凝胶电泳)分离消化的 DNA

中的带后，切下消化物的 DNA 带，并与从其它消化物切下的 DNA 带连接。在上述例子中，因为 BglII 和 BamHI 产生相容的突出端，SfiI 产生仅能以特定的方向转移的不对称的突出端，所以将侧翼于这些位点的 DNA 连接到载体 DNA 上，以产生头到尾方式的串联二聚体。然后这种 DNA 可以转化进微生物。

在这一策略的第二变体中，平端切割限制酶可以替代 BamHI 和 BglII。例如，SmaI 和 EcoRV 可以分别替代 BamHI 和 BglII。然后如上所述进行消化和片段分离。对两种平端和互补/非同裂酶变异相同的这一策略的一个重要特征是，如果需要，族的生理学定相可以精确地保持。可以使用的其它平端切割酶的例子包括 SspI、StuI、ScaI、PmlI、PvuII、Ecl136II、NaeI、EheI、HincII、HpaI、SnaBI、NruI、FspI、DraI、MscI、Bst107I、AluI、Asp 700/XmnI、AviII、BbrPI、Bst1107、Eco47III、DpnI、HaeIII、HindII、NmlI、MluMI、MvnI、RSaI、SwaI、Bsh 1236I、Eco72I、Pall 和 SrfI。

可以采用如实施例 2 描述的简单的生长和稀释实验测定包含大的微生物中合成 α 卫星 DNA 串联族的克隆质粒，例如，可以用传代 50 代和后续的质粒 DNA 结构完整性分析测定结构稳定性。可以通过限制分析和琼脂糖凝胶电泳分析质粒结构。这些克隆极少甚至没有观察到重组，表明定向克隆方案可以用于构建并繁殖在小-F 克隆载体(如 pBAC 108L 载体)和合适的大肠杆菌宿主中的合成 α 卫星族。

本发明的方法可以用于构建任何所需的重复 DNA 单位。例如，可以克隆来自任何真核染色体的 α 卫星 DNA(尤其是人类 DNA)。大串联族高度重复 DNA 的其它例子包括免疫球蛋白 DNA 基因座，异源核染质重复区和端粒。

本发明的定向克隆方法不需要存在多形的限制位点(如克隆内源(非修饰的)族时所需的)。甚至当存在时，这些位点不允许控制族的确切大小。此外，通过利用本发明的方法中使用的高级重复(参见图 1)，接着加倍其大小，整个族的确切的序列是已知的，由于原来的高级的重复的序列是已知的。

当克隆内源族(如内源 α 卫星族)，人们不能确信给定族的精确组成。尤其是，族中非重复 DNA 的间断可以明显影响在大肠杆菌中的稳定性。此外，

对作为基因治疗合适的载体，人们必须知道向这种治疗的接收者所提供的载体的确切的序列。本发明的方法消除了这种疑虑，并且使得技术人员迂回测序天然 α 卫星族，有利于从已知的重复序列从头构建有用的族。

在结构上有差别的高级的重复相对于更为同源的族而言一般显示出在大肠杆菌中增加的结构稳定性。这样，通过利用按照本发明的方法的同源的合成族，可以获得载体中的重复 DNA 的最小限度稳定性的精确测定。

任何所需细菌宿主(其中载体稳定地被保持)可以用作宿主。尤其是当利用 BAC 载体和 BAC 系统时，大肠杆菌是有用的。

合成的 α 卫星族可以以下列方式用于合成的人类染色体的构建：(1) 将合成的 α 卫星族转染到人类或其它哺乳动物细胞系中；(2) 将合成的 α 卫星族与随机克隆的人类 DNA 或者特异性 DNA 片段一起转染到人类或其它哺乳动物细胞系中；(3) 将合成的 α 卫星族与未连接的特异性染色体组分共-转染进入人类或其它哺乳动物细胞系中，所说组分如端粒 DNA、基质连接区和/或提高包含 α 卫星的游离基因 DNA 有丝分裂稳定性的其它染色体座。这些非连接形式的组分的共-转染使得可以转染细胞系，以构建无限量的结构排列，使核分裂最稳定的形式得以保持，而不稳定的形式有时失去。其后可以用标准的方法和过程收获稳定的构造。可以分离在缺少选择压力时显示出有丝分裂稳定性的那些构建体，且其后用于制备基因治疗载体，该载体包含一种或多种治疗有用实体，如基因、核酶或反义转录物。

因此，本发明涉及实质上包含着丝粒、基因组和可有可无的端粒 DNA 的合成或人工哺乳动物染色体。在一个可替性实施方案中，人工染色体是一种环形染色体。在这种情况下，不存在端粒 DNA，由于对复制染色体末端而言它不是必需的。所说染色体至少具有在哺乳动物细胞中提供必需染色体功能的 DNA 序列。

所说的基因组 DNA 是次基因组 DNA 片段，其选自机械剪切片段的限制酶消化片段和体外合成的片段。染色体基因组 DNA 组分可以来源于次基因组片段的混合物(例如限制酶消化物)，或者来源于克隆的片段。

基因组 DNA 的功能是双重的。DNA 表达基因产物或引起基因产物的表达(例如，通过具有调节功能)，所说 DNA 使得可以在细胞中从纯化的 DNA 形成人工染色体，无须在细胞中将纯化的 DNA 整合进内源染色体，人工染

色体也包含着丝粒 DNA 和可有可无的端粒 DNA。基因组 DNA 可以来源于任何有机体，并且可以是任何大小的。

形成合成的哺乳动物染色体组分的基因组 DNA 可以来源于哺乳动物源(除染色体在其中复制的细胞所来源的哺乳动物外)。例如，小鼠基因组 DNA 可以提供给人类细胞，人类基因组 DNA 可以提供给其它哺乳动物细胞。此外，它可以来源于不同于着丝粒或端粒源的源。

另外，基因组 DNA 的功能可以用任何有机体(包括原核有机体)的基因组 DNA 和用体外合成的 DNA(相应于天然存在的序列，部分同源于天然存在的序列或者完全是非同源的)有力地进行例证。

着丝粒 DNA 包括指导或支持动粒形成从而能够使得进行适当的染色体分离的 DNA。在活性功能性着丝粒上的着丝粒 DNA 在有丝分裂期间与 CENP-E 相关联，如免疫荧光或免疫电子显微镜法所说明的。"相关联"意指着丝粒 DNA 和 CENP-E 由荧光原位杂交(FISH)免疫荧光共-定位。

在本发明的一个优选的实施方案中，着丝粒 DNA 是 α 卫星 DNA。然而，任何功能性着丝粒 DNA(尤其是重复 DNA)使得可以用本文描述的方法制造人工染色体。

本发明发明人已建立用于产生大的 α 卫星族的体外方法。以前，没有使大于 200 kb 的在结构上完整的 α 卫星 DNA 定量纯化(对于哺乳动物细胞的转染必需的)的可供使用的方法。通过利用这些方法，控制量的 α 卫星 DNA 可以体外产生。如本文的描述，必要时，通过经验控制连接酶的量、培养时间和 DNA 的浓度，可以变化最终产物的长度。

然而，本发明不限于来源于 α 卫星 DNA 的着丝粒 DNA。由本发明发明人建立的体外方法可以运用于任何如本文所述起作用的着丝粒 DNA，尤其是重复 DNA 的。

此外，对于着丝粒形成，不需要整个 α 卫星重复。这样，着丝粒 DNA 也可以包括 α 卫星衍生物和类似物，例如， α 卫星或有关的卫星 DNA 的子-单体区。

在 alphoid 单体内代表蛋白质结合位点的亚区可以连接在一起，以产生由小的重复单位组成的功能性着丝粒。这一实施方案的功能由小鼠-人类杂交体的数据显示。

在鼠物种 *M. musculus* 中, 小的卫星 DNA 包含 CENP-B 盒, 似乎是 α 卫星 DNA 的功能等同物。有趣的是, 在 *M. musculus* 中, 小的卫星重复单位仅仅 120 bp, 与 CENP-B 盒外的 α 卫星 DNA 没有明显的同源性。尽管在重复大小和序列中不同, 人类染色体在小鼠/人类杂交体中有效分离。这说明丝粒重复单位大小与序列可以变化, 而不消除着丝粒功能。

鼠物种 *M. caroli* 明显缺乏小卫星 DNA (Kipling 等, 分子细胞生物学, 15: 4009-4020 (1995))。在这一物种中, 功能性 α 卫星等价物似乎是 79 bp 卫星序列, 包含 CENP-B 盒 (存在 60 bp 序列, 其 97 % 同源于 79 bp 序列, 但缺乏 CENP-B 盒)。在 *M. musculus* 和 *M. caroli* 杂交中, 两物种的染色体通常在相同的细胞内分离。这表明小卫星和 79 bp 卫星序列在有丝分裂期间被相同的纺锤体识别。这样, 不同的着丝粒重复大小可以是功能性的。

因为 α 卫星, 小卫星和 79 bp 卫星重复是不同大小和功能的, 本身绝对重复大小不是着丝粒 DNA 的功能决定性的。另外, 因为在这些着丝粒重复之间仅有有限的序列同源性, 所以很可能在代表蛋白质结合位点的重复之内的亚区是重要的功能性组分。

这样, 在本发明的一个实施方案中, 在 α 卫星 DNA 之内的着丝粒 DNA 含有亚区。在一个优选的实施方案中, 着丝粒 DNA 是由串联连接由序列 5' aTTCGttggAaaCGGGa 3' (SEQ ID NO: 1) 限定的 CENP-B 盒组成的, 在以上序列中由大写/粗体字母表示的碱基对 CENP-B 结合来说是最重要的, 由小写字母表示的碱基可以由其它碱基替代。

在其它实施方案中, 其它物种 alphoid 等同物用于着丝粒 DNA。人类和其它哺乳动物的染色体已显示出在其它物种的细胞中有效地分离 (如种间肉体细胞杂交体所说明的)。这些杂交体的例子包括小鼠 $\times$ 人类, 仓鼠 $\times$ 人类, 大鼠 $\times$ 人类, 仓鼠 $\times$ 小鼠, 大鼠 $\times$ 小鼠和小鸡 $\times$ 人类。人类染色体在小鸡细胞中分离的能力 (Dieken, E., 等, Nature Genet. 12: 174-182 (1996)) 表明人类着丝粒 DNA 在非哺乳动物物种 (即禽类) 中也是功能性的。

基于从杂交物种杂交体的观察结果, 很明显来源于一个物种的染色体在其它物种中是功能性的。因此, 合成染色体可以在人类细胞中利用其它哺乳动物 (和禽类) 的着丝粒替代 α 卫星 DNA 或者与 α 卫星 DNA 一起产生。相

反， α 卫星 DNA 可以在其它哺乳动物(和禽类)物种中用作着丝粒 DNA 源。

这样，在本发明的另一个实施方案中，将基因组(端粒)DNA 与 *M. musculus* 小卫星 DNA，*Mus caroli* 79 bp 卫星 DNA 或其它哺乳动物的类似的序列一同转染进细胞。在另一个实施方案中，将端粒和基因组 DNA 与来源于禽类(avian)细胞的着丝粒 DNA 一同转染进细胞。

实质上，在有丝分裂期间与 CENP-E 相关联的着丝粒 DNA 体现了本发明的一个方面，即包括使用异源于宿主细胞的着丝粒序列和其它合成染色体组分的一个方面。只要在染色体中的着丝粒序列在有丝分裂期间与 CENP-E 相关联，哺乳动物细胞的功能性染色体就预期与基因组序列和端粒序列无关，并且对那一物质，与特异性着丝粒序列无关。

端粒 DNA 可以来源于保持端粒功能的任何 DNA 序列(来源于任何所需物种)。在哺乳动物和其它脊椎动物中，在染色体末端最丰富的和保守的序列是 TAGGG，这形成长度在 2 到 20 千碱基之间的族。人类端粒 DNA 由约 5 千碱基的重复 TTAGGG 组成的，这种序列的小的伸展在向哺乳动物细胞系引入线性分子后足以引起端粒形成(Huxley, C., 基因, 1: 7-12(1994))。简单的(TTAGGG)族足以提供端粒功能所需的人工染色体。因此，端粒 DNA 包括串联族六聚体 TTAGGG。当需要形成线性染色体时，就包括端粒 DNA。

在 Huxley 等，生物技术 12: 586-590(1994) 中讨论了端粒、着丝粒和复制起点。

本发明也涉及纯化的 DNA 分子，其实质上包含本文所描述的着丝粒、基因组和可有可无的端粒 DNA。

在一个实施方案中，纯化的 DNA 是裸露的 DNA。

在另一个实施方案中，纯化的裸露的 DNA 与一种或多种 DNA 凝聚剂凝聚，为了稳定化它使其抵抗剪切，在转染之前凝聚纯化的 DNA 可以是有利的。通过在转染前凝聚纯化的着丝粒(端粒)和基因组 DNA，其对由转染操作引起的结构损伤具有更强的抵抗力。这样，在本发明的一个实施方案中，纯化的着丝粒(端粒)和基因组 DNA 在转染之前与一种或多种 DNA 凝聚剂凝聚。在这一方面，多聚阳离子已显示出物理凝聚高分子量 DNA，并且保护它抵抗机械剪切(Kovacic 等，核酸研究, 23: 3999-4000(1995); Widom 和 Baldwin, 分子生物学杂志, 144: 431-453(1980); Widom 和 Baldwin,

生物聚合物, 22: 1595-1620(1983))。因此, 在另一实施方案中, 纯化的 DNA 以多聚阳离子化合物凝聚。多聚阳离子的例子包括聚-赖氨酸、聚-精氨酸、精脒、精胺和氯化六胺钴。

在一个可替实施方案中, 本发明包括用蛋白质预包被的 DNA。用 DNA-结合蛋白(如组蛋白, 非组蛋白染色体蛋白质, 端粒结合蛋白和/或着丝粒结合蛋白)包被 DNA 也可以是有利的。这一预包被预期有若干理想的结果。首先, 它导致 DNA 的凝聚, 保护高分子量 DNA 使之免受剪切。第二, 它通过阻断核酶结合到 DNA 上抑制转染 DNA 的核酶降解。第三, 由于以上列出的各蛋白质包含核定位信号, 转染后预包被的 DNA 可以更有效地进入核。通过在转染之前用 DNA 结合蛋白预包被着丝粒(端粒)和基因组 DNA, 我们预期会增加转染和合成染色体形成的效率。

这样, 在另一个本发明的实施方案中, 纯化的着丝粒(端粒)和基因组 DNA 在转染之前用 DNA 结合蛋白包被。DNA-结合蛋白的例子包括组蛋白, 非组蛋白染色体蛋白质, 转录因子, 端粒结合蛋白和/或着丝粒结合蛋白。

DNA-结合蛋白也可以经它们对 DNA 的亲合性鉴别和纯化。例如, 可以在滤膜杂交实验中揭示 DNA 结合, 在所说实验中, 使蛋白质(通常是标记的以利于检测)结合到固定化在滤膜上的 DNA 结合或者反之, 其中 DNA 结合位点(通常是标记的)结合到蛋白质已固定化在其上的滤膜上。这样结合的序列特异性和亲合性用 DNA 保护测定和凝胶阻滞测定揭示。这样的蛋白质的纯化可以用序列-特异性 DNA 亲和色谱进行, 例如采用结合上域的 DNA 衍生化的树脂的柱色谱, DNA 结合蛋白的蛋白酶促降解可以用于揭示保持结合能力的 DNA。

这样, 本发明涉及一种人工哺乳动物染色体, 其是通过用本文描述的纯化的 DNA 转染哺乳动物细胞, 使细胞体外完全重构 DNA 产生的。

这样, 本发明涉及一种人工哺乳动物染色体, 其是通过用纯化的裸 DNA 转染哺乳动物细胞产生的, 该 DNA 实质上包含着丝粒 DNA(端粒 DNA)和基因组 DNA, 如本文所描述的。

这样, 本发明也涉及一种人工哺乳动物染色体, 其是通过用纯化的凝聚 DNA 转染哺乳动物细胞产生的, 该 DNA 实质上包含着丝粒 DNA(端粒 DNA)和基因组 DNA, 如本文所描述的。

本发明也涉及一种人工哺乳动物染色体，其是通过将纯化的包被 DNA 引入哺乳动物细胞产生的，该 DNA 实质上包含着丝粒 DNA (端粒 DNA) 和基因组 DNA，如本文所描述的。

本发明也涉及由体外组合本文所描述的分离的纯化的和基因组 DNA (端粒 DNA) 的方法产生的纯化的 DNA。

本发明也涉及由体外组合本文所描述的分离的纯化的着丝粒 DNA (端粒 DNA) 和基因组 DNA 的方法产生的纯化的凝聚 DNA。另外，单独的 DNA 组分可以预凝聚然后组合。

本发明也涉及由体外组合本文所描述的分离的纯化的着丝粒 DNA (端粒 DNA) 和基因组 DNA 并添加 DNA-结合蛋白的方法产生的纯化的包被 DNA。另外，单独的 DNA 组分可以预包被然后组合。

以上所述的纯化的 DNA 可以包括未连接的着丝粒(端粒)和基因组 DNA。另外，以上所述的纯化的 DNA 也可以包括着丝粒(端粒)和基因组 DNA，其中这些 DNA 中一种或多种是相互连接的。

本发明也涉及包含以上所述的纯化的 DNA 的组合体。该组合体可以包含有利于 DNA 进入细胞的组分。对人工染色体的形成，组合体可以有利于 DNA 摄取进哺乳动物细胞。另外，组合体可以包括有利于 DNA 摄取进细胞的组分，所说的细胞用于繁殖包含所说 DNA 的载体。

本发明也涉及包含以上所描述的 DNA 的载体。该载体可以用于繁殖 DNA，即，在将它们引入到哺乳动物细胞之前扩增以上所述的序列，并形成人工染色体。

因此，本发明也涉及包含以上所描述的 DNA 的载体的组合体。

本发明也涉及包含以上所描述的载体的细胞。

本发明也涉及包含所说的人工染色体的哺乳动物细胞。

本发明也涉及包含以上所述的纯化的 DNA 的哺乳动物细胞。

虽然本发明包括任何哺乳动物细胞，但在本发明的优选的实施方案中，哺乳动物细胞是人类细胞。

在本发明的优选的实施方案中，着丝粒 DNA 是人类 α 卫星 DNA。然而，应当理解， α 卫星 DNA 可以来自任何灵长类。本发明还包括来源于非灵长类哺乳动物的着丝粒 DNA，其中所说的着丝粒 DNA 在有丝分裂期间与 CENP-E

相关联。任何在有丝分裂期间与 CENP-E 相关联的着丝粒 DNA，尤其是重复 DNA (其来源的无关的有机体的) 预期向按照本发明的人工哺乳动物染色体提供功能性的着丝粒序列。这样，人工哺乳动物染色体在人类细胞中起作用，例如，可以包含不是来源于人类而是来源于非人类哺乳动物和甚至来源于非哺乳动物物种 (如禽类) 的着丝粒序列。与 CENP-E 相关联的任何重复 DNA 是潜在的有用的。因此，在本文的教导下，任何着丝粒序列都可以试验作为哺乳动物人工染色体组分的功能。

在具体公开的本发明的实施方案中，着丝粒 DNA 包括大的伸展的 α 卫星族，由重复的端粒序列 (TTAGGG)_n 和由用限制酶 NotI 消化产生的随机基因组片段组成的一个区段。在优选的实施方案中，限制酶将 DNA 消化成片段，其在由 NotI 消化人类基因组 DNA 所产生的片段的范围内，优选地在 10 kb 到 3 μ m 的范围内。所说的限制酶包括但不限于 BamHI、BglI、SalI、XhoI、SfiI、NotI、SrfI、PmeI 和 AscI。

当纯化的 DNA 引入哺乳动物细胞时，这种 DNA 形成一个功能性的合成或人工染色体。这种染色体具有天然存在的哺乳动物染色体的特征。染色体以低的拷贝数存在于细胞中，通常每个细胞一个。染色体是线性的，并包含端粒序列。在有丝分裂期间 CENP-E 与人工染色体相关联，表明功能性动粒的形成。在缺少选择的情况下，染色体是有丝分裂上稳定的。染色体以一种不可检测的整合频率随时间的变化在结构上是稳定的。染色体包含一个或多个转录活性基因。这样，这些染色体不是来源于天然存在的染色体，而是从分离的纯化的 DNA 序列开始体外构建的。

因此，本发明也涉及用于制造人工哺乳动物染色体的方法，该方法包括把以上所述的纯化的 DNA 引入哺乳动物细胞。

可以用大量本领域技术人员已知的方法将 DNA 引入到哺乳动物细胞。这些方法包括但是不限于电穿孔；磷酸钙沉淀法；脂转染；DEAE 葡聚糖、脂质体、受体-介导的胞吞和颗粒运送。染色体或 DNA 也可以用来微注射蛋，胚胎或来自体内的或在体外的细胞。可以用本领域已知的引入技术 (例如脂质体) 用本文描述的染色体或 DNA 转染细胞。在本发明的一个优选的实施方案中，借助于脂转染将纯化的 DNA 引入哺乳动物细胞。

这样，纯化的 DNA 对转染哺乳动物细胞来说是有用的，所说的转染导致

在细胞中从转染的 DNA 形成人工染色体。

DNA 可以在非哺乳动物细胞中单独地繁殖, 或者一种或多种组分连接在一起繁殖。这样, 本发明也涉及连接到用于繁殖的载体上的纯化的 DNA。这样的载体在本领域中是众所周知的, 包括但不限于 pBac108L, p1, pACYC184, pUC19, pBR322, YACs 和粘粒。

本发明也涉及包含人工或合成染色体的哺乳动物细胞。本发明涉及任何哺乳动物染色体或哺乳动物细胞。虽然包括所有的哺乳动物, 但优选的具体例是人类。因此, 本发明的一个优选的实施方案包括包含合成的人类染色体的人类细胞。

所说的 DNA 与染色体已特别地开发来作为基因治疗和其它目的的表达载体。因此, 在本发明的优选的实施方案中, 纯化的 DNA 实质上也是由对于表达所需的基因产物(例如治疗剂)有用的一个或多个 DNA 序列组成。这样, 本发明涉及用于通过把可表达的 DNA 包括在人工染色体中将这种 DNA 引入细胞的方法。DNA 可以是可调节的, 有结构的, 作为基因产物表达的等。在一个优选的实施方案中, DNA 提供基因产物。当转染进哺乳动物细胞时, 在转染后形成的人工染色体具有和表达这些 DNA 序列。

过去 20 年, 重组 DNA 技术已越来越多地用于产生所需的生物材料。编码各种药学上重要的人类基因产物的 DNA 序列已被克隆。这些包括胰岛素, 纤溶酶原激活剂, $\alpha 1$ 抗胰蛋白酶和凝结因子。然而, 本发明包括表达任何和所有所需的药学上和/或生物学上有关的基因产物。

一旦在细胞中, 异源基因产物就在所选择的组织中以产生功能性的基因产物的水平表达。总的共同之处是大多数转染基因的正确组织-特异性表达可以完成。对于正确的组织特异性, 可能重要的是在引入进细胞和形成人工染色体之前除去所有用于克隆兴趣 DNA 序列的载体序列。这样, 兴趣异源基因可以以可控制的方式掺入到人工染色体中, 以便天然存在的序列以其天然存在的构型存在, 并保证组织特异性。

合成染色体可以引入到人类干细胞或骨髓细胞。其它应用对本领域技术人员是显而易见的。

已开发了将载体引入培养的细胞中和引入动物或人类患者组织中的各种方法。用于把载体引入哺乳动物和其它动物细胞的方法包括磷酸钙转

染、DEAE-葡聚糖技术、微注射,脂质体介导技术、基于阳离子脂质的技术、转染、原生质体融合技术、电穿孔等。这些技术对本领域技术人员是公知的,在许多容易得到的出版物中有描述,并有许多综述。一些技术在以下文献中有描述:转录和翻译,实际的方法, Hames , B.D. & Higgins , S.J., 编者, IRL 出版社, 牛津大学 (1984) (其相关技术本文一并参考)和分子克隆,第二版, Maniatis 等, 冷泉港实验室出版社, 冷泉港, NY (1989) (其相关技术本文一并参考)。

在说明书中,参考了各种分子生物学领域技术人员已知的方法。所参考的提出这些已知方法的出版物和其它材料就其相关技术本文一并参考。

提出重组 DNA 技术的一般的原则的一份标准的参考文献是分子克隆, 第二版, Maniatis 等, 冷泉港实验室出版社, 冷泉港, NY (1989)。

定义

所有有关重组体 DNA 技术的术语以专业通用的方式使用,对本领域技术人员是清楚的。

术语"Y α 卫星"和"Ya"可以替换使用,指来源于人类 Y 染色体的 α 卫星 DNA。

术语"17 α 卫星"和"17 α "可以替换使用,指来源于人类染色体 17 的 α 卫星 DNA;

α 卫星 DNA 是存在于人类着丝粒上的串联重复 DNA 序列,并包含大约 170 bp 的碱基单体重复。这种小的重复被组织成显示出对一个或一小组人类染色体特异性的高级单位。

术语"着丝粒的"指被束紧并且是纺锤体的连接位点的染色体区。它在减数分裂和有丝分裂期间对染色体的适当分离来说是必要的,因此是人工染色体的必需组分。着丝粒 DNA 包括指导或支持动粒形成从而能够使得进行适当的染色体分离的 DNA。在活性功能性着丝粒上的着丝粒 DNA 在有丝分裂期间与 CENP-E 相关联,如免疫荧光或免疫电子显微镜法所说明的。"相关联"意指着丝粒 DNA 和 CENP-E 由 FISH 免疫荧光共-定位。

必需染色体功能 在以上的描述和背景中讨论过。这些包括有丝分裂稳定(没有实验选择压力)、实质上 1:1 分离、自我复制的,即,着丝粒、端粒、复制作用的起点。

术语"功能性等同"指起因于不同的 DNA 或者蛋白质序列,但其提供相同的生物功能的遗传功能。

术语"基因产物"指 DNA、RNA、蛋白质或肽。

术语"基因组 DNA"包括一种或多种克隆片段,其来源于限制酶消化物或者其它序列和大小的混合物,例如机械剪切 DNA 或体外合成 DNA。所述 DNA 可以来源于相同的染色体(例如,使用克隆片段时或者当消化纯化的染色体的 DNA 时)或来源于不同的染色体(例如,当基因组限制消化物用于转染时)。

术语"基因组的"指在有机体中天然被发现的基因组 DNA。然而,发明人也认为这种基因组 DNA 的功能可以由其它来源的 DNA(例如在自然界中没有发现的合成 DNA)来完成。这样,如本文所使用的,"基因组"一般也用来指与本文描述的着丝粒(端粒)DNA 一同引入细胞的 DNA。该 DNA 表达基因产物或引起基因产物表达,使得可以在细胞中从纯化的 DNA 形成人工染色体,而无须在细胞中将纯化的 DNA 整合进内源染色体。这样这种 DNA 可以是合成的或来源于任何有机体的,并且可以是任何大小的,只要它含有必需的可表达序列和以上讨论的功能。

因此,除着丝粒 DNA 之外,在本发明中所包括的人工染色体实质上包含这样的 DNA 序列,其表达基因产物或引起基因产物表达,使得可以在细胞中从纯化的 DNA 形成人工染色体,而无须在细胞中将纯化的 DNA 整合进内源染色体。用来提供染色体功能的序列(例如非整合的)和表达序列可以是相同的序列。这样,在发明人考虑的范围内的是可表达的序列也提供了其它功能。另外,提供了染色体功能和表达功能的序列可以是不同的序列和来自不同的来源。

在特别公开实施方案中,基因组 DNA 来源于 NotI 限制性消化物。因此,在一个优选的实施方案中,允许在细胞中从纯化的 DNA 形成人工染色体而无须在细胞中将纯化的 DNA 整合进内源染色体的 DNA 来源于限制性片段,该限制性片段是由用具有 NotI 识别位点(8 个核苷酸)的限制酶消化总基因组 DNA 产生的。然而,完全在发明人所考虑的范围内的是利用天然存在的基因组 DNA 的限制片段和其它片段,其比由 NotI 和可比较的酶产生的那些要小一些。例如,发明人考虑降低 DNA 的大小但保持以上的功能。因此,

在一个非常优选的实施方案中，DNA 被剥离得仅包含对表达一种或多种兴趣基因并且提供使得可以从纯化的 DNA 形成人工染色体而无须将纯化的 DNA 事先整合进内源染色体的功能所必需的 DNA。

这些 DNA 的来源不必相同。这样，表达序列可以来自一种有机体，提供了染色体功能的序列可以来自另一种有机体。此外，一种或两种序列可以是体外合成的，并且不必符合天然存在的序列。在这一方面，序列不必严格地是"基因组"的。对序列的唯一的限制是它们提供以上指出的功能。

术语"异源的"指在天然存在的基因组中尚未发现的 DNA 序列，在其细胞中引入了人工哺乳动物染色体。另外，如果所述的序列被发现，附加的拷贝被认为是"异源"的，因为它们不能以这种形式在天然存在的基因组中发现。如以上所讨论的，异源 DNA 可以同时是所需的表达序列和"基因组 DNA"。"可表达的 DNA 本身不能表达，但使得和引起另一个异源或内源 DNA 序列的表达。例如，在 DNA 是调节的时就是这种情况。

"高级的重复"指自身由较小的(单体)重复单位组成的重复单位。 α 卫星族的基本的组成单位是大约 171 bp alphoid 单体。单体被组织成染色体-特异性高级的重复单位，其也是串联重复的。在给定的高级的重复中的组分单体的数量在小至 2 个(例如人类染色体 1)到大于 30 个(人类 Y 染色体)之间变化。组成型单体相互显示出不同的同源性，从大约 60 %到实质上序列等同。然而，在整个大多数给定的 alphoid 族中，高级的重复保持高的同源性。

术语"哺乳动物染色体"指 DNA 分子或遗传单位，其在哺乳动物细胞中起染色体的作用。

术语"裸露的 DNA"指与任何生物组分(染色体或细胞)组分不相关联的 DNA，其提出以所说的组分与天然存在的染色体相关联，所说的组分例如组蛋白、非组蛋白染色体蛋白质、RNA、转录因子、拓扑异构酶、支架蛋白、着丝粒-结合蛋白和端粒-结合蛋白。这样的 DNA 可以从细胞分离和从非 DNA 染色体组分纯化。另外，这种 DNA 可以是体外合成的。

天然存在的"指出现在自然界中和不是由实验诱导的。

复制起点指 DNA 合成的开始位点。

术语"分离的"指已从细胞移去的 DNA。术语"纯化的"指实质上完全与

细胞的非 DNA 组分分开的分离的 DNA，或者指已体外合成并且实质上完全与可能干涉从 DNA 构建染色体的用于合成的材料分开的 DNA。纯化的 DNA 也可以是 DNA 序列从用以与天然发生联系的 DNA 序列分离的 DNA 序列。

复制子是基因组的片断，在其中 DNA 被复制，并且按定义包含复制起点。

词组 保持所有天然哺乳动物染色体的功能"指染色体作为非整合的构建设没有实验选择压力时稳定地保持在分开的哺乳动物细胞中，至少表明着丝粒，端粒(线性染色体)，复制起点功能和基因表达。

有丝分裂上稳定的 指合成或人工染色体在没有实验选择压力(如药物选择等)下 10 代后保持至少 50%存在于细胞中，最优选地是在 30 代后至少 10%存在于细胞中。优选地，合成染色体在大于 99 %的时间显示出 1:1 分离。

稳定地转化"指包含重复单位的克隆 DNA 族能够在微生物宿主细胞中繁殖至少 50 代，所说的微生物以低于 0.6 %的重组频率(174 kb 族)和低于 0.2 %的重组频率(130 kb 族)生长。当用本发明的方法克隆时，小于 130 kb 的族显示出极少甚至没有重组，

术语 合成的"或"人工的"可以替换使用，"合成"的或"人工染色体"是具有必需染色体功能但不是天然存在的构建体。其已通过把纯化的 DNA 引入细胞产生。因为染色体完全由转染 DNA 组成，所以称它为合成的或人工的。人工或合成染色体以分天然存在的构型存在。

术语"转染"指将核酸引入到细胞中。所引入的核酸在引入的序列、物理构型或拷贝数上在细胞中不是天然的。

端粒指包含经称为端粒酶的核糖核蛋白酶合成的简单的重复 DNA 的染色体末端。所述功能使得线性 DNA 分子的末端可以复制。

如果核酸分子包含编码多肽的序列和表达控制序列，并且如果所说的表达控制序列可操作地连接到编码多肽的核苷酸序列上，则核酸分子(如 DNA 或基因)表达多肽或基因产物，所说的表达控制序列在适当的宿主环境中提供转录、加工和翻译包含在蛋白质产物的 DNA 中的遗传信息。然而，如本文所讨论的，基因产物不限于多肽基因产物而是可以包括 RNA。此外，可以由用作表达载体的人工染色体纠正的遗传缺陷可以是顺式作用影响进一步的基因表达的缺陷。

可操作连接是这样一种连接，其中调节 DNA 序列和所要表达的 DNA 序列

以允许基因表达的方式连接。基因表达所需的调节区的精确性质可以随不同的生物体变化，但一般包括启动子区、涉及转录和翻译起始的 5' 非编码序列如 TATA 盒，CAP 序列，CAAT 序列等。如果需要，可以通过以上所述的方法获得编码蛋白质的基因序列 3' 的非编码区。这个区可以由其转录终止调节序列（如终止和聚腺苷酸化序列）保持。这样，通过保持天然邻接编码蛋白质 DNA 序列的 3' 区，可以提供转录终止信号。在表达宿主细胞中转录终止信号功能上不令人满意时，在宿主细胞中的 3' 功能的可以被替代。

下列实施例不是将本发明限制在所描述的特定的实施方案中，而是用来具体描述实施本发明的某些方式。

实施例

实施例 1

17 号染色体 α 卫星载体的构建

为了利用 BAC 载体在大肠杆菌中克隆扩增 α 卫星 DNA，构建了一系列大小不同的串联 α 卫星族。

以前曾经报道过人类 17 号染色体的高级 α 卫星重复序列的结构(Waye 和 Willard, 分子细胞生物学, 6(9): 3156-3 165(1986))。人类 17 号染色体主要的高级重复序列长度为 2.7 kb, 紧靠 EcoRI 酶切位点, 由 16 个 alphoid 单体组成。用限制性内切酶 EcoRI 消化人的基因组 DNA, 把来源于人 17 号染色体或 Y 染色体的高级重复序列, 即 α 卫星 DNA 分散的结构单位克隆到质粒克隆载体 pACYC184(New England Biolabs)上。用 DNA 序列分析技术证实了克隆的高级重复序列的核苷酸序列。

用聚合酶链反应(PCR)可扩增长度为 2.7Kb 的一个高级重复序列单体, 这样在高级重复序列两端可构建互补的限制性酶切位点(BamHI 和 BglII)。同时要维持高度重复序列的精确长度。扩增此片段的一对引物是:

VB100-5'...gggcgggagatctcagaaaattctttgggatgattgagttg (SEQID NO: 2)
和

VB101-5'...gggcgggatcccttctgtcttctttttataggaagttattt (SEQID NO: 3)

用 BamHI 消化扩增的片断, 用凝胶进行纯化, 再连接到用 BamHI 和 HpaI 消化, 用凝胶纯化的载体 DNA 上, 也就是把修饰的高级重复序列克隆进 BAC

载体 pBAC108L (Bruce Birren 赠与, 加利福尼亚理工学院, Pasadena, CA) 上。将构建的质粒命名为 pBAC-17 α 1。

为了构建合成的 α 卫星 DNA 双体, 用两种不同的酶组合分别将 pBAC-17 α 等份消化: (1) BamHI + SfiI, 或者是 (2) BglIII + SfiI。BglIII + SfiI 消化后, 通过凝胶电泳切下 2.7Kb 的 α 卫星 DNA 片断, 再连接到用 BamHI + SfiI 消化后切下的 pBAC17 α 1ded 片段上。因为 BglIII 和 BamHI 产生相容的突出端, 而 SfiI 产生不对称的突出端, 仅能以特殊的方向连接, 这样 2.7Kb 的片断连接到载体上能构建成头尾相连的串联双体形式。通过电穿孔转入 DH108 细菌菌株。用限制性分析法分析这些克隆, 并且将包含有 17 号染色体的修饰后串联双体排列形式的 α 卫星 DNA 的克隆命名为 pBAC-17 β (图 1)。重复这种技术路线能构建扩展的 α 卫星族, 可含有 4, 8, 16, 32, 48 或 64 个 α 卫星高级重复序列。利用一种相似的技术路线可构建来自于人类其他染色体, 如: Y 染色体的合成族高级重复序列。

构建成的 BAC 载体 (包含有长度为 174Kb 的 α 卫星 DNA) 代表了最大量的目前在大肠杆菌中克隆并扩增的此类 DNA。以前实验曾利用粘粒成功地在大肠杆菌中克隆了大约 40 kb (Willard 等, Prog. Clin. Biol. res. 318: 9-18 (1989))。其它曾利用中等拷贝数量的质粒克隆一系列长度在 171 bp alphoid 单体至 40 kb 范围内的族 (Waye 和 Willard, 核酸研究, 15 (18): 7549-69 (1987))。在本领域的研究中, 发现在大肠杆菌中扩增有最大 alphoid 系列的质粒有高频的重组率。相比之下, 在 pBAC-17 α 64 扩增过程中, 用标准的质粒纯化方法检测未发现重组产物, 因为对克隆载体的研究已显示出 α 卫星 DNA 在微生物中的结构不稳定性, pBAC-17 α 64 的制备中利用凝胶电泳分析有明显重组序列的存在。通过这种方法未发现显著水平的重组质粒。

实施例 2

大肠杆菌中稳定克隆的重复 DNA 的方法

尽管实施例 1 中, 在大肠杆菌中扩增后, 人工 α 卫星族未出现高水平的重组或缺失。但很可能因发生的重组频率相当高, 大量的缺失产物使得任何一种产物都不能被检测出来。因此, 这些构建体的重组率应用高度敏感的方法来测定。除此之外, 因为较长系列的稳定性不如较短系列的, 检测

了几种不同长度的构建体，他们的稳定性如下。

用曾经克隆进 pBAC 108L 的三种不同大小的 α 卫星族进行稳定性实验。将这些构建体 pBAC-17 α 32， pBAC-17 α 48 和 pBAC-17 α 64 含有 α 卫星 DNA 大小分别为 87 kb， 130 kb， 174 kb， 转化进(电穿孔)大肠杆菌株 DH10B(GIBCO BRL)，挑出单一克隆分析。因为可能是转化方法本身可导致 DNA 重排，只有那些经限制性酶消化和电泳后，主要包含全长序列的克隆才能在甘油储存液中保存。

稳定性实验之初，从每个构建体的甘油保存液中取出大肠杆菌，在含有 12.5 μ g/毫升氯霉素的 LB 平板上划线。挑出八个生成的克隆分别培养至饱和，培养基是 5ml 包含 12.5 μ g/毫升氯霉素的 LB(大约 20 代)。从每个克隆里提取质粒 DNA 并纯化的，然后用 BamHI 消化，由脉冲电场凝胶电泳分离。包含任何全长质粒 DNA 的克隆可以说在单细胞期有全长质粒。对没有包含任何可检测的全长质粒的克隆，推测是新化线前重组的结果(如：在甘油储存期间发生)，这些克隆将被排除。从包含一些全长质粒的克隆中随即选出 10 个培养物，稀释一百万倍加入包含 12.5 μ g/ml 氯霉素的新鲜 LB 中，生长至饱和(大约 30 代)。从单细胞培养至饱和大约需生长 50 代。

为了确定 50 代生长中重组质粒的发生频率，将每个饱和培养物划线在 LB 板上，LB 含有 12.5 μ g/毫升氯霉素。单个的菌落在包含 12.5 μ g/毫升氯霉素的 1.5 ml LB 中生长至饱和。饱和后纯化 DNA，并且经限制性酶消化(BamHI)和 PFGE 分析。任何包含可检测到的全长质粒的菌落记数为 50 代实验未重组。相反地，没有包含任何可检测的全长质粒的菌落记数为 50 代实验期间实验重组发生。为了计算每种构建体每一代重组发生的频率，要算出 50 代后重组菌落所占的比率。1 减去这个值等于非重组菌落所占的比率(在 50 代之后)。每代重组菌落发生的比率为 1 减去未重组菌落比率的 50 次方根。总结为下列方程式：

$$X=1-(1-Y)^{1/50}$$

在这里"X 是每代产生的重组菌落比率， Y 是 50 代生长后重组菌落所占的比率。

利用这种方法，计算出含有三种 α 卫星构建体的的重组频率。在 50 代生长之后， 0%(N=9)的 pBAC-17 α 32 菌落重组形成截断式， pBAC-17 α 48 与

pBAC-17 α 64 各有 8.5 % (N=59), 25% (N=84)。每一代重组频率分别为 0.18 %, 0.57 %, 与以前报道的包含较小 alphoid DNA (如: 40kb) 的克隆载体和传代较少 (如: 30 代) 的克隆比较, 它们的重组频率非常低。

结果表明, 利用发明的方法和 BAC 载体, 大到至少 174 kb 长度的 α 卫星族在大肠杆菌中可稳定扩增。每一代, 174kb 和 130kb 系列的重组率分别是 0.18% 和 0.57 %。这样, 以 pBAC-17 α 64 为例, 从单细胞生长 50 代后, 单细胞来的饱和细菌培养物效价约为 1,000, 平均至少 75 % 的细胞包含全长 pBAC-17 α 64。这对用于基因治疗的含有 α 卫星的人工染色体来说, 大量增殖重组率降至此范围是预期的可接受的。

除测定在 pBAC 108L 中 α 卫星 DNA 的重组频率, 也明确了高度重复 α 卫星族的大小与载体中稳定性的相互关系。基于所测的重组频率, BAC 载体中 (假定 50 代后的有 50% 的全长菌落) 同源 α 卫星 DNA 的最小上限保守估计为 200 到 215kb (见图 2)。这是利用计算机程序 Cricket Graph 通过推断算出的。这是 alphoid 容量的最小估计, 因为发现其他系符合数据而预测值比上述的要大。从这种相互关系中, 可估计, 当利用 200-215kb 系列在菌株 DB108 中扩增时, 50 代之后 50% 以上的质粒会是全长的。利用 BAC 载体及特殊的抗重组菌株, 很可能上限值还要变大。而且, 这种估计是基于最大的同源系列。异化序列的稳定性, 包括天然 α 卫星族, 应大于这种估计。

在这里所述的实验表明第一次能在大肠杆菌中稳定克隆, 扩增 大于 50 kb 的 α 卫星族。以前, 曾用粘粒在大肠杆菌中克隆 α 卫星 DNA (Haaf 等, 细胞, 70(4): 68-196 (1992))。除了相对较小的系列 (等于或小于 40 kb), 其他系列未被分析其完整性和稳定性。也有人用 YACs 克隆 α 卫星 DNA (Neil 等, 核酸研究 18(6): 1421-8 (1990))。在这些研究中, α 卫星族的不稳定受到注意, 需要附加的操作, 如琼脂糖凝胶纯化, 才能获得包含主要的全长系列。此外, 利用 YACs 扩增 α 卫星 DNA 有一定的缺点。或许最重要的与 YACs 的拓扑结构有关。总的来说, YACs 是线性的 DNA 分子, 因此, 不能用简单的碱性裂解法纯化 α 卫星构建体, 使其没有酵母的染色体污染。代替它的是必须采用脉冲场凝胶电泳, 这种分离方法经不起放大实验的检验。最后 YACs 的线性拓扑结构使他们在纯化时易被剪切。这里, 已证明了含有 α 卫星的 BACs 可被富集并纯化去除大肠杆菌的染色体 DNA, 同时无实质上的剪切。

以前的研究表明 α 卫星 DNA 是功能性人的着丝粒的重要组分。天然存在的 α 卫星族长度从 230 kb 至几兆碱基不等(Oakey 和 Tyler, Genomics 7(3): 325-30(1990)); (Wevrick 和 Willard, 美国科学院学报, 86(23): 9394-8(1989))。然而, 最近的研究证明 140 kb 小的 α 卫星 DNA 在人类细胞中足以提供着丝粒功能(Brownl. 等, 人类分子遗传学, 3(8): 1227-1237(1994))。这里所述 α 卫星的系列是首次能大规模的, 稳定产生的 α 卫星族, 这样的 α 卫星族大得足以满足功能性人类着丝粒的需要。这些构建体可作为合成的人类染色体的骨架。

实施例 3

Y 染色体 α 卫星载体的构建

除了所用的 DNA 来自人类男性 Y 染色体细胞株 GM07033 外, 构建法与实施例 1 相同, 来自任何正常的人类男性细胞株的 DNA 均可。在人类 Y 染色体上主要的高级 alphoid 重复序列长度为 5.7 kb, 可由旁边的 EcoRI 酶切位点定界。将从 Y 染色体 alphoid 系列中来的 5.7kb 高级重复序列克隆进标准的大肠杆菌克隆载体 pACYC184(New England Biolabs) 中。用 PCR 修饰高度重复序列的末端, 两端分别加上 BamHI 酶切位点及 BglIII 酶切位点, 代替现有的 EcoRI 酶切位点。如前所述, 将修饰后的高级重复序列克隆入 pBAC 108L 克隆载体。

实施例 4

构建杂交 α 卫星载体

除所用的 alphoid DNA 来自两条人类染色体(17 号染色体和人类 Y 染色体)外, 构建法如实施例 1。构建了两个类型的系列。一种类型是简单的交替重复, 即来自 17 号染色体 alphoid DNA 的一个高度重复系列单位与来自 Y 染色体 alphoid DNA 的一个高度重复系列单位相互交替。构建的第二种类型是 17 号染色体 alphoid DNA 的高级重复序列双体与 Y 染色体 alphoid DNA 的高级重复序列单体相互交替。同前面的例子一样, 在每一种类型中, 来自每条染色体的高级重复序列的适当位相都保持在合成杂交连接中。

实施例 5

构建人工哺乳动物的染色体

实验步骤:

DNA 构建体概况

此处所提的构建质粒的方法均为标准的分子生物学技术(Sambrook, J. 等编, "分子克隆", 冷泉港实验室出版社(1989))。从第 17 号和 Y 染色体中克隆 α 卫星高级重复序列如前所述(Wolfe, J. 等, 分子生物学杂志, 182: 477485(1985); Van Bokkelen, G.B 等, "稳定克隆 DNA 大重复单位的方法", 美国专利申请(1995)); Waye & Willard, 分子细胞生物学, 6: 3156-65(1986))。通过加上合适的酶切位点可定向克隆, 成功的在质粒 pBAC108L 中构建了较长的 α 卫星族。Van Bokkelen G. B. 等, "稳定克隆 DNA 大重复单位的方法", 美国专利申请 08/487,989(1995), 1995 年 7 月 7 日申请), 这里一并参考指导克隆长的串联重复序列的文献。

用于实验的质粒

前面已经讲到 pBAC 108L(Shizuya, H. 等, 美国科学院学报, 89: 8794-7(1992))。pVJ10S 是 pBAC108L 修饰形式, 它在多接头中附加有限制性酶切位点, 有一个 β -geo 的表达单位, 此表达单位包含 CMV 早期直接基因启动子与猿猴病毒 SV40 多聚腺苷酸化信号(Seed, B. 自然, 320: 840-2(1987); Seed & Aruffo, 美国科学院学报, 84: 3365-9(1987)), 还包括 β -geo 开放式读码框架(MacGregor G.R. 等, 发展, 121: 1487-96(1995))及 UMS 转录终止序列(Heard J.M. 等, 分子细胞生物学 7: 2425-34(1987); McGeady, M.L. 等, DNA 5: 289-98(1986); Salier & Kurachi, 生物技术, 7: 30-1(1989))。pBACY α 16(92 kb 长的 Y 染色体 α 卫星)是由 16 个相同的克隆得到的高级重复序列首尾相接导入 pBAC108L 形成的。pBAC17 α 32(87 kb 长的 17 号染色体 α 卫星)是由 32 个相同的克隆得到的高级重复序列首尾相接导入 pBAC108L 形成的。pVJ105-Y α 16 是从 pBAC108L 克隆出 α 卫星族再导入 pVJ105 形成的。用 BamHI 和 Sfi I 线性化后, 从 pBAC17 α 32 克隆出 α 卫星族导入 pVJ105 形成 pVJ105-17 α 32, 同上的线性化后, β -geo 的表达方向是朝着 α 卫星族。所有的质粒采用碱性裂解(Sambrook, J. 等编 "分子克隆", 冷泉港实验室出版社(1989))纯化, 其后为琼脂糖凝胶纯化。

多聚化构建 α 卫星族>100 kb

用 BamHI 和 Sfi I 消化 pVJ105Y α 16 后再经脉冲电场凝胶电泳 (PFGE) 凝胶纯化得到 Y 染色体 α 卫星族。其余的 α 卫星 DNA 来自于用 BamHI 和 Bgl II 消化 pBACY α 16 和同上的 PFGE 凝胶纯化后所得的 92kb 片断。分出条带后，将凝胶带在 10mM Tris pH 7.5, 100mM NaCl, 10mM MgCl₂ 中平衡，再在 65 度中融化 5 分钟。两个片段以克分子比率 5: 1 结合 (pBACY α 16 α 卫星片断: pVJ105-Y α 16 片段)。加入 ATP (终浓度 1mM), T4 连接酶 (5 个单位)，同时加 BamHI (40 个单位) 和 Bgl II (40 个单位)，37 °C 温育 4 小时。加入 β 琼脂糖酶 (3 个单位) 37 °C 温育 1 小时。转染入 HT1080 细胞株前放置在冰上 1 小时。构建 17 号 α 卫星 DNA 的扩展 α 卫星族，步骤同上，只是用 pVJ105-17 α 32 与 pBAC17 α 32 分别取代 pVJ105-Y α 16 与 pBAC17 α 32。

高分子量人类基因组 DNA 的准备

HT1080 细胞在四个 150 mM 的平皿长至铺满，用胰蛋白酶乙二胺四乙酸消化后从平皿中取出，用 100 ml PBS 洗涤。高分子量的 DNA 在低凝胶温度的琼脂糖活塞中富集收获 (Sambrook, J. 等编, "分子克隆", 冷泉港实验室出版社 (1989))。大约用 Not I 消化 1 μ g 人基因组 DNA。消化后，70 °C 加热 5 分钟，Not I 被灭活。转染前，琼脂糖块用 3 个单位的 β 琼脂糖酶消化。

端粒 DNA 的准备

通过 PCR 来产生人类端粒 DNA，利用的引物是 42a (5' GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG3'; SEQ ID NO: 4) 和 42b (5' CCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCC3'; SEQ ID NO: 5) (Ijdo, J.W. 等, 核酸研究 19: 4780 (1991))。每个 PCR 反应都是在 1X 的 PCR 缓冲液中进行的，所述缓冲液添加了 250ng 的 42a 和 42b，5 个单位 Taq 聚合酶，250 μ M dNTPs，3.3 μ M MgCl₂ (Gibco BRL)。PCR 反应在 Perkin Elmer 9600 型热循环器中运行 35 个循环，温度变化方式如下：95 °C 20 秒，40 °C 20 秒，72 °C 2 分钟。PCR 后，每个反应物都进行琼脂糖凝胶电泳以纯化端粒 DNA，片断长度 > 1kb。将 DNA 从凝胶上切下，根据制造商的说明利用 Magic Prep 柱纯化 DNA，去除琼脂糖 (promega, WI)。

转染人类细胞

转染以前，用 BamHI 和 Sfi I 消化 pVJ105 Y α 16 和 pVJ10517 α 32；用

BamHI 和 Bgl II 消化 pBac Y α 16 和 pBac Y α 32。用 PFGE 纯化 DNA，在 10mM Tris pH7.5，100 mM NaCl 中平衡后再与端粒 DNA 和/或 Not I 消化的人类基因组 DNA 结合，在某些情况下，用图 3 描述的定向连接方法扩展 α 卫星族。

将每次转染的 DNA 组分合并并轻轻混合。转染物包括 pVJ105 Y α 16 (0.5-1 μ g)，pVJ105 17 α 32 (0.5-1 μ g)，或者是 pVJ105 VK75 (0.5-1 μ g)。在注明之处，转染物还包括纯化的 Y α 16 系列 (0.5-1 μ g)，17 α 32 系列 (0.5-1 μ g)，端粒 DNA (75-250 ng)，人类基因组 DNA (1-3 μ g) 和/或 VK75 片段 (0.5-2 μ g)。添加 1 ml 无血清 α -MEM 培养基 (MediaTech)。再加入 7.5 μ l 脂转染试剂，溶液在室温下温育 5 分钟。依制造商 (Gibco BRL) 的说明，将 DNA 与脂转染试剂混合物加到 2×10^6 个 HT1080 细胞中。37 $^{\circ}$ C 培养 16 小时，去除 DNA: 脂转染物，加入完全培养液。转染后 36 小时，用胰蛋白酶乙二胺四乙酸消化取出细胞，转移到 100 mm 的平皿中，平皿中为完全培养液加 300 μ g/ml G418。选择作用后七天，更换新鲜培养基，其为含有 G418 300 /ml 的完全培养液。选择后 12 天，用无菌克隆环分离单个菌落放到 24 孔板中。每次转染的单个菌落在选择作用下扩增到 100mm 的平皿中。将每种培养物冻存一部分以用作将来的分析，收集大部分培养物用 FISH 技术进行分析。

细胞培养

HT1080 细胞在 α -MEM 培养基中生长 (Gibco BRL, Bethesda, MD)，此培养基又加入了 15% 的胎牛血清 (Hyclone)，青霉素/链霉素和谷氨酰胺。实验中所用的 HT1080 细胞的亚克隆为四倍体。

平板染色

含有合成染色体的细胞平板接种在 6 孔板中，细胞有 10% 汇合。未转染的细胞同样平板接种在板中作为阴性对照。当细胞长到 70 % 汇合时，去掉培养基，用 2ml PBS 洗涤细胞。去掉 PBS 之后，每孔加 1ml 固定液 (2% 甲醛和 0.2% 戊二醛，用 PBS 稀释)，室温温育 4 分钟。移走固定液，立刻用 2ml PBS 清洗细胞。最后，弃去 PBS 洗液，每孔加入 1ml 染色液 (5mM 氰化铁钾，5mM 钾亚铁氰化物，5mM MgCl₂，1 μ g/ml X-半乳糖，用 PBS 稀释)，平板在 37 $^{\circ}$ C 温育 12 小时，用 PBS 洗细胞，利用光学显微镜和相关的成象

硬件将细胞成象(Oncor, Gaithersburg, MD)。

荧光原位杂交

HT1080 细胞在 100 mm 的组织培养平板生长, FISH 检测时将其收获, 根据已确定的方法, 将细胞转移到载玻片上(Verma, R. & Babu, A., 人类染色体, 原理和技术, 第二版, McGraw-Hill, Inc. (1995))。根据制造商(Oncor, Gaithersburg, MD)的说明, 使用染色体特异 α 卫星探针检测在内源染色体和合成染色体的上的 α 的卫星序列。

从包含合成染色体的克隆中测定 Y 染色体的 α 卫星 DNA 的状况

根据确立的方法, 从 HT1080 细胞及其克隆 22-6, 22-7, 22-11, 22-13 和 23-1 收获基因组 DNA(Sambrook, J 等编, 分子克隆, 冷泉港实验室出版社, 冷泉港, NY(1989))。每个克隆大约能回收 10 μ g DNA, 用 EcoRI (50 单位) 和 Pst I (50 单位) 消化过夜。在 0.8% 的琼脂糖胶上样进行电泳, 再转膜到 Nytran 膜上, 与 1kbY 染色体的 α 卫星探针 65°C 杂交过夜, 杂交液为 25% 甲酰胺/10% 葡聚糖/0.5 %SDS/0.5M NaCl/200 μ g /毫升鲑精 DNA。EcoRI 和 PstI 都在内源 Y 染色体 α 卫星高度重复序列中剪切一次, 形成 4kb 和 1.7 kb 的条带。然而, 因为使用的构建 Y 染色体 α 卫星族的方法, 使得 EcoRI 不能切割合成的高度重复系列。因此, EcoRI 和 Pst I 剪切合成的高度重复序列得到 5.7 kb 一条带。因为我们知道内源 Y 染色体 α 卫星族为 1mb 长, 通过测定 5.7 kb 带(合成系列)与 4.0 kb 带(内源系列)的比率, 就可计算出含有合成染色体的细胞中合成 α 卫星 DNA 的数量。不必考虑 1.7 kb 带, 因为在这种杂交条件下它不与探针杂交。然而, 重要的一点是考虑到多数克隆包含 2 个 Y 染色体(因为它们是四倍体)而仅仅一条合成染色体。一种例外情况是 22-13 克隆, 它包含单一 Y 染色体, 看来是二倍体。因此, 对于 22-6, 22-7, 22-11 克隆, 每个细胞的合成的 α 卫星 DNA 的量由下列方程式估计:

$$\frac{5.7 \text{ kb 带的强度}}{4 \text{ kb 带的强度}} \times 1 \text{ mb} \times 2$$

对于 22-13 克隆, 利用下列方程式:

5.7 kb 带的强度

----- × 1mb

4kb 带的强度

免疫荧光

根据确定的程序, 进行抗-CENP 免疫荧光染色(Sullivan & Schwartz, Hum. Mol. Genet. 4: 2189-2197(1995)). 主要步骤如下: HT1080 细胞在组织培养平板中生长至大约 80 %汇合。添加乙酰甲基秋水仙碱至终浓度 40ng/毫升, 37°C 温育 75 分钟。将培养基细心地除去, 同时用胰蛋白酶乙二胺四乙酸温育 3 至 5 分钟将细胞从平板解离。为了中和胰蛋白酶的作用, 加入完全培养基, 制成的细胞悬液用血细胞计数器记数, Jouan CT422 离心机中 1000 转每分钟离心, 弃上清, 缓慢加入低渗溶液(25mMKCl; 0.27%柠檬酸钠), 重悬细胞, 调浓度为 0.6×10^5 细胞/毫升。细胞在室温于低渗溶液中温育 12 分钟。取出 500 μ l 细胞放入细胞漏斗中, 用 Shandon Cytospin 3 离心机以 1900 转/分钟离心 10 分钟。接着将载玻片温育在 10mMTris pH7.7, 120 mMKCl, 20mMNaCl, 0.1 %曲通 100 中, 温育 30 分钟。每个载玻片加上稀释抗体(50 μ l, 稀释 1000 倍, 稀释液为 1mM 三乙醇胺, 25mMNaCl, 0.2 mM 乙二胺四乙酸, 0.5 %曲通 100, 0.1 %BSA), 将盖玻片盖在细胞面上, 37°C 温育 30 分钟, 去掉盖玻片, 在有 KB(10mMTris pH 7.7, 150 mMNaCl, 0.1 %BSA)的 Coplin 罐子中洗涤载玻片, 3X2 分钟。加上 FITC 标记的抗免疫球蛋白(50 μ l, 用 KB1 比 100 稀释)抗体, 细胞面盖上玻璃盖玻片, 37°C 温育 30 分钟, 在有 KB 的 Coplin 罐子中洗涤载玻片, 3X2 分钟。看之前, 将载玻片用 10 μ l DAPI (2 μ g/ml, 溶于抗衰减液中)复染。利用荧光显微镜和成象系统(Oncor, Gaithersburg, MD)收集图像。

有丝分裂稳定性时程

克隆化以后, 在 100mm 的平皿中扩增细胞, 培养基中有 300 μ g/ml 的 G418 存在。当细胞长至 80 %汇合时, 每个克隆收集一个平皿的细胞做 FISH 分析, 方法如前所述。这些细胞做为时程的时间零点。另一块平皿的细胞分

成 16 份在 100mm 的平皿中生长至汇合，培养基为无 G418 的完全培养基，一旦细胞长至汇合，再分成 16 份在无 G418 的完全培养基中生长。这个过程反复重复，时间流程如表 2 所示。在不同的时间点，取出部分培养物做 FISH，分析转染的 α 卫星的存在(17(或 Y α)。对于每个完整的染色体扩展，测定 Y 染色体 α 卫星(或 17 号染色体 α 卫星)信号及它们在染色体上的位置。

结果

哺乳动物的着丝粒是一个复杂的染色体元件，通常认为它由叫做 α 卫星的大块 DNA 组成。阻碍人们阐明着丝粒结构和合成人类染色体发展的一个主要障碍是不能克隆到这群 DNA 的大片断上。最近，已建立了一种能克隆并大规模生产 α 卫星 DNA 的方法，所述 DNA 的长度约至 175kb (Van Bokkelen, G.B. 等，授权"稳定克隆大的 DNA 重复单位的方法"，美国专利申请 08/487,989，1995 年六月六日归档)。同等重要的成就是利用定向克隆技术可构建已知组成和结构的 α 卫星族。

为了促进从裸转染 α 卫星 DNA 形成功能性着丝粒，发明人假定，转染长度大于 175kb 的 α 卫星 DNA 比较有利。在此之前，转染进哺乳动物细胞中的最大的毗连 α 卫星族长度为 120kb (Larin, Z. 等, Hum. Mol. Genet. 3: 689-95 (1994))。为了构建长度大于 175 kb α 卫星 DNA，使用图 3A 所示的定向连接技术。这种体外技术能构建出长度达 736kb 的相邻的，不间断的 Y 染色体 α 卫星族(图 3B, 2-4 道)。在有或无 BssHII 的情况下，连接 VK75 作为对照(图 3B, 5-7 道)。因为连接时 BssHII 末端又再生了一个 BssHII 酶切位点，当在连接反应中有 BssHII 酶时，多体的阶梯被消化为组分单体。在用 BamHI 片段或 BglIII 片段实验中可得到相似的结果(资料未显示)。这证明再剪切的反应是有效的，2, 3 道的序列梯是头尾连接的结果。最后，为了证实不同 α 卫星 DNA 家族的生物学区别，也构建了 17 号染色体 α 卫星 DNA 的扩展系列(未显示数据)。

利用图 4 与实验手册列出的技术，发明人已用长的纯化的卫星族构建成合成染色体。将每个这样的染色体组分共转染，发明人推论细胞将会连接这些组分形成一个功能性染色体。因此，用各种组合的 α 卫星 DNA，端粒 DNA，基因组 DNA 转染进 HT1080 细胞。转染之后，培养物在 G418 的选择下培养 10-14 天。分出单个的菌落，选择的作用下扩增，再收集细胞做 FISH

分析。

稳定转染子的鉴定

如表 1 所示,大多数转染的克隆其 α 卫星 DNA 已经插入到内源染色体中。在许多包括端粒 DNA 的转染中, α 卫星插入现象的高发与观察到的染色体切断相关。以前曾观察到插入后,用端粒 DNA 能有效地切断染色体(Barnett, M. A. 等,核酸研究, 21: 27-36(1993); Brown, K. E. 等, Hum. Mol. Genet. 3: 1227-37(1994); Farr, C. J. 等, EMBO J. 14: 5444-54(1995))。这里,很明显端粒 DNA 与 α 卫星 DNA 一同插入内源染色体,并且引起截短现象。

然而,转染后的细胞亚群中,也观察到含有转染 α 卫星 DNA 的合成染色体(见表 2, 转染 22, 23 及图 5-7)。这些阳性转染在两方面与其它转染不同: 第一,转染之前,用 BamH I 和 Bgl II 在体外预连接 α 卫星 DNA(见图 3A 和 3B)。这就形成了长的,定向的 α 卫星族,长度为 100 kb 至 736 kb。第二,转染中包括 NotI 消化的人类基因组 DNA。包括这两种组分,就提供了形成合成染色体的必需的 DNA 序列。

转染 22, 23 克隆的 FISH 分析显示约 50%的 G418 抗性克隆包含合成的染色体(见表 1)。两次转染后 4/5 的合成染色体包含克隆的 α 卫星 DNA 只能在合成染色体中检测到(见图 4-6)。也就是说,转染 Y 染色体 α 卫星 DNA 后,用 FISH 检测,只能在 Y 染色体和合成染色体上有可检测的信号。同样地,转染 17 号染色体 α 的卫星 DNA 后,用 FISH 检测,只能在 17 号染色体和合成染色体上有可检测的信号。应注意的是,转染 22, 23 都有合成的染色体形成。这证明 Y 染色体和 17 号染色体的 α 卫星都能促进合成染色体的形成。当进一步证明 α 卫星 DNA 是合成染色体的一个重要的组分时,则 α 卫星 FISH 信号包括大多数或所有的合成染色体。

包含合成染色体的细胞中仅有两种例外情况时, α 卫星 DNA(与转染用的 α 卫星 DNA 来自相同的染色体)既不在 Y 染色体上也不在合成染色体上(或者若转染 17 号染色体的 α 卫星,则为 17 号染色体)。第一,在克隆 17-15 中,在 17 号染色体,合成染色体, C 组染色体的末端检测到 17 α 卫星 DNA(见图 7A, 7B)。引人注意的是,这次转染包括未连接 α 卫星和端粒 DNA,但没有入基因组 DNA。一个可能性就是转染 DNA 插入到内源染色体,扩增,再从内源染色体上断裂下来。重要的是注意到如果非 alphoid, 非端粒人类序

列对染色体功能是必须的，那么转染时无人类基因组 DNA 提供其他 DNA 对合成染色体的形成是必须的。在一种情况下，缺乏共转染的人类基因组 DNA 合成染色体也可形成时，发现 α 卫星插入到内源染色体中。这表明基因组 DNA 对合成染色体形成或维持的某些方面是必要的。第二，在 22-11 克隆，发现一条合成染色体和一条 Y 与 14 号染色体的易位体(见图 7C, 7D)。通过脉冲场凝胶电泳，发明人已证明，Y: 14 易位染色体包括内源 Y α 卫星 DNA 和非合成的 Y α 卫星 DNA。因此， α 卫星 DNA 仅能在微型染色体上检测出，在任何内源染色体上检测不出。

合成染色体大小估计

包含合成染色体的细胞中 α 卫星 DNA 的大小约 350kb-2mb(见图 8)。这是通过利用合成与内源 α 卫星族的限制位点的多形性测得的。在 Southern 印迹中，比较合成的 α 卫星带的强度与内源 α 卫星带的强度，测得合成 α 卫星 DNA 与内源 α 卫星 DNA 的比值。由于内源 α 卫星族长度为 1mb(Larin 等, Hum Mol. Genet. 3: 689-95(1994))，而 Y 染色体的拷贝数(克隆 22-6, 22-7, 22-11 是 2 条；克隆 22-13 是 1 条)，与合成染色体的拷贝数量已知(见表 2)，就能估算出合成的 α 卫星 DNA 的大小(见图 7B)，以 kb 为单位。

尽管很难估算出合成染色体的全长，但在某些情况下，用放大 1000 倍的荧光显微镜几乎看不到合成染色体。

合成染色体的结构和拷贝数量

初始的分析是，包含合成染色体的每个克隆几乎没有合成的染色体，大多数情况下，每个细胞仅有一条。这表明合成染色体拷贝数量的调控类似与内源染色体。

除拷贝数量之外，合成染色体与内源染色体还有两个共性。第一，它们包含端粒序列(数据未显示)。这说明这些合成染色体是线性的。第二，在中期染色体中，可清晰看到合成染色体的单体(见图 6, 7)。这说明合成染色体的整体结构类似于内源染色体的。而且，因为通常情况下染色单体紧连于着丝粒，这个结果也说明了合成染色体能执行至少一种着丝粒的功能，吸附姊妹染色单体。

中期 CENP-E 与合成染色体相联系

合成染色体出现在子细胞中(大多数情况下，为单拷贝)说明产生了功能

着丝粒。为了进一步研究，检测了几个合成染色体，测定在中期着丝粒上是否有 CENP-E。以前曾证明功能性和非功能性着丝粒都有 CENP-B (Eamshaw, W.C 等, 染色体, 98: 1-12 (1989)), 因此, 它不能为作为着丝粒活性的标记。由于此原因, 通常强烈识别 CENP-B 的 CREST 抗血清并不是评估着丝粒功能的一种好试剂(曾用于以前的实验, Haaf 等, Larin 等和 Pranovsky 等(以上引用的))。另一方面, 已证明 CENP-E 仅在功能着丝粒才有 (Sullivan & Schwartz, Hum. Mol. Genet. 4:2189-2197 (1995)), 因此, 可用此蛋白质的单特异性抗体来评估着丝粒活性。

与功能着丝粒一致存在的 CENP-E 在克隆 22-11 和 23-1 中出现, 这是目前唯一检测的克隆(见图 9)。进而, 合成染色体着丝粒上的 CENP-E 的量与每一个内源染色体的着丝粒上的相似。有趣的是, 因为并不认为 CENP-E 可直接与着丝粒 DNA 结合, 因此, 它的水平不取决于 α 卫星的量。而是仅仅取决于是否形成一个功能着丝粒。这样, CENP-E 在中期的存在主要说明合成染色体包含有可指导形成着丝粒/动粒复合物的功能着丝粒。

无选择作用合成染色体表现为有丝分裂稳定性

为了确认合成染色体包含一个功能着丝粒, 同时能够正确分隔在子细胞中, 含有合成染色体细胞在无选择作用下生长一段时期。用 FISH 技术分析细胞, 测定含有合成染色体的细胞所占的百分比。无选择作用生长 46 天之后(大约 60 代), 大多数细胞仍然有合成的染色体(见表 2)。在几个克隆中, 仍然是 100% 的细胞有合成染色体。这表明合成染色体有丝分裂性是稳定的, 因此, 证实了这些载体能在体内应用于转染子细胞更正遗传缺陷的想法。

除了测定每一个各合成染色体的分离效率, 这个实验也使我们能评估合成染色体随时间的结构稳定性。每个克隆扫描 50 个铺展染色体, 没有发现合成染色体插入到内源染色体中。而且, 也没有发现其它显著的有关合成染色体的重排。这个结果, 结合高度有丝分裂稳定性, 证明这些合成染色体是作为一个遗传单位, 有许多与内源人类染色体相同的特征。

合成染色体的基因表达

这里描述的合成染色体为细胞基因治疗提供了一种预备的载体。因此, 重要的是确定异源基因是否能有效地在染色体载体上表达出来。

依所述的实验步骤, 通过共转染将 pVJ104-Y α 16 或 pVJ104-17 α 32, 端粒 DNA 和人类基因组 DNA 引入 HT1080 细胞中。每一次转染, 将 β -geo 表达单位与长度至少 100kb 的 α 卫星 DNA 相连。转染之后, 细胞中 α 卫星的位置与 β -geo 基因的位置相同。因此, 含有合成染色体的克隆中, 除 17-15 克隆外, β -geo 基因都仅位于合成的染色体上。

为了测定每个包含合成染色体克隆的 β -geo 表达的水平, 也就是合成染色体的基因表达程度, 按实验步骤所述的方法用 X-gal 平板染色法检测细胞。尽管这种技术比较不敏感(如 β -geo 高水平表达才能检测), 但它可提供大致的表达水平, 及能表达标记基因的细胞百分率。无 G418 选择培养 70 天之后(大约 80 次细胞分裂), 克隆 22-6 至少 50 % 的细胞用此法可检测到 β -geo 的表达(见图 10)。克隆 22-6 无选择作用培养 70 天后, 大约 25% 的细胞有可检测到的 β -geo 的活性(未显示数据)。因为此法的不敏感性及细胞中 β -geo 的低表达, 不能评估其他克隆。

讨论

结果显示裸 DNA 能转染进哺乳动物细胞, 不插入内源染色体, 就可形式一个功能性合成染色体。这些合成染色体有许多正常哺乳动物染色体的特性。第一, 它们在细胞中以低拷贝数存在, 通常每个细胞一条。第二, 合成染色体看来似乎是线性的, 包含端粒序列。第三, 在有丝分裂期间, CENP-E 可结合到合成染色体上表示功能着丝粒的形成。第四, 无选择作用时, 合成染色体表现为有丝分裂稳定。第五, 合成染色体能够装载有转录活性的基因。最后, 合成染色体随时间结构稳定, 具有不可觉察的插入频率。区别于正常的人类染色体, 合成染色体较小, 易修饰, 允许在多种的染色体环境中表达不同基因。

结果显示用体外构建 α 卫星族的方法可达 736kb 长。除了向此处描述的合成染色体提供必要的组分外, 这些结果也证明了构建的 α 卫星的数量具有临床使用性。以前, 没有方法构建长度>200kb, 纯化后有足够的量用于转染哺乳动物细胞, 结构完整的 α 卫星 DNA。

作为对照, 发明人重做了 Haaf 等以前失败的实验。构建一条染色体, 同时伴随有 α 卫星 DNA 插入内源染色体。发生这种情况的转染缺乏附加的基因组 DNA 序列。若没有基因组序列, 这条染色体很可能是由含有 α 卫星 DNA

的内源染色体断裂形成的。除了低效偶而的情况外，对于基因治疗过程来说，这个方法没有用，因为由断裂机制造成的结果使宿主细胞有基因组重排和恶性转化的风险。

总之，发明人证明了通过转染大的 α 卫星族，端粒 DNA 和基因组 DNA 到人类细胞中，能构建有丝分裂稳定性的合成染色体。尽管每个组分看来对形成有效的合成染色体是必要的，很有可能在缺乏基因组(非 α 卫星，非端粒)序列时插入 α 卫星 DNA 后的基因组重排可导致染色体的形成。另一方面，构建这些合成染色体时， α 卫星 DNA 看来似乎绝对需要。这里，通过利用来自 Y 染色体和 17 号染色体的 α 卫星构建合成染色体，发明人证明了 α 卫星 DNA 的来源并不重要。换句话说，来自任何染色体的 α 卫星 DNA 都能用于构建合成染色体。进而，假如人类染色体在各种杂合体中是稳定的，包括小鼠，仓鼠，灵长类，则这些物种的 α 卫星样序列也能用于构建如上所述的合成染色体。

表 1

克隆	Y α	17 α	hg	tel	VK75	特点
C10-7	XX			X		整合体(大染色体的中 Q 带)
C10-10	XX			X		整合体(中等大小的染色体中间臂)
C11-7		X		X		整合型(形成一段 17P ⁺)
C11-19		X		X		无信号
C12-2		XX		X		无信号
C12-3		XX		X		无信号
C12-5		XX		X		整合型(中间臂)
C12-6		XX		X		整合型(非端粒型)
C12-14		XX		X		无信号
C12-16		XX		X		无信号
C13-1				X	X	整合型
C14-1	X			X	X	端粒引导的截短
C15-2	X			X		6 号染色体截短; 从头合成端粒
C15-3	X			X		无信号

C15-4	X			X		整合型 (16 号样染色体 P 臂)
C15-5	X			X		无信号
C15-10	X			X		整合进端着丝粒染色体
C15-12	X			X		整合进 17 号染色体, 着丝粒以下
C15-13	X			X		无信号
C15-21	X			X		无信号
C16-6	XX			X		端粒/端粒着丝粒 p-缢痕
C16-7	XX			X		无信号
C17-2		X		X		无信号
C17-8		X		X		无信号
C17-11		X		X		C 组染色体截短
C17-15		X		X		微染色体, C 组染色体截短
C17-19		X		X		无信号
C19-1	X		X	X		无信号
C19-2	X		X	X		不清楚 (?端粒引导截短)
C21-1			X	X	X	无信号
C22-2	(XX)		X	X		染色体截短 (@19P 端粒从头合成)
C22-3	(XX)		X	X		无信号
C22-4	(XX)		X	X		端粒性/可能为双着丝粒
C22-5	(XX)		X	X		不清楚: 非常小的微体
C22-6	(XX)		X	X		双微体 (多微体)
C22-7	(XX)		X	X		大微体
C22-8	(XX)		X	X		不清楚 (可能有微体)
C22-9	(XX)		X	X		端粒性/长系列
C22-11	(XX)		X	X		小微体
C22-13	(XX)		X	X		大微体
C23-1		(XX)	X	X		小微体

表 1: 转染不同组合的 α 卫星 DNA , 端粒 DNA 和人类基因组 DNA 进入人类细胞的结果。

Y α 和 17 α 分别是 Y 染色体和 17 号染色体 α 卫星的缩写。

hg 是经 NotI 消化后的人类基因组 DNA 的缩写。

Tel 是端粒 DNA 的缩写。

VK75 是来自 X 染色体的一个 75 kb 片段缩写。

X 代表转染中包括此序列。

XX 代表如实验步骤所述，转染中包括附加纯化的 α 卫星 DNA。

(XX) 代表如实验步骤所述，转染前，附加的 α 卫星 DNA 要预连接在 β -geo/ α 卫星构建体上。

表 2

克隆	T=0 天	T=10 天	T=20 天	T=46 天
22-7 (Y)	10 个细胞; 2Y _s	10 个细胞; 2Y _s (9 个细胞)	10 个细胞; 2Y _s (9 个细胞)	10 个细胞; 2Y _s (8 个细胞)
微染色体	1 个微染色体	1 微染色体	1 微染色体	1 微染色体
	100%	100%	100%	100%
22-11 (Y)	20 个细胞; 2Y _s (17)	10 个细胞; 2Y _s (6 个细胞)	10 个细胞; 2Y _s (8 个细胞)	10 个细胞; Y _s (5 个细胞)
微染色体	t (Y; 14) (19 个细胞)	t (Y; 14) (9 个细胞)	4 (Y; 14) (10 个细胞)	1Y (4 个细胞)
	0 微染色体 (17 细胞)	13P' (8-9 个细胞)	13P' (8+细胞)	0Y (1 个细胞)
	1 微染色体 (12 个细胞)	1 个微染色体 (2 个细胞)	0 微染色体* (2 个细胞)	0 微染色体 (5 个细胞)
	2 微染色体 (1 个细胞)	2 个微染色体 (8 个细胞)	1 微染色体 (2 个细胞)	1 微染色体 (2-3 个细胞)
	65%	100%	2 微染色体 (5 个细胞)	2 微染色体 (2 个细胞)
			3 个微染色体 (1 个细胞)	40-50%
			?80+%	
22-13 (Y)	17 个细胞; 1Y	10 个细胞; 1Y	10 个细胞; 1Y	10 个细胞; 1Y
微染色体	1 微染色体	1 微染色体	1 个微染色体	1 微染色体
	100%	. 100%	. 100%	. 100%
23-(17)	11 个细胞; 4-17 _s	10 个细胞; 4-17 _s	10 个细胞; 4-17 _s	10 个细胞; 4-17 _s
微染色体	1 微染色体	1 微染色体	1 微染色体	1 微染色体
	100%	100%	100%	100%

17-15 (17)	17 个细胞: 4-17s (15 个细胞)	N. D.	10 个细胞: 4-17s	N. D.
微染色体	1Cqter (10 个细胞)		1Cqter (6 个细胞)	
	1 微染色体		1 微染色体	
	100%		100%	
22-6 (Y)	9 个细胞: 2Ys	10 个细胞: 2Ys	10 个细胞: 2Ys (7 个细胞)	N. D.
插入信号	1 中着丝粒, 大信号 (2 个细胞)	1 中着丝粒, 大信号 (2 个细胞)	1 中着丝粒, 大信号 (5 个细胞)	
	1D 样的 w/信号 (2 个细胞)	1D 样 w/信号 (2 个细胞)	1E 样 (?) w/信号 (1 个细胞)	
两个伸展 均被切断				

表 2 无选择合成染色体的有丝分裂稳定性

每个克隆按指定的天数无选择作用下生长，一定的培养时间后，收集细胞进行FISH分析并与实验流程所述的适当 α 的卫星探针杂交。图中已指出每一个时间点检测的细胞数，Y染色体数(或23-1及17-15克隆的7号染色体数)，合成染色体数。每一个实验点含有合成染色体的细胞百分数用黑体显示，时间点0，10，20天分别对应0，12，24次细胞分裂。

尽管以上提到的是非常理想的方案，应理解为本发明不受其限制。它会使本领域普通技术人员想到可对这个开放的方案进行多种修改，同时这样的修改在本发明的范围之内。

序 列 表

(1) 一般信息:

(i) 申请人:

- (A) VAN BOKKELEN, Gil B.
- (B) HARRINGTON, John J.
- (C) WILLARD, Huntington F.

(ii) 发明名称: 合成的哺乳动物染色体和其构建方法

(iii) 序列数: 5

(iv) 通讯地址:

- (A) 收信人: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
- (B) 街道: 1100 纽约路; N.W., SUITE 600
- (C) 城市: 华盛顿
- (D) 州: D.C.
- (E) 国家: 美国
- (F) ZIP: 20005-3934

(v) 计算机可读形式:

- (A) 介质类型: 软盘
- (B) 计算机: IBM PC 兼容机
- (C) 操作系统: PC-DOS/MS-DOS
- (D) 软件: Patentln Release#1.0, 版本# 1.30 (EPO)

(vi) 当前申请的数据:

- (A) 申请号: (待定)
- (B) 申请日: 1996 年 6 月 7 日

(vii) 在先申请的数据:

- (A) 申请号: 美国 08/487,989 和美国 08/643,554
- (B) 申请日: 1995 年 6 月 7 日和 1996 年 5 月 6 日

(viii) 律师/代理人信息:

- (A) 姓名: MICHELE. CIMBALA
- (B) 登记号: 33,851

(c) 证书号: 1522.0001PC02

(2) SEQ ID NO: 1 的信息:

(i) 序列特征:

(A) 长度: 17 个碱基对

(B) 类型: 核酸

(C) 链型:

(D) 拓扑结构:

(ii) 分子类型: cDNA

(xi) 序列描述: SEQ ID NO: 1:

ATTCGTTGGA AACGGGA

17

(2) SEQ ID NO: 2 的信息:

(I) 序列特征:

(A) 长度: 41 个碱基对

(B) 类型: 核酸

(C) 链型:

(D) 拓扑结构:

(ii) 分子类型: cDNA

(xi) 序列描述: SEQ ID NO: 2:

GGCGGGAGA TCTCAGAAAA TTCTTTGGGA TGATTGAGTT G 41

(2) SEQ ID NO: 3 的信息:

(I) 序列特征:

(A) 长度: 41 个碱基对

(B) 类型: 核酸

(C) 链型:

(D) 拓扑结构:

(ii) 分子类型: cDNA

(xi) 序列描述: SEQ ID NO: 3:

GGGCGGGATC CCTTCTGTCT TCTTTTATA GGAAGTTATT T 41

(2) SEQ ID NO: 4 的信息:

(I) 序列特征:

(A) 长度: 39 个碱基对

(B) 类型: 核酸

(C) 链型:

(D) 拓扑结构:

(ii) 分子类型: cDNA

(xi) 序列描述: SEQ ID NO: 4:

GGGTTAGGGT TAGGGTTAGG GTTAGGGTTA GGGTTAGGG 39

(2) SEQ ID NO: 5 的信息:

(I) 序列特征:

(A) 长度: 39 个碱基对

(B) 类型: 核酸

(C) 链型:

(D) 拓扑结构:

(ii) 分子类型: cDNA

(xi) 序列描述: SEQ ID NO: 5:

CCCTAACCTT AACCTAACCTT AACCTAACCTT AACCTAACCTT 39

说明书附图

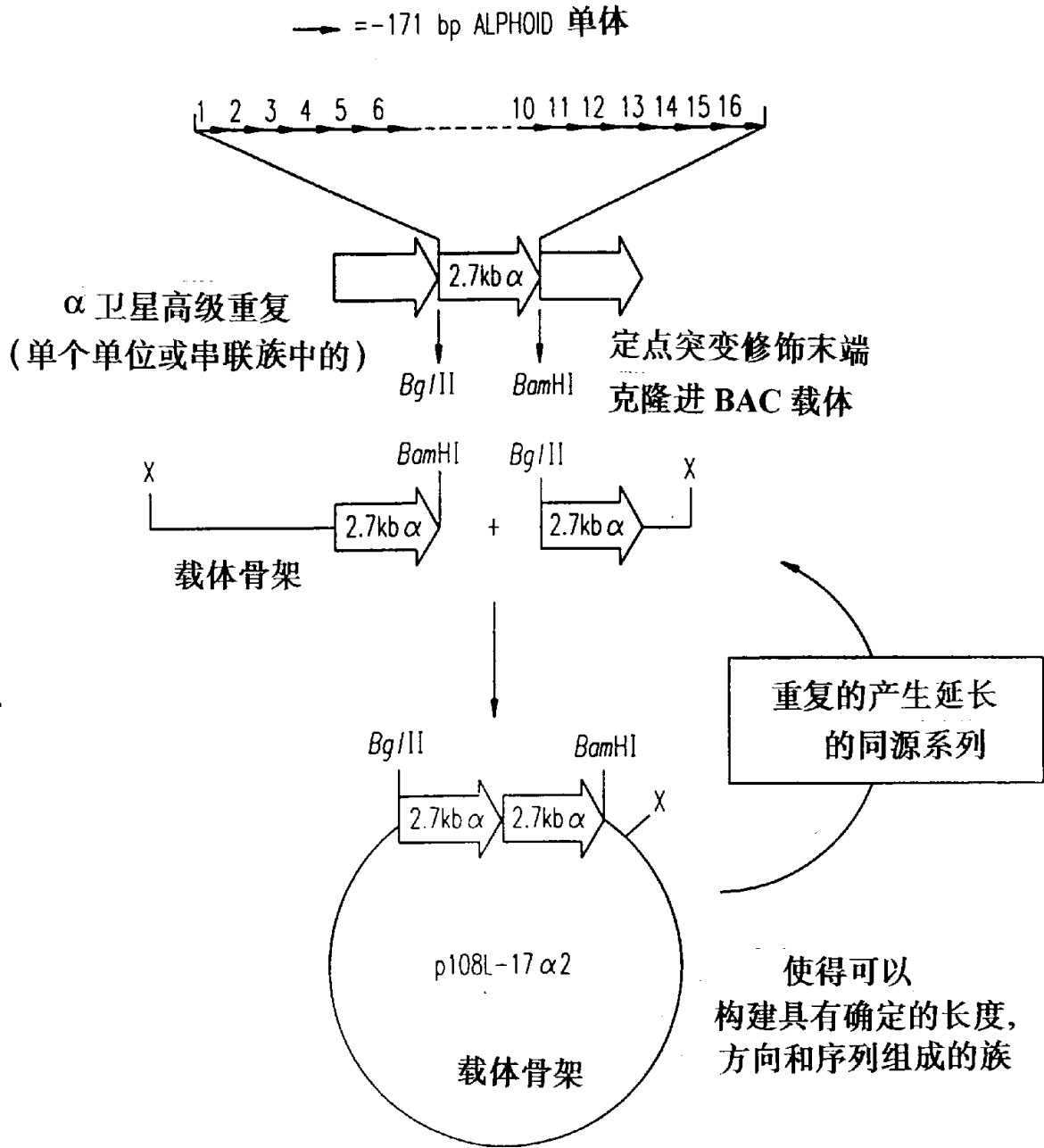


图1

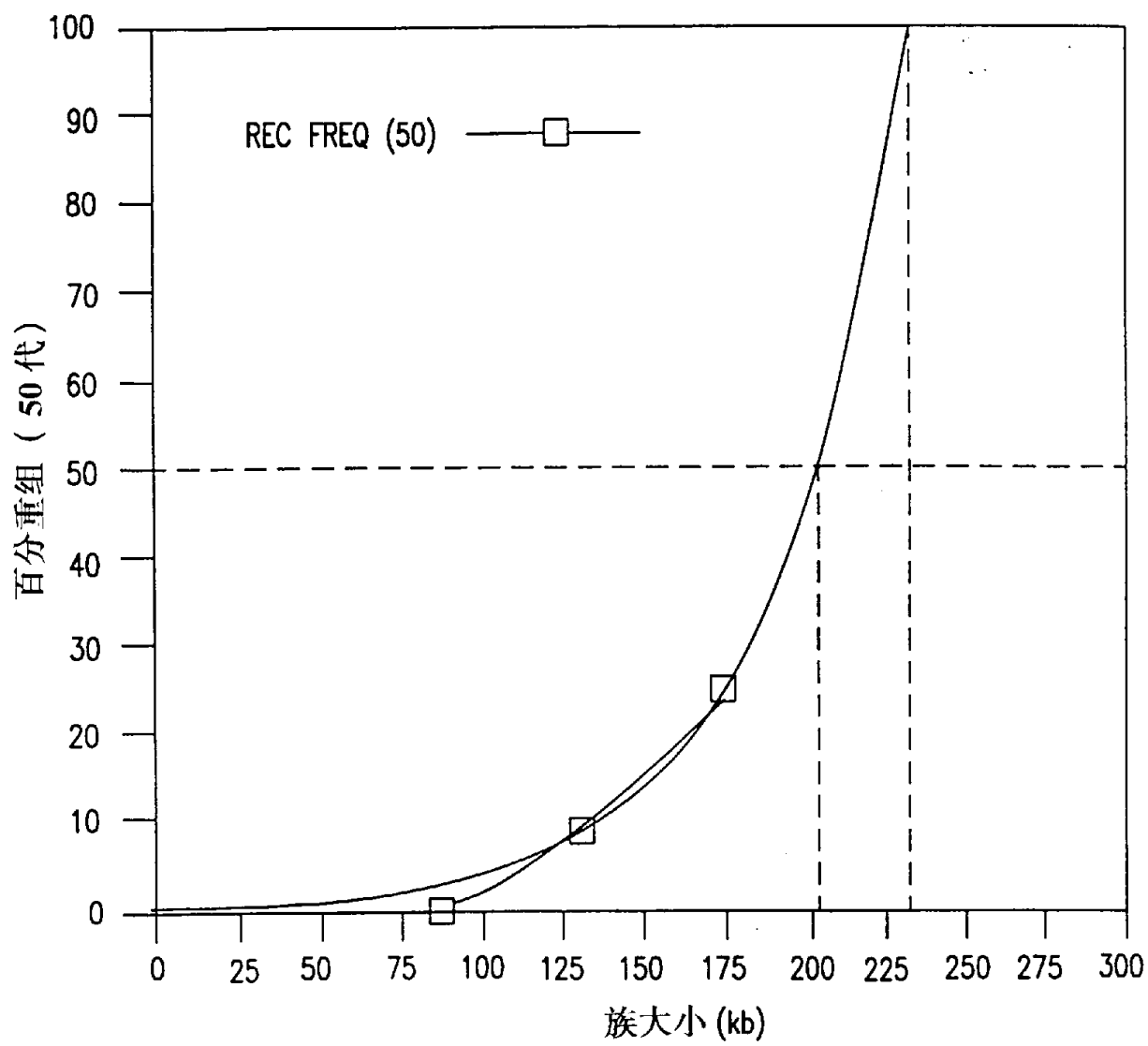


图2

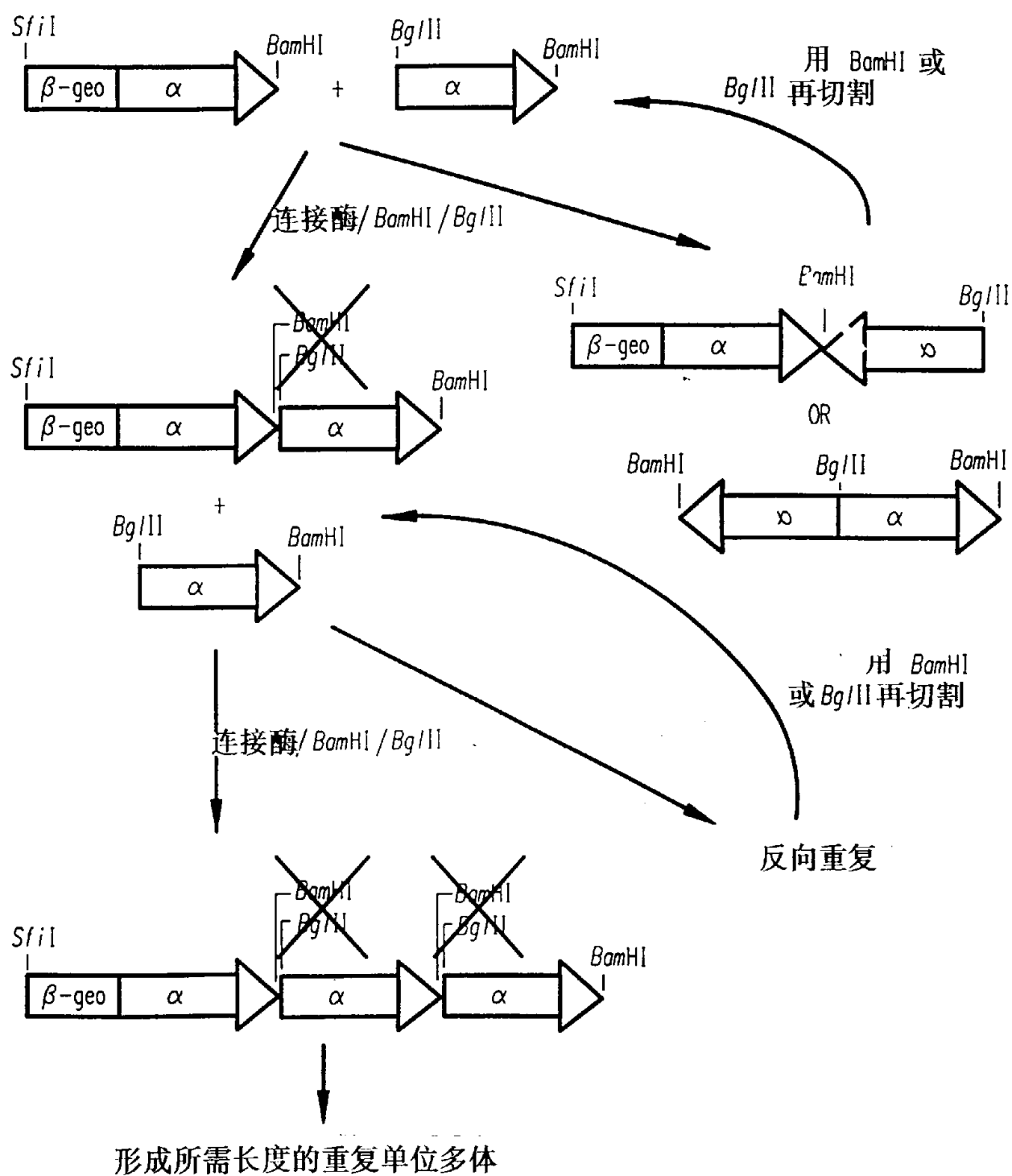


图3A

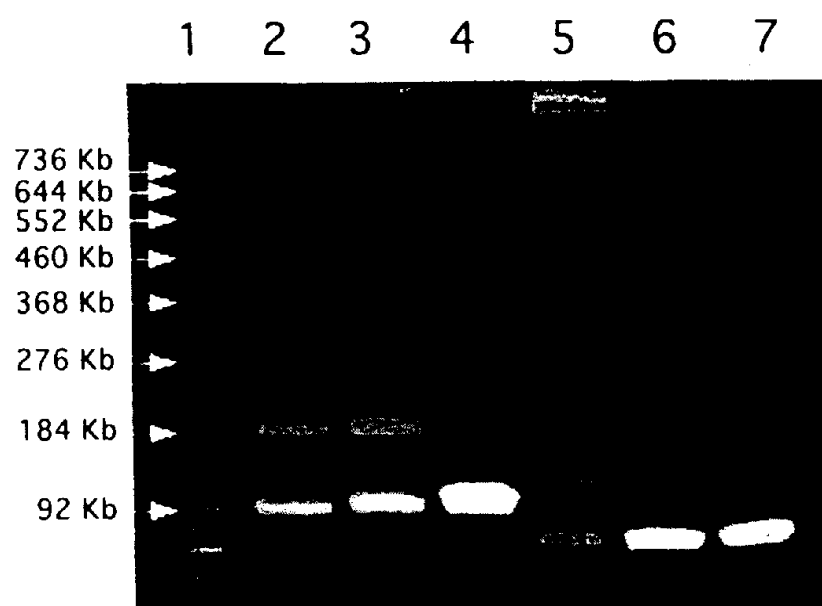


图 3B

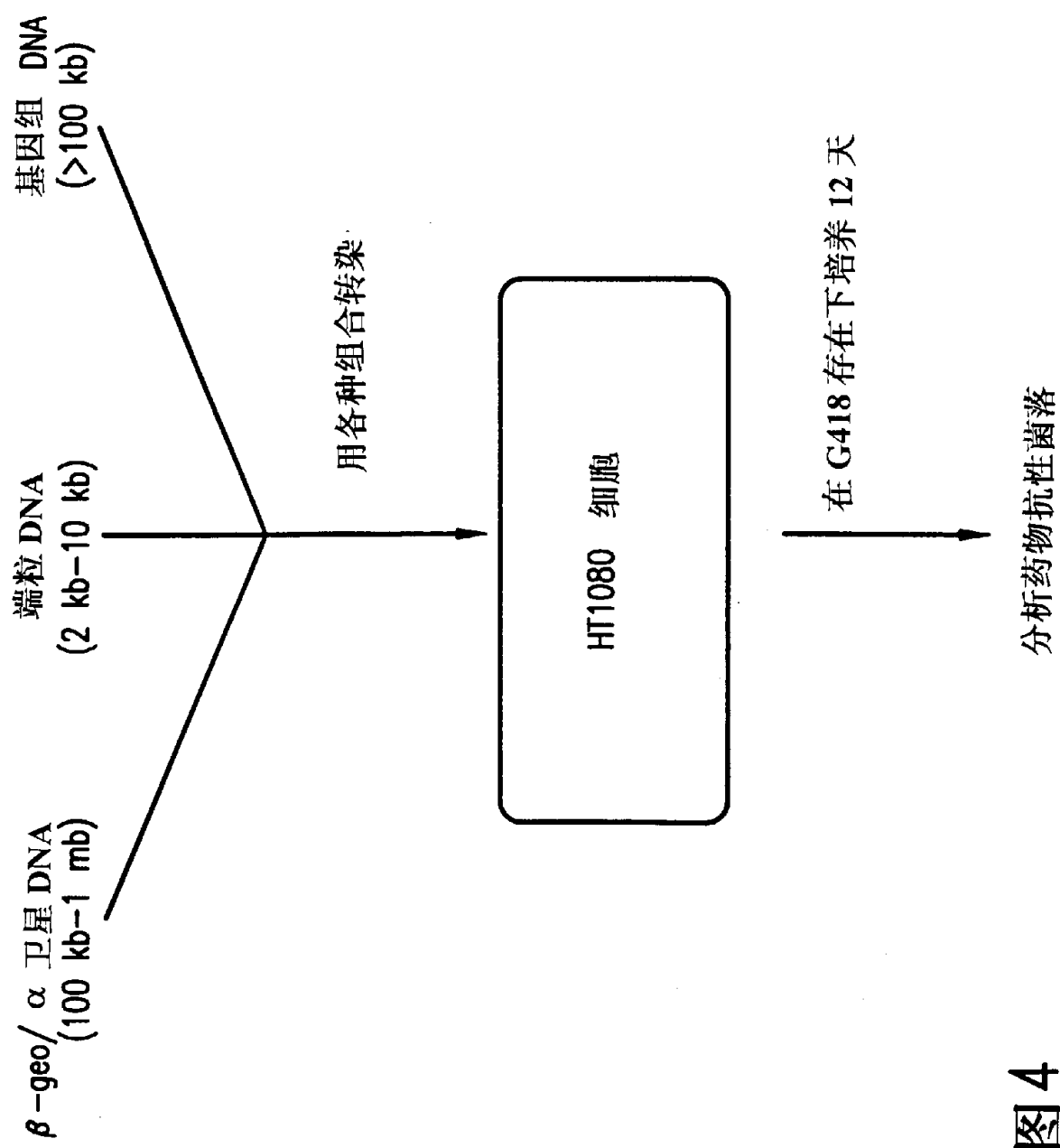


图 4

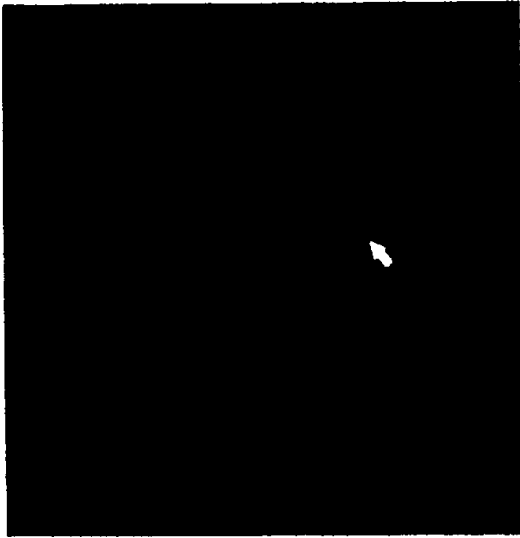


图 5A

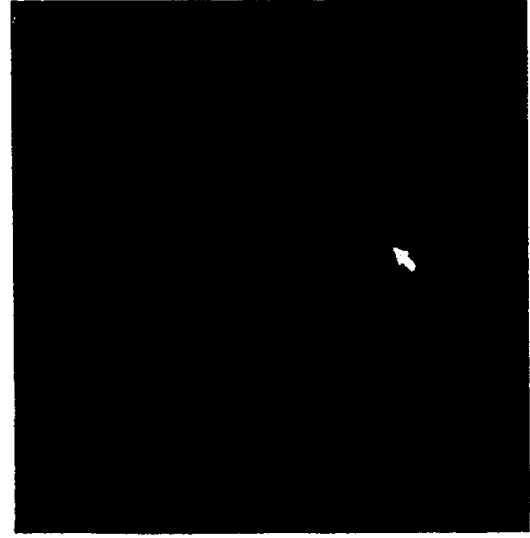


图 5B



图 5C

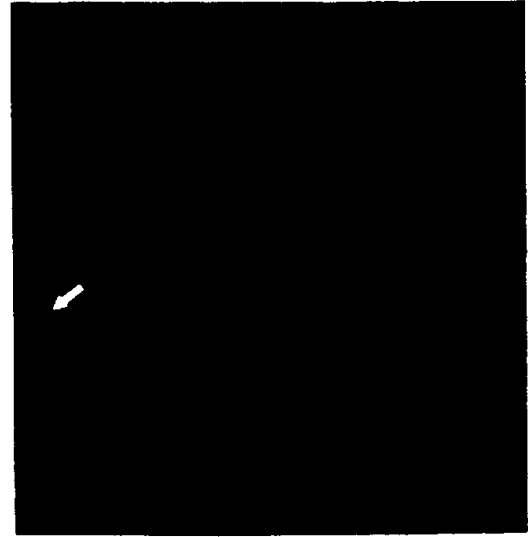


图 5D

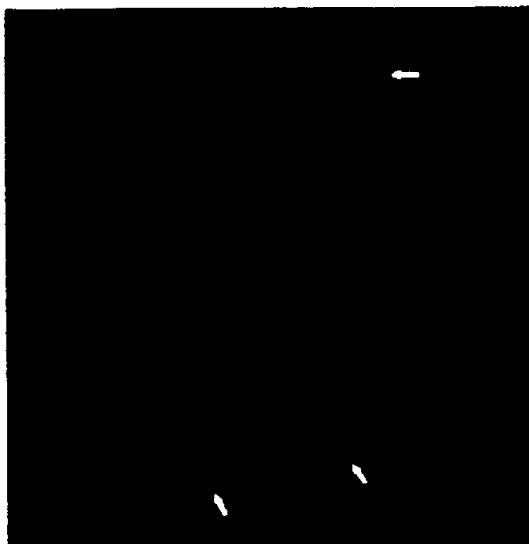


图 6A

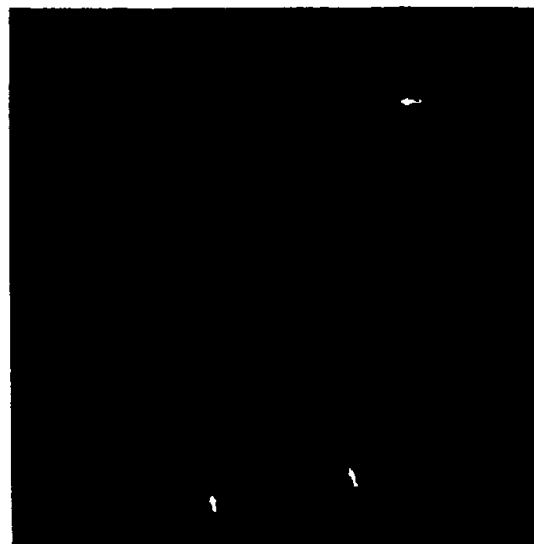


图 6B

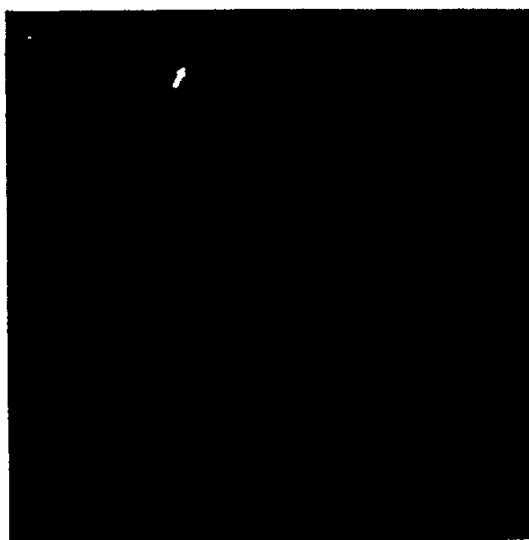


图 6C



图 6D

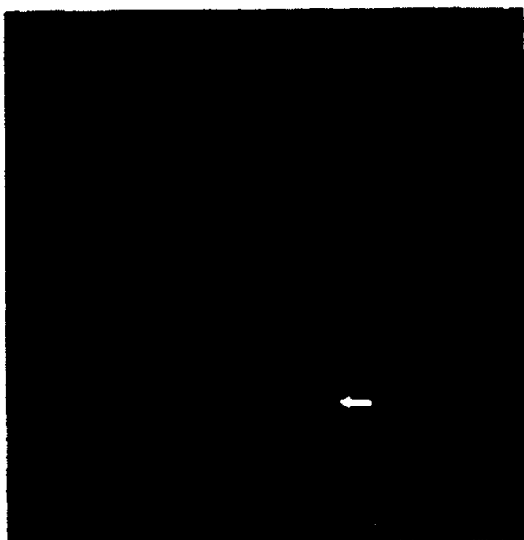


图 7A

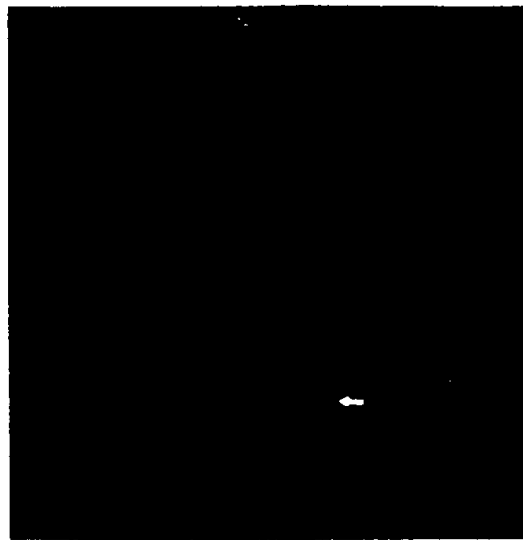


图 7B



图 7C

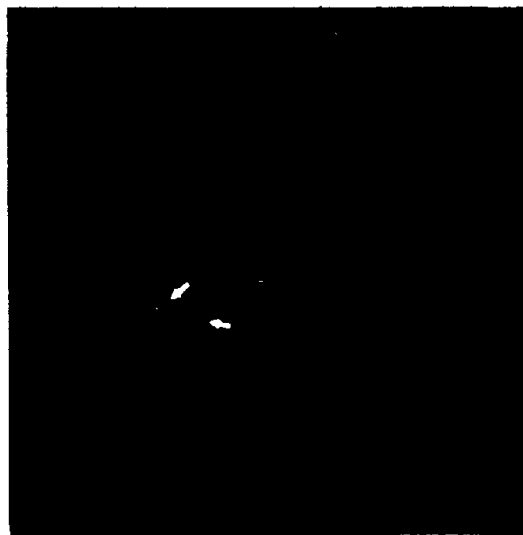


图 7D



图 8A

克隆	转染 DNA	合成 α 卫星 DNA 估计的量
HT1080	$\emptyset$	$\emptyset$
22-6	合成 Y α 卫星	450 kb
22-7	合成 Y α 卫星	1 mb
22-11	合成 Y α 卫星	400 kb
22-13	合成 Y α 卫星	2 mb
23-1	合成 17 α 卫星	$\emptyset$

图8B

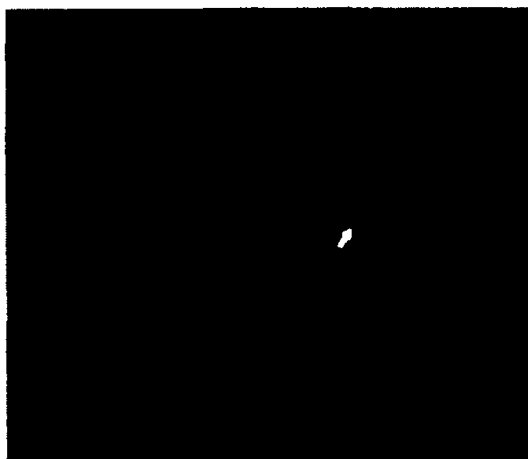


图 9A



图 9B



图 9C



图 9D



图 10A



图 10B