

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200580011637.X

[51] Int. Cl.

*C12Q 1/68 (2006.01)*  
*C12N 1/12 (2006.01)*  
*C07H 21/02 (2006.01)*  
*C12N 5/00 (2006.01)*  
*C07H 21/04 (2006.01)*

[43] 公开日 2007年6月13日

[11] 公开号 CN 1981051A

[22] 申请日 2005.2.17

[21] 申请号 200580011637.X

[30] 优先权

[32] 2004.2.18 [33] US [31] 60/546,075

[86] 国际申请 PCT/US2005/005080 2005.2.17

[87] 国际公布 WO2005/079462 英 2005.9.1

[85] 进入国家阶段日期 2006.10.17

[71] 申请人 克罗莫塞尔公司

地址 美国新泽西

[72] 发明人 K·谢克达 D·J·索查克

J·M·蒙特兹

[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 罗菊华

权利要求书 17 页 说明书 84 页 附图 61 页

[54] 发明名称

使用信号探针的方法和材料

[57] 摘要

本发明涉及使用在与靶序列杂交后可以产生信号的信号探针分离细胞或产生细胞系的方法。利用该信号探针的其它方法包括定量 RNA 表达水平的方法、鉴定活细胞中的遗传重组事件的方法、和使用该分离的细胞产生转基因动物的方法。本发明还涉及蛋白酶探针。本发明提供形成茎环结构、三臂联结结构和哑铃结构的信号探针和蛋白酶探针。

1. 分离表达 RNA 的细胞的方法，包括步骤：  
提供潜在地表达 RNA 的细胞；  
将该细胞暴露于在与所述 RNA 杂交时可以产生可检测信号的信号探针；和  
分离产生该信号的细胞。
2. 分离表达 RNA 的细胞的方法，包括步骤：  
将编码所述 RNA 的 DNA 引入细胞中；  
将该细胞暴露于在与所述 RNA 杂交时可以产生可检测信号的信号探针；和  
分离产生该信号的细胞。
3. 分离表达内源 RNA 的细胞的方法，包括步骤：  
将导致内源 RNA 表达的 DNA 引入细胞中；  
将该细胞暴露于在与所述内源 RNA 杂交时可以产生可检测信号的信号探针；和  
分离产生该信号的细胞。
4. 分离表达 RNA 的细胞的方法，包括步骤：  
将编码所述 RNA 并且还编码标签序列的 DNA 引入细胞；  
将该细胞暴露于在与所述标签序列杂交时可以产生可检测信号的信号探针；和  
分离产生该信号的细胞。
5. 分离表达外源 RNA 和内源 RNA 的细胞的方法，包括步骤：  
将编码所述外源 RNA 的 DNA 引入细胞，其中所述细胞潜在地表达内源 RNA；

将该细胞暴露于在与所述外源 RNA 杂交时可以产生可检测信号的第一信号探针;

将所述细胞暴露于在与所述内源 RNA 杂交时可以产生可检测信号的第二信号探针; 和

分离产生两个信号的细胞。

6. 分离表达两种或多种内源 RNA 的细胞的方法, 包括步骤:

将导致第一内源 RNA 表达的 DNA 引入细胞, 其中所述细胞潜在地表达另外的内源 RNA;

将细胞暴露于在与所述第一内源 RNA 杂交时可以产生可检测信号的第一信号探针;

将所述细胞暴露于在与所述另外的内源 RNA 杂交时可以产生可检测信号的第二信号探针; 和

分离产生两个信号的细胞。

7. 分离表达两种或多种不同 RNA 的细胞的方法, 包括步骤:

提供潜在地表达两种或多种不同 RNA 的细胞;

将该细胞暴露于在与第一 RNA 杂交时可以产生可检测信号的第一信号探针;

将该细胞暴露于在与另外的 RNA 杂交时可以产生可检测信号的第二信号探针; 和

分离产生两个信号的细胞。

8. 分离表达两种或多种不同 RNA 的细胞的方法, 包括步骤:

将编码第一 RNA 的第一 DNA 引入细胞;

将编码另外的 RNA 的另外 DNA 引入所述细胞;

将该细胞暴露于在与所述第一 RNA 杂交时可以产生可检测信号的第一信号探针;

将该细胞暴露于在与所述另外的 RNA 杂交时可以产生可检测信号的

第二信号探针；和  
分离产生两个信号的细胞。

9. 分离表达两种或多种不同内源 RNA 的细胞的方法，包括步骤：  
将导致第一内源 RNA 表达的第一 DNA 引入细胞；  
将导致另外的内源 RNA 表达的另外 DNA 引入所述细胞；  
将该细胞暴露于在与所述第一内源 RNA 杂交时可以产生可检测信号的第一信号探针；  
将该细胞暴露于在与所述另外的内源 RNA 杂交时可以产生可检测信号的第二信号探针；和  
分离产生两个信号的细胞。

10. 分离表达两种或多种不同 RNA 的细胞的方法，包括步骤：  
将编码第一 RNA 且还编码第一标签序列的第一 DNA 引入细胞；  
将编码另外的 RNA 且还编码另外的标签序列的另外 DNA 引入所述细胞；  
将所述细胞暴露于在与第一标签序列杂交时可以产生可检测信号的第一信号探针；  
将所述细胞暴露于在与所述另外的标签序列杂交时可以产生可检测信号的第二信号探针；和  
分离产生两个信号的细胞。

11. 分离包含一个以上拷贝的外源 DNA 的细胞的方法，包括步骤：  
将编码 RNA 且还编码第一标签序列的第一 DNA 引入细胞；  
将编码所述 RNA 且还编码另外的标签序列的另外 DNA 引入该细胞；  
将该细胞暴露于在与所述第一标签序列杂交时可以产生可检测信号的第一信号探针；  
将该细胞暴露于在与所述第二标签序列杂交时可以产生可检测信号的第二信号探针；和

分离产生两个信号的细胞。

12. 权利要求 5-11 之任一项的方法，其中所述暴露步骤同时实施。

13. 权利要求 5-11 之任一项的方法，其中所述暴露步骤相继实施。

14. 权利要求 5-11 之任一项的方法，其中所述两种或多种 RNA 或由所述两种或多种 RNA 编码的蛋白质选自：相同或相关生物学途径中的 RNA 或蛋白质，在彼此上游或下游起作用的 RNA 或蛋白质，对彼此具有调节、激活或阻遏功能的 RNA 或蛋白质，对于功能或活性而言相互依赖的 RNA 或蛋白质，属于相同复合物的成分的 RNA 或蛋白质，以及来自相同蛋白质家族的蛋白质。

15. 权利要求 5-11 之任一项的方法，其中第一信号探针产生的信号与第二信号探针产生的信号不同。

16. 分离包含编码 RNA 的 DNA 构建体的细胞的方法，其中所述 RNA 处在条件型启动子的控制下，所述方法包括步骤：

将编码处于组成型启动子控制下的第一 RNA 的 DNA 构建体引入细胞，其中所述 DNA 构建体还编码处于条件型启动子控制下的第二 RNA，其中所述引入在第二 RNA 不表达的条件下进行；

将细胞暴露于在与所述第一 RNA 杂交时可以产生可检测信号的信号探针；和

分离产生该信号的细胞。

17. 权利要求 12 的方法，其中 DNA 构建体还编码待测 RNA。

18. 分离多个细胞的方法，其中一个细胞子集表达另一细胞子集不表达的 RNA，所述方法包括：

将编码多个 RNA 的多个 DNA 引入细胞，其中至少此多个 RNA 的一个子集彼此不同；

将细胞暴露于多个不同信号探针，其中所述信号探针在与该多个 DNA 编码的一个或多个 RNA 杂交时可以产生可检测信号；和  
分离产生该信号的细胞。

19. 分离多个细胞的方法，其中一个细胞子集表达另一细胞子集不表达的 RNA，所述方法包括步骤：

将导致多个内源 RNA 表达的多个 DNA 引入细胞，其中至少所述多个内源 RNA 的一个子集彼此不同；

将细胞暴露于多个不同信号探针，其中所述信号探针在与该多个内源 RNA 中的一个或多个 RNA 杂交时可以产生可检测信号；和  
分离产生该信号的细胞。

20. 分离多个细胞的方法，其中至少一个细胞子集表达另一细胞子集不表达的 RNA，该方法包括步骤：

将编码多个 RNA 的多个 DNA 引入细胞，其中每一个 DNA 都还编码标签序列；

将所述细胞暴露于在与所述标签序列杂交时可以产生可检测信号的信号探针；和

分离产生该信号的所述细胞。

21. 权利要求 18-20 之任一项的方法，其中所述多个 RNA 形成表达文库。

22. 权利要求 18-20 之任一项的方法，其中所述多个 RNA 或由所述多个 RNA 编码的蛋白质选自：相同或相关生物学途径中的 RNA 或蛋白质，在彼此上游或下游起作用的 RNA 或蛋白质，对彼此具有调节、激活或阻遏功能的 RNA 或蛋白质，对于功能或活性而言相互依赖的 RNA 或蛋白质，

属于相同复合物的成分的 RNA 或蛋白质，以及来自相同蛋白质家族的蛋白质。

23. 权利要求 20 的方法，其中至少所述多个 DNA 的一个子集编码相同标签序列。

24. 分离两种或更多种 RNA 表达文库细胞的方法，包括步骤：

将编码第一 RNA 表达文库的多个 DNA 引入细胞，其中每个 DNA 还编码第一标签序列；

将编码第二 RNA 表达文库的多个 DNA 引入该细胞，其中每个 DNA 还编码第二标签序列；

将该细胞暴露于在与所述第一标签序列杂交时可以产生可检测信号的第一信号探针；

将该细胞暴露于在与所述第二标签序列杂交时可以产生可检测信号的第二信号探针；和

分离产生两个信号的细胞。

25. 权利要求 1-24 之任一项的方法，还包括步骤：培养分离的细胞。

26. 权利要求 1-24 之任一项的方法，还包括步骤：通过培养分离的细胞，产生细胞系或多个细胞系。

27. 权利要求 4、10、11、20 或 24 之任一项的方法，其中标签序列包含多个靶序列，其中每一个靶序列有一个信号探针与其杂交。

28. 权利要求 4、10、11、20 或 24 之任一项的方法，其中标签序列是具有二级结构的 RNA。

29. 权利要求 28 的方法，其中标签序列形成包含茎区、第一茎环区

和第二茎环区的三臂联结结构。

30. 权利要求 4、10、11、20 或 24 之任一项的方法，其中 DNA 编码多个标签序列。

31. 权利要求 4、10、11、20 或 24 之任一项的方法，其中编码所述标签序列的 DNA 的阅读框与编码所述 RNA 的 DNA 的阅读框相符。

32. 权利要求 4、10、11、20 或 24 之任一项的方法，其中编码所述标签序列的 DNA 的阅读框与编码所述 RNA 的 DNA 的阅读框不相符。

33. 权利要求 1-16、18-20 或 24 之任一项的方法，还包括步骤：在提供或引入步骤之前，向细胞添加可以调节所述 RNA、另外的 RNA 或多个 RNA 的表达的化合物。

34. 分离蛋白质表达降低的细胞的方法，包括步骤：

将编码可以降低所述蛋白质的表达的反义 RNA 或 shRNA 的 DNA 引入细胞；

将细胞暴露于在与所述反义 RNA 或 shRNA 杂交时可以产生可检测信号的信号探针；和

分离产生该信号的细胞。

35. 权利要求 2、4、5、8、10、11、18、20 或 24 之任一项的方法，其中所述 DNA 可操作地与条件型启动子连接。

36. 权利要求 35 的方法，其中所述 RNA 对细胞具有致死性或损伤性。

37. 权利要求 2-6、8-11、16-20、24 或 34 之任一项的方法，其中所述 DNA 还编码选择标记，且其中所述方法还包括步骤：在将 DNA 引入

细胞中后但在将所述细胞暴露于信号探针之前，使用选择标记选择细胞。

38. 鉴定可以激活条件型启动子的化合物的方法，包括步骤：  
向通过权利要求 16 的方法分离的细胞添加待测化合物；  
分析处在条件型启动子控制下的第二 RNA 的存在；和  
如果细胞表达第二 RNA，则将该待测化合物鉴定为可以激活此组织特异性启动子的化合物。

39. 获得可以激活条件型启动子的 RNA 的方法，包括步骤：  
获得通过权利要求 17 的方法分离的细胞；  
分析处在条件型启动子控制下的第二 RNA 的存在；和  
获得表达第二 RNA 的细胞。

40. 定量生物学样品中 RNA 的表达水平的方法，包括步骤：  
将生物学样品暴露于在与所述 RNA 杂交时可以产生可检测信号的第一信号探针；  
定量所述生物学样品中该信号的水平；和  
将所述信号水平与所述 RNA 的表达水平相关联。

41. 鉴定可以调节 RNA 的表达的化合物的方法，包括步骤：  
向表达所述 RNA 的细胞添加待测化合物；  
将细胞暴露于在与所述 RNA 杂交时可以产生可检测信号的信号探针；  
将暴露于待测化合物的细胞产生的信号与未暴露于待测化合物的细胞产生的信号进行比较；  
其中与未暴露于待测化合物的细胞产生的信号相比暴露于待测化合物的细胞产生的信号增加或降低，说明该化合物是可以调节所述 RNA 的表达的化合物。

42. 权利要求 41 的方法，其中 RNA 由被引入细胞中的 DNA 编码。
43. 鉴定可以调节 RNA 的表达的 RNA 的方法，包括步骤：  
向包含 RNA 的细胞引入待测 RNA；  
将细胞暴露于在与所述 RNA 杂交时可以产生可检测信号的信号探针；  
将暴露于待测 RNA 的细胞产生的信号与未暴露于待测 RNA 的细胞产生的信号进行比较；  
其中与未暴露于待测 RNA 的细胞产生的信号相比暴露于待测 RNA 的细胞产生的信号增加或降低，说明该待测 RNA 可以调节所述 RNA 的表达。
44. 鉴定活细胞中的遗传重组事件的方法，包括步骤：  
将细胞暴露于在与自重组过的序列转录的 RNA 杂交时可以产生可检测信号的信号探针，  
其中，检测到产生该信号的细胞说明该细胞包含该遗传重组事件。
45. 权利要求 44 的方法，还包括步骤：分离产生该信号的细胞。
46. 通过权利要求 1-24、34 或 45 之任一项的方法获得的细胞。
47. 权利要求 46 的细胞，其中所述细胞在基于细胞的试验中使用。
48. 权利要求 46 的细胞，其中将细胞植入动物体内。
49. 权利要求 46 的细胞，其中细胞是胚胎干细胞。
50. 产生转基因动物或嵌合动物的方法，包括步骤：使用权利要求 49 的胚胎干细胞产生所述转基因动物或嵌合动物。

51. 权利要求 1-45 之任一项的方法, 其中信号探针包含两条分开的核酸或经修饰核酸链, 这两条链形成至少一个相互互补区。

52. 权利要求 51 的方法, 其中分开的两条链从 5' 至 3' 端形成连续的相互互补区, 并且其中两条链具有相同数目的核苷酸。

53. 权利要求 51 的方法, 其中在两条链之间形成相互互补区之后, 一条链的 5' 末端与另一条链不重合, 或者该链的 3' 末端与另一条链不重合, 或者两者兼而有之, 其中不重合是不超过 10 个核苷酸或修饰的核苷酸。

54. 权利要求 51 的方法, 其中核酸是 DNA、RNA 或两者。

55. 权利要求 51 的方法, 其中经修饰的核酸包括肽核酸、化学修饰的 DNA、化学修饰的 RNA 或其组合。

56. 权利要求 51 的方法, 其中经修饰的核酸在糖基团、磷酸二酯键和碱基基团中的一个或多个上被化学修饰。

57. 权利要求 56 的方法, 其中磷酸二酯键被选自以下的化学基团替代 :  
 $--OP(OH)(O)O--$ ,  $-OP(O^-M^+)(O)O--$ ,  $-OP(SH)(O)O--$ ,  
 $-OP(SM^+)(O)O--$ ,  $--NHP(O)_2O--$ ,  $--OC(O)_2O--$ ,  $--OCH_2C(O)_2NH--$ ,  
 $--OCH_2C(O)_2O--$ ,  $--OP(CH_3)(O)O--$ ,  $--OP(CH_2C_6H_5)(O)O--$ ,  $--P(S)(O)O--$   
 和  $--OC(O)_2NH--$ 。

58. 权利要求 56 的方法, 其中磷酸二酯键被  $--OP(SH)(O)O--$ ,  $-OP(SM^+)(O)O--$  或  $--P(S)(O)O--$  替代。

59. 权利要求 56 的方法, 其中化学修饰的 RNA 的 2' 位包含选自以

下的化学基团： $C_1-C_4$  烷氧基、 $OCH_2-CH=CH_2$ 、 $OCH_2-CH=CH-CH_3$ 、 $OCH_2-CH=CH-(CH_2)_nCH_3$  ( $n=0, 1, \dots, 30$ )、卤素、 $C_1-C_6$  烷基和  $OCH_3$ 。

60. 权利要求 56 的方法，其中化学修饰的 RNA 包含 2' -O-甲基取代。

61. 权利要求 51 的方法，其中信号探针包含含有两个化学基团的相互作用对，其中一个化学基团位于一条链的一末端，且另一化学基团位于另一条链的相邻末端。

62. 权利要求 51 的方法，其中信号探针具有两个相互作用对，其中探针的每一端各具有一个相互作用对位于两条链的相邻末端。

63. 权利要求 61 的方法，其中相互作用对选自：荧光团和淬灭剂、化学发光标记物和淬灭剂或加合物、染料二聚体、FRET 供体和受体、收获和发射荧光团、以及酶和该酶的抑制剂或能够可逆地失活该酶的其他分子。

64. 权利要求 1-45 之任一项的方法，其中信号探针包含茎环结构。

65. 权利要求 1-45 之任一项的方法，其中信号探针包含含有两个茎区的哑铃结构。

66. 权利要求 1-45 之任一项的方法，其中信号探针包含形成三臂联结结构的一条链，其中该三臂联结结构包含茎区、第一茎环区和第二茎环区。

67. 权利要求 65 或 66 的方法，其中所述结构的区域之间通过磷酸二酯键或修饰的磷酸二酯键经由各茎区的臂连接。

68. 权利要求 64-66 之任一项的方法，其中所述结构是 DNA、RNA、肽核酸、化学修饰的 DNA、RNA 或其组合。

69. 权利要求 64-66 之任一项的方法，其中所述结构在糖基团、磷酸二酯键和碱基中的一个和多个上被化学修饰。

70. 权利要求 69 的方法，其中磷酸二酯键被选自以下的化学基团替代：  
 $--OP(OH)(O)O--$ ， $-OP(O^-)(O)O--$ ， $-OP(SH)(O)O--$ ，  
 $-OP(S^-)(O)O--$ ， $--NHP(O)_2O--$ ， $--OC(O)_2O--$ ， $--OCH_2C(O)_2NH--$ ，  
 $--OCH_2C(O)_2O--$ ， $--OP(CH_3)(O)O--$ ， $--OP(CH_2C_6H_5)(O)O--$ ， $--P(S)(O)O--$   
 和 $--OC(O)_2NH--$ 。

71. 权利要求 69 的方法，其中磷酸二酯键被 $--OP(SH)(O)O--$ ， $-OP(S^-)(O)O--$ 或 $--P(S)(O)O--$ 替代。

72. 权利要求 69 的方法，其中化学修饰的 RNA 的 2' 位包含选自以下的化学基团： $C_1-C_4$  烷氧基、 $OCH_2-CH=CH_2$ 、 $OCH_2-CH=CH-CH_3$ 、 $OCH_2-CH=CH-(CH_2)_nCH_3$  ( $n=0, 1, \dots, 30$ )、卤素、或  $C_1-C_6$  烷基和  $OCH_3$ 。

73. 权利要求 69 的方法，其中化学修饰的 RNA 具有 2' -O-甲基取代。

74. 权利要求 64-66 之任一项的方法，其中相互作用对选自：荧光团和淬灭剂、化学发光标记物和淬灭剂或加合物、染料二聚体、FRET 供体和受体、捕获和发射荧光团、蛋白水解酶和该蛋白水解酶的抑制剂或能够可逆地失活该酶的其他分子。

75. 权利要求 64-66 之任一项的方法，其中信号探针在链的一端包

含至少两个荧光团,并在链的另一端包含淬灭剂,其中两个荧光团是 FRET 供体和受体对。

76. 包含核酸或修饰的核酸、酶及该酶的抑制剂的探针,其中所述核酸或修饰的核酸含有与靶序列互补的序列和相互互补序列,其中所述探针在与靶序列杂交时可产生可检测的酶活性增加。

77. 权利要求 76 的探针,其中所述酶是蛋白水解酶。

78. 权利要求 76 的探针,其中探针包含两条分开的核酸或经修饰核酸链,这两条链形成至少一个相互互补区。

79. 权利要求 76 的探针,其中酶位于一条链的一个末端,而该酶的抑制剂位于另一条链的相邻末端。

80. 权利要求 78 的探针,其中分开的两条链从 5' 至 3' 端形成连续的相互互补区,并且两条链具有相同数目的核苷酸。

81. 权利要求 78 的探针,其中在两条链之间形成相互互补区之后,一条链的 5' 末端与另一条链不重合,或者该链的 3' 末端与另一条链不重合,或者两者兼而有之,其中不重合是不超过 10 个核苷酸或修饰的核苷酸。

82. 权利要求 78 的探针,其中核酸是 DNA、RNA 或两者。

83. 权利要求 78 的探针,其中经修饰的核酸包括肽核酸、化学修饰的 DNA 或 RNA 或其组合。

84. 权利要求 78 的探针,其中经修饰的核酸在糖基团、磷酸二酯键

和碱基基团中的一个或多个上被化学修饰。

85. 权利要求 84 的探针, 其中磷酸二酯键被选自以下的化学基团替代:  $--OP(OH)(O)O--$ ,  $-OP(O^M)(O)O--$ ,  $-OP(SH)(O)O--$ ,  $-OP(S^M)(O)O--$ ,  $--NHP(O)_2O--$ ,  $--OC(O)_2O--$ ,  $--OCH_2C(O)_2NH--$ ,  $--OCH_2C(O)_2O--$ ,  $--OP(CH_3)(O)O--$ ,  $--OP(CH_2C_6H_5)(O)O--$ ,  $--P(S)(O)O--$  和  $--OC(O)_2NH--$ 。

86. 权利要求 84 的探针, 其中磷酸二酯键被  $--OP(SH)(O)O--$ ,  $-OP(S^M)(O)O--$  或  $--P(S)(O)O--$  替代。

87. 权利要求 84 的探针, 其中化学修饰的 RNA 的 2' 位包含选自以下的化学基团:  $C_1-C_4$  烷氧基、 $OCH_2-CH=CH_2$ 、 $OCH_2-CH=CH-CH_3$ 、 $OCH_2-CH=CH-(CH_2)_nCH_3$  ( $n=0, 1, \dots, 30$ )、卤素、 $C_1-C_6$  烷基和  $OCH_3$ 。

88. 权利要求 84 的探针, 其中化学修饰的 RNA 包含 2' -O-甲基取代。

89. 权利要求 76 的探针, 其包含茎环结构。

90. 权利要求 76 的探针, 其包含含有两个茎区的哑铃结构。

91. 权利要求 76 的探针, 其包含三臂联结结构, 其中该三臂联结结构包含茎区、第一茎环区和第二茎环区。

92. 权利要求 90 或 91 的探针, 其中所述结构的区域之间通过磷酸二酯键或修饰的磷酸二酯键经由各茎区的臂连接。

93. 权利要求 90-92 之任一项的探针, 其中所述结构是 DNA、RNA、

肽核酸、化学修饰的 DNA、RNA 或其组合。

94. 权利要求 90-92 之任一项的探针，其中所述结构在糖基团、磷酸二酯键和碱基中的一个和多个上被化学修饰。

95. 权利要求 94 的探针，其中磷酸二酯键被选自以下的化学基团替代：  
 $--OP(OH)(O)O--$ ， $-OP(O^M)(O)O--$ ， $-OP(SH)(O)O--$ ，  
 $-OP(S^M)(O)O--$ ， $--NHP(O)_2O--$ ， $--OC(O)_2O--$ ， $--OCH_2C(O)_2NH--$ ，  
 $--OCH_2C(O)_2O--$ ， $--OP(CH_3)(O)O--$ ， $--OP(CH_2C_6H_5)(O)O--$ ， $--P(S)(O)O--$   
 和 $--OC(O)_2NH--$ 。

96. 权利要求 94 的探针，其中磷酸二酯键由 $--OP(SH)(O)O--$ ， $-OP(S^M)(O)O--$ 或 $--P(S)(O)O--$ 替代。

97. 权利要求 94 的探针，其中化学修饰的 RNA 的 2' 位包含选自以下的化学基团： $C_1-C_4$  烷氧基、 $OCH_2-CH=CH_2$ 、 $OCH_2-CH=CH-CH_3$ 、 $OCH_2-CH=CH-(CH_2)_nCH_3$  ( $n=0, 1, \dots, 30$ )、卤素、 $C_1-C_6$  烷基和  $OCH_3$ 。

98. 权利要求 94 的探针，其中化学修饰的 RNA 包含 2' -O-甲基取代。

99. 包含编码目的 RNA 和标签序列的 DNA 序列的 DNA 构建体。

100. 包含权利要求 99 的 DNA 构建体的载体。

101. 包含权利要求 99 的 DNA 构建体的细胞。

102. 权利要求 101 的细胞，其中细胞是永生化细胞、原代细胞、干细胞或生殖细胞。

103. 包含至少 1,000 个细胞系的哺乳动物细胞系文库, 其中每个细胞系都包含稳定整合的表达的序列。

104. 包含至少 500 个细胞系的哺乳动物细胞系文库, 其中每个细胞系都包含至少两个稳定整合的序列。

105. 包含至少 50 个细胞系的哺乳动物细胞系文库, 其中每个细胞系都包含至少三个稳定整合的序列。

106. 包含至少 20 个细胞系的哺乳动物细胞系文库, 其中每个细胞系都包含至少 4 个稳定整合的序列。

107. 包含至少 50 个细胞系的哺乳动物细胞系文库, 其中每个细胞系都包含至少 1 个稳定整合的序列, 其中所述细胞系缺乏药物抗性基因。

108. 包含至少 20 个细胞系的哺乳动物细胞系文库, 其中每个细胞系都包含至少 2 个稳定整合的序列, 其中所述细胞系缺乏药物抗性基因。

109. 权利要求 103-108 之任一项的文库, 其中每个细胞系都包含可变文库序列。

110. 权利要求 109 的文库, 其中所述表达文库的可变序列选自: 基因组序列、基因组的非翻译序列、基因组的翻译序列、基因序列、cDNA 序列、EST 序列、寡聚体序列、随机序列、RNA 序列、蛋白质序列、蛋白质域序列、肽序列、内含子序列、外显子序列、标签序列或接头序列, 或其组合或其重组物、或这些未修饰的、经诱变的、随机化的、改组的或重组的序列中的一个或多个。

111. 权利要求 103-108 之任一项的文库, 其中文库是使用具有未知

序列身份的多个不同序列产生的。

112. 权利要求 103-108 之任一项的文库，其中该文库包含多个构建体，其中每个构建体包含具有已知序列身份的序列。

113. 权利要求 111 或 112 的文库，其中所述序列共有序列同源性、功能意义或相关来源。

114. 权利要求 103-108 之任一项的文库，其中文库用于基于细胞的筛选试验中。

115. 鉴定可以增强信号探针对细胞中靶的检测的化合物的方法，包括步骤：

将信号探针引入包含靶序列的细胞中，其中信号探针在与靶序列杂交时产生可检测信号；

将细胞暴露于待测化合物；和

检测细胞产生的该信号，

其中与未暴露于待测化合物的细胞产生的信号相比，暴露于待测化合物的细胞产生的信号增加说明该待测化合物是可以增强信号探针对细胞中靶的检测的化合物。

116. 鉴定可以介导或提高信号探针向细胞中的引入的化合物的方法，包括步骤：

在待测化合物存在下将细胞暴露于信号探针，其中细胞包含靶序列且其中信号探针在与靶序列杂交时产生信号；和

检测细胞产生的该信号，

其中与未暴露于待测化合物的细胞相比，暴露于待测化合物的细胞产生的信号增加说明该待测化合物是可以介导或提高信号探针向细胞中的引入的化合物。

## 使用信号探针的方法和材料

本申请要求 2004 年 2 月 18 日提交的美国临时申请 60/546,075 的优先权，该临时申请的公开内容完整地并入此处作为参考。

### 背景技术

识别和报道特定核酸序列的存在的核酸探针已经用于主要在体外反应中检测特定核酸。见例如美国专利 5,925,517，该专利并入此处作为参考。一类探针经过设计而具有发夹形状的结构，其中中间一段核苷酸序列与靶序列互补，而两末端包含短相互互补序列。见例如，Tyagi 和 Kramer, *Nature Biotechnology*, 14, 303-308 (1996)，并入此处作为参考。该茎环形状的探针的一个末端与荧光团共价结合，而另一末端与淬灭部分共价结合。当探针处于具有杂交的两末端的天然状态时，荧光团和淬灭剂接近以致产生相对低的荧光或基本不产生荧光。茎环探针在与其靶核酸杂交时发生构象改变，导致从荧光团产生的荧光出现可检测的变化。研究者已经使用发夹形状的探针在活细胞中原位显现信使 RNA (Matsuo 1998, *Biochim. Biophys. Acta* 1379: 178-184)。

### 发明概述

本发明提供用于分析或分离细胞或产生细胞系的包含新信号探针 (signaling probes) 的方法和组合物，其中所述细胞和细胞系表达一种或多种 RNA。该方法基于检测探针在与靶序列杂交时所产生的信号。所述 RNA 可以通过 DNA 构建体引入细胞中，或者细胞可能被怀疑内源性表达该 RNA。DNA 构建体可以进一步编码标签序列，而信号探针与该标签序列互补。一个实施方案中，分离的细胞或产生的细胞系功能性地不表达一种或多种预定的蛋白质或 RNA，或者降低地表达一种或多种预定的蛋白质或 RNA。本发明还提供使用根据所述方法分离的细胞产生转基因

因动物的方法。

本发明提供在用于定量一种或多种 RNA 转录物的表达水平的方法中使用的信号探针。此外，该信号探针可以在用于鉴定调节至少一种预定 RNA 的转录的化合物或 RNA 序列的方法中使用。另一个实施方案中，信号探针在用于鉴定活细胞中的遗传重组事件的方法中使用。信号探针包含一条或多条核苷酸链，其中信号探针包含与目的靶核酸（例如 RNA）互补的核苷酸，并且其中信号探针还包含含有两个结构部分（moieties）的相互作用对。信号探针的结构使得当该信号探针未与靶序列杂交时该相互作用对的这两个结构部分的物理位置将造成无信号产生或造成背景信号产生。当信号探针与靶序列杂交时，这两个组分部分将导致信号产生。或者，信号探针的这些结构部分可以使得当探针未与靶杂交时产生特定的信号而在探针与靶序列杂交时产生不同的信号。相互作用对的这两个结构部分可以附着于信号探针的一条或多条链的一个或多个末端。或者，所述结构部分可以掺入到信号探针的一条或多条链的内部。信号探针的核苷酸也可以被修饰。

本发明还提供蛋白酶探针。一个实施方案中，信号探针或蛋白酶探针包含两条分开的核酸链或修饰的核酸链，其中这两条链的一个或多个部分相互退火，并且一条链的至少一个末端与另一条链的末端相邻。所述核酸可以是 DNA 或 RNA。对于具有两条分开的链的信号探针，在一个实施方案中，一条链至少在一个末端具有淬灭部分（quencher moiety），而另一条链至少在所述相邻末端具有荧光团。对于蛋白酶探针，一个实施方案中，一条链至少在一个末端具有蛋白水解酶，而另一条链至少在所述相邻末端具有该蛋白水解酶的抑制剂。

另一实施方案中，信号探针或蛋白酶探针经设计而包含至少一个相互互补区和至少一个非互补区。一个实施方案中，至少一个非互补区可以经设计而形成环区。在一个实施方案中，设计探针以形成至少一个茎环结构。另一实施方案中，探针形成哑铃结构或三臂联结结构（three-arm junction structure）。一个实施方案中，信号探针具有至少一个荧光团和至少一个淬灭部分各分别位于链的一个末端。一个实施方案中，蛋白

酶探针具有蛋白水解酶和该蛋白水解酶的抑制剂各分别位于链的一个末端。

另一实施方案中，信号探针或蛋白酶探针是化学修饰的。糖-磷酸二酯类型主链、2' OH 和嘌呤或嘧啶碱基中的一个或多个被修饰。一个实施方案中，脱氧核糖主链被肽核酸替代。

一个实施方案中，标签序列是结构 RNA，即，该 RNA 具有二级结构，优选地三臂联结结构。一个实施方案中，标签序列包含根据图 42 A、B 或 C 的结构或序列。本发明还提供包含至少一个编码至少一种目的 RNA 和标签序列的 DNA 的 DNA 构建体。本发明还提供包含该 DNA 构建体的载体和细胞。

在其它实施方案中，提供：

1. 分离表达至少一种 RNA 的细胞的方法，包括步骤：
  - a) 向细胞中引入编码至少一种 RNA 的至少一种 DNA；
  - b) 将所述细胞暴露于至少一种在与所述至少一种 RNA 杂交时产生可检测信号的信号探针；和
  - c) 分离产生该信号的所述细胞。
2. 段 1 的方法，还包括步骤：通过使所述分离的细胞生长，产生表达所述至少一种 RNA 的细胞系或多个细胞系。
3. 分离表达两种或两种以上 RNA 中的至少一种 RNA 的细胞的方法，包括步骤：
  - a) 将编码第一 RNA 的第一 DNA 引入细胞；
  - b) 将编码至少一种第二 RNA 的至少一种第二 DNA 引入所述细胞；
  - c) 将所述细胞暴露于至少一种在与所述第一 RNA 杂交时产生可检测信号的第一信号探针；
  - d) 将所述细胞暴露于至少一种在与所述至少一种第二 RNA 杂交时产生可检测信号的第二信号探针；和
  - e) 分离在所述信号探针与它们相应的 RNA 杂交时产生所述信号中的至少一种的细胞。
4. 段 3 的方法，还包括步骤：通过使所述分离的细胞生长，产生表

达所述两种或两种以上 RNA 中的至少一种 RNA 的一个细胞系或多个细胞系。

5. 分离多个细胞的方法，其中所述细胞中的至少一部分细胞表达至少一种不同的 RNA，所述方法包括步骤：

a) 将编码多种 RNA 的多种 DNA 引入细胞，其中至少一部分所述细胞中引入了编码至少一种不同的 RNA 的至少一种不同的 DNA；

b) 将所述细胞相继地或同时地暴露于多种信号探针，其中所述信号探针在与所述多种 RNA 杂交时产生可检测的信号；和

c) 分离产生该信号的所述细胞。

6. 段 5 的方法，还包括步骤：通过使所述分离的细胞生长，产生多个表达至少一种不同的 RNA 的细胞系。

7. 分离表达至少一种 RNA 的细胞的方法，包括步骤：

a) 将至少一种编码所述至少一种 RNA 和至少一种标签序列的 DNA 引入细胞；

b) 将所述细胞暴露于至少一种信号探针，其中所述探针在与标签序列杂交时产生可检测的信号；和

c) 分离产生该信号的所述细胞。

8. 段 7 的方法，还包括步骤：通过使所述分离的细胞生长，产生一个或多个表达所述至少一种 RNA 的细胞系。

9. 分离表达两种或两种以上 RNA 中的至少一种 RNA 的细胞的方法，包括步骤：

a) 将编码第一 RNA 和至少一种第一标签序列的第一 DNA 引入细胞；

b) 将至少一种编码至少一种另外的 RNA 和至少一种第二标签序列的第二 DNA 引入所述细胞，其中所述第二标签序列与第一标签序列相同或不同；

c) 将所述细胞暴露于至少一种第一信号探针，其中所述探针在与第一标签序列杂交时产生可检测信号；

d) 将所述细胞暴露于至少一种第二信号探针，其中所述探针在与第二标签序列杂交时产生可检测信号；和

e) 分离在所述信号探针与它们相应的 RNA 杂交时产生所述信号中的至少一种信号的细胞。

10. 段 9 的方法, 还包括步骤: 通过使所述分离的细胞生长, 产生表达所述两种或两种以上 RNA 中的至少一种 RNA 的一个细胞系或多个细胞系。

11. 段 3 或 9 的方法, 其中所述第一 RNA 的所述步骤与所述至少一种另外的 RNA 的相应步骤同时地或相继地实施。

12. 段 3 或 9 的方法, 其中所述两种或两种以上 RNA 或由所述两种或两种以上 RNA 编码的蛋白质选自: 相同或相关生物学途径中的 RNA 或蛋白质、在彼此上游或下游起作用的 RNA 或蛋白质、相互具有调节、激活或阻遏功能的 RNA 或蛋白质、对于功能或活性而言相互依赖的 RNA 或蛋白质、形成复合物的 RNA 或蛋白质、来自蛋白质家族的蛋白质。

13. 分离过表达至少一种 RNA 的细胞的方法, 包括步骤:

a) 向细胞中引入编码所述至少一种 RNA 和至少一种第一标签序列的至少一种第一 DNA; 和编码所述至少一种 RNA 和至少一种第二标签序列的至少一种第二 DNA, 其中第一和第二 DNA 构建体的引入相继地或同时地进行, 其中第一和第二标签序列相同或不同;

b) 将所述细胞暴露于至少一种在与所述至少一种第一标签序列杂交时产生可检测信号的第一信号探针, 和至少一种在与所述至少一种第二标签序列杂交时产生可检测信号的第二信号探针; 和

c) 分离在所述信号探针与它们相应的 RNA 杂交时产生所述信号中的至少一种信号的细胞。

14. 段 13 的方法, 还包括步骤: 通过使所述分离的细胞生长, 产生过表达所述 RNA 的一个细胞系或多个细胞系。

15. 段 3、9 和 13 的方法, 其中第一信号探针产生的信号与第二信号探针产生的信号不同。

16. 段 13 的方法, 其中所述第一信号探针的所述步骤与所述第二信号探针的相应步骤同时或相继地进行。

17. 段 3、9 和 13 之任一的方法, 其中所述第一 DNA 和所述第二 DNA

位于相同构建体或不同构建体上。

18. 分离多个细胞的方法，其中所述细胞中的至少一部分表达至少一种不同的 RNA，所述方法包括步骤：

a) 将编码多种 RNA 和至少一种标签序列的多种 DNA 引入细胞中，其中所述细胞中的至少一部分细胞被引入了编码至少一种不同 RNA 的至少一种不同的 DNA；

b) 将所述细胞暴露于至少一种在与所述标签序列杂交时产生可检测信号的信号探针；和

c) 分离产生该信号的所述细胞。

19. 段 18 的方法，其中所述多种 RNA 形成至少一个表达文库。

20. 段 18 的方法，还包括步骤：通过使所述分离的细胞生长，产生多个表达至少一种不同的 RNA 的细胞系。

21. 分离两种或两种以上 RNA 表达文库细胞中的至少一种的方法，包括步骤：

a) 向细胞中引入编码至少一种第一 RNA 表达文库和至少一种第一标签序列的 DNA；

b) 向细胞中引入编码至少一种第二 RNA 表达文库和至少一种第二标签序列的 DNA，其中所述第二标签序列与第一标签序列相同或不同；

c) 将所述细胞暴露于至少一种第一信号探针，其中所述第一信号探针在与所述至少一种第一标签序列杂交时可产生可检测信号；

d) 将所述细胞暴露于至少一种第二信号探针，其中所述第二信号探针在与所述至少一种第二标签序列杂交时可产生可检测信号，其中来自第一信号探针的可检测信号与来自第二信号探针的可检测信号相同或不同；和

e) 分离产生所述信号中的至少一种信号的所述细胞。

22. 段 21 的方法，还包括步骤：通过使所述分离的细胞生长，产生所述两种或两种以上 RNA 表达文库中的至少一种。

23. 段 5 或 18 的方法，其中所述多种 RNA 或由所述多种 RNA 编码的蛋白质选自：相同或相关生物学途径中的 RNA 或蛋白质、在彼此上游或

下游起作用的 RNA 或蛋白质、相互具有调节、激活或阻遏功能的 RNA 或蛋白质、对于功能或活性而言相互依赖的 RNA 或蛋白质、形成复合物的 RNA 或蛋白质、来自蛋白质家族的蛋白质。

24. 段 18 的方法，其中所述多种 DNA 中的至少一部分编码相同标签序列。

25. 段 7、9、13、18 或 21 的方法，其中标签序列包含多个待被至少一种信号探针所识别的靶序列。

26. 段 7、9、13、18 和 21 之任一的方法，其中标签序列是结构 RNA。

27. 段 25 的方法，其中标签序列形成三臂联结结构。

28. 段 27 的方法，其中茎区包含 8-9 个碱基对，第一茎环区包含 4-6 个碱基对，第二茎环区包含 13-17 个碱基对。

29. 段 27 的方法，其中三个臂的茎区还包含非互补区。

30. 段 27 的方法，其中茎区和第一茎环区均还包含一个错配区，第二茎环区还包含 2-7 个错配或凸起区。

31. 段 27 的方法，其中茎区之间的连接具有总共 8-12 个核苷酸。

32. 段 26 的方法，其中标签序列选自图 42 A、B 和 C 中显示的序列。

33. 段 32 的方法，其中标签序列形成由图 42A、B 或 C 中所示序列预测的能量上更有利的结构。

34. 段 27 的方法，其中靶序列是从第一茎环区茎 3' 侧的全部或部分至第一和第二茎环区之间的连接、至第二茎环区茎 5' 侧的全部或部分的区域。

35. 段 7、9、13、18 和 21 之任一的方法，其中 DNA 的至少一部分编码多个相同的标签序列。

36. 段 35 的方法，其中 DNA 的至少一部分编码至多 50 个的相同标签序列。

37. 段 7、9、13、18 和 21 之任一的方法，其中编码所述标签序列的 DNA 与编码所述 RNA 的 DNA 阅读框相符。

38. 段 7、9、13、18 和 21 之任一的方法，其中编码所述标签序列的 DNA 与编码所述 RNA 的 DNA 阅读框不相符。

39. 分离表达至少一种 RNA 的细胞的方法, 包括步骤:

- a) 提供潜在地表达所述至少一种 RNA 的细胞;
- b) 将所述细胞暴露于至少一种信号探针, 其中所述探针在与所述至少一种 RNA 杂交时可产生可检测信号;
- c) 分离产生该信号的所述细胞。

40. 段 39 的方法, 还包括步骤: 通过使所述分离的细胞生长, 产生表达所述至少一种 RNA 的细胞系或多个细胞系。

41. 段 39 的方法, 其中所述细胞还潜在地表达一种或多种另外的 RNA, 所述方法还包括步骤:

- a) 将所述细胞暴露于至少一种另外的信号探针, 其中所述探针在与所述至少一种另外的 RNA 杂交时可产生可检测信号; 和
- b) 分离产生该信号的细胞。

42. 段 41 的方法, 还包括步骤: 通过使所述分离的细胞生长, 产生表达所述至少一种 RNA 和另外的 RNA 的细胞系或多个细胞系。

43. 段 1、3、5、7、9、13、18 和 21 之任一的方法, 还包括步骤: 在步骤 a) 之前向细胞添加化合物, 其中所述化合物调制或调节所述 RNA、另外的 RNA 或多种 RNA 的表达。

44. 段 43 的方法, 其中化合物诱导所述 RNA、另外的 RNA 或多种 RNA 的表达。

45. 段 5 或 18 的方法, 其中有关所述多种 RNA 的所述步骤同时或相继地进行。

46. 分离表达至少一种外源 RNA 和至少一种内源 RNA 的细胞的方法, 包括步骤:

- a) 将编码所述至少一种外源 RNA 的 DNA 引入细胞, 其中所述细胞潜在地表达至少一种内源 RNA;
- b) 将所述细胞暴露于至少一种第一信号探针, 其中所述探针在与所述至少一种外源 RNA 杂交时可产生可检测信号;
- c) 将所述细胞暴露于至少一种第二信号探针, 其中所述探针在与所述至少一种内源 RNA 杂交时可产生可检测信号; 和

d) 分离在所述信号探针与它们的相应 RNA 杂交时产生至少一种所述信号的细胞。

47. 段 46 的方法, 还包括步骤: 通过使所述分离的细胞生长, 产生表达所述至少一种外源 RNA 或所述至少一种内源 RNA 或两者的细胞系或多个细胞系。

48. 段 46 的方法, 其中有关所述外源 RNA 的所述步骤与有关所述内源 RNA 的相应步骤同时地或相继地进行。

49. 段 46 的方法, 其中所述第二信号探针与第一信号探针产生不同的信号。

50. 段 46 的方法, 其中内源 RNA 和外源 RNA 或由所述内源和外源 RNA 编码的蛋白质选自: 相同或相关生物学途径中的 RNA 或蛋白质、在彼此上游或下游起作用的 RNA 或蛋白质、相互具有调节、激活或阻遏功能的 RNA 或蛋白质、对于功能或活性而言相互依赖的 RNA 或蛋白质、形成复合物的 RNA 或蛋白质、来自蛋白质家族的蛋白质。

51. 段 1、3、5、7、9、13、18、21 和 39 之任一的方法, 其中 RNA 包含编码蛋白质的信使 RNA、编码肽的 RNA、反义 RNA、siRNA、tRNA、结构 RNA、核糖体 RNA、hnRNA 和 snRNA 中的一个或多个。

52. 段 51 的方法, 其中所述蛋白质选自: 细胞表面定位的蛋白质、分泌蛋白和细胞内蛋白。

53. 分离过表达至少一种第一蛋白质并且功能性地不表达至少一种第二蛋白质或减少地表达至少一种第二蛋白质的细胞的方法, 包括步骤:

a) 向所述细胞中引入至少一种第一 DNA 和至少一种第二 DNA, 其中所述第一 DNA 编码至少一种编码所述至少一种第一蛋白质的 RNA 以及至少一种第一标签序列; 所述第二 DNA 编码所述至少一种 RNA 以及至少一种第二标签序列; 其中所述第一和第二标签序列不同;

b) 向细胞中引入至少一种编码至少一种反义 RNA 或 siRNA 的 DNA, 其中所述反义 RNA 或 siRNA 结合或干扰所述至少一种第二蛋白质的 mRNA 转录物;

c) 将所述细胞暴露于至少一种在与所述至少一种第一标签序列杂

交时产生可检测信号的第一信号探针，和至少一种在与所述至少一种第二标签序列杂交时产生可检测信号的第二信号探针；

d) 将所述细胞暴露于至少一种在与所述至少一种反义 RNA 或 siRNA 杂交时产生可检测信号的信号探针；和

e) 分离在所述信号探针与它们的相应 RNA 杂交时产生至少一种所述信号的细胞。

54. 段 53 的方法，还包括步骤：产生过表达至少一种第一蛋白质并且功能性地不表达至少一种第二蛋白质或减少地表达至少一种第二蛋白质的细胞系或多个细胞系。

55. 段 53 的方法，其中有关所述第一蛋白质的所述步骤与有关所述第二蛋白质的相应步骤同时地或相继地进行。

56. 段 1、3、5、7、9、13、18、21、39 和 53 之任一的方法，其中所述 DNA 与条件型启动子可操作地连接。

57. 段 56 的方法，其中启动子是诱导型或阻遏型启动子，并且在步骤 (a) 之前向细胞添加最小量的诱导物或充足量的阻遏物。

58. 段 57 的方法，其中 RNA 是反义 RNA 或 siRNA。

59. 段 57 的方法，其中 RNA 对细胞而言是致死性的或破坏性的。

60. 段 1、3、5、7、9、13、18、21、39 和 53 之任一的方法，还包括步骤：在向细胞中引入 DNA 后但在将所述细胞暴露于所述信号探针之前，选择细胞。

61. 段 1、3、5、7、9、13、18、21、39 和 53 之任一的方法，其中至少一种 DNA 还编码至少一种药物抗性标记，并且所述方法还包括步骤：选择对至少一种药物具有抗性的细胞，其中所述药物是所述标记所赋予的抗性所针对的药物。

62. 分离包含 DNA 构建体的细胞的方法，其中所述 DNA 构建体编码在组织特异性启动子控制下的 RNA 序列，所述方法包括步骤：

a) 向细胞中引入至少一种 DNA 构建体，所述 DNA 构建体编码在组成型启动子控制下的至少一种第一 RNA 序列并编码在组织特异性启动子控制下的至少一种第二 RNA 序列；

b) 将所述细胞暴露于至少一种在与所述第一 RNA 序列杂交时产生可检测信号的信号探针; 和

c) 分离产生所述信号的所述细胞。

63. 段 62 的方法, 还包括步骤: 产生包含 DNA 构建体的细胞系或多个细胞系, 其中所述 DNA 构建体编码在组织特异性启动子控制下的 RNA 序列。

64. 段 62 的方法, 其中组织特异性启动子控制选择标记基因的表达。

65. 段 62 的方法, 其中选择标记基因是药物抗性基因或可检测的蛋白质的基因。

66. 鉴定可激活组织特异性启动子的化合物的方法, 包括步骤:

a) 向分离自段 62 的细胞添加化合物;

b) 通过选择标记鉴定细胞;

c) 将该化合物鉴定为可激活所述组织特异性启动子的化合物。

67. 分离包含 DNA 构建体的细胞的方法, 其中所述 DNA 构建体编码至少一种待检 RNA 序列和在组织特异性启动子控制下的 RNA 序列, 所述方法包括步骤:

a) 向细胞中引入至少一种 DNA 构建体, 该构建体编码在组成型启动子控制下的至少一种待检 RNA 序列、在相同或不同的第二组成型启动子控制下的至少一种第二 RNA 序列、和在组织特异性启动子控制下的至少一种第三 RNA 序列;

b) 将所述细胞暴露于至少一种在与所述第二 RNA 序列杂交时产生可检测信号的信号探针; 和

c) 分离产生所述信号的所述细胞。

68. 段 67 的方法, 还包括步骤: 产生包含 DNA 构建体的细胞系或多个细胞系, 所述 DNA 构建体编码至少一种待检 RNA 序列和在组织特异性启动子控制下的 RNA 序列。

69. 段 67 的方法, 其中待检 RNA 序列来自表达文库。

70. 段 67 的方法, 其中组织特异性启动子控制选择标记基因的表达。

71. 段 67 的方法, 其中选择标记基因是药物抗性基因或可检测的蛋

白质的基因。

72. 鉴定可激活组织特异性启动子的待检 RNA 序列的方法，包括步骤：

- a) 通过选择标记鉴定段 67 的分离的细胞；
- b) 鉴定可激活所述组织特异性启动子的待检 RNA 序列。

73. 段 1、3、5、7、9、13、18、21、39 和 53 之任一的方法，还包括步骤：

i) 将所述分离的细胞暴露于在与相应的 RNA 杂交时产生可检测信号的信号探针；

ii) 确定所述分离的细胞是否表达该相应的 RNA；或者定量信号水平以确定该相应 RNA 的表达水平。

74. 从段 1、3、5、7、9、13、18、21、39 和 53 之任一的方法获得的分离的细胞，其中所述细胞应用于基于细胞的试验中。

75. 从段 1、3、5、7、9、13、18、21、39 和 53 之任一的方法获得的分离的细胞，其中所述细胞可植入动物中。

76. 根据段 1、3、5、7、9、13、18、21、39 和 53 之任一产生表达 RNA 的转基因动物的方法，包括利用胚胎干细胞或可以植入动物体的细胞实施段 1、3、5、7、9、13、18、21、39 和 53 之任一的步骤，确定所述干细胞或细胞的生存力，和使用所述活的胚胎干细胞或细胞产生所述转基因动物。

77. 段 1、3、5、7、9、13、18、21 和 39 的方法，其中 RNA 是反义 RNA 或 siRNA，并且在所述分离的细胞中至少一种预定的蛋白质由于其 mRNA 转录物被该反义 RNA 结合或者被 siRNA 干扰而导致在功能上是无效的或者具有降低的表达水平。

78. 段 77 的方法，其中所述预定的蛋白质是基因产物的可变剪接形式。

79. 制备转基因动物的方法，其中至少一种预定的蛋白质功能性地不表达或者其表达降低，所述方法包括利用胚胎干细胞或可植入动物的细胞实施段 77 的步骤，确定所述干细胞或细胞的生存力，和使用所述活

的胚胎干细胞或细胞产生所述转基因动物。

80. 定量生物学样品中至少一种 RNA 转录物的表达水平的方法，包括步骤：

- a) 将所述生物学样品暴露于在与所述 RNA 转录物杂交时产生可检测信号的第一信号探针；
- b) 定量所述生物学样品中该信号的水平；和
- c) 将所述信号水平与所述至少一种 RNA 转录物的所述水平相关联。

81. 段 80 的方法，其中所述生物学样品是细胞样品、组织样品或来源于其的制备物。

82. 段 80 的方法，其中所述 RNA 转录物是编码蛋白质的信使 RNA、编码肽的 RNA、反义 RNA、siRNA、tRNA、结构 RNA、核糖体 RNA、hnRNA 和 snRNA 中的一种或多种。

83. 段 80 的方法，其中所述生物学样品被固定。

84. 段 80 的方法，其中使用第二信号探针在所述生物学样品中定量至少一种第二 RNA 转录物的表达水平，其中所述第二信号探针在与所述第二 RNA 转录物杂交时可产生可检测信号。

85. 鉴定调节至少一种预定的 RNA 的转录的化合物的方法，包括步骤：

- a) 向外源地或内源地表达所述预定的 RNA 的细胞添加化合物；
- b) 将所述细胞暴露于在与所述至少一种预定的 RNA 杂交时产生可检测信号的至少一种信号探针；
- c) 定量所述细胞中该信号的水平；
- d) 鉴定与未添加化合物的细胞中的信号相比具有增加或降低的信号水平的细胞；和
- e) 鉴定调节所述至少一种预定 RNA 的转录的化合物。

86. 段 85 的方法，其中根据段 1 或 7 的方法分离外源地表达所述预定 RNA 的细胞。

87. 段 85 的方法，其中 DNA 构建体包含启动子或操纵基因并编码阻遏蛋白、增强子 (enhancer) 或调节转录的序列。

88. 鉴定调节至少一种预定 RNA 的转录的 RNA 序列的方法, 包括步骤:

a) 向细胞中引入至少一种待测 RNA 序列, 其中所述待测 RNA 序列潜在地调节外源或内源表达的至少一种预定 RNA 的转录;

b) 将所述细胞暴露于在与所述至少一种预定 RNA 杂交时产生可检测信号的至少一种信号探针;

c) 定量所述细胞中该信号的水平;

d) 与无待测 RNA 序列的细胞的信号相比, 鉴定具有增加或降低的信号水平的细胞; 和

e) 鉴定调节所述至少一种预定 RNA 的转录的待测 RNA 序列。

89. 段 88 的方法, 其中外源表达所述预定 RNA 的细胞是根据段 1 或 7 的方法分离的。

90. 段 88 的方法, 其中使用 RNA 表达文库作为待测 RNA 序列。

91. 鉴定活细胞中的遗传重组事件的方法, 包括步骤:

a) 将细胞暴露于在与如下 RNA 序列杂交时产生可检测信号的信号探针, 所述 RNA 序列选自: 自重组后的序列转录的 RNA 序列和自未发生重组的序列转录的 RNA 序列;

b) 鉴定表达所述 RNA 序列的所述细胞。

92. 段 1、3、5、7、9、13、18、21、39、53、62、67、73、80、85、88 和 91 之任一的方法, 其中信号探针包含两条分开的核酸或经修饰核酸的链, 所述两条链形成至少一个相互互补区。

93. 段 92 的方法, 其中所述分开的两条链从 5' 至 3' 端形成连续的相互互补区, 并且这两条链具有相同数目的核苷酸。

94. 段 92 的方法, 其中在两条链之间形成相互互补区后, 一条链的 5' 端与另一条链不重合 (offset), 或者该链的 3' 端与另一条链不重合, 或者两者兼而有之, 其中所述不重合 (offset) 为不超过 10 个的核苷酸或修饰的核苷酸。

95. 段 92 的方法, 其中两条链在每一端形成 5 或 6 个连续碱基对的相互互补区。

96. 段 92 的方法, 其中两条链在序列上是不同的。
97. 段 92 的方法, 其中所述链具有 30 个以上的核苷酸或修饰的核苷酸。
98. 段 92 的方法, 其中核酸是 DNA、RNA 或两者。
99. 段 92 的方法, 其中修饰的核酸包括肽核酸、化学修饰的 DNA 或 RNA、或其组合。
100. 段 92 的方法, 其中修饰的核酸在糖基团、磷酸二酯键和碱基基团中的一个或多个上发生化学修饰。
101. 段 100 的方法, 其中磷酸二酯键由选自以下的化学基团替代:  
--OP(OH)(O)O--, -OP(O<sup>M+</sup>)(O)O--, -OP(SH)(O)O--, -OP(S<sup>M+</sup>)(O)O--,  
--NHP(O)<sub>2</sub>O--, --OC(O)<sub>2</sub>O--, -OCH<sub>2</sub>C(O)<sub>2</sub>NH--, --OCH<sub>2</sub>C(O)<sub>2</sub>O--,  
--OP(CH<sub>3</sub>)(O)O--, --OP(CH<sub>2</sub>C<sub>6</sub>H<sub>5</sub>)(O)O--, --P(S)(O)O--和--OC(O)<sub>2</sub>NH--。
102. 段 101 的方法, 其中磷酸二酯键由 --OP(SH)(O)O--,  
-OP(S<sup>M+</sup>)(O)O--或--P(S)(O)O--替代。
103. 段 100 的方法, 其中化学修饰的 RNA 的 2' 位包含选自以下的化学基团: C<sub>1</sub>-C<sub>4</sub> 烷氧基、OCH<sub>2</sub>-CH=CH<sub>2</sub>、OCH<sub>2</sub>-CH=CH-CH<sub>3</sub>、OCH<sub>2</sub>-CH=CH-(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub> (n=0, 1, ... 30)、卤素、C<sub>1</sub>-C<sub>6</sub> 烷基和 OCH<sub>3</sub>。
104. 段 100 的方法, 其中化学修饰的 RNA 包含 2' -O-甲基取代。
105. 段 92 的方法, 其中信号探针包含至少一个含有两个化学基团的相互作用对, 并且一个化学基团位于一条链的一个末端, 而另一化学基团位于另一链的相邻末端。
106. 段 92 的方法, 其中信号探针具有两个相互作用对, 其中探针的每一端都在两条链的相邻末端具有一个相互作用对。
107. 段 92 的方法, 其中相互作用对选自: 荧光团和淬灭剂、化学发光标记物和淬灭剂或加合物、染料二聚体、和 FRET 供体和受体、蛋白水解酶和该蛋白水解酶的抑制剂或能够可逆地失活该酶的其他分子。
108. 段 92 的方法, 其中相互作用对是荧光团和淬灭剂, 并分离发荧光的细胞。
109. 段 92 的方法, 其中信号探针包含至少两个相同或不同的荧光

团。

110. 段 109 的方法，其中信号探针包含至少两个荧光团，所述荧光团是 FRET 供体和受体对、或收获 (harvester) 和发射 (emitter) 荧光团。

111. 段 108 的方法，其中分离所述发荧光的细胞的步骤采用荧光激活细胞分选仪、荧光显微镜或荧光计实施。

112. 段 1、3、5、7、9、13、18、21、39、53、62、67、73、80、85、88 和 91 之任一的方法，其中信号探针包含茎环结构。

113. 段 112 的方法，其中茎区形成 4 至 6 个连续的碱基对。

114. 段 112 的方法，其中茎环结构包含至少一个含有两个化学基团的相互作用对，且在链的每一端有一个化学基团，其中茎区包含通过非互补区连接的两个相互互补区，与相互作用对邻近的相互互补区形成 5 至 6 个碱基对，与环区邻近的相互互补区形成 4 至 5 个碱基对。

115. 段 114 的方法，其中非互补区是单链环区或错配区。

116. 段 112 的方法，其中茎环结构包含至少一个含有两个化学基团的相互作用对，且在链的每一端有一个化学基团，其中茎区包含通过两个非互补区连接的三个相互互补区，与相互作用对邻近的第一相互互补区形成 4 至 5 个碱基对，第二相互互补区形成 2 至 3 个碱基对，而与环区邻近的第三相互互补区形成 2 至 3 个碱基对。

117. 段 116 的方法，其中非互补区是单链环区和错配区中的一种或多种。

118. 段 1、3、5、7、9、13、18、21、39、53、62、67、73、80、85、88 和 91 之任一的方法，其中信号探针包含哑铃结构。

119. 段 118 的方法，其中哑铃结构包含一个具有 4 个连续碱基对的茎区和一个具有 3 个连续碱基对的茎区。

120. 段 118 的方法，其中两个茎区经由茎区的一条臂通过磷酸二酯键或修饰的磷酸二酯键连接。

121. 段 118 的方法，其中两个茎区经由茎区的一条臂通过 1 或 2 个核苷酸或修饰的核苷酸连接。

122. 段 1、3、5、7、9、13、18、21、39、53、62、67、73、80、85、88 和 91 之任一的方法，其中信号探针包含三臂联结结构。

123. 段 122 的方法，其中三臂联结结构包含至少一个含有两个化学基团的相互作用对，且在链的每一个末端有一个化学基团，其中与相互作用对邻近的茎区形成 3 至 4 个连续碱基对，第一茎环结构的茎区形成 4 至 5 个连续碱基对，第二茎环结构的茎区形成 2 至 3 个连续碱基对。

124. 段 122 的方法，其中三个区通过磷酸二酯键或修饰的磷酸二酯键经由茎区的臂连接。

125. 段 122 的方法，其中三个区通过 1 或 2 个核苷酸或修饰的核苷酸经由茎区的臂连接。

126. 段 112、118 和 122 之任一的方法，其中茎环结构、哑铃结构或三臂联结结构具有 30 个以上的核苷酸或修饰的核苷酸。

127. 段 112、118 和 122 之任一的方法，其中所述结构是 DNA、RNA、肽核酸、化学修饰的 DNA、RNA 或其组合。

128. 段 112、118 和 122 之任一的方法，其中所述结构在糖基团、磷酸二酯键和碱基中的一个或多个上被化学修饰。

129. 段 128 的方法，其中磷酸二酯键被选自以下的化学基团替代： $-\text{OP}(\text{OH})(\text{O})\text{O}-$ 、 $-\text{OP}(\text{O}^{\text{M}^-})(\text{O})\text{O}-$ 、 $-\text{OP}(\text{SH})(\text{O})\text{O}-$ 、 $-\text{OP}(\text{S}^{\text{M}^-})(\text{O})\text{O}-$ 、 $-\text{NHP}(\text{O}_2)\text{O}-$ 、 $-\text{OC}(\text{O})_2\text{O}-$ 、 $-\text{OCH}_2\text{C}(\text{O})_2\text{NH}-$ 、 $-\text{OCH}_2\text{C}(\text{O})_2\text{O}-$ 、 $-\text{OP}(\text{CH}_3)(\text{O})\text{O}-$ 、 $-\text{OP}(\text{CH}_2\text{C}_6\text{H}_5)(\text{O})\text{O}-$ 、 $-\text{P}(\text{S})(\text{O})\text{O}-$ 和 $-\text{OC}(\text{O})_2\text{NH}-$ 。

130. 段 129 的方法，其中磷酸二酯键被  $-\text{OP}(\text{SH})(\text{O})\text{O}-$ 、 $-\text{OP}(\text{S}^{\text{M}^-})(\text{O})\text{O}-$ 和 $-\text{P}(\text{S})(\text{O})\text{O}-$ 替代。

131. 段 128 的方法，其中化学修饰的 RNA 的 2' 位包含选自以下的化学基团： $\text{C}_1\text{-C}_4$  烷氧基、 $\text{OCH}_2\text{-CH=CH}_2$ 、 $\text{OCH}_2\text{-CH=CH-CH}_3$ 、 $\text{OCH}_2\text{-CH=CH-(CH}_2)_n\text{CH}_3$  ( $n=0, 1, \dots, 30$ )、卤素、 $\text{C}_1\text{-C}_6$  烷基和  $\text{OCH}_3$ 。

132. 段 128 的方法，其中化学修饰的 RNA 具有 2' -O-甲基取代。

133. 段 112、118 和 122 之任一的方法，其中相互作用对选自：荧光团和淬灭剂、化学发光标记物和淬灭剂或加合物 (adduct)、染料二聚体 (dye dimer)、和 FRET 供体和受体、蛋白水解酶和该蛋白水解酶的抑

制剂或能够可逆地失活该酶的其他分子。

134. 段 133 的方法，其中相互作用对是荧光团和淬灭剂，且分离发荧光的细胞。

135. 段 134 的方法，其中分离所述发荧光的细胞的步骤使用荧光激活细胞分选仪、荧光显微镜或荧光计进行。

136. 段 112、118 和 122 之任一的方法，其中信号探针在链的一个末端包含至少两个荧光团，且在链的另一个末端包含淬灭剂，其中所述两个荧光团是 FRET 供体和受体对。

137. 包含核酸或经修饰的核酸及至少一个蛋白水解酶和至少一个该蛋白水解酶的抑制剂的探针，其中所述核酸或经修饰的核酸含有与靶序列互补的序列以及相互互补的序列，其中所述探针在无所述靶序列存在的试验条件下具有特征性的蛋白水解活性，该活性的水平是所述蛋白水解酶及其抑制剂相互作用程度的函数；且其中在存在过量的所述靶序列的条件下，所述靶互补序列与靶序列的杂交增加所述特征性蛋白水解活性的水平。

138. 段 137 的探针，其中蛋白水解酶是位点特异的或靶特异的蛋白酶。

139. 段 137 的探针，其中所述蛋白水解酶抑制剂是肽或化合物。

140. 段 137 的探针，其中所述蛋白水解酶和所述蛋白水解酶的抑制剂选自：氨肽酶和抑氨肽酶肽、胰蛋白酶样半胱氨酸蛋白酶和抗蛋白酶、氨肽酶和苯丁抑制素、胰凝乳蛋白酶样半胱氨酸蛋白酶和胰凝乳蛋白酶抑制剂、氨肽酶和 diprotin A 或 B、羧肽酶 A 和 EDTA、弹性蛋白酶样丝氨酸蛋白酶和 elastinal、及嗜热菌蛋白酶或氨肽酶 M 和 1,10-二氮杂菲。

141. 段 137 的探针，其包含两条分开的核酸或经修饰核酸的链，这两条链形成至少一个相互互补的区。

142. 段 141 的探针，其中蛋白水解酶位于一条链的一个末端，而蛋白水解酶抑制剂位于另一条链的相邻末端。

143. 段 141 的探针，其中所述两条分开的链从 5' 至 3' 端形成连续的相互互补区，且两条链具有相同数目的核苷酸。

144. 段 141 的探针, 其中在两条链之间形成相互互补区后, 一条链的 5' 端与另一条链不重合, 或者该链的 3' 端与所述另一条链不重合, 或者两者兼而有之, 其中所述不重合为不超过 10 个的核苷酸或修饰的核苷酸。

145. 段 141 的探针, 其中两条链在每一端形成 5 或 6 个连续碱基对的相互互补区。

146. 段 145 的探针, 其中两条链在序列上是不同的。

147. 段 141 的探针, 其中所述链具有 30 个以上的核苷酸或修饰的核苷酸。

148. 段 141 的探针, 其中核酸是 DNA、RNA 或两者。

149. 段 141 的探针, 其中经修饰的核酸包括肽核酸、化学修饰的 DNA 或 RNA 或者其组合。

150. 段 141 的探针, 其中经修饰的核酸在糖基团、磷酸二酯键和碱基基团中的一个或多个上被化学修饰。

151. 段 150 的探针, 其中磷酸二酯键被选自以下的化学基团替代:  $--OP(OH)(O)O--$ 、 $--OP(O^-M^+)(O)O--$ 、 $--OP(SH)(O)O--$ 、 $--OP(S^-M^+)(O)O--$ 、 $--NHP(O)_2O--$ 、 $--OC(O)_2O--$ 、 $--OCH_2C(O)_2NH--$ 、 $--OCH_2C(O)_2O--$ 、 $--OP(CH_3)(O)O--$ 、 $--OP(CH_2C_6H_5)(O)O--$ 、 $--P(S)(O)O--$ 和 $--OC(O)_2NH--$ 。

152. 段 150 的探针, 其中磷酸二酯键被  $-OP(SH)(O)O--$ 、 $-OP(S^-M^+)(O)O-$ 或 $--P(S)(O)O--$ 替代。

153. 段 150 的探针, 其中化学修饰的 RNA 的 2' 位包含选自以下的化学基团:  $C_1-C_4$  烷氧基、 $OCH_2-CH=CH_2$ 、 $OCH_2-CH=CH-CH_3$ 、 $OCH_2-CH=CH-(CH_2)_nCH_3$  ( $n=0, 1, \dots, 30$ )、卤素、 $C_1-C_6$  烷基和  $OCH_3$ 。

154. 段 150 的探针, 其中化学修饰的 RNA 包含 2' -O-甲基取代。

155. 段 137 的探针, 其包含茎环结构。

156. 段 155 的探针, 其中茎区形成 5 至 6 个连续的碱基对。

157. 段 155 的探针, 其中茎环结构在该链的两个末端分别包含至少一个蛋白水解酶和所述蛋白水解酶的抑制剂, 其中茎区包含经由非互补区连接的两个相互互补区, 邻近链末端的相互互补区形成 5 至 6 个碱基

对，邻近环区的相互互补区形成 4 至 5 个碱基对。

158. 段 157 的探针，其中非互补区是单链环区或错配区。

159. 段 155 的探针，其中茎环结构在该链的两个末端分别包含至少一个蛋白水解酶和所述蛋白水解酶的抑制剂，其中茎区包含经由两个非互补区连接的三个相互互补区，邻近链末端的第一相互互补区形成 4 至 6 个碱基对，第二相互互补区形成 2 至 3 个碱基对，邻近环区的第三相互互补区形成 2 至 3 个碱基对。

160. 段 159 的探针，其中非互补区是单链环区和错配区中的一种或多种。

161. 段 137 的探针，其包含哑铃结构。

162. 段 161 的探针，其中哑铃结构包含一个具有 4 个连续碱基对的茎区和一个具有 3 个连续碱基对的茎区。

163. 段 161 的探针，其中两个茎区经由茎区的一条臂通过磷酸二酯键或修饰的磷酸二酯键连接。

164. 段 161 的探针，其中两个茎区经由茎区的一条臂通过 1 或 2 个核苷酸或修饰的核苷酸连接。

165. 段 137 的探针，其包含三臂联结结构。

166. 段 165 的探针，其中三臂联结结构在该链的两个末端分别包含至少一个蛋白水解酶和所述蛋白水解酶的抑制剂，其中邻近链末端的茎区形成 3 至 4 个连续碱基对，第一茎环结构的茎区形成 4 至 5 个连续碱基对，和第二茎环结构的茎区形成 2 至 3 个连续碱基对。

167. 段 165 的探针，其中三个区通过磷酸二酯键或修饰的磷酸二酯键经由茎区的臂连接。

168. 段 165 的探针，其中三个区通过 1 或 2 个核苷酸或修饰的核苷酸经由茎区的臂连接。

169. 段 155、161 或 165 的探针，其中茎环结构、哑铃结构或三臂联结结构具有 30 个以上的核苷酸或修饰的核苷酸。

170. 段 155、161 或 165 的探针，其中该结构是 DNA、RNA、肽核酸、化学修饰的 DNA、RNA，或其组合。

171. 段 155、161 或 165 的探针，其中该结构在糖基团、磷酸二酯键和碱基中的一个或多个上被化学修饰。

172. 段 177 的探针，其中磷酸二酯键被选自以下的化学基团替代： $--OP(OH)(O)O--$ 、 $--OP(O^M)(O)O--$ 、 $--OP(SH)(O)O--$ 、 $--OP(S^M)(O)O--$ 、 $--NHP(O)_2O--$ 、 $--OC(O)_2O--$ 、 $--OCH_2C(O)_2NH--$ 、 $--OCH_2C(O)_2O--$ 、 $--OP(CH_3)(O)O--$ 、 $--OP(CH_2C_6H_5)(O)O--$ 、 $--P(S)(O)O--$ 和 $--OC(O)_2NH--$ 。

173. 段 171 的探针，其中磷酸二酯键被  $--OP(SH)(O)O--$ 、 $--OP(S^M)(O)O--$ 或 $--P(S)(O)O--$ 替代。

174. 段 171 的探针，其中化学修饰的 RNA 的 2' 位包含选自以下的化学基团： $C_1-C_4$  烷氧基、 $OCH_2-CH=CH_2$ 、 $OCH_2-CH=CH-CH_3$ 、 $OCH_2-CH=CH-(CH_2)_nCH_3$  ( $n=0, 1, \dots, 30$ )、卤素、 $C_1-C_6$  烷基和  $OCH_3$ 。

175. 段 171 的探针，其中化学修饰的 RNA 具有 2' -O-甲基取代。

176. 包含至少一种编码至少一种目的 RNA 和标签序列的 DNA 的 DNA 构建体，其中标签序列形成三臂联结结构，且茎区包含 8-9 个碱基对，第一茎环区包含 4-6 个碱基对、第二茎环区包含 13-17 个碱基对。

177. 段 176 的 DNA 构建体，其中三臂的茎区还包含非互补区。

178. 段 176 的 DNA 构建体，其中茎区和第一茎环区均还包含一个错配区，第二茎环区还包含 2-7 个错配或凸出区域。

179. 段 176 的 DNA 构建体，其中茎区之间的连接具有总共 8-12 个核苷酸。

180. 包含至少一种编码至少一种目的 RNA 和标签序列的 DNA 的 DNA 构建体，其中标签序列选自图 42A、B 和 C 中所述的序列。

181. 包含至少一种编码至少一种目的 RNA 和标签序列的 DNA 的 DNA 构建体，其中标签序列形成通过图 42A、B 或 C 中所述的序列预测的在能量上更为有利的结构。

182. 包含段 176 至 181 之任一的 DNA 构建体的载体。

183. 包含段 182 的载体或段 176 至 181 之任一的 DNA 构建体的细胞。

184. 段 183 的细胞，其选自永生化细胞、原代细胞、干细胞和生殖细胞。

185. 段 183 的细胞, 其选自 HeLa 细胞、NIH3T3 细胞、HEK293 细胞和 CHO 细胞。

186. 稳定的哺乳动物细胞系的文库, 其包含至少 10,000 个细胞系, 每个细胞系包含至少一个稳定整合的表达的序列。

187. 稳定的哺乳动物细胞系的文库, 其包含至少 500 个细胞系, 每个细胞系包含至少两个稳定整合的序列。

188. 稳定的哺乳动物细胞系的文库, 其包含至少 50 个细胞系, 每个细胞系包含至少三个稳定整合的序列。

189. 稳定的哺乳动物细胞系的文库, 其包含至少 20 个细胞系, 每个细胞系包含至少 4 个稳定整合的序列。

190. 段 186-189 之任一的文库, 其中细胞系还包含药物抗性基因。

191. 稳定的哺乳动物细胞系的文库, 其包含至少 50 个细胞系, 每个细胞系包含至少一个稳定整合的序列, 其中所述细胞系缺少药物抗性基因。

192. 稳定的哺乳动物细胞系的文库, 其包含至少 20 个细胞系, 每个细胞系包含至少两个稳定整合的序列, 其中细胞系缺少药物抗性基因。

193. 段 186-192 之任一的文库, 其中每个细胞系都包含可变的文库序列。

194. 段 193 的文库, 其中所述表达文库的可变序列选自: 基因组序列、基因组的非翻译序列、基因组的翻译序列、基因序列、cDNA 序列、EST 序列、寡聚物(oligo)序列、随机序列、RNA 序列、蛋白质序列、蛋白质域序列、肽序列、内含子序列、外显子序列、标签序列、或接头序列, 或其组合或其重组物, 或者这些未修饰的、经诱变的、随机化的、改组的或重组的序列中的一种或多种。

195. 段 186-192 之任一的文库, 其中使用至少一个由具有未知序列身份的遗传序列组成的汇合物或混合物产生文库。

196. 段 195 的文库, 其中所述具有未知身份的序列共有序列同源性、功能意义(functional significance)或相关来源。

197. 段 186-192 之任一的文库, 其中文库是单独合成的构建体的集

合，所述构建体包含具有已知身份的特定序列。

198. 段 197 的文库，其中所述特定序列共有序列同源性、功能意义或相关来源。

199. 段 186-192 之任一的文库，其中所述表达序列处于组成型启动子的控制之下。

200. 段 186-192 之任一的文库，其中所述表达序列处于选自以下的条件型启动子的控制之下：诱导型、阻遏型、组织特异性、时限型(temporal)或热休克型启动子。

201. 段 186-192 之任一的文库，其中文库用于基于细胞的筛选试验。

202. 段 201 的文库，其中基于细胞的筛选试验针对所述文库中的所有细胞系平行地进行。

203. 段 201 的文库，其中基于细胞的筛选试验针对所述文库中的部分细胞系进行。

204. 核酸或修饰的核酸分子，其包含根据图 8 至 24 和 41 的 FP1 至 FP18 之任一的序列。

205. 段 204 的核酸或修饰的核酸分子，其还包含相互作用对。

206. 段 205 的核酸或修饰的核酸分子，其中相互作用对选自：荧光团和淬灭剂、化学发光标记物和淬灭剂或加合物、染料二聚体、和 FRET 供体和受体、蛋白水解酶和该蛋白水解酶的抑制剂或能够可逆地失活该酶的其他分子。

207. 核酸或修饰的核酸分子，其可以与根据图 42 D1、2 或 3 的序列杂交。

208. 段 207 的核酸或修饰的核酸分子，其还包含相互作用对。

209. 段 208 的核酸或修饰的核酸分子，其中相互作用对选自：荧光团和淬灭剂、化学发光标记物和淬灭剂或加合物、染料二聚体、和 FRET 供体和受体、蛋白水解酶和该蛋白水解酶的抑制剂或能够可逆地失活该酶的其他分子。

210. 包含图 42 A、B 或 C 中的序列的核酸或修饰的核酸分子。

211. 包含图 42 D1 或 D2 中的序列的核酸或修饰的核酸分子。

本发明提供分离表达 RNA 的细胞的方法，包括步骤：提供表达该 RNA 的细胞，将细胞暴露于在与该 RNA 杂交时可以产生可检测信号的信号探针，和分离产生该信号的细胞。一个实施方案中，所述 RNA 是内源性 RNA。另一实施方案中，由引入细胞中的 DNA 导致该内源性 RNA 的表达，例如，该 DNA 包含增强子或启动子或可以诱导内源性 RNA 表达的其它序列。例如，该 DNA 可以编码诱导 RNA 表达的蛋白质，例如转录因子。在再一实施方案中，所述 RNA 由引入细胞中的核酸编码。引入细胞中的该 DNA 还可以编码标签序列，其中信号探针可以任选地靶向该标签和/或 RNA 序列。标签序列可以与 RNA 的开放阅读框相符或不相符。一个实施方案中，该方法包括检测外源或异源 RNA 及内源 RNA 的表达。可以实施这些方法以同时检测多种 RNA 和/或标签。这些 RNA 可以是相同的或不同的 RNA。在利用该方法通过使用一种以上的信号探针检测一种以上的 RNA 的那些实施方案中，由不同信号探针产生的信号可以彼此相同或者可以彼此不同。

本发明还提供分离包含一个以上拷贝的外源或异源 DNA 的细胞的方法。该方法包括步骤：引入编码 RNA 和标签序列的 DNA 并引入编码相同 RNA 和不同标签序列的 DNA；将细胞暴露于可以针对两个标签产生可检测信号的信号探针；和分离产生两个信号的细胞。

在具有一个以上的暴露步骤的本发明任一种方法中，一个或多个暴露步骤（例如，将细胞暴露于信号探针的步骤）可以同时或相继地实施。

在本发明任一方法中，所述 RNA 或蛋白质可以是处在相同或相关的生物学途径中的 RNA 或蛋白质，在彼此上游或下游起作用，相对于彼此具有调节、激活或阻遏功能、对于功能或活性而言相互依赖、是同一复合物的成分、是相同蛋白质家族的成员，等等。

本发明提供分离包含 DNA 构建体的细胞的方法，所述 DNA 构建体编码处于条件型启动子控制下的第一 RNA，所述方法包括步骤：将编码处于组成型启动子控制下的 RNA 的 DNA 构建体引入细胞，其中该 DNA 构建体还编码处于条件型启动子控制下的第二 RNA，其中该引入在第二 RNA

不表达或低水平表达的条件下进行；将细胞暴露于在与第一 RNA 杂交时可以产生可检测信号的信号探针；和分离产生该信号的细胞。该 DNA 构建体还可以编码待测 RNA，其中例如该待测 RNA 是可变的，例如来源于表达文库。可以使用这些细胞获得或鉴定能够激活驱动所述第二 RNA 表达的条件型启动子的 RNA 或化合物。

本发明还提供分离多个细胞的方法，其中一个细胞子集表达另一细胞子集不表达的 RNA，所述方法包括步骤：将编码多种 RNA 的多种 DNA 引入细胞，其中所述多种 RNA 的至少一个子集彼此不同，将细胞暴露于多种不同的信号探针，其中所述信号探针在与该多种 DNA 所编码的一种或多种 RNA 杂交时产生可检测信号，和分离产生该信号的细胞。所述细胞子集可以是一个细胞或一个以上的细胞。一个实施方案中，所述 DNA 不编码 RNA，但是导致内源性 RNA 的表达，例如，该 DNA 包含增强子或启动子序列，例如，该 DNA 包含诱导内源 RNA 表达的序列。例如，该 DNA 可以编码蛋白质，例如诱导 RNA 表达的转录因子。该 DNA 还可以编码标签序列，并且信号探针可以靶向该标签序列和/或 RNA 序列。在一个实施方案中，该多种 RNA 形成表达文库。另一实施方案中，所述 DNA 的至少一个子集编码相同标签序列。

本发明还提供分离两个或两个以上 RNA 文库细胞的方法，包括步骤：将编码第一 RNA 表达文库的 DNA 引入细胞，其中每一个 DNA 还编码第一标签序列，将编码第二 RNA 表达文库的 DNA 引入细胞，其中每一个 DNA 还编码第二标签序列，将细胞暴露于可以在与第一标签杂交时产生可检测信号的第一信号探针和可以在与第二标签序列杂交时产生可检测信号的第二信号探针，和分离产生两个信号的细胞。该方法还可以对编码其它 RNA 表达文库的 DNA 实施，其中一个使用第三标签，等。

本发明任一方法都可以通过将化合物加入细胞并分析由信号探针所产生的信号的改变（增加或降低）而用于鉴定调节 RNA 或多种 RNA 表达的化合物。

本发明提供降低蛋白质表达的方法，包括步骤：将编码可以降低该蛋白质表达的反义 RNA 或 shRNA 的 DNA 引入细胞，使细胞暴露于可以在

与该反义 RNA 或 shRNA 杂交时产生可检测信号的第一信号探针，和分离产生该信号的细胞。该方法还可以包括步骤：使细胞暴露于可以在与编码该蛋白质的 RNA 杂交时产生可检测信号的第二信号探针，其中缺乏来自该第二信号探针的信号将指示该蛋白质的表达被降低。还可以使用其它方法，例如使用特异结合该蛋白质的抗体，使用功能试验，例如分析该蛋白质的已知生物学活性，等等，来检查蛋白质表达的降低。在该方法的一个实施方案中，省去使细胞暴露于第一信号探针的步骤。任何使用 siRNA 的方法也都可以使用 shRNA 进行。

在本发明任何方法中，向细胞中引入的 DNA 可以可操作地与条件型启动子连接。一个实施方案中，DNA 所编码的 RNA 对细胞是致死性的或损伤性的。DNA 还可以包含可以用于选择包含该 DNA 的细胞的选择标记。

本发明还提供定量生物学样品中 RNA 的表达水平的方法，包括步骤：将生物学样品暴露于在与该 RNA 杂交时可以产生可检测信号的第一信号探针，定量生物学样品中该信号的水平，并将此信号水平与该 RNA 的表达水平关联起来。

本发明还提供鉴定调节 RNA 表达的化合物的方法，包括：将待测化合物加入表达该 RNA 的细胞，使细胞暴露于可以在与该 RNA 杂交时产生可检测信号的信号探针，并将暴露于待测化合物的细胞所产生的信号与未暴露于该待测化合物的细胞所产生的信号进行比较，其中与前一细胞所产生的信号相比，前一细胞所产生的信号的增加或降低指示该化合物调节该 RNA 的表达。一个实施方案中，该 RNA 由引入细胞中的 DNA 编码。一个实施方案中，化合物是 RNA 或蛋白质。

本发明提供鉴定活细胞中的遗传重组事件的方法，包括步骤：将细胞暴露于可以在与自重组序列转录产生的 RNA 杂交时产生可检测信号的信号探针，其中检测到产生该信号的细胞将指示该细胞包含所述遗传重组事件。

本发明还提供通过这些方法中的任何方法产生的细胞。可以培养这些细胞，并且可以将其用于产生细胞系或多个细胞系。这些细胞可以用于多种目的，例如，用于基于细胞的试验中或将细胞植入动物、非人动

物或哺乳动物。该细胞可以是胚胎干细胞、原代细胞、生殖细胞或干细胞。该细胞还可以是永生化的细胞。该细胞可以是内皮细胞、表皮细胞、间充质细胞、神经细胞、肾细胞、肝细胞、造血细胞或免疫细胞。该细胞可以是真核细胞、原核细胞、哺乳动物细胞、酵母细胞、植物细胞、人细胞、灵长类动物细胞、牛细胞、猪细胞、猫细胞、啮齿类动物细胞、有袋类动物的细胞、鼠细胞或其它细胞。

标签序列可以包含多个靶序列，其中每一个靶序列与一个信号探针杂交。标签序列可以是具有二级结构的 RNA。该结构可以是三臂联结结构。DNA 可以包含多个标签序列。标签序列可以转录成与该 DNA 所编码的 RNA 相同的 RNA，或者标签序列可以转录成分开的 RNA。本发明还提供包含编码 RNA 和标签序列的 DNA 序列的 DNA 构建体。该标签序列可以是此处所述的任一标签序列。本发明还提供包含该 DNA 构建体的细胞和载体。

本发明还提供包含至少 1,000 个、至少 800 个、至少 600 个、至少 500 个、至少 400 个、至少 200 个、至少 100 个或至少 50 个细胞系的哺乳动物细胞系文库，其中每个细胞系包含稳定整合的表达的序列。还提供包含至少 500 个、至少 400 个、至少 300 个、至少 200 个、至少 200 个、至少 100 个、至少 50 个细胞系的哺乳动物细胞系文库，其中每个细胞系包含至少两个稳定整合的序列。还提供包含至少 100 个、至少 50 个、至少 25 个、至少 10 个细胞系的哺乳动物细胞系文库，其中每个细胞系包含至少三个稳定整合的序列。还提供包含至少 50 个、至少 25 个、至少 20 个、至少 10 个细胞系的哺乳动物细胞系文库，其中每个细胞系包含至少 4 个稳定整合的序列。在这些细胞系中的稳定整合的序列还可以缺乏选择标记，例如药物抗性基因。该稳定整合的序列可以具有已知的或未知的序列身份。这些序列可以共有序列同源性、功能意义或相关来源。这些文库可以用于多种目的，例如基于细胞的筛选试验。

本发明还提供鉴定化合物的方法，其中所述化合物可以增强使用信号探针对细胞中的靶进行的检测，所述方法包括步骤：将信号探针引入包含靶序列的细胞中，其中信号探针在与靶序列杂交时产生可检测信号，

将细胞暴露于待测化合物，和检测细胞所产生的信号，其中与未暴露于待测化合物的细胞产生的信号相比，暴露于待测化合物的细胞产生的信号的增加指示该待测化合物是可以增强使用信号探针对细胞中的靶进行的检测的化合物。

本发明还提供鉴定化合物的方法，其中所述化合物可以介导或提高信号探针向细胞中的引入，所述方法包括步骤：在待测化合物存在下将细胞暴露于信号探针，其中所述细胞包含靶序列并且其中信号探针在与该靶序列杂交时可以产生信号；和检测所述细胞产生的信号，其中与未暴露于该待测化合物的细胞相比，暴露于该待测化合物的细胞所产生的信号的增加指示，该待测化合物是可以介导或提高信号探针向细胞中的引入的化合物。

本发明的这些和其它方面将从以下详述中彰显。

### 附图简述

图 1 和 2 描述具有分开的两条链的信号探针或蛋白酶探针。相互作用的化学基团显示为椭圆。这些不同的椭圆构成本发明的一个实施方案，其中黑色椭圆指示淬灭剂部分，而白色和灰色椭圆指示不同的荧光团。

图 3、4 和 5 描述经设计而具有茎环结构的信号探针或蛋白酶探针。相互作用化学基团显示为椭圆。这些不同的椭圆构成本发明的一个实施方案，其中黑色椭圆指示淬灭剂部分，而白色椭圆指示荧光团。

图 6 描述具有三臂联结结构的信号探针或蛋白酶探针。相互作用化学基团显示为椭圆。这些不同的椭圆构成本发明的一个实施方案，其中黑色椭圆指示淬灭剂部分，而白色椭圆指示荧光团。

图 7 描述具有哑铃 (dumbbell) 结构的信号探针或蛋白酶探针。相互作用化学基团显示为椭圆。这些不同的椭圆构成本发明的一个实施方案，指示其中黑色椭圆指示淬灭剂部分，而白色椭圆指示荧光团。

图 8 显示荧光探针 FP1 的序列和预测的天然构象。FP1 序列包含被设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基以下划线表示。A 图显示使用根据 *Nucleic Acids Res.* 31: 3429-3431 (2003) 的 DNA 折

叠程序预测的该序列的结构。B图显示根据 oligoanalyzer 3.0 软件(可以获自 <http://biotools.idtdna.com/analyzer/oligocalc.asp>)预测的 FP1 序列的自身二聚体化。在 A 图和 B 图中,侧翼碱基加有灰色阴影,白色和黑色椭圆分别指示荧光团和淬灭剂部分。

图 9 显示荧光探针 FP2 的序列和预测的天然构象。该序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基以下划线表示。A 图显示使用根据 Nucleic Acids Res. 31: 3429-3431 (2003) 的 DNA 折叠程序预测的该序列的结构。可能所有的或部分的阴影区形成 Watson-Crick 碱基对,由此形成三臂连接。B 图显示根据 oligoanalyzer 3.0 软件(可以获自 <http://biotools.idtdna.com/analyzer/oligocalc.asp>)预测的 FP2 序列的自身二聚体化。在 A 图和 B 图中,侧翼碱基加有灰色阴影,白色和黑色椭圆分别指示荧光团和淬灭剂部分。

图 10 显示荧光探针 FP3 的序列和预测的天然构象。FP3 序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基以下划线表示。该图显示根据 oligoanalyzer 3.0 软件(可以获自 <http://biotools.idtdna.com/analyzer/oligocalc.asp>)预测的 FP3 序列的自身二聚体化。侧翼碱基加有灰色阴影,白色和黑色椭圆分别指示荧光团和淬灭剂部分。

图 11 显示荧光探针 FP4 的序列和预测的天然构象。FP4 序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基以下划线表示。该图显示了使用根据 Nucleic Acids Res. 31: 3429-3431 (2003) 的 DNA 折叠程序预测的该序列的结构。侧翼碱基加有灰色阴影,白色和黑色椭圆分别指示荧光团和淬灭剂部分。

图 12 至 13 显示荧光探针 FP5 至 FP6 的序列。这些序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基加有下划线。

图 14 显示荧光探针 FP7 的序列和预测的天然构象。FP7 序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基以下划线表示。显示了根据 oligoanalyzer 3.0 软件(可以获自 <http://biotools.idtdna.com/analyzer/oligocalc.asp>)预测的 FP7 序列的自身二聚体化。侧翼碱

基加有灰色阴影，白色和黑色椭圆分别指示荧光团和淬灭剂部分。

图 15 显示荧光探针 FP8 的序列和预测的天然构象。FP8 序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基以下划线表示。A 图显示使用根据 *Nucleic Acids Res.* 31: 3429-3431 (2003) 的 DNA 折叠程序预测的该序列的结构。B 图显示根据 oligoanalyzer 3.0 软件(可以获自 <http://biotools.idtdna.com/analyzer/oligocalc.asp>) 预测的 FP1 序列的自身二聚体化。在 A 图和 B 图中，侧翼碱基加有灰色阴影，白色和黑色椭圆分别指示荧光团和淬灭剂部分。

图 16 至图 20 显示荧光探针 FP9 至 FP13 的序列。这些序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基加有下划线。

图 21 显示荧光探针 FP14 的序列和预测的天然构象。FP14 序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基以下划线表示。该图显示了使用根据 *Nucleic Acids Res.* 31: 3429-3431 (2003) 的 DNA 折叠程序预测的该序列的结构。侧翼碱基加有灰色阴影，白色和黑色椭圆分别指示荧光团和淬灭剂部分。

图 22 至图 24 分别显示荧光探针 FP15 至 FP17 的序列和预测的天然构象。这些序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基加有下划线。A 图显示使用根据 *Nucleic Acids Res.* 31: 3429-3431 (2003) 的 DNA 折叠程序预测的该序列的结构。B 图显示根据 oligoanalyzer 3.0 软件(可以获自 <http://biotools.idtdna.com/analyzer/oligocalc.asp>) 预测的 FP15 序列的自身二聚体化。在 A 图和 B 图中，侧翼碱基加有灰色阴影，白色和黑色椭圆分别指示荧光团和淬灭剂部分。

图 25 显示在 FACS 过程中相对于荧光强度所观察到的细胞数目。图 25A 显示暴露于信号探针 FP1 的对照 NIH3T3 细胞的曲线。FP1 中使用荧光团 Alexafluor633。对照 NIH3T3 细胞未用质粒转染，不含靶序列。图 25B 显示暴露于信号探针 FP1 的经转染的 NIH3T3 细胞的曲线。该转染的 NIH3T3 细胞含有编码目的 RNA 和图 42A 中所示标签 1 序列的 DNA。图 25C 是图 25A 和 25B 的重叠。图 25C 显示 FACS 可以区分以编码靶序列的质粒

转染的细胞和未转染的对照细胞。

图 26 显示在 FACS 过程中相对于荧光强度所观察到的细胞数目。图 26A 显示暴露于信号探针 FP2 的对照 NIH3T3 细胞的曲线。FP2 中使用荧光团 Alexafluor680。对照 NIH3T3 细胞未用质粒转染，不含靶序列。图 26B 显示暴露于信号探针 FP2 的经转染的 NIH3T3 细胞的曲线。该转染的 NIH3T3 细胞含有编码目的 RNA 和图 42B 中所示标签 2 序列的 DNA。图 26C 是图 26A 和 26B 的重叠。图 26C 显示 FACS 可以区分以编码靶序列的质粒转染的细胞和未转染的对照细胞。

图 27 显示在 FACS 过程中相对于荧光强度所观察到的细胞数目。图 27A 显示暴露于信号探针 FP3 的对照 NIH3T3 细胞的曲线。FP3 中使用荧光团荧光素。对照 NIH3T3 细胞未用质粒转染，不含靶序列。图 27B 显示暴露于信号探针 FP3 的经转染的 NIH3T3 细胞的曲线。该转染的 NIH3T3 细胞含有编码目的 RNA 和图 42C 中所示标签 3 序列的 DNA。图 27C 是图 27A 和 27B 的重叠。图 27C 显示 FACS 可以区分以编码靶序列的质粒转染的细胞和未转染的对照细胞。

图 28 显示在 FACS 过程中相对于荧光强度所观察到的细胞数目。图 28A 显示暴露于信号探针 FP1 的对照 HeLa 细胞的曲线。FP1 中使用荧光团荧光素。对照 HeLa 细胞未用质粒转染，不含靶序列。图 28B 显示暴露于信号探针 FP1 的经转染的 HeLa 细胞的曲线。该转染的 HeLa 细胞含有编码反向 vav RNA 的 DNA。图 28C 是图 28A 和 28B 的重叠。图 28C 显示 FACS 可以区分以编码靶序列的质粒转染的细胞和未转染的对照细胞。

图 29 显示在 FACS 过程中相对于荧光强度所观察到的细胞数目。图 29A 显示暴露于信号探针 FP8 的对照 HeLa 细胞的曲线。FP8 中使用荧光团荧光素。对照 HeLa 细胞未用质粒转染，不含靶序列。图 29B 显示暴露于信号探针 FP8 的经转染的 HeLa 细胞的曲线。该转染的 HeLa 细胞含有编码反向 vav RNA 的 DNA。图 29C 是图 29A 和 29B 的重叠。图 29C 显示 FACS 可以区分以编码靶序列的质粒转染的细胞和未转染的对照细胞。

图 30 显示在 FACS 过程中相对于荧光强度所观察到的细胞数目。图 30A 显示暴露于信号探针 FP5 的对照 HeLa 细胞的曲线。FP5 中使用荧光

团荧光素。对照 HeLa 细胞未用质粒转染，不含靶序列。图 30B 显示暴露于信号探针 FP5 的经转染的 HeLa 细胞的曲线。该转染的 HeLa 细胞含有编码反向 vav RNA 的 DNA。图 30C 是图 30A 和 30B 的重叠。图 30C 显示 FACS 可以区分以编码靶序列的质粒转染的细胞和未转染的对照细胞。

图 31 显示在 FACS 过程中相对于荧光强度所观察到的细胞数目。图 31A 显示暴露于信号探针 FP9 的对照 HeLa 细胞的曲线。FP9 中使用荧光团荧光素。对照 HeLa 细胞未用质粒转染，不含靶序列。图 31B 显示暴露于信号探针 FP9 的经转染的 HeLa 细胞的曲线。该转染的 HeLa 细胞含有编码反向 vav RNA 的 DNA。图 31C 是图 31A 和 31B 的重叠。图 31C 显示 FACS 可以区分以编码靶序列的质粒转染的细胞和未转染的对照细胞。

图 32 显示经药物选择的、用编码反向克隆的部分 vav 序列(称作 r-vav)和药物抗性基因的表达质粒转染的 HeLa 细胞的荧光图像。将细胞暴露于荧光探针(FP)，这些荧光探针被设计为识别 r-vav 中的相同靶序列(5' GTTCTTAAGGCACAGGAAGTGGGA 3')。使用荧光显微镜和设计用于检测来自 Fam 的荧光的滤光器，获得这些图像。在此使用的所有 FP 除了 FP1 使用荧光素标记之外，其余均使用 FAM 标记。图 A、B、C、D 各暴露于 FP10、FP11、FP12 和 FP13。

图 33 显示用编码 RNA 和标签序列(该标签序列设计为被 FP15 所识别)的构建体转染的细胞(图 A、B、D)或用编码相同 RNA 但不编码该标签序列的对照构建体转染的细胞(图 C 或 E)的荧光图像。所用标签序列被称作 6-5、6-7 或 6B10，其分别含有 1、2 和 3 个拷贝的、FP15 所设计识别的靶序列。用包含设计为被 FP15 识别的标签序列的构建体转染的细胞(图 A、B 和 D)显示出比对照细胞(图 C 或 E)强的荧光。

图 34 显示用编码 RNA 和标签序列 6CA4(该标签序列设计为被 FP16 所识别)的构建体转染的细胞(图 A)或用编码相同 RNA 但不编码该标签序列的对照构建体转染的细胞(图 B)的荧光图像。用包含标签序列的构建体转染的细胞(图 A)显示出比对照细胞(图 B)强的荧光。

图 35A 显示被选择用于形成标签序列的部分(带下划线) vav DNA 序列的反向互补序列(r-vav DNA)。该序列被克隆至设计为表达 r-vav mRNA

的表达质粒中。粗体所示序列是用于某些荧光探针的靶序列。图 35B 显示图 35A 中带下划线的序列，在此这些序列已经组合形成标签序列(标签 1 序列)。

图 36A 显示使用 *Nucleic Acids Res.* 31: 3429-3431 (2003) 中的 RNA 折叠程序预测的部分 r-vav RNA 的结构。图 36B 是预测的图 35B 中所示标签 1 序列的结构。阴影指示经设计用于被一些荧光探针识别的靶序列。

图 37A、B 和 C 显示预测的图 42 中所示标签 1、2 和 3 序列的结构。这些结构彼此相似但给出不同的被荧光探针所识别的序列。该预测使用 *Nucleic Acids Res.* 31: 3429-3431 (2003) 中的 RNA 折叠程序产生。

图 38 显示在存在靶或对照寡聚物序列时自溶液中的 FP 发射出的荧光信号。用 UV 照射样品并拍照。在此使用的所有 FP 均掺入了荧光素。每个管子都含有由 5  $\mu$ l 20  $\mu$ M FP 原液、1.5  $\mu$ l 25mM MgCl<sub>2</sub>、8  $\mu$ l 20  $\mu$ M 寡聚物和 1.5  $\mu$ l 水组成的总共 16  $\mu$ l，其中最终镁浓度为大约 2.34mM。在此使用了 FP1 和 FP18，两者在合成时在序列的碱基之间包含硫连接 (sulfur linkages)，设计用于分别识别靶寡聚物 T0-FP1 和 T0-FP18。FP1 针对靶寡聚物 1 的序列 (T0-FP1 5' GTTCTTAAGGCACAGGAAGTGGGA3')，FP18 针对靶寡聚物 FP18 的序列 (T0-FP18 5' TCCCAGTTCCTGTGCCT TAAGAAC 3')。T0-FP1 和 T0-FP18 的序列彼此反向互补。T0-FP18 具有不被 FP1 靶向的序列，充当 FP1 的对照寡聚物。T0-FP1 具有不被 FP18 靶向的序列，充当 FP18 的对照寡聚物。

在各组中，管中的组成如下所示：

管子	FP	寡聚物
1	FP18	T0-FP18
2	FP18	T0-FP1
3	FP1	T0-FP18
4	FP1	T0-FP1

该图显示了所有测试的 FP 通过在含有具有所靶向的序列的寡聚物

的管中比在含有具有非被靶向的序列的寡聚物的对照管中释放出更大的信号而均特异地报道靶序列的存在。在含有靶序列的寡聚物存在下含有 FP 的管子以星号指示。

图 39 显示在靶或对照寡聚物序列存在下自溶液中的 FP 释放的荧光信号。用 UV 照射样品并拍照。在此使用的所有 FP 均掺入了 FAM。图 A、B、C 各显示其中加入了相同 FP 的 4 个管子。每个管子含有总共  $10\ \mu\text{l}$ ，其中在 PBS 中含有  $2\ \mu\text{l}$   $20\ \mu\text{M}$  FP 原液和  $1\ \mu\text{l}$   $100\ \mu\text{M}$  寡聚物原液补加至  $4\text{mM}$   $\text{MgCl}_2$ 。在每一个图中，管 1 不含寡聚物原液而含有  $1\ \mu\text{l}$  水，管 2 含有寡聚物 T0-M1，管 3 含有寡聚物 T0-M2，管 4 含有寡聚物 T0-M3。

以下列出在每一个图中测试的 FP 以及包含设计用于被该 FP 识别的序列的寡聚物 (oligo)。

图	FP	含有靶序列的寡聚物
A	FP4	T0-M3 (管 4)
B	FP6	T0-M2 (管 3)
C	FP7	T0-M3 (管 4)

该图显示，所有的被测 FP 均通过在含有具有所靶向的序列的寡聚物的管子中比在不含寡聚物或含有具有非被靶向的序列的寡聚物的对照管中释放出更大的信号而特异地报道靶序列的存在。在含有靶序列的寡聚物存在下含有 FP 的管子以星号标出。

以下列出了 T0-M1、T0-M2 和 T0-M3 从 5' 至 3' 方向的序列：

T0-M1:

TTTCTCTGTGATCCGGTACAGTCCTTCTGCGCAGGTGGACAGGAA  
GGTTCTAATGTTCTTAAGGCACAGGAAGTGGACATCTGGGCCCG  
GAAAGCCTTTTTCTCTGTGATCCGGTACAGTCCTTCTGCGCAGGT  
GGACAGGAAGTTCTAATGTTCTT

T0-M2:

TTTAACTGATGGATGGAACAGTCCTTCTGCGCAGGTGGACAGCTT  
GGTTCTAATGAAGTTAACCCTGTCGTTCTGCGACATCTGGGCCCG  
GAAAGCGTTTAACTGATGGATGGAACAGTCCTTCTGCGCAGGTGG

ACAGCTTGGTTCTAATGAAGTT

TO-M3:

GTAAAGTCAGACATCCGGTACAGTCCTTCTGCGCAGGTGGACAGG

AAGGTTCTAATGTTCTATAGGGTCTGCTTGTGCTCATCTGGGCC

CGGAGATGCGTAAAGTCAGACATCCGGTACAGTCCTTCTGCGCAG

GTGGACAGGAAGGTTCTAATGTTCTAT

图 40 显示在靶或对照寡聚物序列存在时自溶液中的 FP 释放的荧光信号。样品以 UV 进行照射并拍照。在此使用的所有 FP 均掺入了 FAM。在图 A、B 和 C 中分别测试 FP1、2 和 3，其各被设计以分别识别掺入到标签 (tag) 1、2 和 3 中的相关靶序列 (如图 42 所示)。在含有靶序列的寡聚物存在下含有 FP 的管子以星号标出。

图 A、B、C 的操作方案与图 39 的方案一致，且样品如下所述：

图	FP	含有靶序列的寡聚物
A	FP1	TO-M1 (管 2)
B	FP2	TO-M2 (管 3)
C	FP3	TO-M3 (管 4)

图 41 显示荧光探针 FP18 的序列和预测的天然构象。该序列包含设计为与靶序列互补的碱基和额外的侧翼碱基。侧翼碱基加有下划线。A 图显示根据 Nucleic Acids Res. 31: 3429-3431 (2003) 的 DNA 折叠程序预测的该序列的结构。B 图显示根据 oligoanalyzer 3.0 软件 (可获自 <http://biotools.idtdna.com/analyzer/oligocalc.asp>) 预测的 FP2 序列的自身二聚体化。图 A 和 B 中，侧翼碱基加有灰色阴影，白色和黑色椭圆分别指示荧光团和淬灭剂部分。

图 42A、B 和 C 显示被荧光探针识别的三个标签序列。图 42A、B 和 C 中，靶序列以粗体显示，其也显示在图 42D 中。第一个序列 (标签 1, 42A) 与图 35B 中所示序列相同。接着的两个序列 (标签 2, 42B 和标签 3, 42C) 是标签 1 的变化版本。图 D 中以下划线指出靶 2 和靶 3 与靶 1 的序列区别。为了弥补在所示部分中发生的变化，在剩余的标签序列中实施了额外的序列改变。

图 43 显示自标签 1 序列进行的标签 2 序列的设计。A-F 显示在该设

计过程中进行的相继碱基改变。

图 44 显示自标签 1 序列进行的标签 3 序列的设计。A-F 显示在该设计过程中进行的相继碱基改变。

图 45 显示根据本发明方法分离的细胞是活的。图 A 显示在细胞用分别编码带有标签 1、2 和 3 的目的 RNA 的三个 DNA 构建体转染后使用 FACS 分离的细胞。细胞经过药物选择后暴露于 FP1、2 和 3 并分离。这些细胞与未转染任何所述 DNA 构建体的对照细胞相比，三个探针中每一个探针的荧光强度高于背景强度。细胞直接通过 FACS 分别单独地接种在的 96 孔板的孔中，其中的一个在分离后立即拍照。图 B 显示在细胞贴壁于孔表面上之后，即，1 小时之后的相同细胞。图 C 显示在细胞经历过细胞分裂后，即第二天的相同细胞。

三个图每一个都显示出，尽管之前不知道在将细胞暴露于这些探针时使用的试剂有何影响，但根据本发明方法分离的细胞仍是存活的。这些影响包括危害细胞质膜和可能地进一步在 FACS 期间给细胞造成高的压力。图 A 显示未发现细胞膜被危害，图 B 和 C 进一步说明，由于细胞能够贴壁于培养皿表面并分裂（两者均是活细胞的性质），故细胞是活的。

图 46 显示与对照细胞相比，用 mcon 1 转染的 293T 细胞的 FACS 分析结果。

图 47 显示与对照细胞相比，用 mcon 2 转染的 293T 细胞的 FACS 分析结果。

图 48 显示与对照细胞相比，用 mcon 3 转染的 293T 细胞的 FACS 分析结果。

图 49 显示与对照细胞相比，用 mcon 4 转染的 293T 细胞的 FACS 分析结果。

图 50 显示与对照细胞相比，用 mcon 5 转染的 293T 细胞的 FACS 分析结果。

图 51 显示与对照细胞相比，用 mcon 6 转染的 293T 细胞的 FACS 分析结果。

图 52 显示与对照细胞相比，用 mcon 7 转染的 293T 细胞的 FACS 分

析结果。

图 53 显示与对照细胞相比, 用 mcon 8 转染的 293T 细胞的 FACS 分析结果。

图 54 显示与对照细胞相比, 用 mcon 9 转染的 293T 细胞的 FACS 分析结果。

图 55 显示与对照细胞相比, 用 mcon 10 转染的 293T 细胞的 FACS 分析结果。

图 56 显示与对照细胞相比, 用 mcon 11 转染的 293T 细胞的 FACS 分析结果。

图 57 显示与对照细胞相比, 用 mcon 12 转染的 293T 细胞的 FACS 分析结果。

图 58 显示与对照细胞相比, 用 mcon 13 转染的 293T 细胞的 FACS 分析结果。

图 59 显示与对照细胞相比, 用 mcon 14 转染的 293T 细胞的 FACS 分析结果。

图 60 显示与对照细胞相比, 用 mcon 15 转染的 293T 细胞的 FACS 分析结果。

### 发明详述

除非另行定义, 否则本文所用所有技术和科学术语均具有本发明所属技术领域的普通技术人员所通常理解的含义。本文提及的所有出版物和其它参考文献均完整地并入作为参考。万一冲突, 则由本说明书(包括定义)控制。

在有关探针的上下文中使用的术语“邻近”或“相邻”指允许相互作用对彼此发生功能性相互作用的接近状态。例如, 该接近状态允许荧光团被淬灭剂部分(moiety)淬灭或部分地淬灭, 或者允许蛋白酶抑制剂抑制或部分地抑制蛋白酶。对于目前已知的荧光团和淬灭剂, 两者发生相互作用所需的距离是大约 20-100 Å。

术语“碱基对”指 Watson-Crick 碱基对。

术语“凸起区 (bulge region)”指一个未形成碱基对的核苷酸或修饰的核苷酸的单链区。凸起的核苷酸可以介于两个相互互补区之间(例如, 图 4A)。

术语“哑铃结构”指具有两个茎环结构的构象的核酸链或经修饰核酸链, 其中这两个茎环结构通过来自两个茎区每一个的臂的末端连接在一起(例如, 图 7)。该连接可以是非互补区、或是修饰的或未修饰的磷酸二酯键。

术语“相互作用对”指两个在彼此邻近时发生功能性相互作用的化学基团, 而且, 相对于该化学基团在相互作用时产生的背景信号或无信号而言这两个化学基团在彼此不邻近时产生可检测的信号, 或者相对于该化学基团相互作用时产生的信号而言这两个化学基团在彼此不邻近时产生不同的信号。相互作用对包括但不限于荧光团和淬灭剂、化学发光标记物和淬灭剂或加合物、染料二聚体和 FRET 供体及受体、或其组合。信号探针可以包含一种以上的相互作用对。例如, 波长转换 (wavelength-shifting) 信号探针具有均与淬灭剂发生相互作用的第一荧光团和第二荧光团, 这两个荧光团是 FRET 供体和受体对。

术语“环区”指由一个以上未碱基配对的核苷酸或修饰的核苷酸组成的单链区(例如, 图 4B 和图 23A)。环也可以介于相互互补区之间(图 8A)。

术语“信号探针 (signaling probe)”指含有与靶核酸序列互补的序列和至少一个相互互补区且还含有至少一个相互作用对的探针。当信号探针未与其靶序列结合时, 相互作用对的结构部分彼此邻近以致不产生或几乎不产生信号或产生不同的信号。当信号探针与靶序列结合后, 相互作用对的结构部分不再彼此邻近, 并产生可检测的信号或者产生与探针在未结合状态时所产生的信号不同的信号。一个实施方案中, 信号探针是包含荧光团和淬灭剂部分的产荧光的 (fluorogenic) 或发荧光的 (fluorescent) 探针, 在与靶序列结合后产生荧光的改变。相互作用对的结构部分可以附着在信号探针的末端或者可以附着在核酸序列内。可以掺入信号探针序列内部的所述结构部分的实例包括淬灭剂: dabcy1 dT,

BHQ2 dT 和 BHQ1 dT, 以及荧光团: 荧光素 dT, Alexa dT 和 Tamra dT。

术语“蛋白酶探针”指包含与靶序列互补的序列和至少一个相互互补区并还包含至少一个蛋白水解酶和至少一个该蛋白水解酶的抑制剂或能够可逆地失活该酶的其他分子的探针。当探针未与靶序列杂交时, 蛋白水解酶和该蛋白水解酶的抑制剂的接近使得两者发生相互作用, 从而抑制蛋白水解活性。在探针与靶序列杂交时, 蛋白水解酶与其抑制剂被分开, 从而激活该蛋白水解酶。蛋白水解酶和抑制剂可以共价地或非共价地附着于探针。

术语“错配区”指核酸分子或修饰的核酸分子中的双链区, 其中碱基或修饰的碱基不形成 Watson-Crick 碱基配对(例如, 图 4B 和 C)。错配区介于两个碱基配对的区域之间。该双链区可以通过非氢键键合或通过氢键键合以形成 Hoogsteen 碱基对等, 或者两种情况兼而有之。

术语“相互互补区”指核酸分子或修饰的核酸分子中形成 Watson-Crick 碱基对的区域。

术语“非互补区”指核酸分子或修饰的核酸分子中不形成 Watson-Crick 碱基对的区域。例如可以设计非互补区使之具有凸起的核苷酸、单链环、5' 或 3' 端突出的核苷酸或者错配区。

术语“茎区”指核酸分子或修饰的核酸分子中具有至少两个 Watson-Crick 碱基对的区域。例如, 茎区可以经设计而具有一个以上的通过非互补区连接的相互互补区, 或者形成连续的相互互补区。

术语“茎环区”指具有单链环序列的核酸分子或修饰的核酸分子, 其中该单链环序列的侧翼是一对 5' 和 3' 寡核苷酸或修饰的寡核苷酸臂(例如, 图 4)。该 5' 和 3' 臂形成茎区。

术语“三臂联结结构”指具有茎区、第一茎环区和第二茎环区构象的核酸或经修饰核酸链, 其中所述三个区域通过茎区的臂连接在一起(例如如图 6)。第一茎环区位于第二茎环区的 5'。这三个区域可以通过非互补区、磷酸二酯键或修饰的磷酸二酯键或其组合连接。

## 信号探针

### 相互作用对

信号探针可以具有一个以上相互作用对，或者具有不同的相互作用对。一个实施方案中，信号探针是产荧光的探针。一个实施方案中，产荧光的探针在其未杂交的状况下不释放荧光或释放背景水平的荧光，而在与其靶结合时释放荧光或者释放高于背景水平的荧光。多种荧光团可以用于增加信号或提供不同颜色范围的荧光。多种淬灭剂可以用于在缺少靶序列时降低或消除信号。淬灭剂的实例包括但不限于 DABCYL、EDAC、铯、溴化对二甲苯-二吡啶鎓、铈和金纳米颗粒。荧光团的实例包括但不限于磺基若丹明 101 (sulforhodamine 101)、吡啶、5-(2'-氨基乙基) aminoaphthaline-1-磺酸 (EDANS)、得克萨斯红、曙红、及 Bodipy 和 Alexa Fluor 350、Alexa Fluor 405、Alexa Fluor 430、Alexa Fluor 488、Alexa Fluor 500、Alexa Fluor 514、Alexa Fluor 532、Alexa Fluor 546、Alexa Fluor 555、Alexa Fluor 568、Alexa Fluor 594、Alexa Fluor 610、Alexa Fluor 633、Alexa Fluor 635、Alexa Fluor 647、Alexa Fluor 660、Alexa Fluor 680、Alexa Fluor 700、Alexa Fluor 750、别藻蓝蛋白、氨基香豆素、Bodipy-FL、Cy2、Cy3、Cy3.5、Cy5、Cy5.5、羧基荧光素 (FAM)、Cascade Blue、APC-Cy5、APC-Cy5.5、APC-Cy7、香豆素、ECD (Red613)、荧光素 (FITC)、六氟荧光素 (HEX)、羟基香豆素、丽丝胺若丹明 B、荧光黄、甲氧基香豆素、Oregon 绿 488、Oregon 绿 514、Pacific 蓝、PE-Cy7 缀合物、PerC、PerCP-Cy5.5、R-藻红素 (PE)、若丹明、若丹明绿、若丹明红-X、四氟荧光素 (TET)、TRITC、四甲基若丹明、得克萨斯红 X、TRITC、XRITC 和 Quantum dots。见例如，Tyagi 等，*Nature Biotechnology* 16: 49-53 (1998) 和 Dubbertret 等，*Nature Biotechnology*, 19: 365-370 (2001)，并入此处作为参考。

本发明还提供波长转换的信号探针。一个实施方案中，探针的一个末端至少具有收获荧光团和发射荧光团，该探针的邻近末端至少具有淬灭剂部分。见例如 Tyagi 等，*Nature Biotechnology* 18, 1191-1196 (2000)，并入此处作为参考。一个实施方案中，收获荧光团和发射荧光团位于同一末端，其中发射荧光团位于远端，淬灭剂部分位于收获荧光

团相对的末端。发射荧光团可以通过少数核苷酸的间隔臂与收获荧光团分开。在单色光源的波长范围内，收获荧光团强烈地吸收。在缺少靶序列时，两个荧光团均被淬灭。在靶存在时，探针在发射荧光团的发射范围内发荧光。发射光谱的这种迁移是由吸收的能量通过荧光共振能量转移方式从收获荧光团向发射荧光团的转移所致。与含有不能有效地从单色光源吸收能量的荧光团的信号探针相比，此类信号探针可以提供更强的信号。一个实施方案中，收获荧光团是荧光素，发射荧光团是6-羧基若丹明6G、四甲基若丹明或得克萨斯红。

另一实施方案中，探针的一个末端至少具有荧光团F1，而邻近的另一末端至少具有另一荧光团F2。选择这两个荧光团，以便当它们紧靠时可以发生荧光共振能量转移 (fluorescence resonance energy transfer, FRET)。当探针未与其靶序列结合时，以F1的吸收光谱带 (absorption band) 激发时，F1的荧光被F2淬灭，观察到F2的荧光。当探针与其靶序列结合时，FRET被降低或消除，F1的荧光增加而F2的荧光减弱或消失。可以监测荧光强度的这种差异并可以计算F1荧光和F2荧光之间的比率。由于在荧光团-淬灭剂系统中有时会观察到残余的荧光，因此对于定量检测靶序列，该系统可能更为有利。见，Zhang等，*Angew. Chem. Int. Ed.*, 40, 2, pp. 402-405 (2001)，并入此处作为参考。FRET供体-受体对的实例分别包括但不限于香豆素基团和6-羧基荧光素基团。

一个实施方案中，信号探针包含发光标记物和加合物 (adduct) 对。加合物与发光标记物的相互作用可以减弱从该标记物产生的信号。见Becker和Nelson，美国专利5,731,148，并入此处作为参考。

另一实施方案中，信号探针至少包含染料二聚体。当探针与靶序列结合时，来自染料的信号与来自二聚体构象的染料的信号不同。

再一实施方案中，相互作用对可以是酶和该酶的抑制剂，例如核酸酶和核酸酶抑制剂、激酶和激酶抑制剂、蛋白酶和蛋白酶抑制剂、磷酸酶和磷酸酶抑制剂、caspase和caspase抑制剂、或核酶和核酶抑制剂、或抗原和可以特异地与该抗原结合的抗体 (这使得探针所检测的靶可以

被穿梭运输至抗原的特定细胞位置(例如,神经元的突触等))。

信号探针和蛋白酶探针或其它探针的构象

### 双链结构

本发明提供包含至少两条分开的核酸链的信号探针或蛋白酶探针或其它探针,其中所述至少两条核酸链经设计能够彼此退火或形成至少一个相互互补区。一条链的至少一个末端与另一条链的末端邻近(图1)。核酸可以是DNA、RNA、修饰的DNA或RNA。两条链可以是相同的链,形成自身二聚体(图8B)。这些链也可以具有不同的序列。

这两条分开的链可以设计成是完全互补的或者包含互补区和非互补区。一个实施方案中,两条分开的链被设计成彼此完全互补。一个实施方案中,两条链在每一端各形成4至9、5至6、2至10、10至40、或40至400个连续碱基对的相互互补区(见例如,图8、9、15、22、24或41)。这些链可以含有5-7、8-10、11-15、16-22、30个以上、3-10、11-80、81-200、或200个以上的核苷酸或修饰的核苷酸。两条链可以具有相同或不同数目的核苷酸(图2)。例如,一条链可以比另一条链长(图2C)。一个实施方案中,一条链的5'端与另一条链不重合,或者该链的3'端与另一条链不重合,或者两者兼有,其中所述不重合为不超过10个、不超过20个、或不超过30个的核苷酸或修饰的核苷酸。

与靶序列杂交的区域可以在互补区中、一条链或两条链的非互补区中或其组合。相同的信号探针可以靶向一个以上的靶核酸序列。这一个或多个靶可以在同一或不同的序列上,它们可以与设计用于结合靶的探针部分完全地互补或者至少足够互补。一个实施方案中,两条链在每一端各形成相互互补区,靶互补序列位于不是末端相互互补区的区域中(图8B)。

一个实施方案中,具有至少两条分开的链的信号探针是产荧光的探针。一个实施方案中,一条链至少在一个末端具有淬灭剂部分,在另一链的邻近末端具有荧光团(图1)。一个实施方案中,一条链的5'和3'末端都具有相同或不同的荧光团,另一条链的5'和3'末端都具有相同

或不同的淬灭剂部分(图 1B 和 2A)。一个实施方案中,一条链的 5' 末端具有荧光团, 3' 末端具有淬灭剂, 而另一条链的 3' 末端具有相同或不同的淬灭剂部分, 5' 末端具有相同或不同的荧光团(图 1C 和 2B)。

对于蛋白酶探针, 一个实施方案中, 一条链至少在一个末端具有蛋白水解酶, 在另一链的邻近末端具有该蛋白水解酶的抑制剂。一个实施方案中, 一条链的 5' 和 3' 末端都具有蛋白水解酶, 而另一链的 5' 和 3' 末端都具有该蛋白水解酶的抑制剂。一个实施方案中, 一条链的 5' 末端具有蛋白水解酶, 3' 末端具有该蛋白水解酶的抑制剂, 而另一链的 3' 末端具有该蛋白水解酶的抑制剂, 5' 末端具有蛋白水解酶。

### 茎环结构

另一实施方案中, 信号探针或蛋白酶探针或其它探针是包含至少一个相互互补区和至少一个非互补区的核酸或修饰核酸链。一个实施方案中, 探针形成茎环结构。茎区可以是相互互补的, 或者包含相互互补区和非互补区(图 4)。例如, 茎区可以具有未碱基配对的凸起核苷酸(图 4)。茎区还可以在 5' 或 3' 端含有未碱基配对的突出核苷酸(图 3B 和 3C)。

当茎区完全互补时, 茎区可以包括 3-4、5-6、7-8、9-10、2-6、7-10、或 11-30 个碱基对(见例如图 3 和 5)。环区可以含有 10-16、17-26、27-36、37-45、3-10、11-25 或 25-60 个核苷酸。一个实施方案中, 茎区形成 4-10、4 或 5 个连续碱基对(见例如图 23)。

一个实施方案中, 茎环结构包含至少一个含有两个化学基团的相互作用对, 分别地一个化学基团位于该链的一个末端。一个实施方案中, 信号探针在该链的两个末端各分别至少具有荧光团和淬灭剂部分(图 3、4 和 5)。蛋白酶探针在该链的两个末端各分别至少具有蛋白水解酶和该蛋白水解酶的抑制剂。

一个实施方案中, 茎区包含两个通过非互补区连接的相互互补区, 与相互作用对邻近的相互互补区形成 5 至 9 个碱基对, 与环区邻近的相互互补区形成 4 至 5 个碱基对(图 8、15、21、22、24 或 41)。一个实施方案中, 非互补区是单链环区(图 8)、错配区(图 15)或两者。另一实施

方案中，茎区包含通过两个非互补区连接的三个相互互补区，邻近相互作用对的第一相互互补区形成 4 至 5 个碱基对，第二相互互补区形成 2 至 3 个碱基对，邻近环区的第三相互互补区形成 2 至 3 个碱基对。

在茎环结构中，与靶序列互补的区域可以在一个或多个茎区或环区或两者中。茎中与靶杂交的区域可以在相互互补区、非互补区或两者中。一个实施方案中，靶互补序列在单链环区中。一个实施方案中，除邻近相互作用对的茎区以外的区域是靶互补序列(图 8)。同一探针可以靶向一个以上的靶核酸序列。这一个或多个靶可以在同一或不同的序列上，它们可以与设计用于结合靶的探针部分实现完全地互补或者至少足够互补。

茎长度的增加可能增加信号探针在其闭合构象时的稳定性，由此可以增加可检测信号的信噪比。将这些信号探针暴露于细胞可以在对于细胞仍旧安全的稍微升高的温度下进行，之后返回正常温度。在较高的温度下，信号探针将打开并在它们的靶存在的情况下与靶结合。一旦冷却后，未与靶结合的信号探针将在茎的增加的稳定性的帮助下恢复其闭合状态。类似地，可以使用其它力来达到相同的结果，例如被认为可以松弛碱基配对的 DMSO。

### 三臂联结结构

另一实施方案中，信号探针或蛋白酶探针或其它探针是形成三臂联结结构的核酸链(图 6A 和 6B)。在此结构中，茎区和两个茎环区连接形成三向联结(junction)。这些茎区可以含有 2-5、7-9、10-12 个碱基对。茎环区的环可以含有 3-7、8-10、11-13 个核苷酸或修饰的核苷酸。

一个实施方案中，三臂联结结构包含至少一个含有两个化学基团的相互作用对，分别地一个化学基团位于该链的一端。一个实施方案中，探针在该链的两个末端各分别至少具有荧光团和淬灭剂部分。蛋白酶探针在该链的两个末端各分别至少具有蛋白水解酶和该蛋白水解酶的抑制剂。

一个实施方案中，与相互作用对邻近的茎区形成 3 至 4、或 3 至 6

个连续碱基对，第一茎环结构的茎区形成 4 至 5 个连续碱基对，第二茎环结构的茎区形成 2 至 3 个连续碱基对(图 9)。一个实施方案中，这三个区域通过磷酸二酯键或修饰的磷酸二酯键经由茎区的臂连接在一起。一个实施方案中，这三个区域通过 1 或 2 个核苷酸或修饰的核苷酸经由茎区的臂连接在一起。

茎中与靶杂交的区域可以在相互互补区、非互补区或两者中。一个实施方案中，靶互补序列在单链环区中。一个实施方案中，除邻近相互作用对的茎区以外的区域是靶互补序列(图 9)。同一探针可以靶向一个以上的靶核酸序列。这一个或多个靶可以在相同或不同的序列上，并且它们可以与设计用于结合靶的探针部分完全地互补或至少足够互补。

### 哑铃结构

另一实施方案中，信号探针或蛋白酶探针或其它探针是形成哑铃形结构的核酸链(图 7 或 11)。该结构是通过两个茎区的一条臂连接在一起的两个茎环区。茎区可以含有 3-5、7-9、10-12 个碱基对。茎环区的环可以含有 5-7、8-10、11-13 个核苷酸。一个实施方案中，哑铃结构具有 3 个连续碱基对的一个茎区以及 4 个连续碱基对的一个茎区。一个实施方案中，两个茎区通过 1 或 2 个核苷酸或修饰的核苷酸连接。另一实施方案中，两个茎区通过磷酸二酯键或修饰的磷酸二酯键连接。一个实施方案中，茎环结构、哑铃结构或三臂联结结构具有 30 个以上的核苷酸或修饰的核苷酸。

一个实施方案中，该信号探针在该链的两个末端各分别至少具有荧光团和淬灭剂部分。蛋白酶探针在该链的两个末端各分别至少具有蛋白水解酶和该蛋白水解酶的抑制剂。

茎中与靶杂交的区域可以在相互互补区、非互补区或两者中。一个实施方案中，靶互补序列在单链环区中。一个实施方案中，靶互补序列是非此两个茎区的区域。同一探针可以靶向一个以上的靶核酸序列。这一个或多个靶可以在相同或不同的序列上，并且它们可以与设计用于结合靶的探针部分完全地互补或至少足够互补。

可从本领域中获得 DNA 或 RNA 折叠程序以预测给定的核酸或修饰的核酸的构象。此类折叠程序包括但不限于 Nucleic Acids Res. 31: 3429-3431 (2003) 中描述的程序和可在 <http://biotools.idtdna.com/analyzer/oligoalc.asp> 获得的 oligoanalyzer 3.0 软件, 并它们并入本文作为参考。此类折叠程序常常预测出多个能量上更有利的结构。在其它实施方案中, 本发明包括通过折叠程序预测的探针 FP1-18 的能量更有利结构(图 8-24 和 41)。如果通过自由能测量构象的能量, 则较低的自由能数值(负数)指示该构象在能量上更有利。

### 信号探针和蛋白酶探针或其它探针的化学修饰

本发明还提供经过化学修饰的信号探针和蛋白酶探针或其它探针。可以修饰糖-磷酸二酯型主链、2' OH 和碱基中的一个和多个。磷酸二酯键的替代包括但不限于  $--OP(OH)(O)O--$ ,  $--OP(O^M)(O)O--$ ,  $--OP(SH)(O)O--$ ,  $--OP(S^M)(O)O--$ ,  $--NHP(O)_2O--$ ,  $--OC(O)_2O--$ ,  $--OCH_2C(O)_2NH--$ ,  $--OCH_2C(O)_2O--$ ,  $--OP(CH_3)(O)O--$ ,  $--OP(CH_2C_6H_5)(O)O--$ ,  $--P(S)(O)O--$  和  $--OC(O)_2NH--$ 。M 是无机或有机阳离子。主链也可以是肽核酸, 其中脱氧核糖磷酸主链被假肽主链替代。肽核酸描述在 Hyrup 和 Nielsen, Bioorganic & Medicinal Chemistry 4: 5-23, 1996, 和 Hydig-Hielsen 及 Godskesen, WO 95/32305 中, 特此将这两篇文献并入此处作为参考。

糖的 2' 位包括但不限于 H、OH、 $C_1-C_4$  烷氧基、 $OCH_2-CH=CH_2$ 、 $OCH_2-CH=CH-CH_3$ 、 $OCH_2-CH=CH-(CH_2)_nCH_3$  ( $n=0, 1, \dots, 30$ )、卤素(F, Cl, Br, I)、 $C_1-C_6$  烷基和  $OCH_3$ 。 $C_1-C_4$  烷氧基和  $C_1-C_6$  烷基可以是或者可以包括直链、支链或环状基团。

核苷酸的碱基可以是腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶、尿嘧啶、肌苷、或具有修饰的前述碱基中的任何一个。修饰的碱基包括但不限于 N4-甲基脱氧鸟苷、脱氮杂或氮杂嘌呤及嘧啶。环氮例如腺嘌呤的 N1、鸟嘌呤的 N7、胞嘧啶的 N3 可以被烷基化。嘧啶碱基可以在 5 或 6 位被

取代，嘌呤碱基可以在 2、6 或 8 位被取代。见例如 Cook, WO 93/13121; Sanger, Principles of Nucleic Acid Structure, Springer-Verlag, 纽约 (1984), 并入此处作为参考。

常规核苷酸的衍生物是本领域熟知的，并包括例如具有不同类型的糖的分子。糖的 04' 位可以用 S 或 CH<sub>2</sub> 替代。例如，核苷酸碱基识别序列可以具有通过连接部分连在一起的环丁基部分，其中这些环丁基部分在其上连接有杂环碱基。见例如 Cook 等，国际公布 W094/19023 (特此并入此处作为参考)。

可以用于促进探针向细胞中递送的其它探针化学修饰包括但不限于胆固醇、转导肽 (例如 TAT、penetratin 等)。

## 方法

本发明方法基于信号探针在与活细胞中的靶 RNA 序列杂交时产生可检测信号的能力。该产生的信号应比对照细胞中产生的信号 (例如背景荧光) 可检测地高。因此，对照细胞根本不产生荧光是必需的。一个实施方案中，该方法用于检测或定量 RNA。一个方法是分离表达至少一种 RNA 的细胞或产生表达至少一种 RNA 的细胞系。在涉及分离细胞的本发明任何方法中，均可以培养该细胞并且还可以培养该细胞以产生细胞系。将编码 RNA 或 RNA 及标签序列的 DNA 构建体引入细胞中。该 DNA 构建体可以整合在细胞基因组中的不同位置。也可以实现在一个或多个特定位置的整合。然后，将转染的细胞暴露于信号探针，该信号探针在与所述靶 RNA 或标签序列结合时产生可检测信号。分离产生可检测信号的细胞。可以通过本领域任何方法分离和培养细胞，例如，可以单个地或成批地分离和接种细胞。可以通过使分离的细胞生长而产生细胞系。

本发明任何方法均可以使用选择标记来进行。尽管药物筛选 (或使用任何其它适宜的选择标记来筛选) 不是必需步骤，但可以用于从转染的细胞群体中富集稳定转染的细胞，条件是转染的构建体经设计可以赋予药物抗性。如果在转染后过快地使用信号探针进行筛选，则一些阳性细胞可能仅被瞬时地或不稳定地转染。然而，这可以通过充分的细胞传代来

最小化，充分的细胞传代可以使得不稳定转染的细胞中转染的质粒被稀释或发生丢失。一些稳定整合的质粒可能不产生任何对应于克隆的 cDNA 插入物的 RNA。其它可能产生不可以被信号探针检测到或不可以被信号探针有效地检测到的 RNA。

本发明 RNA 可以具有一种或多种以下不同角色：编码蛋白质、融合蛋白质、与蛋白质融合的肽、输出信号、输入信号、细胞内定位信号或其它信号（可以与蛋白质或肽融合）的信使 RNA；反义 RNA、siRNA、结构 RNA、细胞 RNA（包括但不限于例如核糖体 RNA、tRNA、hnRNA、snRNA）；随机 RNA、对应于 cDNA 或 EST 的 RNA；来源于不同物种的 RNA、对应于寡核苷酸的 RNA、对应于整个细胞、组织、或生物体的 cDNA 制备物的 RNA；与其它核酸、蛋白质、其它细胞成分或药物分子具有某些结合活性的 RNA；可以掺入各种大分子复合物中的 RNA；可以影响一些细胞功能的 RNA；或不具有上述功能或活性但可以被细胞表达的 RNA；对应于病毒或外源 RNA、接头 RNA、或连接一个或多个 RNA 的序列的 RNA；或充当标签的 RNA，或者上述任何一或多个 RNA 的未修饰的、诱变的、随机化的或改组的序列的组合或重组。RNA 可以由组成型启动子或条件型启动子（包括但不限于诱导型、阻遏型、组织特异性、热休克型、发育型、细胞谱系特异性、或时间型启动子，或者上述任何一个或多个启动子的未修饰的或诱变的、随机化的或改组的序列的组合或重组）控制。

一个实施方案中，信号探针是产荧光的探针。荧光细胞分选仪或相关技术可以用于发荧光的探针以鉴定和/或分离在一个或多个波长下显示出某一种或多种水平的荧光的细胞。能够可靠地和有效地检测活细胞中的 mRNA 及其它 RNA 的能力使得可以使用它们基于期望的特征，例如通过使用荧光激活细胞分选仪 (FACS)，鉴定和，如果期望的话，分离细胞。FACS 技术目前允许每秒分选多达 70,000 个细胞。在不足 2 分钟内可以分选 5,000,000 个细胞。

### 1. 产生表达蛋白质的细胞系

在制备细胞系时涉及到的最冗长乏味的步骤中有一些步骤可以通过

应用本文所述的信号探针而去除。一个实施方案中，在用编码期望的基因的 DNA 构建体转染后，向这些细胞中引入设计用于识别目的基因的信使的发荧光探针。该步骤可以在使用选择标记进行筛选（例如，药物筛选，条件是该转染的 DNA 构建体也编码药物抗性）后进行。转录所述基因的细胞将发荧光。之后的 FACS 分析导致发荧光的细胞的分离，然后可以使该细胞生长以产生表达所选基因的细胞系。

一个实施方案中，设计信号探针使之与编码目的蛋白的 RNA 的一部分或者其 5' 或 3' 非翻译区的部分互补。如果设计用于识别目的信使 RNA 的信号探针能够检测内源性存在的靶序列，则通过将内源靶序列的比例与转染细胞所产生的靶序列的比例相比较，将使得分选仪能够区分这两类细胞。目的基因可以用标签序列加上标签，可以设计信号探针使之识别该标签序列。标签序列可以和基因的信使 RNA 的蛋白质编码部分读框相符或读框不相符，这取决于是否希望给产生的蛋白质加上标签。

此外，目的基因在任何给定的细胞中的表达水平可能变化。这可能由多种能够影响 RNA 表达水平的因素导致，所述因素包括但不限于经转染进入细胞中的 DNA 量或拷贝数、任何所致的 DNA 在基因组中的整合的位点和 DNA 的完整性以及在基因组整合后从 DNA 导致的表达。可以应用 FACS 评价表达水平和差别地选择表达此相同基因的各个细胞。

## 2. 产生下调基因的细胞系

有几项研究描述了如下细胞系的制备，所述细胞系不表达编码蛋白质的 RNA，而表达基因或基因部分的反义 RNA。这些方法旨在降低给定细胞中特定 RNA 或蛋白质的量。除了此处信号探针被设计用于检测的 RNA 是反义 RNA 外，上述用于产生表达蛋白质的细胞系的步骤在此同样适用而且实质上是相同的。

不是所有制备稳定转染的反义 RNA 表达细胞系的尝试都导致其中被靶向的蛋白质的表达受到充分影响的细胞系。此困难使得不太值得去从事此类细胞系的制备。考虑到此处所述的方法的容易性，故可以容易地

测定众多不同遗传序列在其产生活性的（即有效的）反义表达细胞系的能力方面的有效性。然后可以分析这些以确定哪些显示出适当表达谱，在其中可以分析所靶向的基因的下调。

RNA 干扰是一种也旨在降低特定基因的转录水平的可选方法。可以使用化学合成的 siRNA 或编码短 siRNA 的 DNA 构建体瞬时地诱导 RNA 干扰，或者如果制备了适当地表达短 siRNA 的稳定细胞则可以稳定地诱导 RNA 干扰。也可以使用此处所述的方法，根据细胞中此类 siRNA 的表达来分析或分离细胞或细胞系。

此外，使用此处所述方法获得表达可以诱导 RNA 干扰的 RNA 的细胞之所以重要是因为这将帮助克服目前在此类细胞中遇到的一些复杂问题。例如，尽管 RNA 干扰的实施旨在降低所靶向的 RNA 的转录水平而且可能对其它 RNA 的转录水平造成下游影响，但已经证明其也降低不是目的靶标的 RNA 的转录水平。这些非目的靶标的身份根据用于诱导 RNA 干扰的 RNA 的序列而变化。由于结果可能被此类非需要的后果危害，这造成了 RNA 干扰在细胞中使用时的复杂性。因为本文所述方法能够有效地产生多个稳定细胞，其中每一个细胞分别表达例如用于诱导对同一基因的 RNA 干扰的不同 RNA 序列，故可以在细胞中测试这些序列中每一个序列对其它基因转录水平的影响。对这些信息的分析可以用于辨别降低了转录水平但所述转录水平的降低并非由所靶向的 RNA 的表达水平的降低导致的 RNA。由于已经成功地利用 RNA 干扰克服了与使用反义 RNA 进行特异下调时有关的困难，RNA 干扰已经快速地流行起来。我们的方法可以帮助确定用于实现更特异的 RNA 干扰的具有降低的非特异活性的最有效序列。由于本发明方法也提供了可用于筛选最有效和特异的反义序列的方法，它们也是鉴定有效反义 RNA 的方法。

### 3. 基于细胞表面定位的抗原区分细胞

免疫学家等人已经长期使用 FACS 分选细胞。一般地，该方法基于用差异标记的探针（通常是荧光团标记的抗体探针）标记定位在细胞表面的蛋白质来进行。例如，可以产生阳性地表达细胞表面定位蛋白的细胞。

该方法最常在设计用于保存细胞的完整性和维持细胞的生存力的条件下使用。

根据本发明，为了检测细胞表面定位蛋白质的存在，使信号探针靶向编码该目的蛋白质的 mRNA。在不破坏细胞生存力的情况下，通过转染将信号探针引入细胞中。然后，使用细胞分选仪分离阳性评分的细胞。此外，如果使用信号探针的组合，且其中每一种信号探针分别地靶向一种目的蛋白质的 mRNA，则可以差异地标记每一种探针。如果细胞中具有靶种类数量比单次应用 FACS 能够检测的靶种类数量大，则可以实施多轮分选以分选细胞。

此处所述方法也使得能够分析或分选具有其它细胞 RNA（例如，编码细胞内部定位的蛋白质或从细胞分泌的蛋白质的 mRNA，或者不编码蛋白质的 RNA）的细胞。因此，可以将前述一种或多种细胞 RNA 作为靶标来检测，包括编码无法被常用抗体探针接近的蛋白质的 RNA 或者编码尚未开发出相应探针的蛋白质的 RNA。这些蛋白质可以包括膜结合的、跨膜的、膜锚定的、细胞质的、或核质的蛋白质。

#### 4. 分析细胞中特定 RNA 的表达和定量细胞中 RNA 表达的水平

如果存在引入细胞中的荧光探针的靶 RNA，则细胞发荧光。该信息可以通过使用荧光显微镜（包括共聚焦的、激光扫描的和其它类型的显微镜）和 FACS 定性地评价，并且该信息也可以通过这些方法之一量化。例如，代替在组织切片上实行原位逆转录聚合酶链式反应 (RT-PCR) 来确定特定 RNA 的表达模式，可以使用信号探针来实施相同实验。而且，通过组合使用分别靶向特定 RNA 的发出不同荧光的产荧光探针，可以在一个步骤中检测几种目的 RNA 的存在或量。可以在经验确定能够限制非靶特异性信号产生的温度下，实施 RNA 的检测（在固定的样品中）。建立使用核酸探针检测靶的最佳温度条件，属于常规操作。（组合进行抗原的定位与细胞和组织中 RNA 的检测）

使用固定的细胞或组织切片，利用免疫细胞化学可以描述出所识别的蛋白质抗原的定位，而使用靶向特定 RNA 的信号探针，可以在相同样

品中对目的 RNA 作出联合定位。已经证实，靶向 RNA 的发荧光探针可以在固定的细胞中发挥功能。

### 5. 产生表达多种 RNA 或蛋白质的细胞系

使用本发明的方法，可以非常快速地产生表达任何数量 RNA 或蛋白质的稳定转染细胞系，甚至无需在多种选择药物的混合物存在下（或者使用其它选择试剂）维持这些细胞。在基因转染和任选地药物筛选后，将信号探针的组合（每一种信号探针分别靶向一种蛋白质的信使）引入细胞。通过设计每一种产荧光探针的靶互补序列，使该序列仅与一种基因的 mRNA 杂交或与该信使可能连接的标签序列杂交，由此设计每一种信号探针以使之识别仅一种基因编码的 mRNA。一个实施方案中，然后通过 FACS 分选细胞。通过针对一种或多种信号进行选择，可以在一次应用中产生多个细胞系。

有可能需要产生表达多种目的 RNA 的细胞系，其中所述目的 RNA 的种类数目大于单次应用 FACS 所可能鉴定的种类数目。例如，具有过表达被认为参与特定复合物形成或参与生物学途径的所有蛋白质和 RNA 序列的细胞系，将提供大量的信息。例如，在相同或相关生物学途径中的 RNA 或蛋白质、在彼此上游或下游起作用的 RNA 或蛋白质、对彼此具有调节、激活或阻遏功能的 RNA 或蛋白质、对于功能或活性而言相互依赖的 RNA 或蛋白质、形成复合物或彼此结合的 RNA 或蛋白质、或者享有同源性（例如，序列、结构或功能同源性）的 RNA 或蛋白质。如果所需 RNA 的数目大于应用一次 FACS 所能分析的数目，则为了达到此目的，可以使用已经表达这些 RNA 中一些 RNA 之组合的细胞作为将转染编码其它 RNA 的其它构建体的宿主细胞，重复上述步骤。可以多轮次实施所述方法以获得表达所有或部分所需 RNA 的细胞。

当待表达的多种 RNA 都被克隆在赋予细胞对相同药物的抗性的构建体中时，一个实施方案中，可以使用 FACS 分离表达所有期望 RNA 的细胞。在这些序列稳定整合在基因组中的情况下，这些细胞被期望不丢失任何序列的表达。然而，可能地，一种或多种序列发生丢失。如果这样，可

以增加培养这些细胞的培养基中的选择药物或筛选试剂的浓度，从而使这种可能性更小。或者，可以使用各赋予对不同药物的抗性的多个构建体，并在适宜药物的混合物中维持细胞。此外，也可以对待稳定转染至细胞中的构建体进行选择，使得其中一个子集编码针对一种药物的抗性基因，而另一个子集编码针对另一种药物的抗性基因。

此外，如果细胞系中的一些细胞丧失了目的 RNA 的表达，则作为一种手段，可以重复上述用于分离该细胞系的最初实验并获得新的细胞。或者，可以通过 FACS 分析所述细胞的混合物，以便重新分离表达所有期望 RNA 的细胞。这是一个非常有用的程序，因为它再次产生细胞，由该细胞可以得到与最初选择的细胞系具有相同遗传构成的细胞系。

通过上述方法可以产生无限制供应的以下细胞，所述细胞表达任何蛋白质和 RNA 序列的组合，并可以在实际无限种分析方法中使用。再一可能是过表达的蛋白质可能对细胞有毒，而正如之后将讨论的，这种可能性可以容易地解决。

由于可以容易地从不再表达所有期望 RNA 的细胞中重新分离出表达所有期望 RNA 的细胞，这使得可以在无药物存在下或在最小浓度的药物存在下维持细胞系。此处所述方法也使得能够对前面产生的细胞或细胞系重新应用信号探针。例如，以便确定细胞是否就其最初分离时所针对的 RNA 中的任何一种或多种而言仍是阳性的以及阳性的程度如何。

## 6. 产生显著地过表达一种或多种 RNA 或蛋白质的细胞系

对于待高度过表达的每一基因，例如，首先将该同一基因的两个或多个序列克隆至 DNA 构建体中，该 DNA 构建体可以任选地还赋予药物抗性或其它选择标记。对每个基因的多个序列分别地就每一个序列进行设计以包括编码不同标签序列的序列。一个实施方案中，在用 DNA 构建体转染细胞和随后进行筛选后，将产荧光的探针引入细胞中，其中所述探针每一种仅靶向一种标签序列并被差异荧光标记，之后使用细胞分选仪分离具有这些探针的阳性信号的细胞。这些细胞已经在基因组中整合了这些被差异地加标签的序列每一个至少一个拷贝，由此从拷贝数目增加

的、实质上相同的目的序列获得了目的序列的表达。目的序列可以整合在细胞基因组的不同位置。该方法可以与 FACS 联用以挑出针对每一种荧光团的信号分别具有最强信号评分的那些细胞。所述不同标签中一部分标签或全部标签可以相同，这样可以使用针对此共同序列的共同信号探针在细胞中检测所有的不同标签。

#### 7. 产生表达多种反义 RNA 的细胞系

如下构建产生多种反义信使的稳定转染细胞系。这些反义信使可以靶向 mRNA 或其它 RNA。可以选择以不同水平表达任何一种转染的反义序列的细胞。通过多轮重复的稳定转染，可以容易地选出如下细胞，所述细胞将容易地产生表达无限多种 RNA 的反义信使的稳定转染细胞系。

当然，也可以通过此处所述方法制备表达反义 RNA 之外的 RNA 的细胞。此 RNA 包括但不限于 mRNA、rRNA、siRNA、shRNA、其它结构 RNA 如 hnRNA、tRNA 或 snRNA、具有 RNA 干扰活性的 RNA、充当标签的 RNA 等中的一种或多种。

#### 8. 产生细胞系文库

用形成表达文库的 DNA 构建体转染多个细胞。DNA 序列表达文库可以包括本领域已知的任何类型或其混合物，包括但不限于从例如整个生物体、组织、细胞或细胞系产生的 cDNA 或 EST 文库，以及合成文库，包括但不限于编码肽的序列或寡核苷酸。类似地，也可以使用未明确地称为表达文库的 DNA 构建体文库。例如，DNA 构建体文库可以包括可能具有调节功能的 DNA 序列，例如启动子、阻遏物或增强子元件，并且这些 DNA 序列可以是组成型的、诱导型的或阻遏型的。同样，DNA 构建体文库可以包含可以发生转录的大片段 DNA，例如完整的或部分的基因座位或基因组 DNA。可以对这些 DNA 构建体表达文库之任一个进行完全或部分地诱变、随机化、重组、改组、改变或以其任何组合的方式进行处理。此外，任何这些类型的文库都可以还包含至少一个标签，所述标签被表达并可以用作信号探针的靶标。

可以制备特定类型的蛋白质的 DNA 序列表达文库，并用于产生细胞系文库。例如，可以将蛋白质激酶的 cDNA 克隆至包含标签序列的表达构建体中。可以获得稳定表达不同激酶的细胞并用于产生限于表达激酶的细胞系的细胞系文库。可以选择序列的类型以满足其它的应用，例如药物筛选。

细胞系文库也可以表达具有序列同源性、属于相同蛋白质家族或属于一个功能家族的蛋白质类型，并还可以表达在给定蛋白质途径或复合物或系统中依据其作用定义的蛋白质。例如，在筛选可能和各种 HIV 蛋白质结合的化合物的药物筛选中，可以制备其中的每一个细胞系分别表达不同的 HIV 蛋白质的细胞系文库，并用于筛选药物化合物。

可以使用细胞系文库分析具有期望活性的分泌肽或蛋白质。首先，使用表达文库(肽、蛋白质、EST、cDNA 等的表达文库)制备细胞系文库，其中所述表达文库还编码以符合所述肽/蛋白质的阅读框的方式翻译的输出信号。表达的肽/蛋白质被分泌至生长培养基中。假如存在恰当设计的试验，然后可以确定这些肽对特定活性的影响。可以收集来自细胞系的组织培养上清液并施加于待测细胞以分析该活性。

此外，可以例如通过用表达文库转染细胞，并任选地首先选择转染了表达文库的细胞(例如，使用针对包括在表达文库中的标签的信号探针)，然后将细胞暴露于针对一种或多种目的 RNA 的一种或多种信号探针，从而制备细胞系文库。这使得可以在细胞中表达各种 RNA 并确定这些 RNA 中哪个导致了一种或多种目的 RNA 的下游转录上调或下调。

#### 9. 产生功能性敲除了一种或多种蛋白质的细胞系

本发明方法提供了在培养细胞中实现功能性敲除的手段。通过制备从多个座位表达实质上相同的针对独特 RNA 的反义 RNA 或对独特的 RNA 序列具有 RNA 干扰活性的 siRNA 的细胞，可以产生任何一种目的蛋白质被功能性敲除的细胞系。使用 siRNA 的本发明任何方法也可以使用 shRNA 实施。例如，可以向细胞中转染多种构建体，其中每一种构建体分别编码针对特定基因的反义 RNA 或者编码对该基因具有 RNA 干扰活性的

siRNA, 或者编码两者。在此, 每一种反义 RNA 序列的区别仅在于每一种被具有独特标签序列的核苷酸序列加上标签。可以选择那些表达一种或多种或所有的被差异地加标签的反义 RNA 的细胞。类似地, 可以通过检测与每种 siRNA 或 shRNA 连接的标签, 分别确定每一种具有 RNA 干扰活性的 siRNA 或 shRNA 的存在。因为 FACS 可以用于定量如前所述的荧光, 故此性质使得可以选出最强地表达任何一种或多种反义序列的细胞。可以分离由于如下原因导致的、所靶向的 RNA 表现出期望的表达水平或者几乎不或完全不表达的细胞, 其中所述原因是任何种类数量或组合的被表达 RNA 的表达, 其中所述被表达的 RNA 起着降低所靶向的 RNA 的表达水平的作用。

重要的是, 可以在此方法中使用靶向相同基因的不同反义序列和具有 RNA 干扰活性的 siRNA 中的一种或多种。例如, 对于其表达使用信号探针和 FACS 进行选择反义 RNA, 可以对其中的一些进行设计使之靶向该基因的信使 RNA 的特定区域, 而设计其它的反义 RNA 使之靶向相同信使的其它可选部分。为了产生目的蛋白质被功能性敲除的细胞系, 用编码相似或不同的反义 RNA 或 siRNA 的许多种遗传序列稳定转染细胞系, 其中所述相似或不同的反义 RNA 或 siRNA 对相同目的基因具有 RNA 干扰活性, 其中所述遗传序列的种类数量是对于产生显示出无可检测的目的蛋白质表达水平或者显示出可接受的低表达水平的细胞系而言所必需的数量。

此外, 通过重复上述步骤, 并靶向任何种类的待被反义或 siRNA 功能性敲除的序列, 可以产生多种蛋白质被功能性敲除或者具有降低的表达水平的细胞系。例如, 为了研究蛋白质复合物的功能, 可以敲除或降低构成该复合物的蛋白质中的一种或所有或任何组合的表达水平。

#### 10. 产生一种或多种基因的仅选定的可变剪接形式被功能性敲除的细胞系

一个基因的不同剪接形式常常翻译成具有不同功能的蛋白质。使用本发明方法, 可以产生一种或多种蛋白质的仅选定的可变剪接形式被功

能性敲除或表达水平被降低的细胞系。例如，通过设计反义或 siRNA 使之仅仅靶向基因的信使 RNA 中那些希望从细胞中消除的可变剪接形式，可以将目的基因中除了有意义的可变剪接信使之外的所有可变剪接 RNA 功能性地敲除，或者将它们的表达水平充分地降低至期望水平。

#### 11. 产生表达一种或多种 RNA 或蛋白质而另一种或多种蛋白质被功能性敲除的细胞系

例如，对于给定的一组被认为彼此相互作用的蛋白质，可以通过产生如下稳定转染的细胞系来研究它们的相互作用，在所述细胞系中通过细胞表达反义或 siRNA (或 shRNA)，目的蛋白质中的一种或多种被功能性敲除或者具有降低的表达水平。然后可以研究细胞中剩余的蛋白质的功能，但是可能更有意义的是，可以通过进一步操作该细胞使之过表达剩余目的蛋白质中的一种或多种来进一步改变此细胞。此外，可以在细胞中过表达或消除或降低其它蛋白质的表达。还有可能的是，过表达某些蛋白质或功能性敲除某些蛋白质可能对细胞是致死的。这是下面将要解决的问题。

可以使用类似方法，通过引入在转录调节中具有直接或间接作用的 DNA，随机地上调或下调基因。例如，可以向细胞中转染 DNA 序列，包括但不限于启动子、增强子或阻遏物序列、或具有某种导致一种或多种 RNA 的转录水平获得调节的其它结合或功能活性的序列之一或其组合。这些元件的活性可以是组成型的、诱导型的或阻遏型的。可以使用针对一种或多种特定 RNA 的信号探针鉴定或分离这些 RNA 的表达水平已经被增加或降低的细胞。通过使用所述方法，一些 DNA 序列的稳定整合可能随机地充分开启或关闭自遗传座位的转录。可以筛选这些细胞系文库，选择特定基因被有效地过表达或者敲除的细胞。以此方式可以选出一种或多种 RNA 具有期望水平的细胞。如果必要的话，可以实施该程序多轮次，以分离多种目的 RNA 具有期望的表达谱的细胞。

#### 12. 产生转基因小鼠

对于一些目的，研究培养的细胞是不够的。然而，上述方法学也适于操作胚胎干细胞。可以根据上述程序获得如下胚胎干细胞，所述胚胎干细胞可以表达多种 RNA 或蛋白质或者可以作为其中的多种蛋白质或多种蛋白质的可变剪接形式中的一个子集被功能性地敲除的细胞起作用，等等。然后此胚胎干细胞可以用作产生转基因动物的基础。

根据所述方法分离的细胞可以直接地植入生物体，或者，可以将它们的细胞核移植入其它受体细胞，然后将这些受体细胞植入。一种用途可以是产生转基因动物，而其它的用途可以包括但不限于引入可以合成或分泌细胞产物的细胞进入生物体和引入为了在生物体中执行期望作用而被工程化改造的细胞。

### 13. 产生诱导型稳定转染的细胞系

某些蛋白质或 RNA 在细胞中的过表达或表达缺乏可能是致死性的或损伤性的。然而，研究过表达毒性蛋白质或 RNA 的细胞，或者蛋白质或 RNA（没有该蛋白质或 RNA，细胞将无法存活）被功能性敲除的细胞，可能是至关重要的。为此，可以制备稳定转染的细胞，在该细胞中对细胞具有此类破坏性作用的选定 RNA 在诱导型或条件性启动子的控制下。为了分离此细胞系，一个实施方案中，首先最小地诱导该转染的和任选地经过药物筛选的细胞，以实现该诱导型基因的转录，然后在向细胞中转染了设计用于识别适宜 RNA 的信号探针后，对细胞进行 FACS 分析。维持所获得的细胞以便仅在需要时诱导和转录该毒性 RNA。

诱导型系统对于毒性 RNA 表达以外的应用而言也可能是有利的。例如，可以在同步化细胞系的细胞周期中于某一点上诱导稳定转染入细胞中的遗传序列表达。或者，如果一组稳定转染的一个或多个遗传序列的表达产物被认为对另一组遗传序列的表达产物起作用，那么有意义的是，将第一组遗传序列克隆在一种诱导型启动子控制下，而将第二组遗传序列克隆在第二诱导型启动子控制下。通过改变诱导，可以在任一组遗传序列的表达产物缺乏或存在的情况下研究另一组遗传序列所编码的表达产物。一般地，按上述方法中所述掺入细胞中的 DNA 序列均可以放置在

具有期望活性的启动子或其它转录调节物的控制下。例如，可以使用诱导型、组织特异性、时间特异性或时限性(temporal)启动子、增强子或阻遏物，以及由于细胞的或细胞外的信号而被调节、激活或阻遏的调节元件，其中所述信号包括但不限于化合物或化学药品、其它细胞、蛋白质、肽、激素、信号传递分子、细胞分泌的因子、来自生物体、组织或细胞的整个提取物或分级分离的提取物、或环境样品中的一种或多种。在这些情况下，首先将细胞暴露于适当水平的调节转录的药剂，之后再将其暴露于信号探针。

#### 14. 检测活细胞中的遗传重组事件和随后分离未重组的或以不同方式重组的细胞

与使用信号探针检测经历了在产生稳定细胞系时所涉及的重组事件的细胞类似，可以使用信号探针从活细胞混合物中检测和分离那些经历了其它特定重组事件的细胞。此相同的原则可以用于分析例如，VDJ 重组、易位和病毒基因组整合。

在细胞重组事件中，例如，以基因组 DNA 的一个序列交换了另一序列。如果编码发生重组事件的区域的 DNA 序列转录成 RNA，则可以通过设计用于识别自该未重组的 DNA 序列转录的 RNA 或自重组后的序列转录的 RNA 的信号探针，检测该事件是否存在。该试验也可以使用荧光显微镜(或者能够定量所得信号的其它仪器)实施。如果希望从未重组的细胞中分离出重组的细胞，可以对细胞实施 FACS 并进行分选。此外，可以使用 FACS，基于细胞中是否存在多个重组事件来分选细胞。

#### 15. 基于表达的 RNA 分选细胞

使用本文所述的信号探针将允许基于编码内部定位的、细胞表面定位的或分泌性的蛋白质的 RNA 以及可能存在于该细胞中的其它 RNA 在细胞中的表达来分选细胞。例如，从一个混和的细胞群体出发，可以通过设计识别产生内部定位的目的蛋白质的 mRNA 的信号探针，分离表达这些目的蛋白质的细胞。将这些信号探针转染至细胞混合物中，并在适当的

情况下可以使用 FACS 分选细胞。可以实施多轮分选。

此外,研究者也可能例如对分离表达一种或多种特定蛋白质的 mRNA 的细胞或者表达可以响应于给定的添加因子而发生转录的 RNA 的细胞感兴趣,或者对确定何种添加因子诱导或阻遏一种或多种蛋白质或 RNA 的转录感兴趣。可以以此方式测试的添加因子可以包括但不限于如下之一或多种:核酸、蛋白质、肽、激素、信号传递分子、化学化合物、无机或有机化学药品、细胞、来自或衍生自生物体、组织或细胞的完整提取物或分级分离的提取物、从细胞或生物体纯化或分离的产物、来自环境或其它来源的样品。

为了分离响应于例如细胞因子而被诱导以表达一种或多种特定 RNA 的细胞,首先利用细胞因子诱导细胞混合物,然后用信号探针进行转染,其中每一种信号探针分别被设计为识别将产生其中一种目的蛋白质的 mRNA。在一个实施方案中,然后使用 FACS 分离就目的 mRNA 而言被评定为阳性的细胞。在一个备选实施方案中,也可以分析病毒感染的细胞以检测例如细胞中特定基因的表达。或者,可以将一组化合物或一个化合物文库,例如化学化合物文库、或 RNA 表达文库,施用于细胞,以确定,如果存在的话,何种化合物或化合物的混合物或 RNA 诱导、阻遏或调节一种或多种特定蛋白质或 RNA 的转录。

可以利用上文所述方法学检测存在具有一种或多种以下不同作用的 RNA 的阳性细胞:编码蛋白质、融合蛋白质、与蛋白质融合的肽、输出信号、输入信号、细胞内定位信号或其它信号(可以与蛋白质或肽融合)的信使 RNA;反义 RNA、siRNA、形成与 siRNA 具有相似活性的发夹结构的短 RNA;结构 RNA、细胞 RNA,包括但不限于例如核糖体 RNA、tRNA、hnRNA、snRNA;随机 RNA、对应于 cDNA 或 EST 的 RNA;来自不同物种的 RNA、对应于寡核苷酸的 RNA、对应于整个细胞、组织、生物体的 cDNA 制备物的 RNA;与其它核酸、蛋白质、其它细胞成分或药物分子具有某种结合活性的 RNA;可以掺入各种大分子复合物中的 RNA;可以影响某种细胞功能的 RNA;或不具有上述功能或活性但可以由细胞表达的 RNA;对应于病毒的或外源的 RNA 的 RNA、接头 RNA、或连接一个或多个 RNA 的序

列；或者充当标签的 RNA；或者以上任一种或多种 RNA 的未修饰的、经诱变的、随机化的或改组的序列的组合或重组。RNA 可以在组成型或条件型启动子的控制下，所述条件型启动子包括但不限于诱导型、阻遏型、组织特异性、或时限性启动子，或者以上任一种或多种的未修饰的、经诱变的、随机化的或改组的序列的组合或重组。上述 RNA 的表达可以通过向细胞中引入 DNA 构建体、载体，或者通过递送导致这些 RNA 表达的核酸的其它递送方法而引起。

#### 16. 使用蛋白酶探针在体内检测核酸和随后选择细胞

本发明也涉及新的蛋白酶探针形式，与信号探针不同，该蛋白酶探针在与其靶核酸结合时显示出蛋白水解活性。此蛋白水解活性可以用于检测目的，此外，如果细胞中存在靶核酸，也可以用于在细胞中降解特定蛋白质序列。例如，特异地切割病毒蛋白质的蛋白酶，可以在病毒序列的转录被激活，例如在潜伏感染中，而且例如该蛋白酶探针靶向该病毒信使时被激活。

另一方面，本发明涉及产生蛋白水解活性的一元杂交探针 (unitary hybridization probe)，在此称作蛋白酶探针。此蛋白酶探针也包含与靶 RNA 互补的核苷酸或修饰的核苷酸以及相互互补的核苷酸或修饰的核苷酸。这些蛋白酶探针以和前述探针相似的方式运作，但与信号探针和其靶核酸序列相互作用时荧光信号的产生或改变不同，该蛋白酶探针在靶存在的情况下转变为具有蛋白水解活性。

在以上方法中可以用蛋白酶探针替代信号探针，从而产生新的可能方案。在蛋白酶探针转染表达该蛋白酶探针所识别的 RNA 的细胞后，该蛋白酶探针与其靶杂交。因为在杂交状态下蛋白酶不再位于其蛋白酶抑制剂的附近，这造成该蛋白酶的激活。其中存在该蛋白酶探针的靶而且该靶被识别的细胞将被损伤，并由此得以被选择。不表达该 mRNA 的细胞或多或少地不受影响。相反地，可以设计蛋白酶探针使之催化如下蛋白水解活性，该蛋白水解活性可以刺激或以其它方式有益地影响细胞生长或存活力，或者可以使得其中存在蛋白酶探针的靶且该靶被识别的细胞

具备生长优势。

此外，蛋白酶探针可以用于各种其它用途。蛋白酶的活性可以容易地进行测定，而且在特定核酸靶序列存在下活性蛋白酶不仅可以用于检测目的还可以用于治疗目的，例如，当特定基因，如与细胞转化、癌发生、异常增生等有关的基因转录时，该蛋白酶探针所递送至的细胞将发生蛋白水解作用从而不能存活。例如，假设一个其中的一些细胞被特定病毒感染的细胞混合物，可以向细胞中引入靶向特定病毒 mRNA 的蛋白酶探针。带有此 mRNA 的细胞激活其所包含的蛋白酶探针的蛋白水解活性，并且该活性将破坏这些细胞。

优选地，蛋白水解酶抑制剂是肽或小的化学物质，但是也可以使用其它分子，包括但不限于金属和金属螯合物，以便在酶与抑制剂相互作用时提供对酶的可逆抑制。有用的蛋白水解酶及蛋白水解酶抑制剂对的实例包括但不限于氨肽酶和抑氨肽酶肽、胰蛋白酶样半胱氨酸蛋白酶和抗蛋白酶、氨肽酶和苯丁抑制素、胰凝乳蛋白酶样半胱氨酸蛋白酶和胰凝乳蛋白酶抑制剂、氨肽酶和 diprotin A 或 B、羧肽酶 A 和 EDTA、弹性蛋白酶样丝氨酸蛋白酶和 elastinal、及嗜热菌蛋白酶或氨肽酶 M 和 1,10-二氮杂菲 (1,10-phenanthroline)。

此外，也可以使用掺入了其它相互作用对的探针，其中相互作用对的一个成员具有期望的活性，而另一成员在探针未与靶结合时起着抑制或降低该活性的作用。当探针与其靶结合时，由于相互作用对的抑制性成员不再位于具有期望活性的成员的附近，探针的活性显示出来。

基于前面的描述，可以实施以下方法。

分离表达至少一种 RNA 的细胞的方法，包括步骤：

- a) 向细胞中引入编码所述至少一种 RNA 的 DNA；
- b) 将所述细胞暴露于在与所述至少一种 RNA 杂交时产生可检测信号的至少一种信号探针；和
- c) 分离产生所述信号的所述细胞。

该方法还可以包括步骤：使分离的细胞生长以产生表达该 RNA 的细胞系。当 DNA 构建体整合在转染细胞的基因组的不同位置时，则可以产

生多个细胞系。除非使所转染的构建体在基因组中的整合定向于基因组的特定位置，否则转染被认为将随机发生，这样每一个阳性细胞都可能互不相同，这就存在多个不同的细胞系，对于作为选择这些细胞系所针对的 RNA 而言，所有的细胞系都是阳性的。如果将 DNA 构建体引入细胞，例如永生化细胞、原代细胞、干细胞和生殖细胞或细胞系的混和群体中，则也可以产生多个细胞系。细胞也可以来自任何已经建立的细胞系，包括但不限于 HeLa、HEK 293T、Vero、Caco、Caco-2、MDCK、COS-1、COS-7、K562、Jurkat、CHO-K1、Huvec、CV-1、HuH-7、NIH3T3、HEK293、293、A549、HepG2、IMR-90、MCF-7、U-2 OS 或 CHO。任选地，DNA 构建体还可以编码至少一种药物抗性标记或其它选择标记，并且该方法还可以包括步骤：在步骤 a) 后使用选择标记选择细胞。分离的细胞可以单独地培养或者合并在一起。无论细胞何时分离，是否在用了一个或多个构建体或一个或多个表达文库进行转染之后，都可以将分离的细胞彼此分开进行培养或者将之合并在一起。

可以进一步操作分离的细胞以表达第二 RNA。以同时或相继方式，额外的步骤包括：用编码第二 RNA 的第二 DNA 构建体转染细胞或细胞系；将所述细胞暴露于在与所述第二 RNA 杂交时产生可检测信号的第二信号探针；和分离显示出所述 RNA 和第二 RNA 中的至少一个或两者的信号的细胞。第一信号探针可以与第二信号探针产生相同或不同的信号，例如，它们可以具有相同或不同的荧光团。通过同时或相继地重复这些步骤，可以提供表达两种以上 RNA 的细胞或细胞系。第二 DNA 构建体也可以含有相同或不同的药物抗性或其它选择标记。如果第一和第二药物抗性标记相同，则可以通过增加药物水平，进行同时选择。使用形成表达文库的 DNA 构建体，通过以同时或相继的方式重复以上步骤，可以产生多个细胞系，其中至少一部分细胞表达不同的 RNA。

本发明公开了相关的方法，其中使用与所转染的基因连接的标签序列作为信号探针的靶标，该方法的一个用途是允许选择可能难于（例如，当信号探针可检测密切相关的 RNA 种类时）在背景下鉴定出其 RNA 的细胞。由此，提供分离表达至少一种 RNA 的细胞的方法，该方法包括步骤：

- a) 向细胞中引入编码所述至少一种 RNA 和至少一种标签序列的 DNA;
- b) 将所述细胞暴露于在与标签序列杂交时产生可检测信号的至少一种信号探针; 和
- c) 分离产生该信号的所述细胞。

该方法除了所用信号探针被设计用于识别标签序列而非所述 RNA 之外, 与前述方法基本相同。此方法与前述方法相比的一个优点是, 仅需要对应于不同标签序列的数量的少量信号探针, 即可以制备出大量表达一种或多种 RNA 的不同细胞系。任选地, DNA 构建体还可以编码至少一种药物抗性或其它选择标记, 而且该方法还可以包括步骤: 在步骤 a) 后选择对至少一种药物或其它选择性试剂具有抗性的细胞, 其中所述药物或选择性试剂标记是所述标记所赋予的抗性所针对的药物或选择性试剂。分离的细胞可以单独地培养或者合并在一起。无论细胞何时分离, 是否在用一个或多个构建体或一个或多个表达文库进行转染之后, 都可以将分离的细胞彼此分开进行培养或者将之合并在一起。

标签序列指作为待通过信号探针检测的 RNA 的一部分表达的核酸序列。通过设计信号探针使之包括与标签的序列互补的部分, 信号探针可以靶向该标签。可以用于本发明中并可以对它制备信号探针的标签序列的实例包括但不限于表位标签(包括但不限于 HA(流感血凝素蛋白)、myc、his、蛋白质 C、VSV-G、FLAG 或 FLU)的 RNA 转录物。这些和其它标签序列是本领域技术人员已知的, 而且典型地对应于如下氨基酸序列, 该序列可以掺入所表达的蛋白质产物中并常常基于可获得用于报道其存在的强力抗体或蛋白质检测剂而被选用。本文所述标签序列不意在仅仅指可以用于在氨基酸水平上修饰被标签的 RNA 编码的蛋白质产物的序列、或者可以通过使用相应的抗体或蛋白质检测剂而用于帮助对任何此类修饰的蛋白质产物的随后检测的序列。在本文中, 标签序列至少提供被信号探针识别的独特核酸序列。已经描述了使用信号探针检测各种 RNA。这些 RNA 均可以用作标签。构建体中编码标签序列的 DNA 部分可以与 DNA 构建体中编码所述至少一种 RNA 的蛋白质编码区的部分具有相符或不相

符的阅读框。因此，对于用信号探针检测而言，标签序列不必须被翻译。

标签序列可以包含多个重复序列，其中所述序列被设计为信号探针的靶位点。此类标签序列将提供多个信号探针靶位点。由此，可以结合更大量的信号探针。这将增加能够从信号探针所检测的任何一个核酸分子产生的总信号，并由此增加信噪比。

除了探针的靶序列外，标签还可以包含一个或多个如下额外序列(称作“辅助者”或“辅助序列”)，所述额外序列被设计、鉴定或选择来改善对表达该标签的细胞的检测。例如，辅助序列可以具有多种效果，包括但不限于影响标签的折叠、定位、或二级、三级或四级结构，其中这些效果中的任何一个或其组合起着改善或增加细胞中靶序列的检测的作用。辅助序列可以影响标签的折叠或结构，由此使得靶序列可以以更容易被接近以发生探针结合的方式呈现出来。这可以由改变的碱基配对引起，或者可以由辅助者与蛋白质或其它细胞的或引入的成分的结合相互作用而引起。辅助者可以对包含它们的序列的折叠或结构起到稳定作用，或使之更活跃。此外，辅助者对于包含它们的序列在探针结合之前、之时或之后就降解而言可以起到稳定作用，此效果可以直接由序列折叠或结构的改变引起，或者由蛋白质或细胞的或引入的成分与辅助序列的结合引起。此外，辅助者可以通过例如增加转录起始或前进的效率、减少转录的提前终止或增加转录后加工的效率，而增加包含它们的序列的转录。

不论辅助者如何发挥其作用，都可以使用功能性方法鉴定辅助者，并且对于任何给定的标签和相应的探针，可以存在不同的辅助序列。服务于多组标签和相应探针的辅助序列也可以通过功能加以鉴定。

例如可以通过构建包含基因序列和标签序列的表达文库，其中将可变序列插入基因和标签序列之间，来检测可变序列以鉴定何者可以充当辅助者。如果基因具有终止密码子，则可以将可变序列插在终止密码子的下游。此外，也可以将其它的可变序列插在不同的位点。接着，将表达文库引入细胞中，随后通过引入指向靶序列的信号探针，检查细胞。检测显示出与对照相比具有增加的信号的细胞(在此，对照信号是对照

细胞或引入了对照表达构建体（例如包含基因及标签但不包含额外的可变序列的表达构建体）的细胞表现出的信号）。可以例如通过 FACS 分离此细胞，并且可以分离它们所代表的可变序列并如果期望的话，对其进行进一步表征。该方法可以用于检测或鉴定可以充当特定的基因、标签及所用相应信号探针的辅助序列的序列。对于任何包含至少一个相应信号探针的靶序列的序列，可以使用基本上相同的方法寻找辅助序列（即，包含可变序列和靶序列但是不含例如基因的表达文库，或者包含本身含有靶序列的基因和可变序列但不含标签的表达文库，均可以用于检测适宜的辅助者）。可以使分离的细胞生长，并可以产生细胞系。

对于以此方式检测或鉴定的辅助序列，其益处可能在于其对于所使用的序列环境（sequence context）（即，对于特定的标签、基因、靶序列或相应探针）是特异的。更广泛地有益于各种序列环境的更通用的辅助序列可以通过例如利用上述方法来实验确定。例如，可以重复多轮可变序列的选择，其中可以在每一轮分离充当辅助序列的可变序列，然后在随后的轮次中测试。在此情况下，例如，通过构建表达载体进行每一轮测试，其中使用自前一轮分离的序列构建表达构建体，该表达构建体包含与第一轮表达构建体中所用不同的基因或标签或靶序列。给定多样的序列环境，还可以使用所述方法验证辅助序列的通用性。鉴定通用的或万用的辅助序列，可能有助于帮助检测细胞中多种多样的序列。

可变序列的一个来源可以是基因组序列。可以用限制性酶消化基因组序列产生不同大小的片段，可以获得这些片段用于克隆以产生表达文库。

为了最佳地呈现标签序列以实现信号探针的结合，可以选择或设计标签序列使之显示出一定数量的预测的或实验确定的二级结构。照此，标签序列至少包含被至少一个信号探针靶向的序列，标签序列还可以包含并非选择或设计用于直接地与信号探针相互作用的其它序列。可以使用核酸折叠预测算法，根据序列的结构适应性，设计潜在的标签序列。见例如，*Nucleic Acids Res.* 31: 3429-3431 (2003)，特此并入作为参考。对于一个给定序列，核酸折叠预测算法常常预测出多个能量上最有

利的结构。或者，可以分析代表可变核酸序列的序列文库，例如包括但不限于消化的基因组 DNA，以确定或鉴定哪些序列起着帮助信号探针检测标签序列的作用。这是一种鉴定或分离标签序列的功能性方法。这可以通过构建包含至少共同序列（经选择或设计而被信号探针所识别的序列）和至少可变序列的表达文库来实现。可以用此表达文库转染细胞，然后将细胞暴露于信号探针，并且可以通过 FACS 分离阳性最高的细胞。然后可以分离这些细胞所代表的可变序列。例如，可以通过 PCR 技术，之后通过克隆扩增产物，直接地从分离的细胞扩增这些序列。或者，可以裂解分离的细胞以释放对应于在分离的细胞中表达的可变序列的 DNA 构建体，可以从所得物分离和扩增这些构建体或载体。例如，当构建体是质粒载体的情况，可以使用裂解物转化感受态细菌细胞，之后使用细菌宿主分离和扩增该质粒。对于并入了多个重复序列单元的标签序列，由于重复单元和/或并入了该标签序列的分子中存在的其它序列之间的潜在相互作用，每一个单元可能不一定采取相同的结构。对于任何给定的标签序列，可以通过其序列环境影响其结构。

一个实施方案中，标签序列来源于反向 vav RNA (reverse vav RNA)。一个实施方案中，标签序列形成三臂联结结构(图 36)。一个实施方案中，标签序列的长度为 10-100 个、80-100 个、90-100 个、80-120 个、100-2Kb、2Kb-15Kb 核苷酸。一个实施方案中，靶序列是从第一茎环区的茎 3' 侧的全部或部分至第一茎环区和第二茎环区之间的连接处、至第二茎环区的茎 5' 侧的全部或部分的区域(图 36)。

一个实施方案中，茎区包含 8-9 个碱基对，第一茎环区包含 4-6 个碱基对，第二茎环区包含 13-17 个碱基对(图 37)。一个实施方案中，三臂的茎区还包含非互补区。一个实施方案中，茎区和第一茎环区的茎还包含一个错配区，第二茎环区的茎还包含 2-7 个错配或凸起区域。一个实施方案中，茎区之间的连接具有总共 8-12 个核苷酸(图 37)。一个实施方案中，标签序列包含根据图 42 A、B 或 C 的结构或序列。另一实施方案中，标签序列具有从图 42 A、B 或 C 的序列预测的能量更有利结构。

本发明还提供上述包含编码至少一种目的 RNA 和标签序列的至少一

种 DNA 的 DNA 构建体。本发明还提供包含该 DNA 构建体的载体和细胞。

在一个进一步的实施方案中，可以对细胞系进行操作以表达至少一种第二 RNA；所述步骤还包括用编码第二 RNA 和第二标签序列和任选地第二选择标记，例如药物抗性标记的第二 DNA 构建体转染细胞系；任选地，选择转录该第二标记的细胞；将细胞暴露于在与第二标签序列杂交时产生可检测信号的第二信号探针；和分离显示出第一和第二标签序列两者或其中至少一个的信号信号的细胞。在两种 RNA 的情况下，DNA 序列中编码第二标签序列的部分也可以与 DNA 序列中编码第二 RNA 的蛋白质编码区的部分在阅读框上相符或不相符。第二 RNA 可以与第一 RNA 同时或相继地进行转染。如果同时实施该方法，并且两个构建体使用了相同选择标记，例如药物抗性标记，则可以使用更高水平的药物或适当选择性试剂来选择表达两个构建体的细胞。此外，可以通过重复上述步骤，在细胞系中提供两种以上的 RNA。

使用形成表达文库的 DNA 构建体，以同时或相继的方式重复上述步骤，可以产生多个细胞系。一个实施方案中，表达文库使用单一一种标签序列，将细胞暴露于与该标签序列互补的相同信号探针。另一实施方案中，不同的表达文库可以使用不同的标签序列。一个实施方案中，每一个细胞系表达一种 RNA。另一实施方案中，每一个细胞系表达一种以上的 RNA。这可以通过用多种 DNA 构建体同时地或相继地转染细胞来实现。通过引入较高浓度的用于转染的 DNA 构建体，可以增加获得稳定转染了多种 DNA 构建体的细胞的可能性。或者，为了获得具有多种 DNA 构建体的细胞，起始细胞可以已经被第一 DNA 构建体转染。此外，也可以使用多个不同的表达文库转染细胞。每个文库可以并入不同的标签序列，此标签序列使用相应的信号探针进行检测。每个表达文库都可以包含选择标记，例如药物抗性基因。

分离的细胞可以单独地培养或进行合并。通过培养单独分离的细胞或合并的细胞，可以产生细胞群体。可以通过培养单独分离的细胞，产生细胞系。可以将单细胞系或多个细胞系分开培养或将其合并。如果细胞系的合并物产生期望的活性，则可以进一步对其进行分级分离直到鉴

定出具有该效果的一个细胞系或一组细胞系。这可以使维持大量的细胞系变得更为容易，而无需分开单独维持各个细胞系。

本发明提供产生过表达 RNA 的细胞系的另一方法，包括步骤：

a) 向细胞中引入编码所述 RNA 和第一标签序列的第一 DNA；和至少一个编码所述 RNA 和第二标签序列的第二 DNA，其中第一和第二标签序列是不同的；

b) 将所述细胞暴露于至少一种在与所述第一标签序列杂交时产生可检测信号的信号探针和至少一种在与所述第二标签序列杂交时产生可检测信号的信号探针；和

c) 分离显示出至少一种所述信号探针的信号细胞。

该方法还可以包括步骤：使分离的细胞生长以产生表达或过表达所述 RNA 的细胞系。如果 DNA 构建体整合在转染细胞的基因组的不同位置，则可以产生多个细胞系。除非使所转染的构建体的基因组整合定向于基因组的特定位置，否则整合被认为将随机地发生，这样每一个阳性细胞都可能彼此不同，由此将存在多个不同的细胞系，对于选择这些细胞系所基于的 RNA 而言，所有这些细胞系都是阳性的。任选地，DNA 构建体还可以编码至少一种选择标记，例如药物抗性标记，并且该方法可以进一步包括步骤：在步骤 a) 之后使用选择标记选择细胞，例如选择对至少一种如下药物的抗性，其中所述标记赋予针对此药物的抗性。无论何时分离细胞，是否在使用一个或多个构建体或一个或多个表达文库进行转染后，都可以将分离的细胞彼此分开培养，或将其合并。

一个实施方案中，细胞表达反义或 siRNA 或 shRNA 或蛋白质。此外，可以使用经过操作而表达特定的（一种或多种）蛋白质的细胞作为构建表达蛋白质和反义 RNA 分子的细胞的起点。当然，可以使用表达反义或 siRNA 或 shRNA 分子的细胞作为使用本文方法添加编码其它蛋白质的其它 RNA 的起点。也可以同时转染编码蛋白质的 RNA 和反义或 siRNA 或 shRNA 分子、以及相应的信号探针和，如果期望的话，标签序列。本文包括上述方法的各种组合。

同样地，本发明还提供分离表达至少一种 RNA 的细胞的方法，包括

步骤:

- a) 提供被怀疑表达所述至少一种 RNA 的细胞;
- b) 将所述细胞暴露于在与所述至少一种 RNA 杂交时产生可检测信号的至少一种信号探针;
- c) 分离产生该信号的所述细胞。

使用在与其靶 RNA 杂交时产生可检测信号的第二信号探针, 该方法也可以用于鉴定还表达第二 RNA 的细胞, 分离具有第一和第二信号探针之一或两者的荧光的细胞。同时表达两种以上的 RNA 也是可以实现的。无论何时分离细胞, 是否在使用一个或多个构建体或一个或多个表达文库进行转染后, 都可以将分离的细胞彼此分开培养, 或者将之合并。

本发明还提供分离表达至少一种外源 RNA 和一种内源 RNA 的细胞的方法, 包括步骤:

- a) 向细胞中引入编码所述至少一种外源 RNA 的 DNA, 其中所述细胞潜在地表达至少一种内源 RNA;
- b) 将所述细胞暴露于在与所述至少一种外源 RNA 杂交时产生可检测信号的至少一种第一信号探针;
- c) 将所述细胞暴露于在与所述至少一种内源 RNA 杂交时产生可检测信号的至少一种第二信号探针, 其中所述第二信号探针与第一信号探针产生不同的信号; 和
- d) 分离在所述信号探针与它们相应的 RNA 杂交时产生至少一种所述信号的细胞。

以上两种方法还可以包括步骤: 通过使所述分离的细胞生长, 产生表达所述至少一种外源 RNA 和至少一种内源 RNA 的细胞系或多个细胞系。

这些方法可用于表达蛋白质, 例如细胞表面定位的蛋白质、细胞内蛋白质、分泌性蛋白质或其它蛋白质的 RNA。这些方法无需使用针对蛋白质本身的探针, 利用探针检测蛋白质本身可能更为困难或将影响细胞使得无法实施进一步的实验。使用多种信号探针(其种类的数量可多达分离所用技术可同时检测的数量), 可以鉴定一种以上的编码蛋白质的 RNA。任选地, DNA 构建体还可以编码至少一种选择标记, 例如药物抗性

标记,该方法可以进一步包括步骤:在步骤 a) 后使用选择标记选择细胞,例如选择对至少一种药物具有抗性的细胞,其中所述药物为所述标记赋予的抗性所针对的药物。分离的细胞可以单独地进行培养或合并在一起。无论何时分离细胞,是否在使用一个或多个构建体或一个或多个表达文库进行转染之后,都可以将分离的细胞彼此分开培养或者将其合并。

对于以上方法,一个实施方案中,细胞是可植入动物体中的。一个实施方案中,信号探针是产荧光的探针,表达所述 RNA 的细胞发荧光。分离发荧光的所述细胞可以使用荧光激活细胞分选技术或可以用于基于荧光分离细胞的任何技术来实现。一个实施方案中,使用两种信号探针靶向 RNA 或标签序列。第一探针的荧光团可以与第二探针的荧光团相同或不同。在两种不同荧光团的情况下,它们可以具有相似或不同的发射波长。目前,FACS 技术可以允许在分选过程中检测最多 7 种不同的荧光团。可以同时重复以上方法以获得最多 7 种不同蛋白质的表达。如果期望的话,然后可以使用表达 7 种不同蛋白质的细胞系作为根据该方法引入更多蛋白质的起点。随着 FACS 技术的进步,将能够解析更大数量的信号,由此可以在一次 FACS 应用中选择具有更多种 RNA 的细胞。

自然地,可以使用上述方法定量生物学样品中至少一种 RNA 转录物的表达水平,包括步骤:

- a) 将生物学样品暴露于在与所述 RNA 转录物杂交时产生可检测信号的第一信号探针;
- b) 定量生物学样品中该信号的水平; 和
- c) 将该信号水平与所述至少一种 RNA 转录物的水平相关联。

生物学样品可以是细胞样品、组织样品或来源于其的制备物; 这些样品可以是冷冻的和/或固定的,例如用甲醛、戊二醛,或任何数量的对使用信号探针检测 RNA 不产生干扰的已知细胞固定剂固定。细胞样品制备物包括但不限于亚细胞细胞器或区室、线粒体、细胞的细胞器、亚细胞部分、细胞的质膜、细胞内或细胞外物质、膜制备物、从这些中的任何一种或多种中得到的核酸制备物、存在于这些中的任何一种或多种中的任何病毒或其它病毒或生物体的制备物,或这些中的一种或多种的任

何组合。

对于产荧光探针的实施方案，荧光可以通过荧光显微镜或荧光激活的细胞分选技术进行定量。使用在与第二 RNA 转录物杂交时发荧光的第二信号探针，同时可以测定其它 RNA 种类的数量（即，定量）。以上方法可以与利用产荧光报道分子检测细胞内事件、状态或组成的试验同时使用。此类荧光试验包括但不限于 TUNEL、凋亡、坏死、Ca<sup>2+</sup>/离子流、pH 流、免疫荧光、细胞器标记、细胞粘着、细胞周期、DNA 含量、以及用于检测蛋白质与蛋白质、蛋白质与 DNA、蛋白质与 RNA 的相互作用的试验。可以被荧光标记从而用于这些试验的试剂包括但不限于蛋白质（用荧光分子标记的或自身发荧光的蛋白质）；荧光代谢指示剂（C12 刃天青、CFES 用于细胞分裂）；荧光底物或副产物；被荧光标记的凝集素；荧光化学药品；笼状 (caged) 荧光化合物；荧光核酸染料；荧光聚合物、脂质、氨基酸残基和核苷酸/ side 类似物。

将反义或 siRNA 分子引入细胞中对于从该细胞中功能性消除或降低一种或多种蛋白质或 RNA 的水平是有用的。遵循以上方法，提供分离细胞或产生细胞的方法，其中所述细胞功能性地不表达至少一种预定的蛋白质或 RNA 或者其表达降低，所述方法包括步骤：在所述细胞中提供多种针对所述预定蛋白质或 RNA 的反义或 siRNA，这些反义或 siRNA 每一种均根据前述方法提供，其中所述多种反义或 siRNA 与所述至少一种预定蛋白质或 RNA 的基本上全部或足够水平的 mRNA 转录物结合。预定蛋白质可以是基因产物的可变剪接形式。

遵循相似的方式，提供产生转基因动物的方法，所述转基因动物是至少一种预定蛋白质或 RNA 的功能性地不表达的突变体，或者所述转基因动物以降低的水平表达所述至少一种预定蛋白质或 RNA，所述方法包括使用胚胎干细胞实施上文所述的步骤，和使用所述活的胚胎干细胞产生所述转基因动物。

同样地，还提供分离细胞或产生细胞系的方法，其中所述细胞或细胞系功能性地不表达或者降低地表达至少一种蛋白质或 RNA 并过表达至少一种其它的蛋白质或 RNA，所述方法包括在相同细胞上实施本文的方

法。以相似的方式，可以提供产生如下细胞系的方法，所述细胞系表达在诱导型启动子或者具有某种可以导致转录水平调节的其它结合或功能活性的序列控制下的致死性反义或 siRNA。这可以通过实施此处的方法而实现，其中转染步骤在存在最小量的诱导剂或化合物的情况下进行。

因此，本发明提供分离细胞的方法，在所述细胞中至少一种第一蛋白质过表达而至少一种第二蛋白质功能性地不表达或者其表达降低，所述方法包括步骤：

a) 向细胞中引入编码至少一种 RNA 和至少一种第一标签序列的至少一种第一 DNA，其中所述至少一种 RNA 编码所述至少一种第一蛋白质；并引入编码所述至少一种 RNA 和至少一种第二标签序列的至少一种第二 DNA 构建体，其中所述第一和第二标签序列不同；

b) 向细胞中引入编码至少一种反义 RNA 或 siRNA 的至少一种 DNA，其中所述至少一种反义 RNA 或 siRNA 可结合或干扰所述至少一种第二蛋白质的 mRNA 转录物；

c) 将所述细胞暴露于在与所述至少一种第一标签序列杂交时产生可检测信号的至少一种第一信号探针，和在与所述至少一种第二标签序列杂交时产生可检测信号的至少一种第二信号探针；

d) 将所述细胞暴露于在与所述至少一种反义 RNA 或 siRNA 杂交时产生可检测信号的至少一种信号探针；和

e) 分离在所述信号探针与它们的相应 RNA 杂交时产生至少一种所述信号的细胞。

再一实施方案中，本发明提供鉴定化合物的方法，所述化合物调节至少一种预定 RNA 的转录，所述方法包括步骤：

a) 向细胞添加单个化合物或一组化合物；

b) 将所述细胞暴露于在与所述至少一种预定 RNA 杂交时产生可检测信号的至少一种信号探针；

c) 定量所述细胞中该信号的水平；

d) 鉴定与未加化合物的细胞的信号相比具有增加或降低的信号的细胞；和任选地

e) 鉴定调节所述至少一种预定 RNA 的转录的化合物。

一个实施方案中, 所述预定 RNA 由细胞基因组编码。一个实施方案中, 所述预定 RNA 由步骤 a) 之前转染至细胞中的 DNA 构建体编码。一个实施方案中, 将转染细胞暴露于设计用于识别所述 RNA 的信号探针, 并且该细胞表达所述 RNA。一个实施方案中, DNA 构建体包含启动子或操纵基因并编码阻遏蛋白、增强子或调节转录的序列。一个实施方案中, 预定的 RNA 与标签序列连接, 信号探针在与标签序列杂交时产生可检测信号。

另一实施方案中, 本发明提供鉴定调节至少一种预定 RNA 的转录的 RNA 序列的方法, 包括步骤:

a) 向细胞中引入至少一个编码待测 RNA 序列的构建体, 其中该待测 RNA 潜在地调节所述至少一种预定 RNA 的转录;

b) 将所述细胞暴露于在与所述至少一种预定 RNA 杂交时产生可检测信号的至少一种信号探针;

c) 定量所述细胞中该信号的水平;

d) 选择与不具有待测 RNA 序列的细胞的信号相比具有增加或降低的信号; 和任选地

e) 鉴定调节所述至少一种预定 RNA 的转录的待测 RNA 序列。

可以分离和培养这些细胞以产生细胞系。这些细胞可以分开培养或合并在一起。转录的调节可以是对下游预定 RNA 的上调或下调。一个实施方案中, 所述预定 RNA 由基因组编码。所述预定 RNA 由在步骤 a) 之前转染至细胞中的 DNA 构建体编码。一个实施方案中, 将转染的细胞暴露于设计用于识别所述 RNA 的信号探针, 并且该细胞表达所述 RNA。一个实施方案中, 预定 RNA 与标签序列连接, 信号探针在与标签序列杂交时产生可检测信号。一个实施方案中, 使用 RNA 序列表达文库鉴定调节转录的 RNA 序列。一个实施方案中, 待测 RNA 序列与标签序列连接。步骤 e) 可通过在步骤 a) 之后将细胞暴露于在与所述 RNA 序列或标签序列杂交时产生可检测信号的至少一种信号探针, 并之后进行步骤 b) 来获得便利。由指向待测 RNA 序列或其标签序列的信号探针产生的信号可以与指向预

定 RNA 的信号探针产生的信号不同，因此可以将细胞同时暴露于不同的信号探针。

本文中还提供鉴定活细胞中的遗传重组事件的方法，包括步骤：

a) 将细胞暴露于在与 RNA 序列杂交时产生可检测信号的信号探针，其中所述 RNA 选自：从重组后的序列转录的 RNA 和从未重组的序列转录的 RNA；和

b) 检测表达所述 RNA 序列的所述细胞。

可以通过 FACS 或荧光显微镜检测和/或分选细胞。

通过参考以下作为本发明示例提供的非限制性实施例，可以更好地理解本发明。提供以下实施例的目的在于更充分地说明本发明的优选实施方案。它们不应当以任何方式理解为限制本发明的宽范围。与本文所述方法和材料相似或等价的方法和材料也可以用于实施或检测本发明。

17. 鉴定可以将信号探针引入细胞或可以提高信号探针向细胞中的引入或可以增强靶的检测的试剂或化合物

可以通过测试各种化学药品，包括但不限于蛋白质、脂质、聚合物、提取物或化合物或它们的混合物，鉴定可以用于将信号探针引入细胞中的试剂。该化学药品可以是气态、液态或固态。所述鉴定可以通过将化学药品与信号探针以从 1:1,000,000,000 至 1,000,000,000:1 的各种比例在各种不同溶剂（包括但不限于有机和水性溶剂、缓冲溶液或介质或这些的任何组合）中混和来实现，其中混合物被温育不同的时间长度（从 1 分钟至 48 小时），并在从低于 0 摄氏度至高于 100 摄氏度的各种温度条件下并在从无至轻柔摇动或摆动至涡旋或持续抽吸的各种程度的搅动或混和下实施该温育，而且在亮处或暗处或在各种其它环境条件下实施该温育。接着，将混合物施用于细胞，使用荧光显微镜、荧光平板读数器技术或 FACS 或其它荧光检测方法，分析探针的递送。这可以例如在 96 孔、384 孔或 1536 孔板上或在与之之后的高通量分析相容的珠子上进行。注意，当使用珠子时，可以使用 FACS 分析或分离这些珠子，并且如果珠子还被编号的话，可以使用分离的珠子确定用于处理它的混合物的身份。

所用的细胞可以是活的或固定的，并可以以结合在固相表面（包括但不限于组织培养板或珠）上的方式提供，或者它们可以处于悬浮状态。在添加混合物前可以使用各种缓冲液或溶液洗涤细胞。一旦向细胞施用了试剂后，可以以各种时间长度在各种程度的搅动或混和下在亮处或暗处或在各种其它环境条件下以各种温度（所有均在上面进行了描述）温育反应物，其中如果例如维持细胞的生存力的话，可以对这些参数加以限制。在温育后，可以直接分析细胞或在分析前如上所述进行洗涤。

以上方法可以使用信号探针或具有组成型活性的探针（例如在产荧光的探针的情况下，可以使用缺乏淬灭剂的探针）实施。此外，待测的每个混合物都可以使用一种以上的细胞样品检查。例如，可以使用已知包含所用信号探针的靶的细胞和对照细胞，检查混合物。这多种细胞可以混和在一起或分开给予。优选的混合物将是导致增加的信噪比的混合物，例如与对照细胞相比在包含靶的细胞中导致增加的信号的混合物。

类似地，除了向细胞添加混合物之外，可以实施基本上相同的方法。例如，可以首先将待测混合物施加于测试室，例如 96 孔、384 孔或 1536 孔板的孔或珠，接着可以施加细胞。细胞可以直接地加入板中的混合物中，或者可以在通过例如干燥、加热或蒸发加工混合物后加入细胞。

以相关的方式，可以检查化合物增强信号探针对细胞中的靶的检测的能力。在此情况下，首先将信号探针引入细胞中，其中所述细胞可以包括已知包含靶的细胞和对照细胞。接着，使用上述可变条件和参数，可以分析上述化学药品提高信号的能力。注意，在此情况下使用的化学药品不与信号探针混和。优选的化学药品是经检测或鉴定可以增加信噪比的化学药品，例如与对照细胞相比在包含靶的细胞中导致增加的信号的化学药品。使用此方法鉴定的化学药品可以通过多种机制起作用，包括但不限于增加信号探针向细胞质中的递送、影响靶的折叠或结构从而提高探针对其的检测（例如通过提高靶的可接近性）、或者降低非特异的背景信号（例如，通过降低与细胞外表面结合的信号探针的量）。

具有增强信噪比的作用、可以用于将信号探针引入细胞中或者可以用于增强靶的检测的化学药品，也可以通过使用待测信号探针和对照信

号探针来确定，其中将待测信号探针和对照信号探针都施用于已知包含靶的细胞并分析与对照信号探针相比化学药品是否增加使用待测信号探针时的信号。

### 实施例 1

一般方案。原材料：可以将信号探针引入不表达来自 DNA 构建体的任何 RNA 的细胞中，或者可以使用信号探针检测 DNA 构建体编码的 RNA 信使。将信号探针引入这两类细胞中的方法是相同的。以下方案仅需要待被分析的细胞可以彼此分离并可接受 FACS 分析。

1) 如本发明详述中更彻底描述的，可以使用信号探针与 FACS 结合，基于细胞中特定 RNA 的表达或缺乏表达，分选组织的细胞。为此，首先通过标准且成熟建立的方法例如匀浆以及进一步的化学处理，将细胞彼此分离。然后，可以根据以下操作方案，将适宜的信号探针引入该细胞中。

2) 其次，可以使用信号探针选择表达已经转染进入细胞群体中的 DNA 构建体所编码的特定 RNA 的细胞。为此，首先用编码期望 RNA 的一个或多个 DNA 构建体转染细胞培养物。然后，如本发明详述中更详细地描述的，可以制备信号探针以识别这些 RNA。DNA 构建体转染细胞可以通过多种多样的方法实现，所述方法包括但不限于使用自制试剂或获自生物技术公司 (Qiagen, Promega, Gene Therapy Systems, Invitrogen, Stratagen 等) 的试剂盒按照厂商的说明书进行。选择 DNA 构建体以便每一个构建体都可以赋予对抗生素的抗性。在将这些 DNA 构建体转染至细胞中并让细胞短期恢复 (通常 24 小时) 后，向细胞施用适宜的抗生素，以便仅有被 DNA 构建体赋予了抗生素抗性的细胞存活。取决于细胞的类型以及所用的抗生素，这一般需要 3 至 4 天，有时更长。

结果是，一部分细胞保留下来，并且这些细胞全部都对抗生素具有抗性，但仅有其中一小部分细胞表达目的 RNA。为了选择表达期望 RNA 的细胞，可以实施如下的方案。

## 实施例 2: 使用信号探针选择细胞

1) 将信号探针转染至细胞中: 必须设计信号探针使之通过与 RNA 中的内源序列杂交或通过添加至天然 RNA 序列上的标签杂交, 而识别期望的 RNA。信号探针的设计在本发明详述中已经作了详细阐述。

可以通过类似于将 DNA 构建体转染至细胞中的各种方法, 实施转染。由于一些细胞对一些转染方法比对其它方法有更好的反应, 故应当基于使用的细胞类型选择所用的方法。转染应当根据所用转染试剂的制造商的说明书来实施, 并且可能需要进行优化。优化可能包括用化学药品处理以增加待转染的物质向细胞或细胞质中的递送。

取决于所用的细胞和转染试剂, 可以在悬浮细胞上或在生长于固相表面的细胞上实施信号探针向细胞中的转染。

2) 在产荧光探针转染进入细胞中后, 可以对细胞进行 FACS 分析。FACS 可以用于分选就所用的任何一种或多种产荧光探针而言为阳性的细胞。也可以使用 FACS, 基于产荧光探针的信号强度来分选细胞, 由此允许研究者选择以不同水平表达 RNA 的细胞。

## 实施例 3

产生表达一种或多种 RNA 的细胞系

在 FACS 选择后, 可以将阳性评分的细胞维持在适宜的培养基中 (见本发明详述中的更详细描述)。这些细胞将产生表达目的 RNA 的细胞系。

信号探针的浓度: 所用信号探针的浓度取决于几个因素。例如, 必须考虑细胞中待检 RNA 的丰度以及信号探针对此 RNA 的可接近性。例如, 如果待检 RNA 以极低量存在或者其存在于不易接近的 RNA (由于该 RNA 的三维折叠或由于蛋白质与该 RNA 的结合) 部分中, 则在此应当比待检 RNA 大量存在以及信号探针识别的位点可以被充分接近的情况使用更多的信号探针。使用的信号探针的确切量必须针对每一应用经验性地确定。

这可以通过将不同量的信号探针引入不同的细胞组中并选择背景荧光低且信号高的条件 (不是全部而是一些细胞被评定为信号探针阳性的条件)。

#### 实施例 4: 将细胞暴露于信号探针

可以使用各种方法, 将附着于或部分附着于或停留于表面或存在于溶液中的细胞暴露于 FP, 以便将分子引入细胞中, 所述方法包括但不限于本领域已知的方法, 例如显微注射、机械剪切力如涡旋或混和、穿过针、或细胞装载技术包括刮擦、使用诸如某些抗生素等试剂或去污剂或者试剂或溶剂的组合进行透化处理、或者使用各种不同化学性质的(例如基于脂质体的、化学的或基于蛋白质的)转染试剂、或者组合使用这些方法中的任一种或多种。

以下给出了一个描述使用基于脂质的转染试剂的更详细样品操作方案:

##### 1. 细胞的制备

在将细胞暴露于 FP 之前(细胞接种方法 1)或同一天(细胞接种方法 2)将细胞接种在组织培养孔中, 或者在细胞暴露于 FP 的同一天将细胞转移至微量离心管(细胞接种方法 3)。

所有三种制备方法都成功地获得了应用。细胞接种方法 1 允许细胞有充足的时间附着在培养孔的表面上; 而细胞接种方法 2 允许对停留在板上或以不同程度附着在板上的细胞进行加工, 这取决于在对细胞作进一步加工之前允许经过的时间长度; 细胞接种方法 3 允许细胞在被转移至管子中后立即被处理从而使得细胞没有任何时间停留或附着在任何表面上, 但是也可以在经过给定的一段时间后实施处理。

使用缓冲液例如无血清培养基或 PBS (但也可以使用其它缓冲液), 洗涤细胞一次或多次。一般地, 缓冲液包括不同毫摩尔浓度的  $MgCl_2$ 。根据暴露的方法, 可以省去此洗涤步骤。

##### 2. 制备 FP 试剂和细胞的暴露

使用商业可获得的转染试剂和厂商描述的操作方案, 制备向细胞制备物添加的 FP。厂商的操作方案教导多个参数需要针对多个变量, 包括所用的细胞类型、应当在何种汇合度或浓度处理细胞、引入何种特定分子、以何种比例和绝对量组合各种试剂、以及各步骤的实施或温育持续

时间、可以省去哪些步骤、及其它变量，进行经验性地确定。在厂商操作方案中，尽管提供某些范围，但方案也提示为了成功使用它们的试剂，这些范围可能被超越或者其它参数可能必须进行优化。一般地，使用了位于厂商方案中提出的初始范围内的这些条件的参数，将细胞暴露于 FP。

一般地，将 FP 加入含有无血清培养基的管中，并使用厂商提议的体积和浓度将转染试剂加入也含有无血清培养的第二管中。都按照厂商的操作方案，将各管中的内容物混和、合并和温育一定长度的时间。接着，将 FP 施用于细胞制备物，在各种温育期之后分析这些细胞。

如下是图 32 中将细胞暴露于 FP 的操作方案：

在细胞暴露于 FP 的前一天，按大约  $1 \times 10^5$  个细胞/ml 以及每孔 0.5ml，将细胞接种在 24 孔板中。所用细胞是使用编码 r-vav 的表达构建体转染并经过药物选择的细胞。

通过在一个试管中在  $50 \mu\text{l}$  至  $200 \mu\text{l}$  含有 1 至 4mM  $\text{MgCl}_2$  的无血清培养基中温育 0.625 至  $2.5 \mu\text{l}$  20  $\mu\text{M}$  FP 原液，以及在第二管中具有相同浓度  $\text{MgCl}_2$  的等体积无血清培养基中温育 0.625 至  $2.5 \mu\text{l}$  Tfx50 (Promega)，制备 FP 试剂。混和各管中的内容物，然后合并并在室温温育 15 至 45 分钟。

用补充了以上所用相同浓度的  $\text{MgCl}_2$  的无血清培养基洗涤细胞一次或多次。将 FP 制备物施用于细胞，任选地可以根据所加 FP 制备物的体积，加入额外的无血清培养基加  $\text{MgCl}_2$ ，以覆盖细胞。

在组织培养箱中孵育细胞 3 至 5 小时，然后进行分析，任选地可以在观察前加入 DMSO。DMSO 可以帮助信号探针递送进入细胞中或帮助待检靶标的折叠或呈现。为了增加信号探针向细胞或细胞质的递送，或者为了提高信号探针对靶的检测，也可以使用其它溶剂或化学药品。可以通过测定溶剂和化学药品是否在表达靶的细胞中比在不表达靶的细胞中（在此，两种细胞都暴露于信号探针）导致增加的递送效率或导致增加的靶检测信噪比，而检查溶剂和化学药品的合意性或适宜性。

除了向细胞（其中所述细胞被转染或未被转染并通过药物针对编码含有靶序列 6CA4 的 RNA 的表达构建体进行了选择）添加 F16 外，图 34

的方案基本如针对图 32 所述的一样实施。

在图 33(A、B、C)中将细胞暴露于 FP 的方案如下所述：

将以大约  $2.5 \times 10^5$  个细胞/ml 接种在补加 4mM  $MgCl_2$  的 PBS (PBS+4) 中的大约 50 至 100  $\mu$ l 细胞，接种在 96 孔板的孔中，在经过充足时间以允许细胞停落在板表面上之后但在细胞散布之前，将细胞暴露于 FP。

通过在一个管中在 100  $\mu$ l PBS+4 中混和 2.5  $\mu$ l 20  $\mu$ M FP 原液，制备 FP 试剂，在第二管中向 100  $\mu$ l PBS+4 中加入 7  $\mu$ l Lipofectamine (Invitrogen)。混和各管的内容物，并室温温育溶液 30 分钟，之后将之合并、混和并室温再温育 15 分钟。向各孔中分别加入 50  $\mu$ l FP 制备物。任选地用无血清培养基或其它缓冲液洗涤细胞。在组织培养箱中温育细胞 2 至 3 小时后分析细胞。

除了向含有 FP 的管子中加入 4  $\mu$ l Plus Reagent (Invitrogen)外，图 33(D、E)中将细胞暴露于 FP 的方案与图 32 的基本相同。

#### 实施例 5: 通过 FACS 分析和分离细胞

如上所述将细胞暴露于 FP 后，处理细胞以便如果细胞贴壁的话使之脱离表面并使之彼此分离(例如使用胰蛋白酶，但也可以使用其它酶的或非酶的方法)，然后将细胞用于 FACS。

在暴露于 FP 后，从细胞中除去含有 FP 的溶液，并将胰蛋白酶直接施加于细胞上。在此步骤前可以使用各种缓冲液或试剂洗涤细胞，包括设计用于除去任何 FP 试剂或可能在细胞暴露于 FP 的过程中与细胞表面结合的其它试剂的试剂。将细胞重悬在缓冲液(例如含有血清和镁的培养基)中，通过例如吸管抽吸进一步分散细胞。

然后根据标准 FACS 方法用 FACS 分析细胞。对于分析细胞，通过使用 FP 检查靶序列的存在从而比较对照细胞和潜在表达靶序列的细胞的荧光强度，可以确定经过此过程的细胞的背景荧光，由此确定潜在表达靶序列的细胞是否显示出任何改变的荧光。可以基于此信息，将细胞单个地或成批地分离。例如，现有技术允许将期望的细胞直接分离至 96 孔的独特孔中，或者可以获得多个期望的细胞。

通过在 FACS 中增加或降低荧光强度的阈值，可以分离显示出不同荧光信号水平的细胞。可能希望将阈值设定得非常高以高度确保细胞确实地为阳性，或者如果在获得群体中阳性细胞的纯度不是关键的，例如在简单地希望从分离的细胞中富集阳性细胞的情况下，可以设定较低的阈值。

#### 实施例 6: 确定表达的稳定性

可以监测一种或多种基因在细胞中随时间的表达水平。例如，假定一个细胞系，该细胞系来源于使用 FP 作为序列表达的阳性细胞（其中所述 FP 被设计以识别该序列）而被分离的细胞，则可以将该细胞和对照细胞暴露于 FP 并测定两种细胞的荧光强度的比例以确定由 FP 释放的信号。可以随着时间的推进重复该过程，比例的改变将反映出 FP 所检测的序列的相对表达水平的改变。该方法可以使用多种 FP 进行。

#### 实施例 7: 设计标签序列

为了产生具有相同或相似的预测结构（尤其是就最直接地参与信号探针结合的区域而言）的序列以便信号探针的靶序列可以对于结合而言以相似的方式呈现，使用标签 1 序列产生两个额外的不同序列，其中每一个序列都可以被独特的信号探针识别。对产生的两个不同标签序列（标签 2 和标签 3）之每一个，首先改变标签 1 的最初序列中的被靶向的序列，对经过预测与该被改变的序列发生相互作用的一些其它的碱基进行补偿性改变，以便在这些同样的位置保留相互作用（图 43 和 44）。获得新序列的预测结构，如果该结构与标签 1 的预测结构不紧密地相匹配，则使用该结构预测哪些其它改变将是必须的。重复实施该过程直到获得与标签 1 具有相似的预测结构的新序列。对这些序列的改变包括碱基替代、缺失和添加。

#### 实施例 8: 在升高的温度下暴露信号探针

预测 FP17 在相互作用对附近形成 7 个碱基对的相互互补区。对于

FP17, 为了使之在靶寡聚物存在下温育时比在对照寡聚物存在下温育时产生更强的荧光信号, 需要升高的温度。例如, 短暂地将它们放入 90 至 95°C 的热水中。在此使用的管子中含有总共 16  $\mu$ l, 由 5  $\mu$ l 20  $\mu$ M FP 原液、1.5  $\mu$ l 25mM  $MgCl_2$ 、8  $\mu$ l 20  $\mu$ M 寡聚物和 1.5  $\mu$ l 水组成, 具有大约 2.34mM 的最终镁浓度。如上述, 所用靶寡聚物是 T0-FP1, 所用对照寡聚物是 T0-FP18。

#### 实施例 9: 化学修饰的信号探针

基于 FP1 探针序列, 合成了 15 种不同的化学修饰的探针。这些探针均具有相同的序列, 指向相同的靶序列。将探针引入 293T 细胞中。这些细胞表达包括 FP1 的靶序列的标签序列。使用 FACS 分析来自这些细胞的荧光。这 15 种不同探针如下所述。FACS 分析结构显示在图 46-60 中。结果显示, 所有 15 种探针均能够检测表达靶序列的细胞。所有这些探针除了化学修饰外在浓度、递送方式、荧光团、淬灭剂和序列上是相同的。

探针	修饰	序列	注释
Mcon 1	2'-氨基-dA 和 5'-甲基-dC	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	C=5-甲基 dC A=2-氨基 dA
Mcon 2	2'-5' 连接的寡核苷酸	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	粗体=2', -5' 连接
Mcon 3	胞嘧啶 Arabanoside (Ara-C)	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	粗体=Ara-C
Mcon 4	间隔区亚磷酸酰胺 9	GCCAGSTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	S=间隔区亚磷酸酰胺 9
Mcon 5	2' -脱氧-2' -氟-RNA (2' -F-RNA)	GCCAG <u>UcccAGUuuccuGuGccuuAAGAAcCTCGC</u>	下划线=硫代磷酸酯连接 小写=2' -F-RNA
Mcon 6	2' -脱氧-2' -氟-RNA (2' -F-RNA)	GCCAG <u>UCCCAGU</u> TcGIGcGcIuAAGAA <u>cCTCGC</u>	下划线=硫代磷酸酯连接 小写=2' -F-RNA
Mcon 7	2-氨基-A	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	A=2-氨基-A
Mcon 8	2' -0-甲基-5-甲基-C (2' -OMe-5-Me-C)	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	C = 2'-OMe-5-Me-C
Mcon 9	闭锁核酸 (locked nucleic acid, LNA)	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	粗体=LNA
Mcon 10	闭锁核酸 (LNA)	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	粗体=LNA
Mcon 11	硫代磷酸酯连接	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	下划线=硫代磷酸酯连接
Mcon 12	硫代磷酸酯连接	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	下划线=硫代磷酸酯连接
Mcon 13	2' -0-甲基-RNA (2' -OMe-RNA)	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	粗斜体=2-0-甲基-RNA
Mcon 14	2' -0-甲基-RNA (2' -OMe-RNA)	GCCAGTCCCAGTTCCTGTGCCTTAAGAA <u>CCCTCGC</u>	粗斜体=2-0-甲基-RNA
Mcon 15	C-5 丙炔基胞嘧啶类似物	GCCAG <u>UcccAGUuuccuGuGccuuAAGAAcCTCGC</u>	小写=C5-丙炔类似物

注解: 所有探针具有 5' Cy5 和 3' BHQ-3

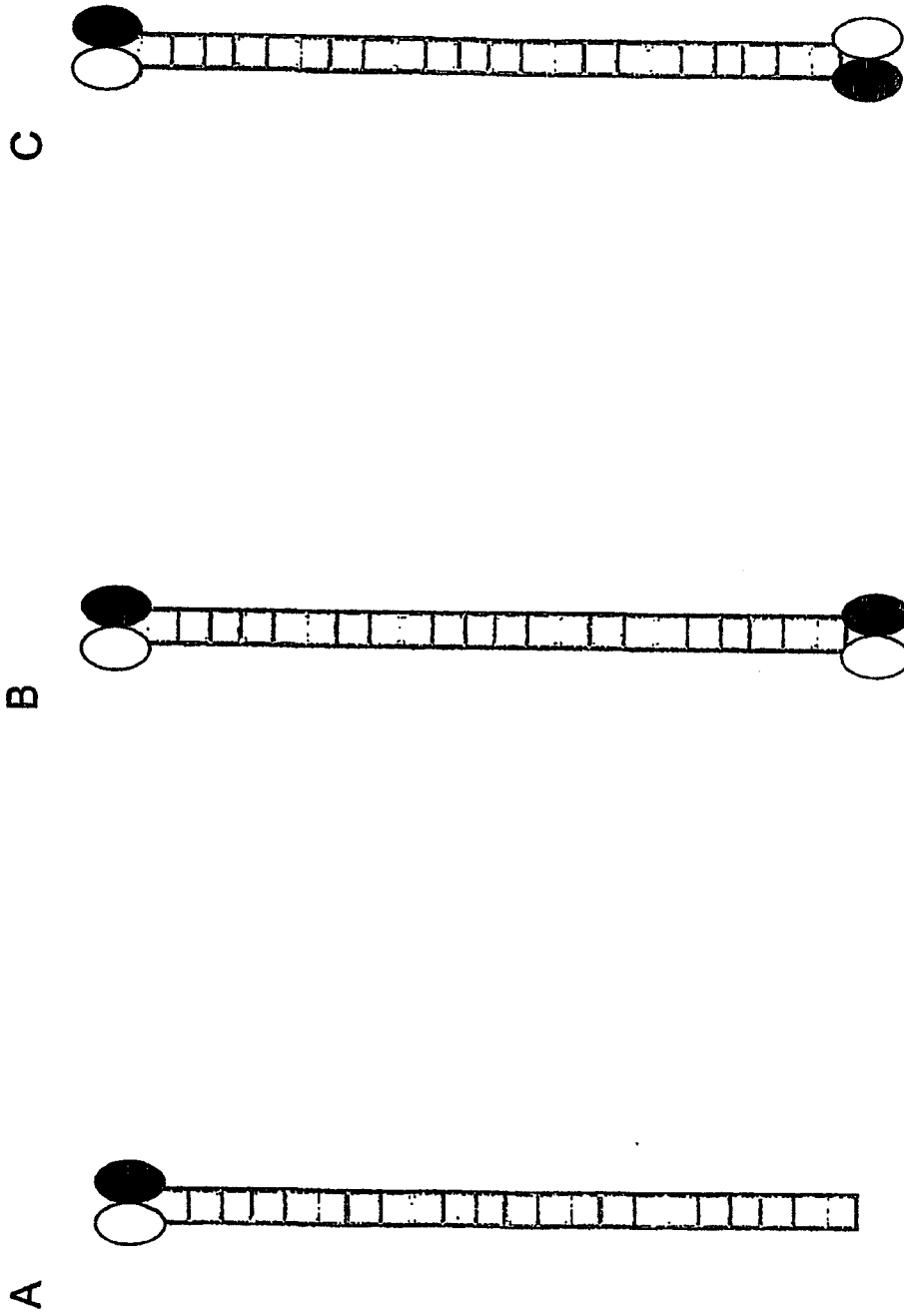


图1

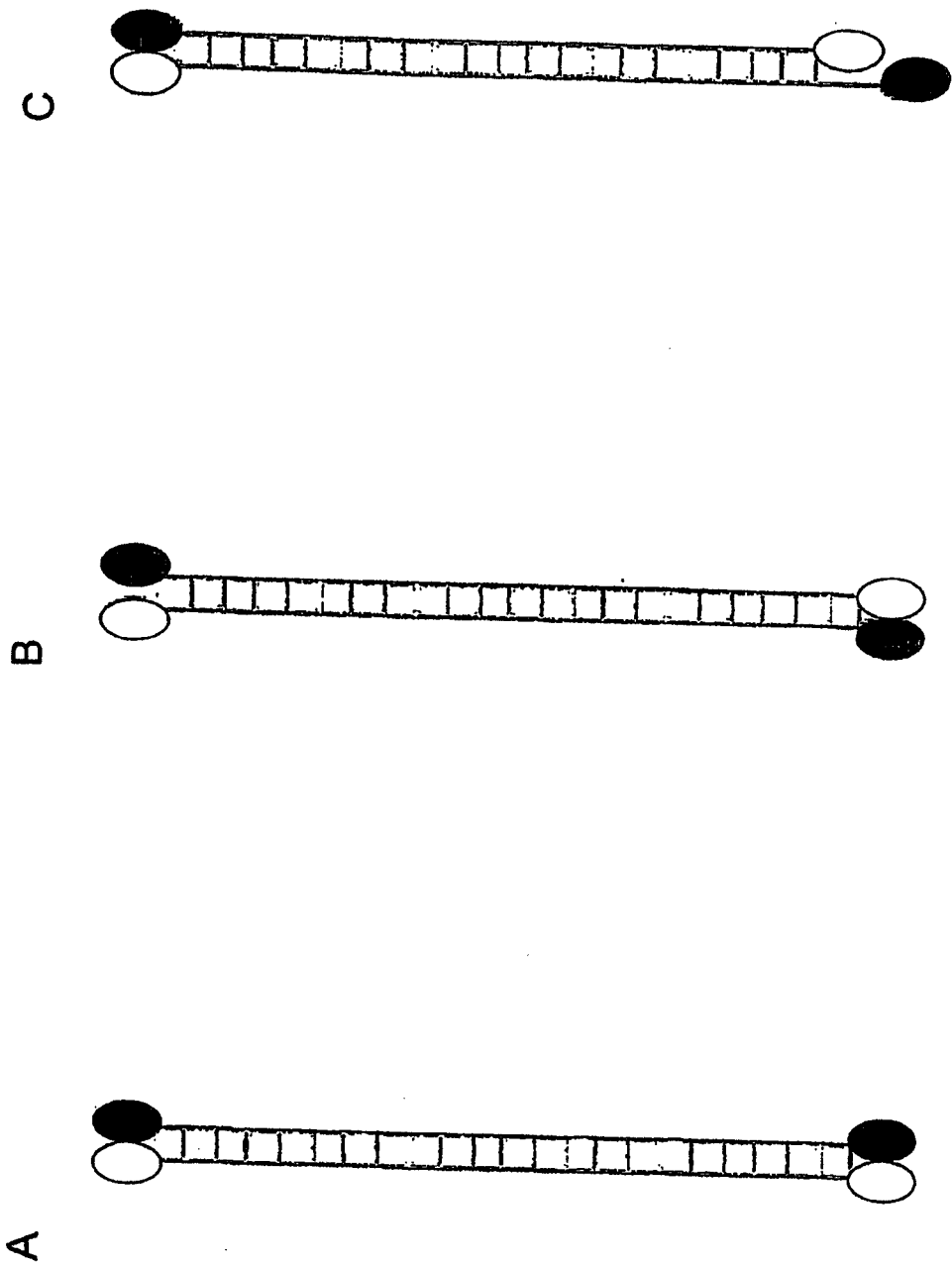


图2

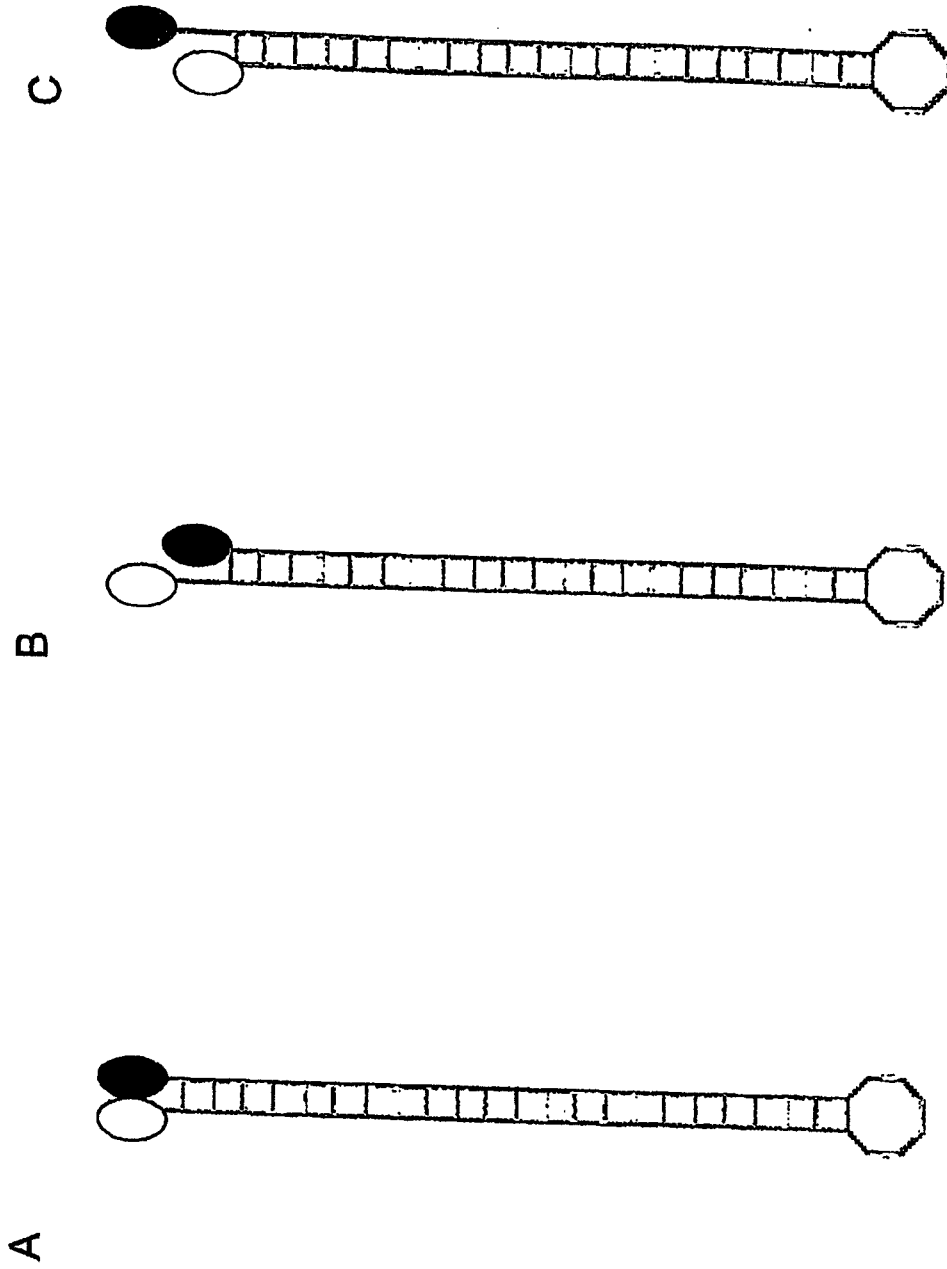


图3

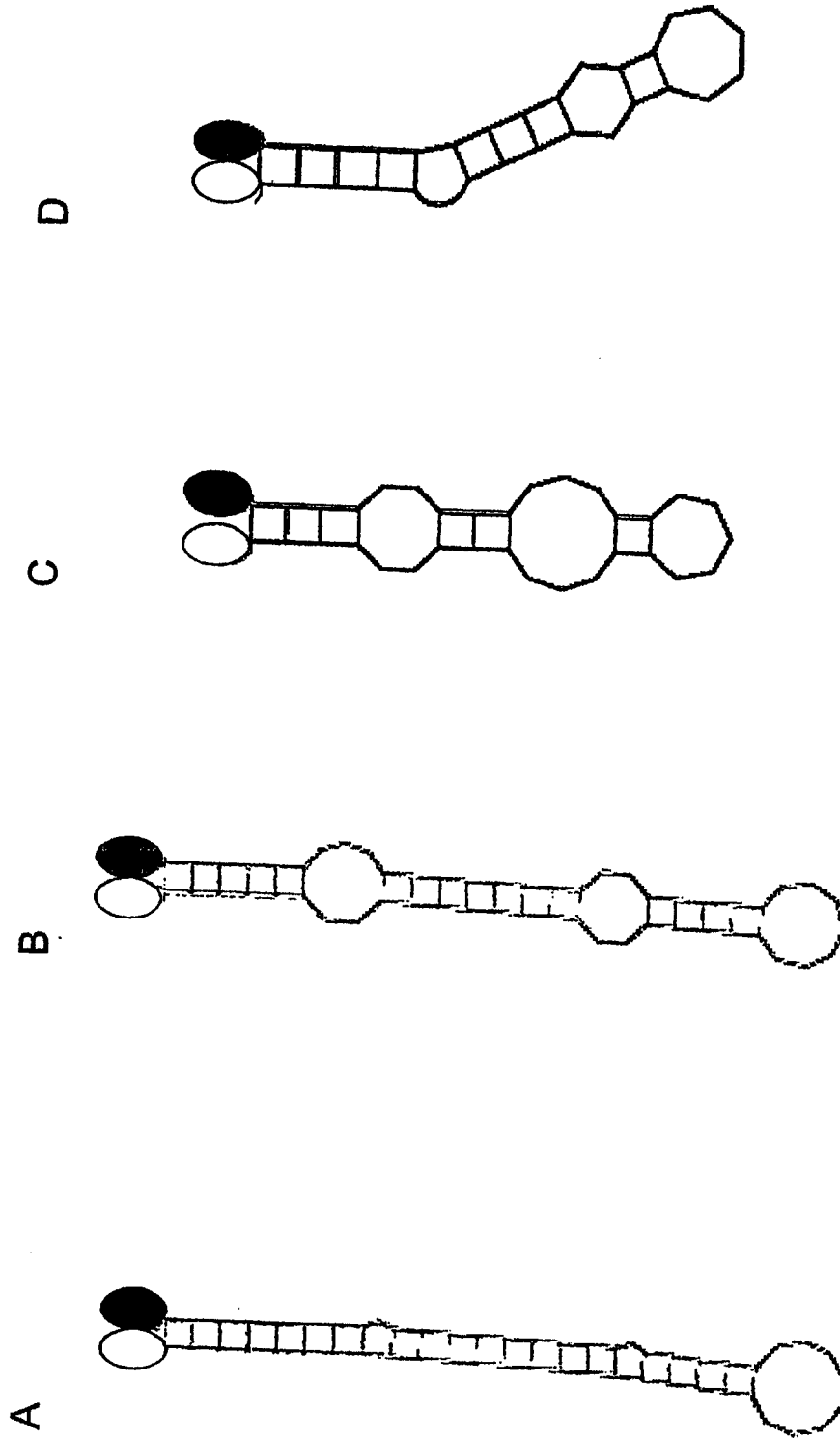


图4

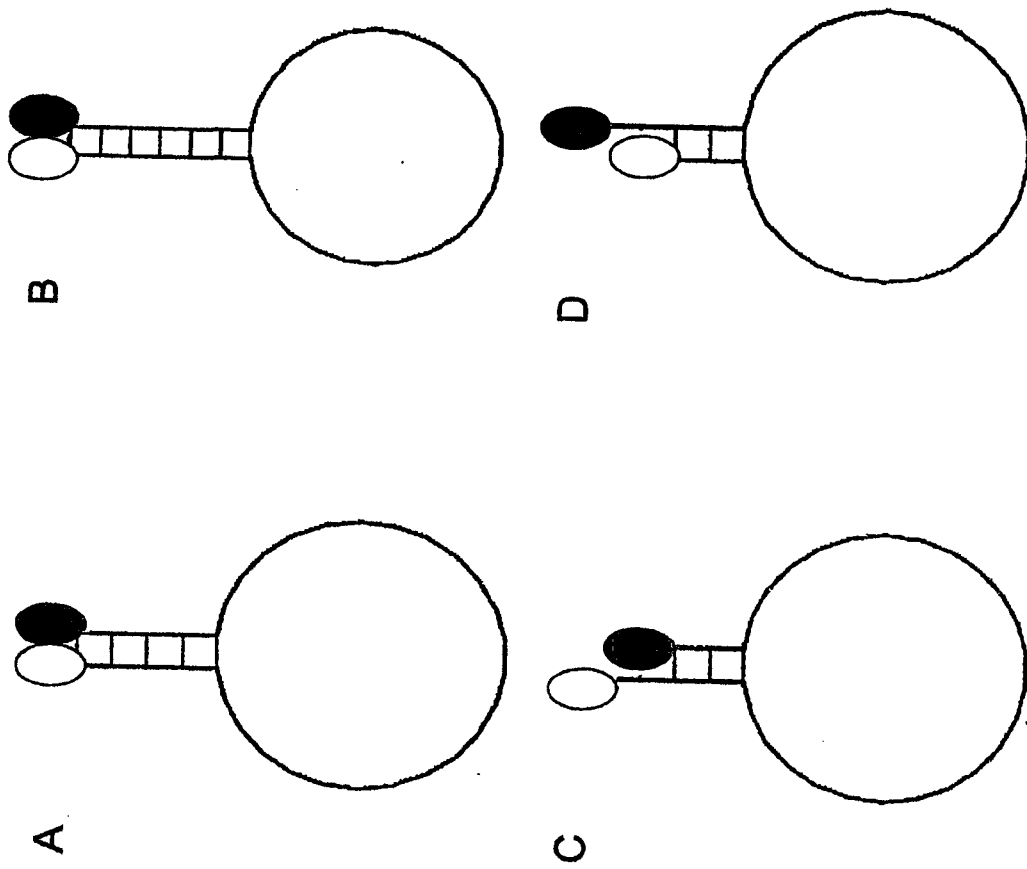
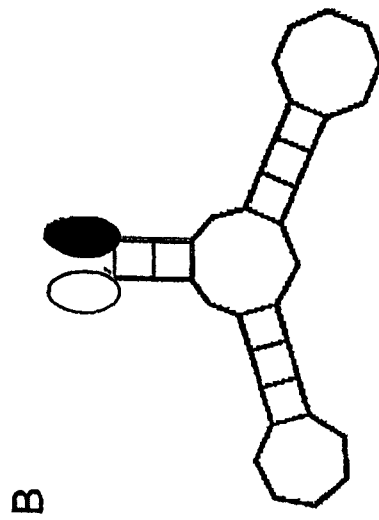
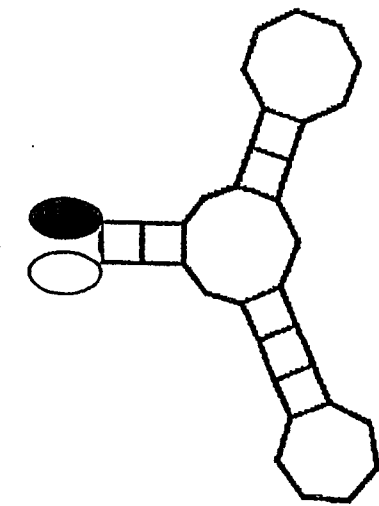


图5



B



A

图6

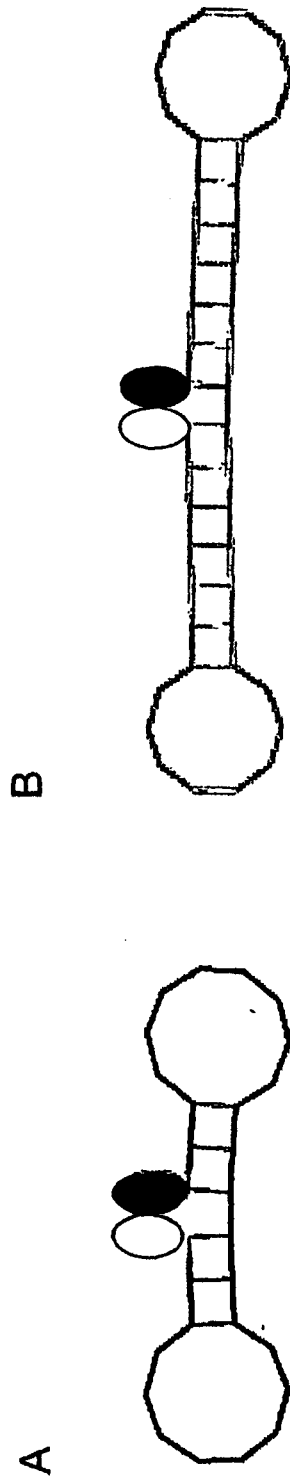


图7



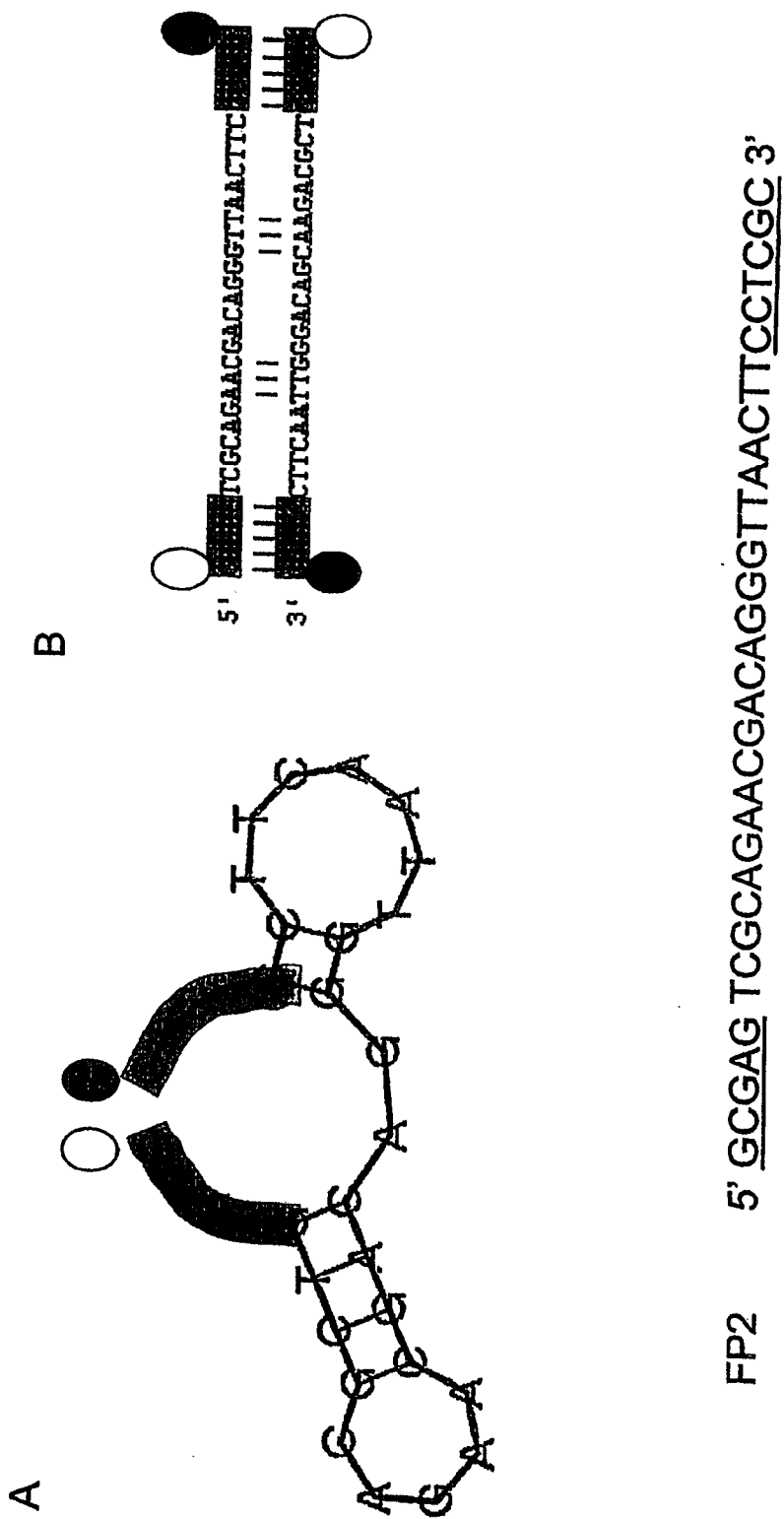
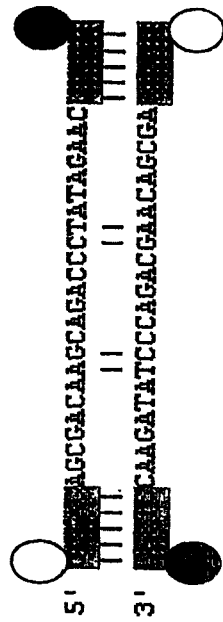
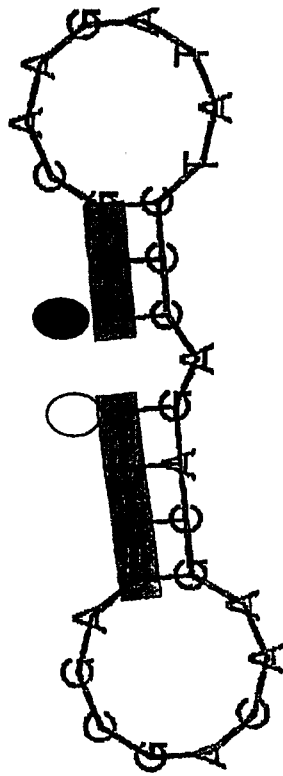


图9



FP3 5' GCGAG AGCGACAAGCAGACCCTATAGAAC CTCGC 3'

图 10



FP4 5' CTGC AGCGACAAGCAGACCCTATAGAAC GGG 3'

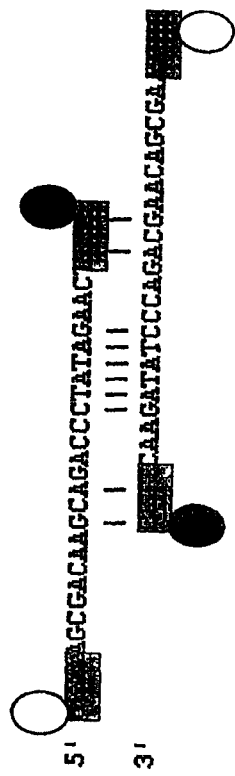
图 11

FP5 5' ggaa TCCCAGTTCCTGTGCCCTTAAGAAC ttcc 3'

图 12

FP6 5' ggaa TCGCAGAACGACAGGGTTAACTTC ttcc 3'

图 13



FP7 5' ggaa AGCGACAAGCAGACCCTATAGAAC tcc 3'

图 14



FP9 5' ggaa TCCCAGTTCCTGTGCCCTTAAGAACattagtttcc 3'

图16

FP10 5' gtgat TCCCAGTTCCTGTGCCCTTAAGAAC atcac 3'

图 17

FP11 5' ggac TCCAGTTCCTGTGCCCTTAAGAAC gtcc 3'

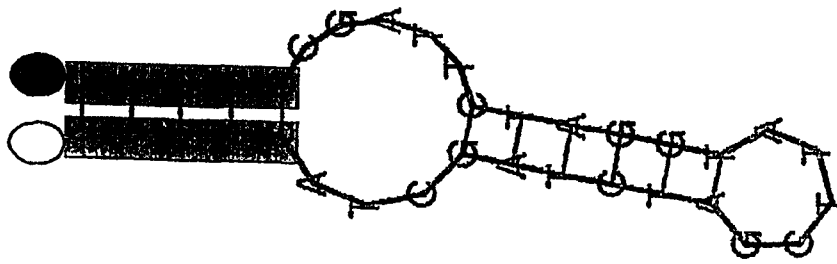
图18

FP12 5' ggg TCCCAGTTCCTGTGCCCTTAAGAAC ccc 3'

图19

FP13 5' gcg TCCAGTTCCTGTGCCCTTAAGAAC cgc 3'

图 20



FP14 5' gcgag ATCGATCTAGCTTATGGATCCTTAGC CTCGC 3'

图 21

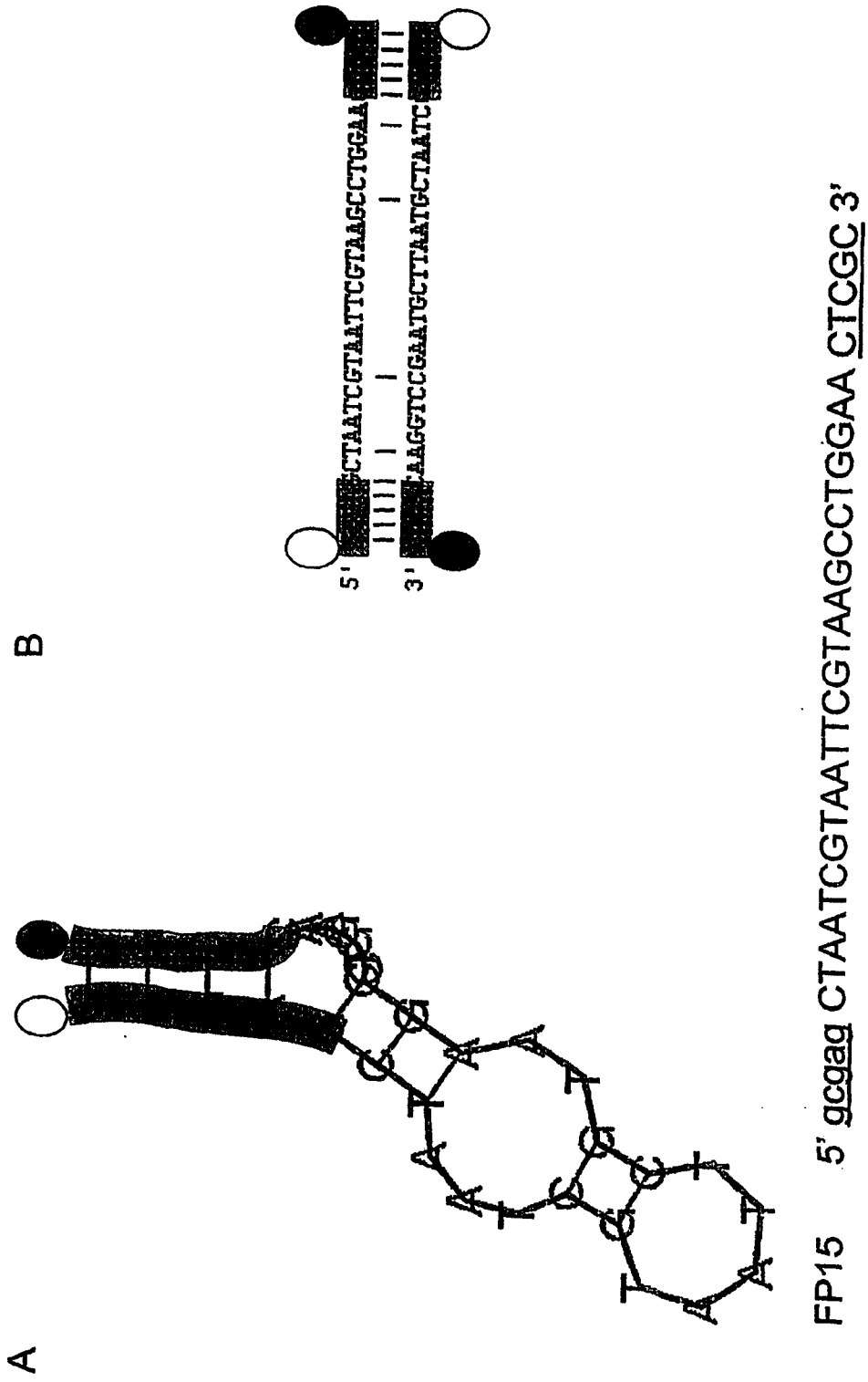


图 22

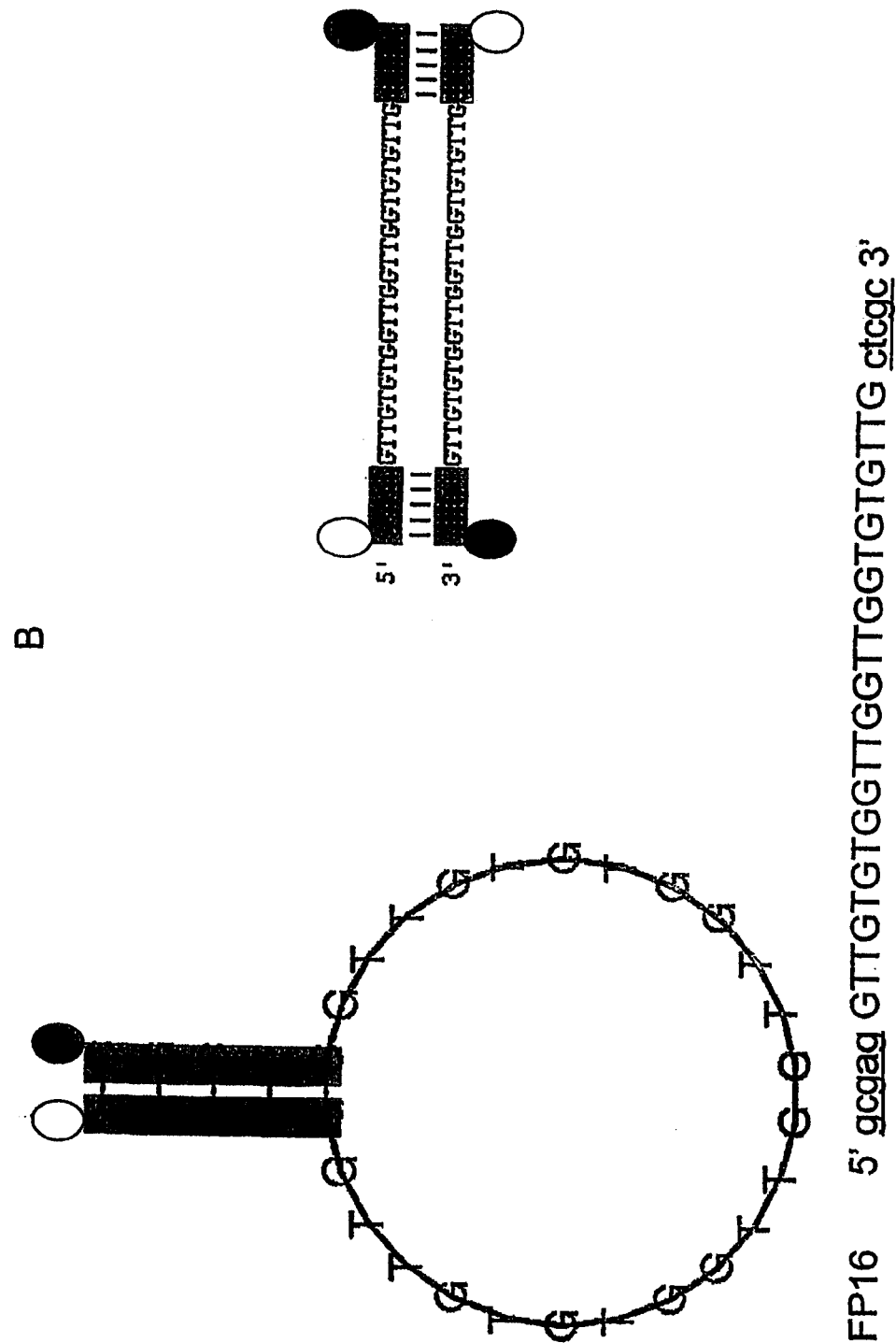


图 23

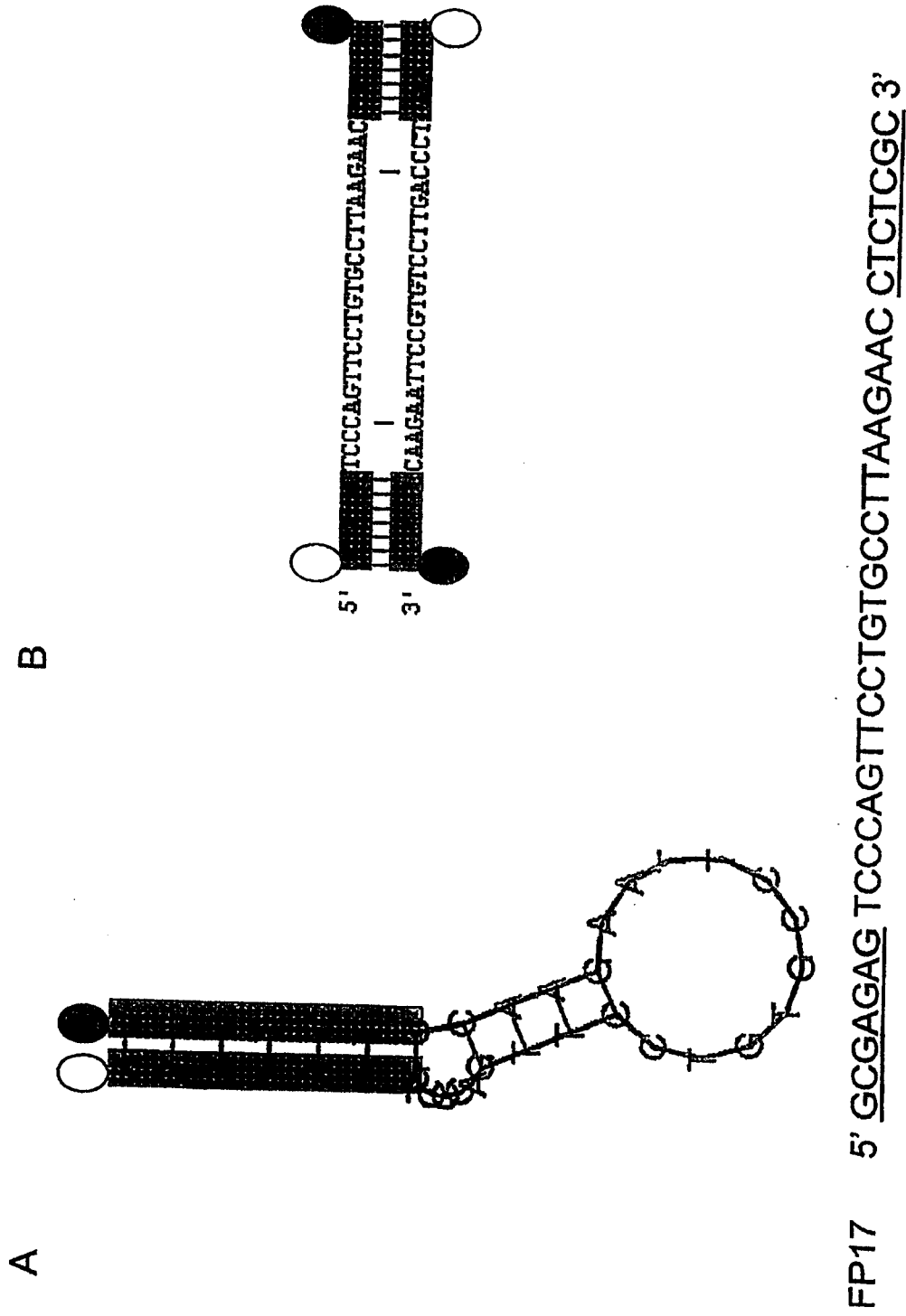


图 24

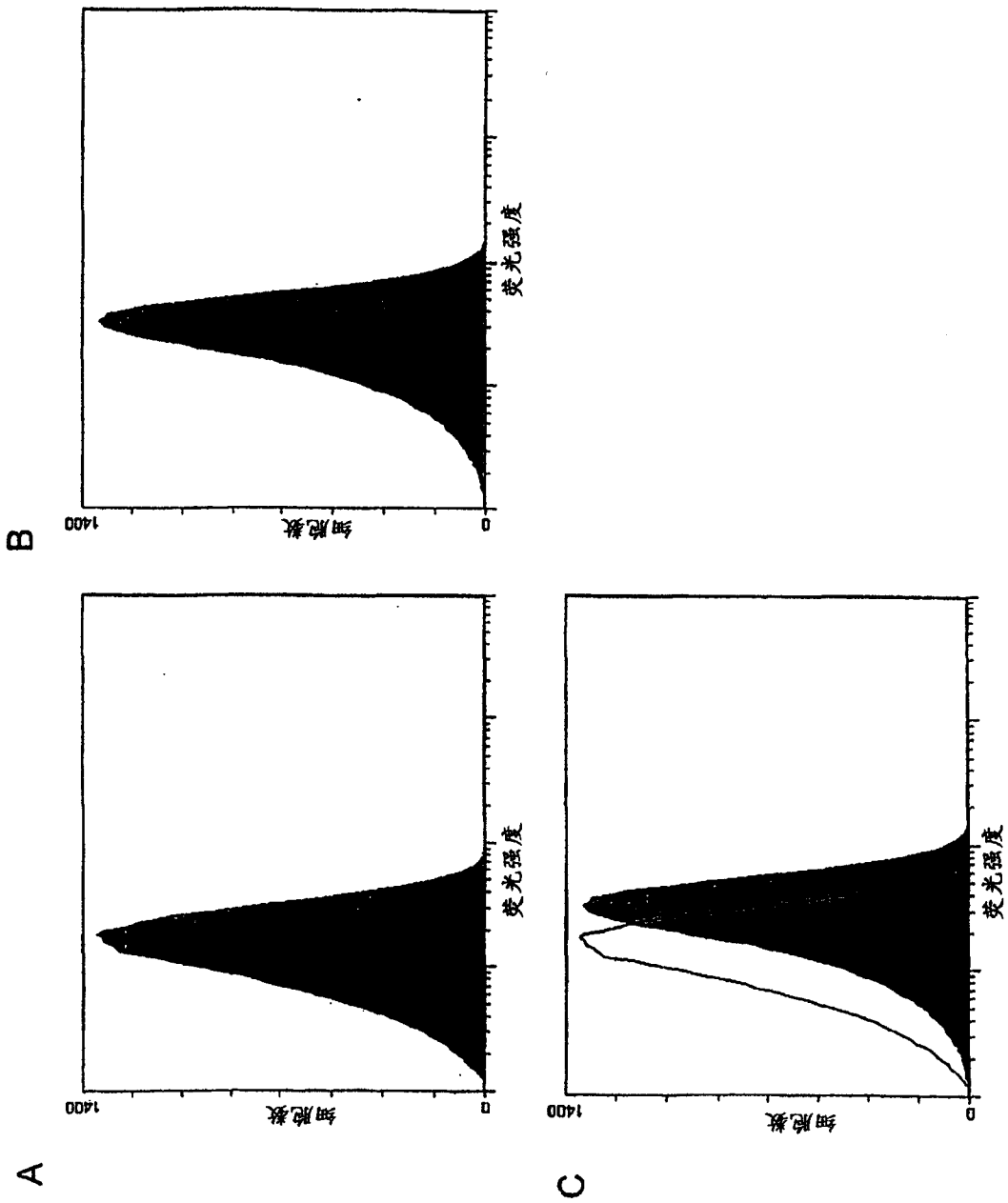


图 25

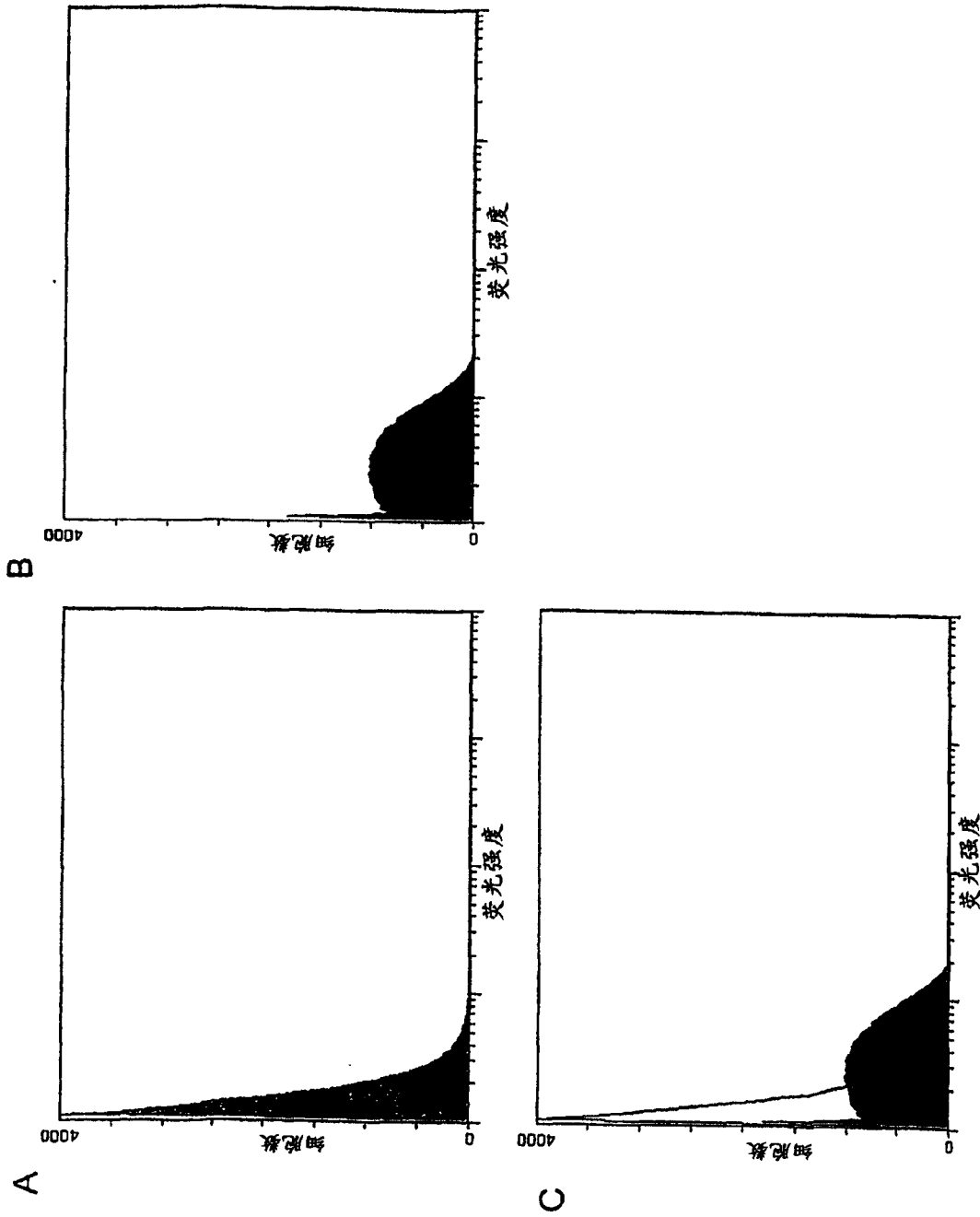


图 26

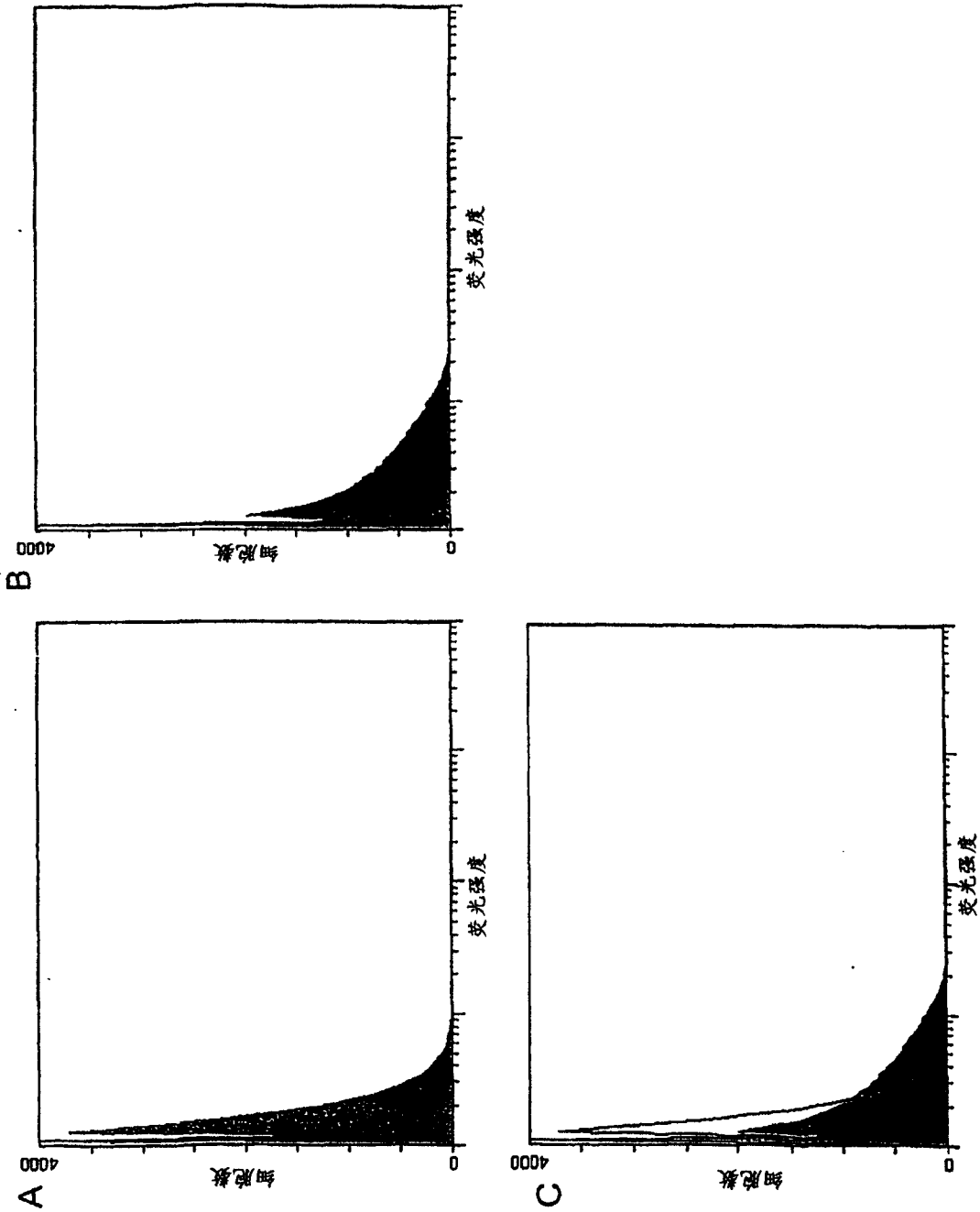


图 27

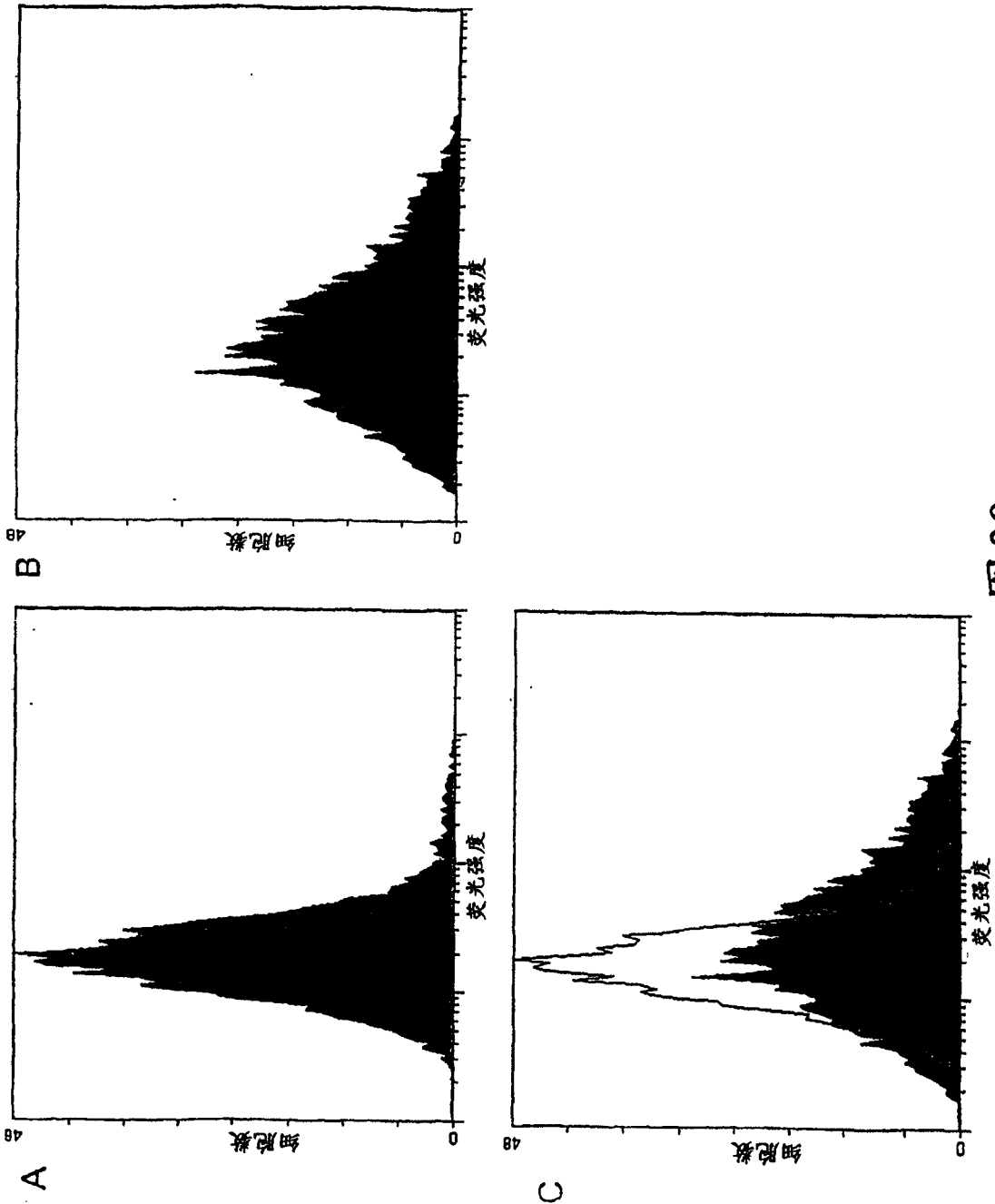


图 28

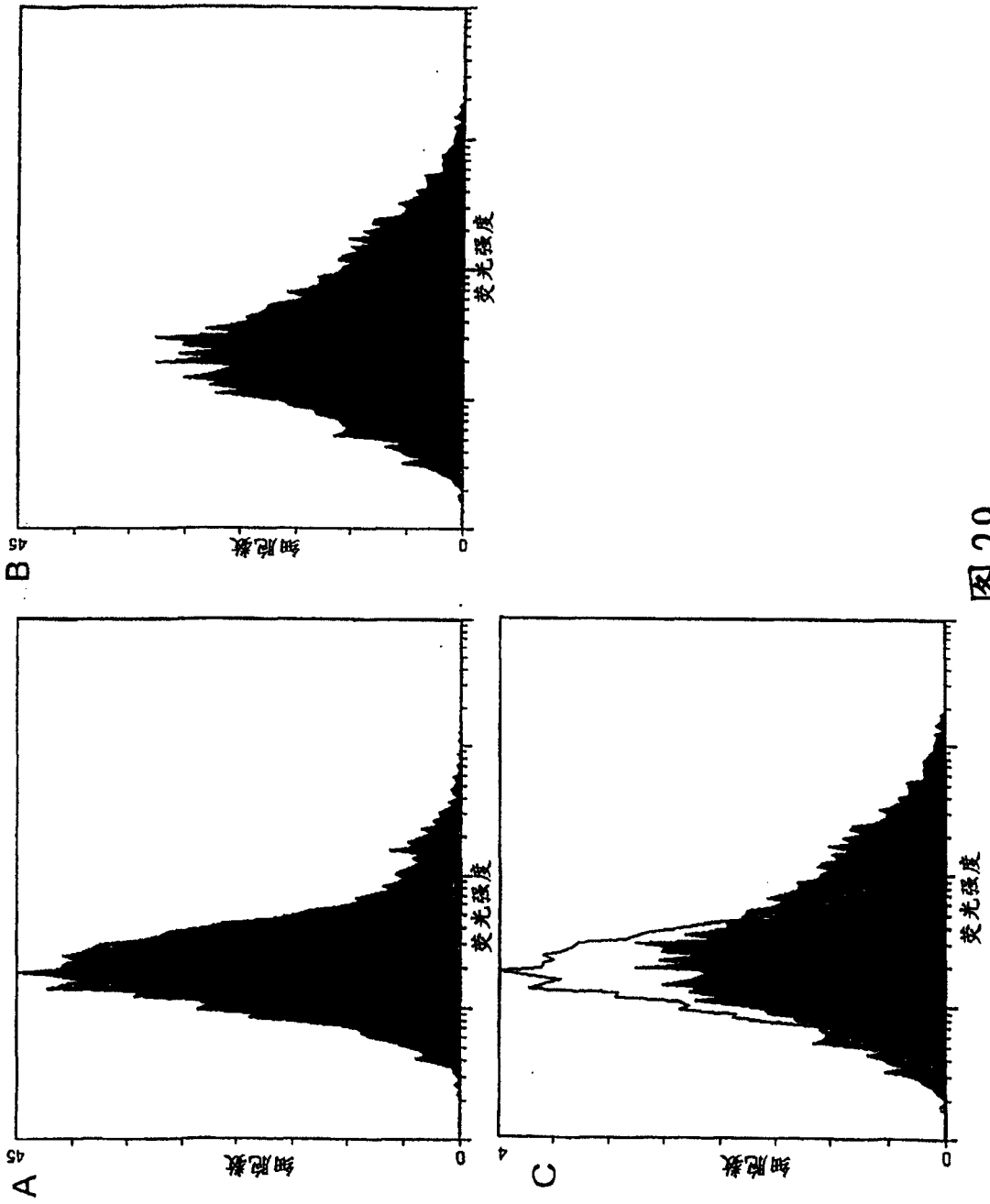


图 29

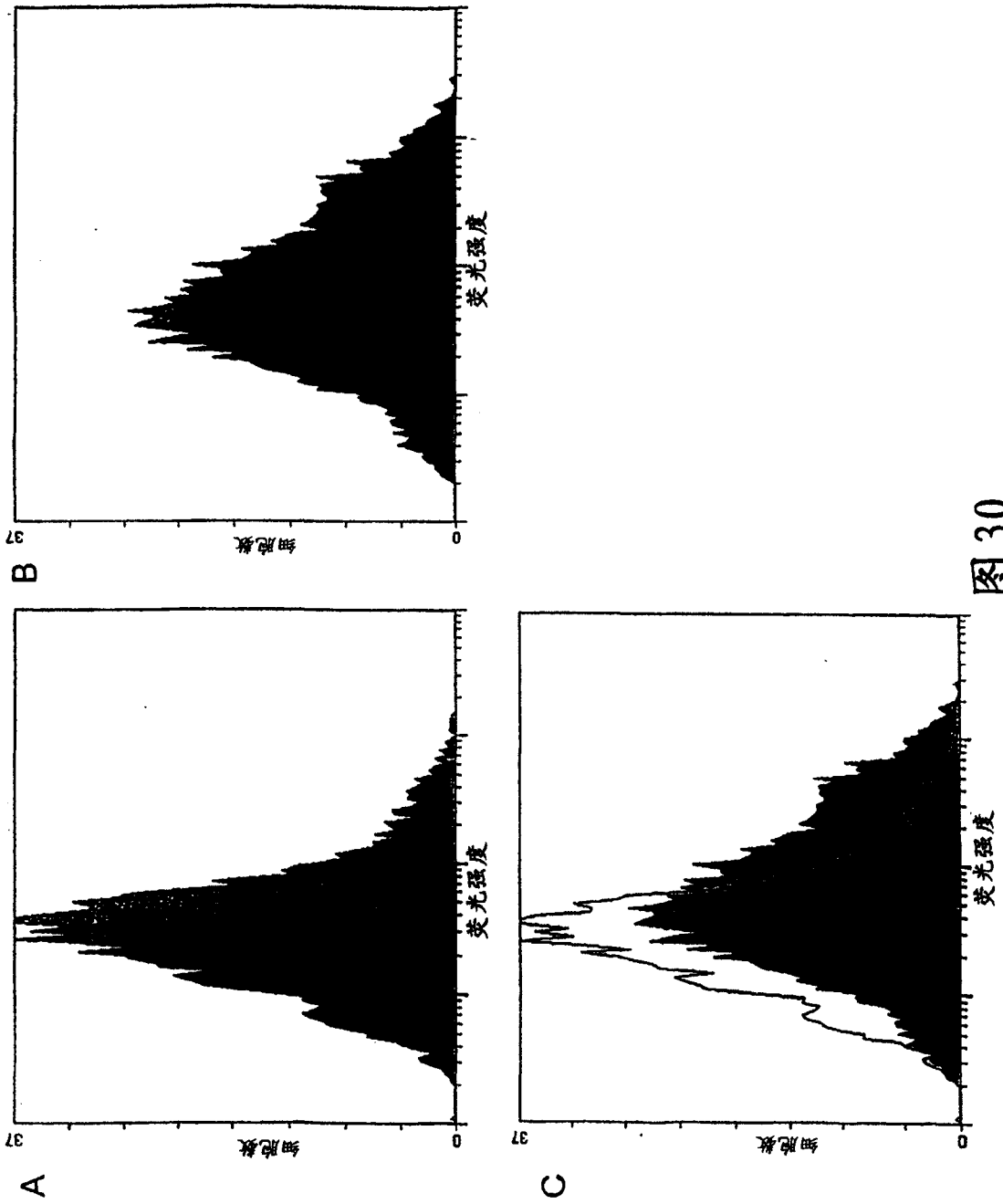


图 30

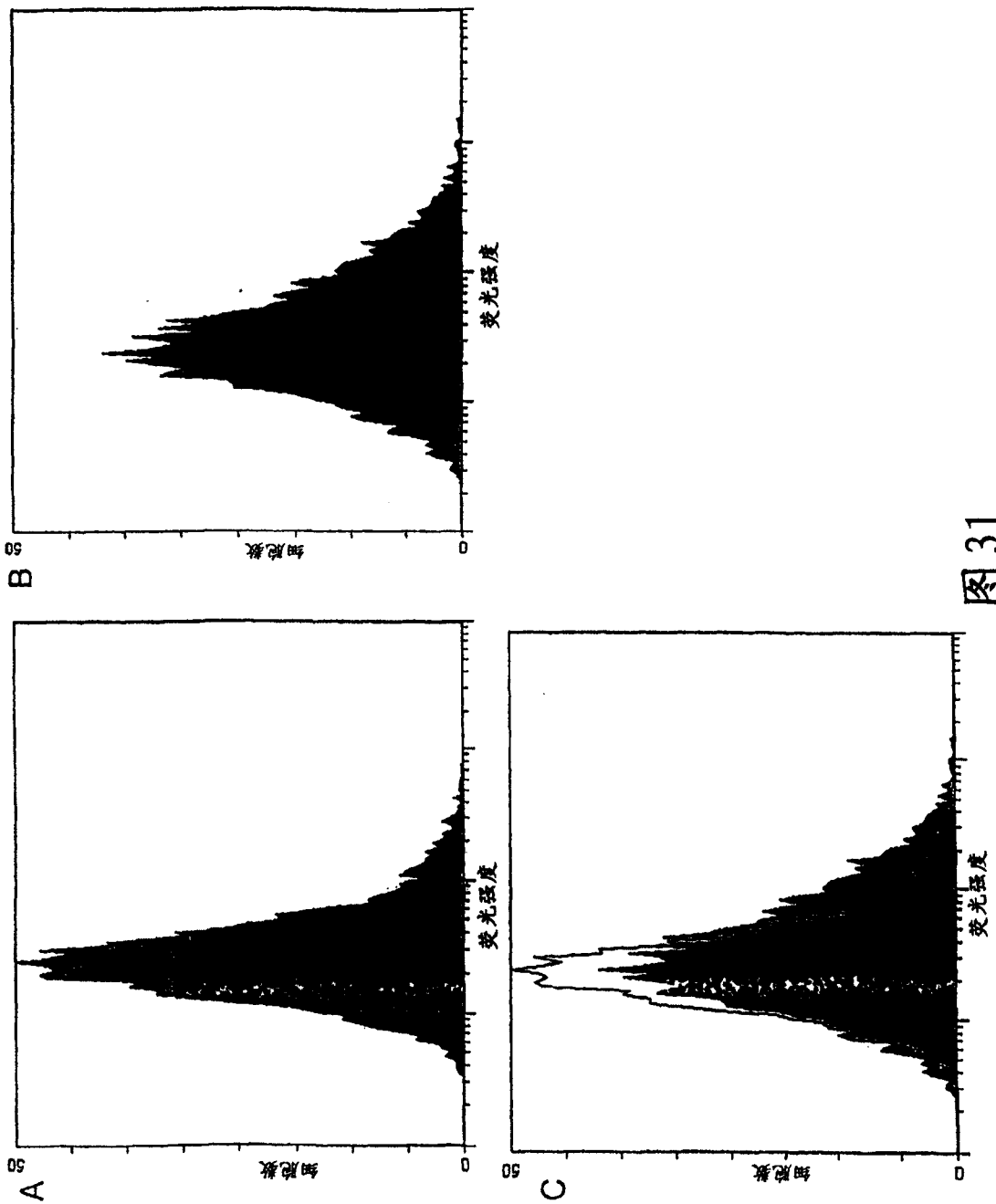


图 31

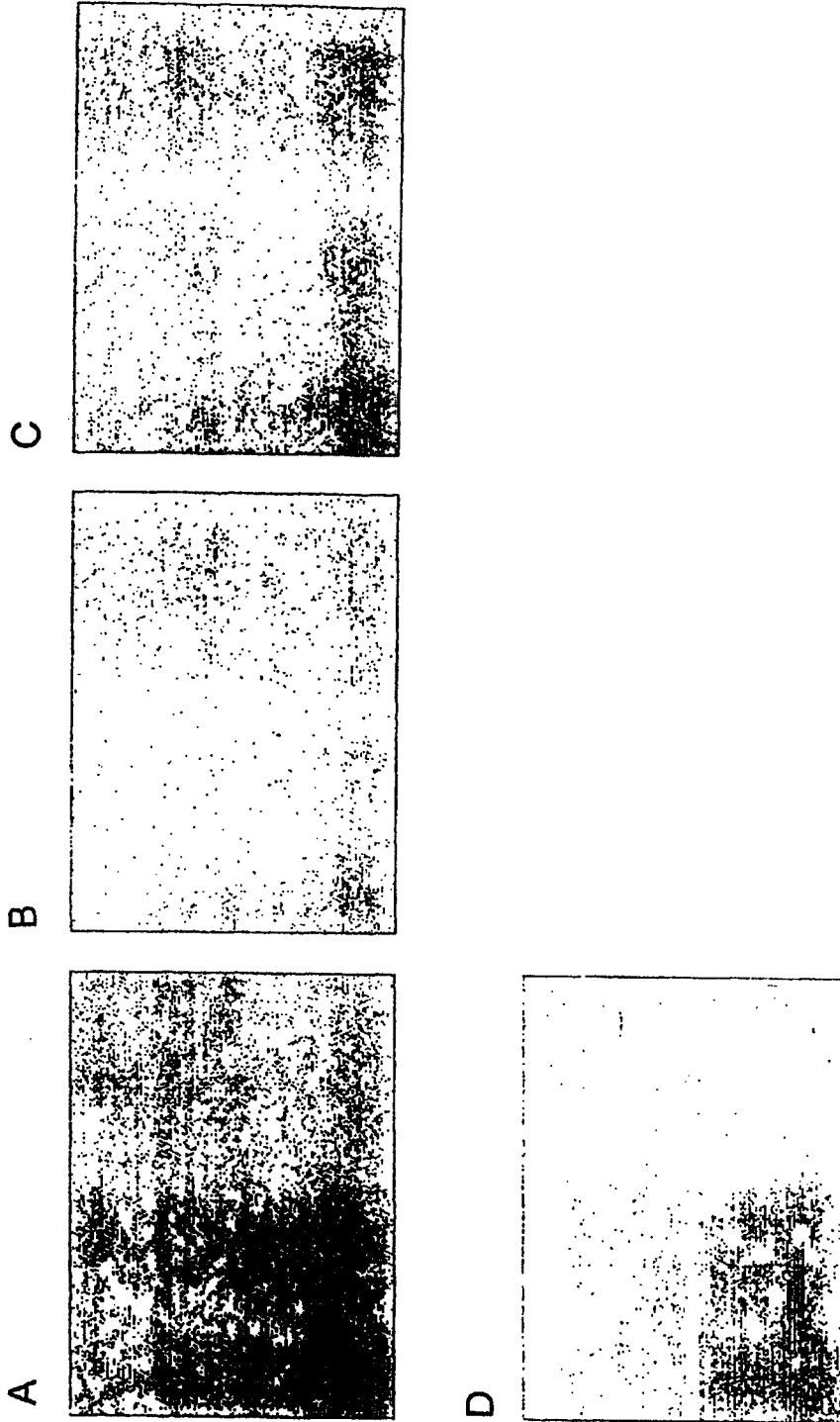


图32

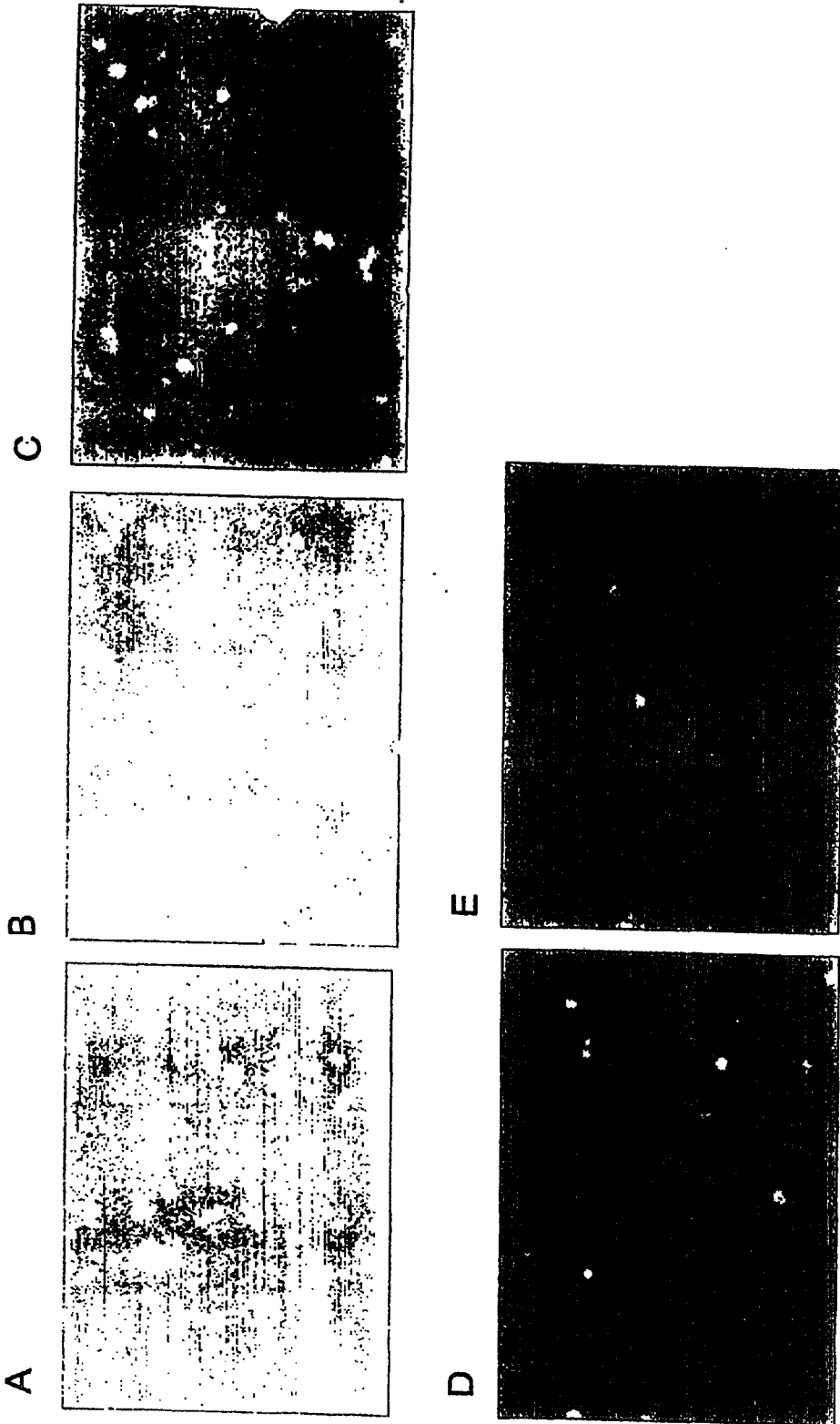


图 33

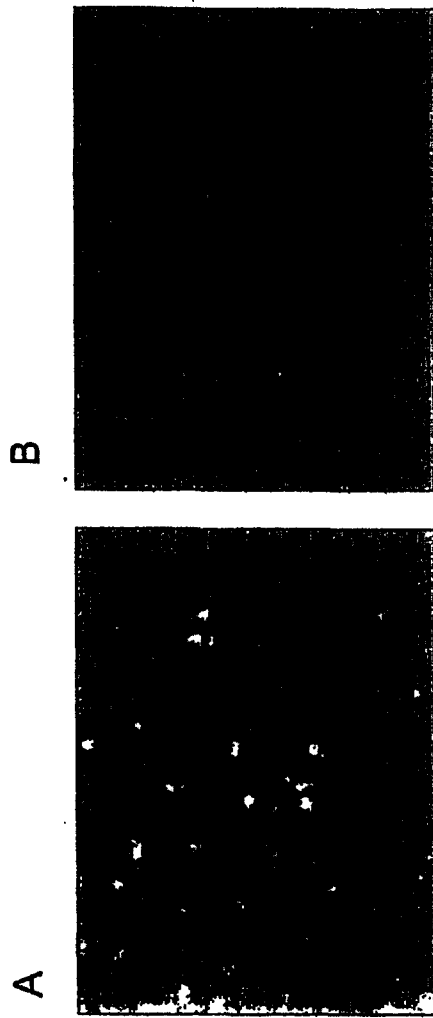


图 34



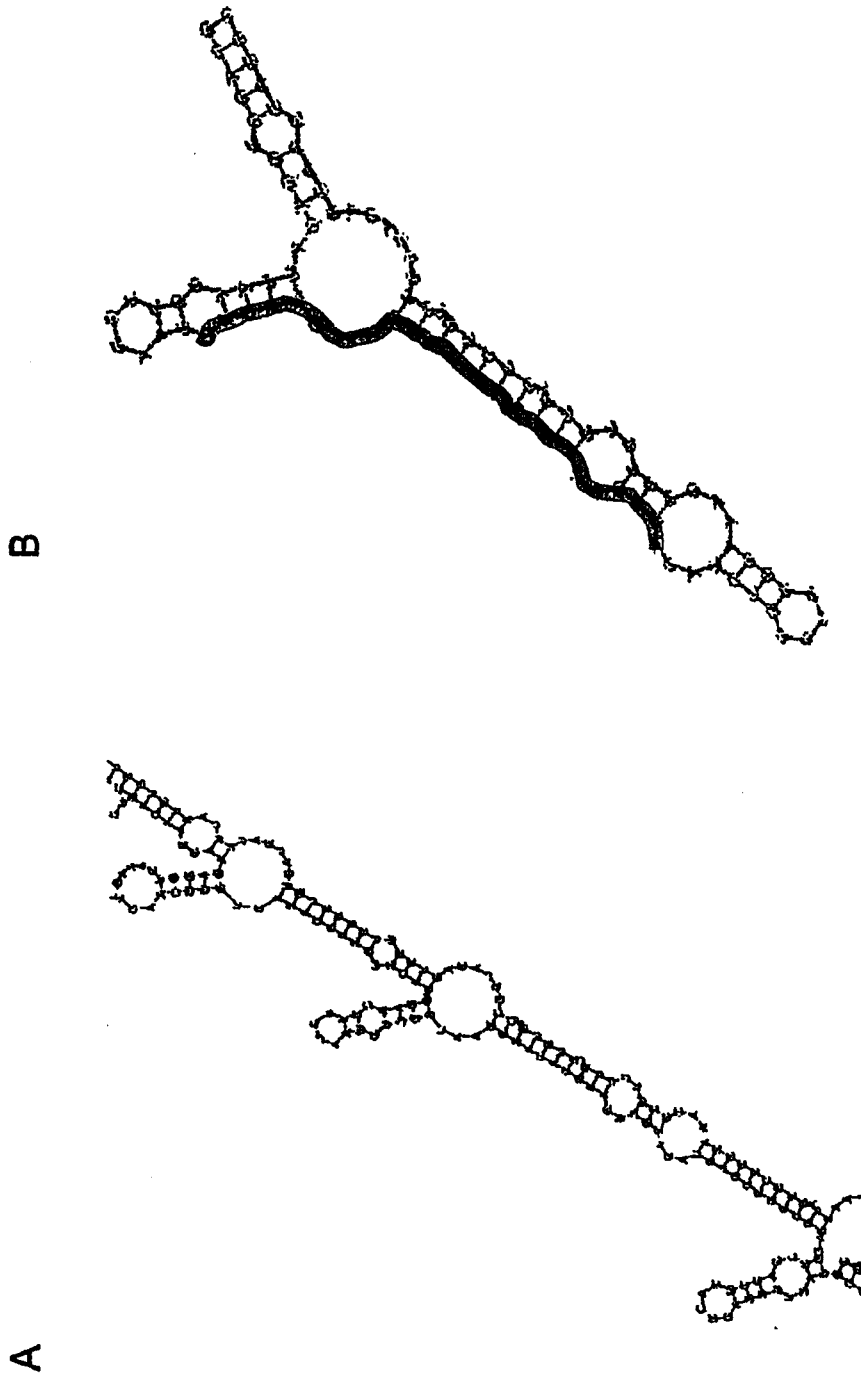


图 36

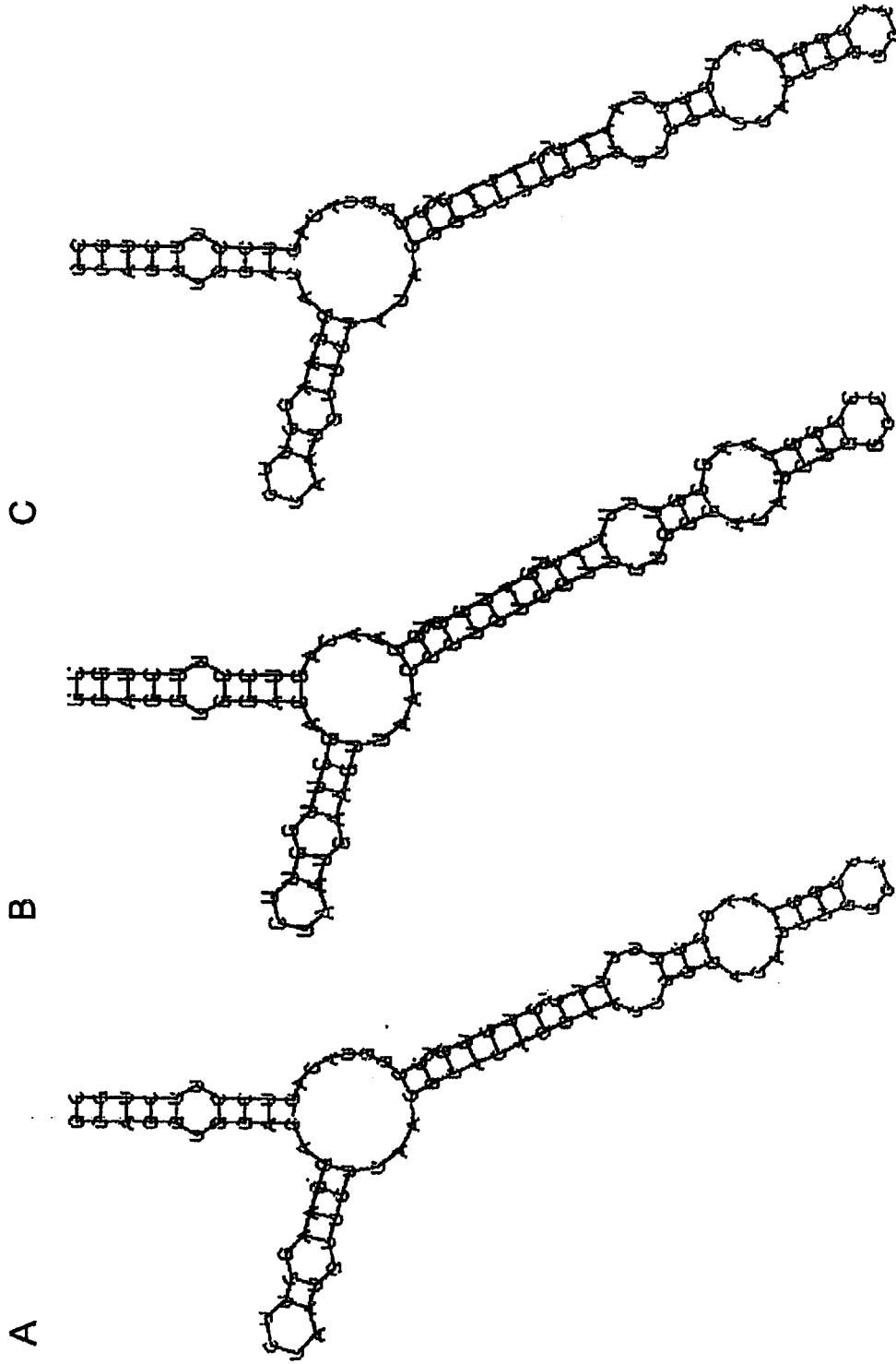


图 37

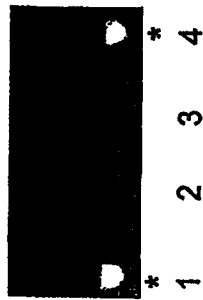


图 38

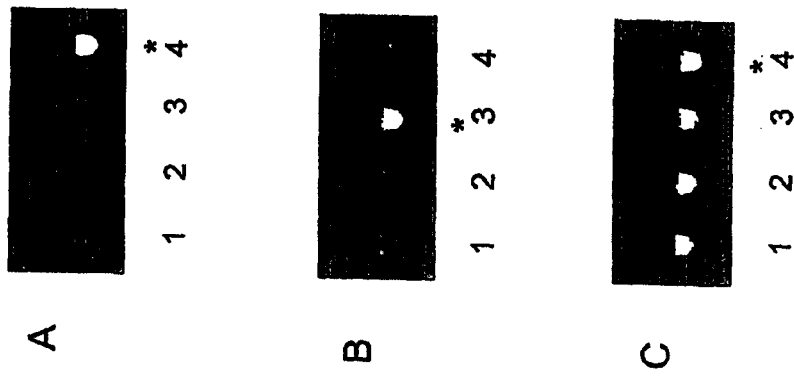


图 39

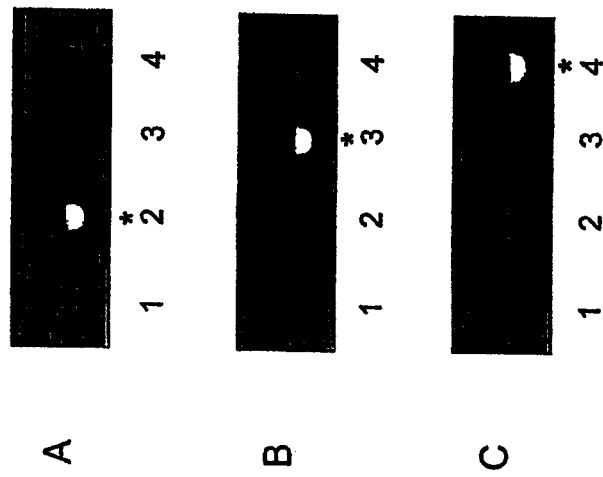


图40



A GCAGGTGGACAGGAAGGTTCTAATGTTCTTAAGGCACA  
**GGAACTGGGAC**ATCTGGGCCCGGAAAGCCTTTTCTCT  
 GTGATCCGGTACAGTCCTTCTGC

B GCAGGTGGACAGCTTGGTTCTAATGAAGTTAACCCCTGT  
**CGTTCTCGAC**ATCTGGGCCCGGAAAGCGTTTAACUGA  
 TGGAUUGGAACAGTCCTTCTGC

C GCAGGTGGACAGGAAGGTTCTAATGTTCTATAGGGTCT  
**GCTTGTGCTC**ATCTGGGCCCGGAGATGCGT  
 AAAGTCAGACATCCGGTACAGTCCTTCTGC

D 1 GTTCTTAAGGCACAGGAACTGGGA  
 2 GAAGTTAACCCIGTCGTTCTGGGA  
 3 GTTCTATAGGGICIGCTIGTCGCI  
 与靶1有50%差异;  
 与靶3有66%差异  
 与靶1有50%差异

图 42

**A** 起始  
 GCAGGTGGACAGGAAGTTCTAATGTTCTTAAGGCACAGGAACCTGGACATCTGGGCCCGGAAAGCCCTTTTCTCTGTGATCCGGTACAGTCCTTC  
 TGC

**B** 1  
 GCAGGTGGACAGCCTGGTTCTAATGAAGTTAACCTGCTGCGACATCTGGGCCCGGAAAGCGTTTAACTGACACATGGGTACAGTCCTTC  
 TGC

**C** 2  
 GCAGGTGGACAGCTTGGTTCTAATGAAGTTAACCCCTGTGTTCTGGACATCTGGGCCCGGAAAGCGTTTAACTGACA]GATGGGTACAGTCCT  
 TCTGC

**D** 3  
 GCAGGTGGACAGCTTGGTTCTAATGAAGTTAACCCCTGTGTTCTGGACATCTGGGCCCGGAAAGCGTTTAACT[[T]]GATGGGTACAGTCCTTC  
 TGC

**E** 4  
 GCAGGTGGACAGCTTGGTTCTAATGAAGTTAACCCCTGTGTTCTGGACATCTGGGCCCGGAAAGCGTTTAACTGATGG[[AT]]GGTACAGTCCT  
 TCTGC

**F** 5  
 GCAGGTGGACAGCTTGGTTCTAATGAAGTTAACCCCTGTGTTCTGGACATCTGGGCCCGGAAAGCGTTTAACTGATGGATGGAAACAGTCCTTCCTG  
 C

与前面的序列相比的序列修饰:

1. 下划线=碱基改变
2. [单括号]=切除的碱基
3. [[双括号]]=添加的碱基
4. [[[双括号]]]=添加的碱基
5. 下划线=碱基改变

图 43

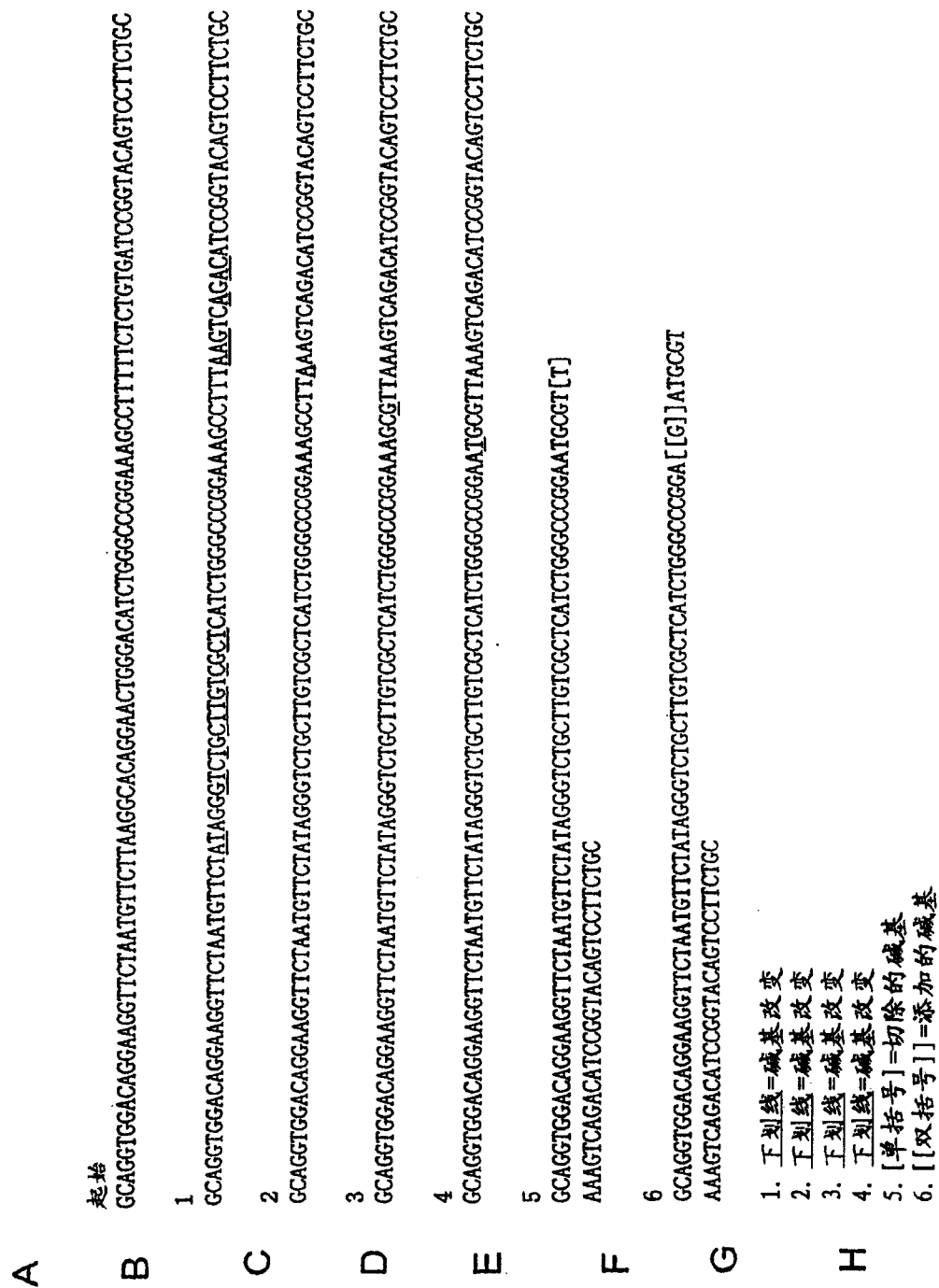


图 44

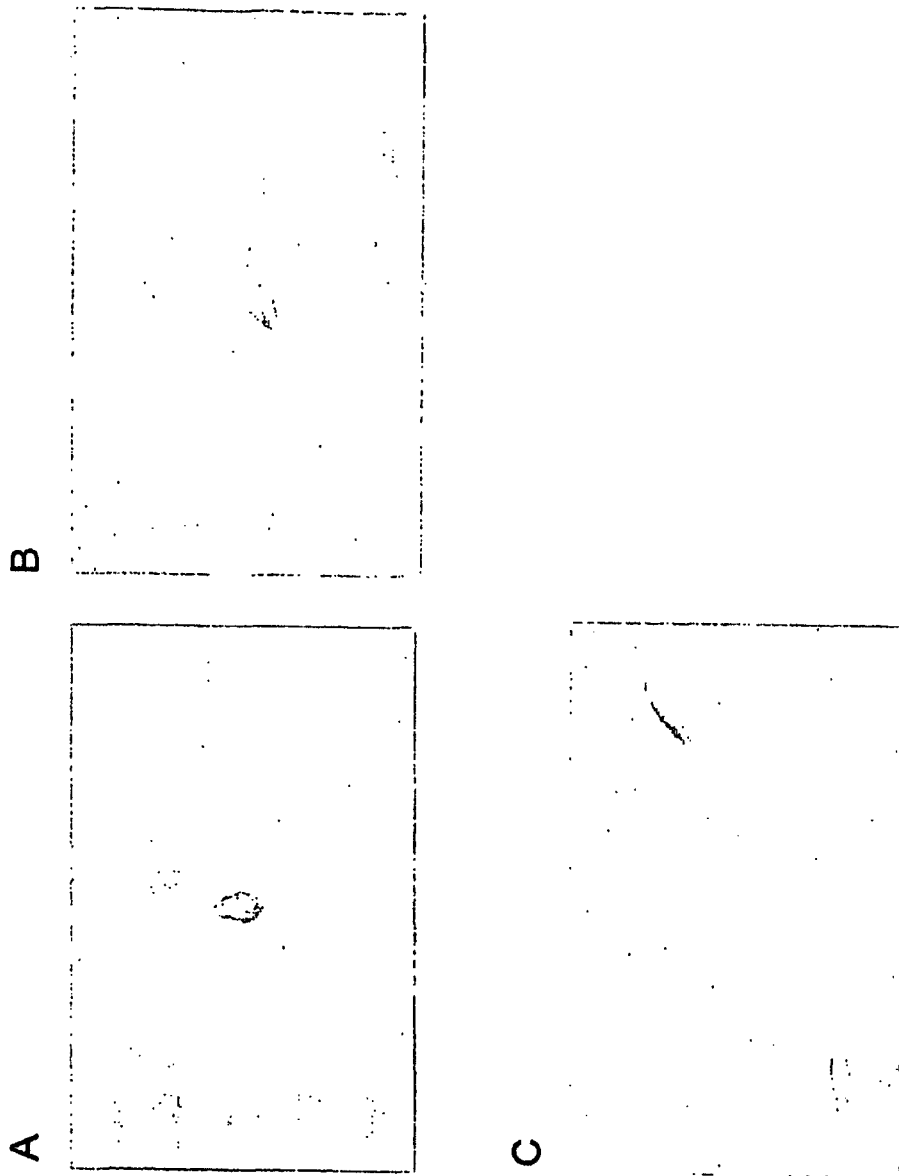


图45

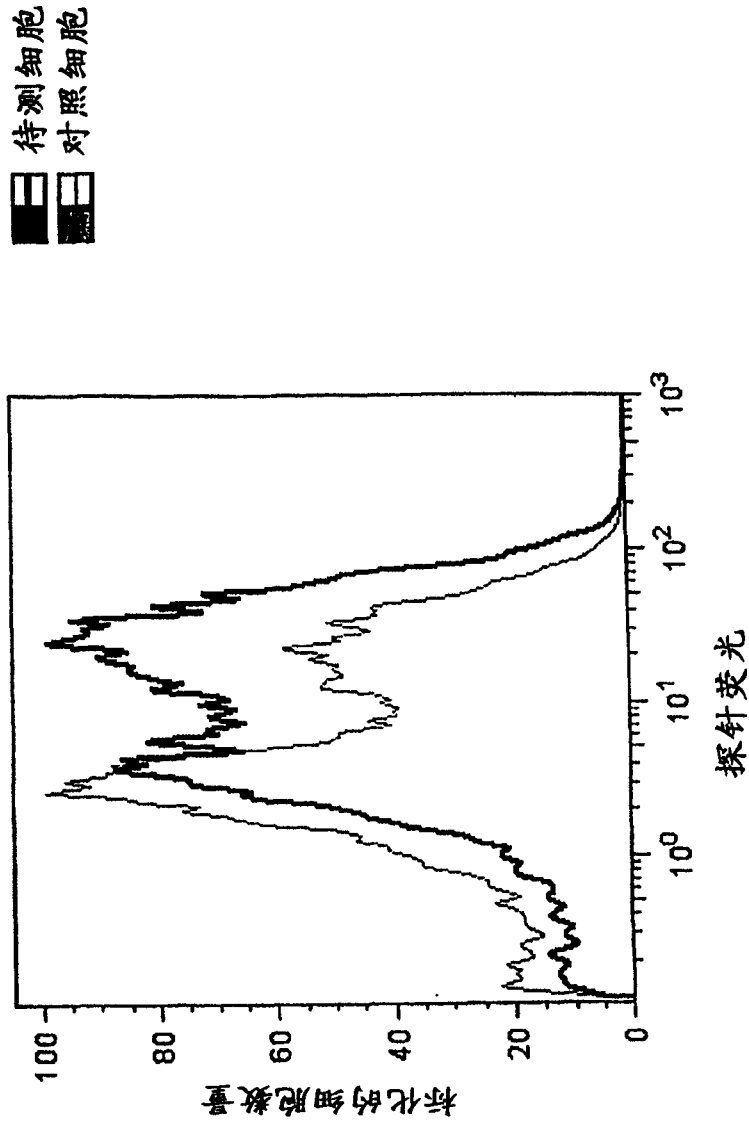


图 46 mcon1

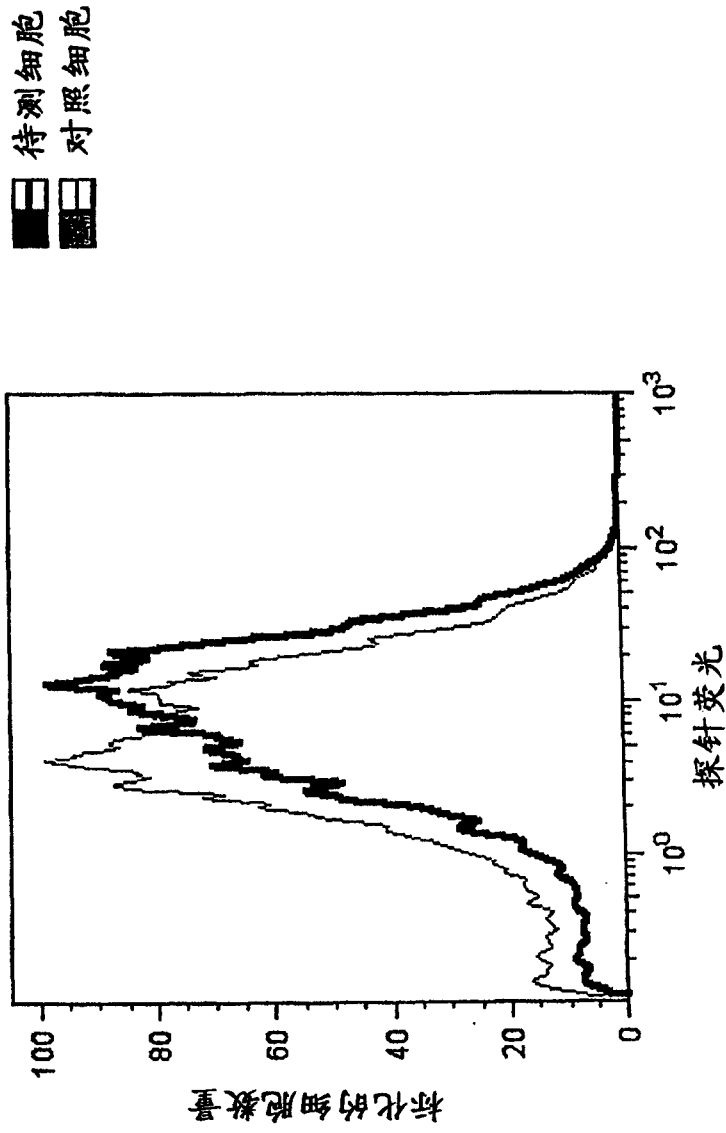


图 47 mcon2

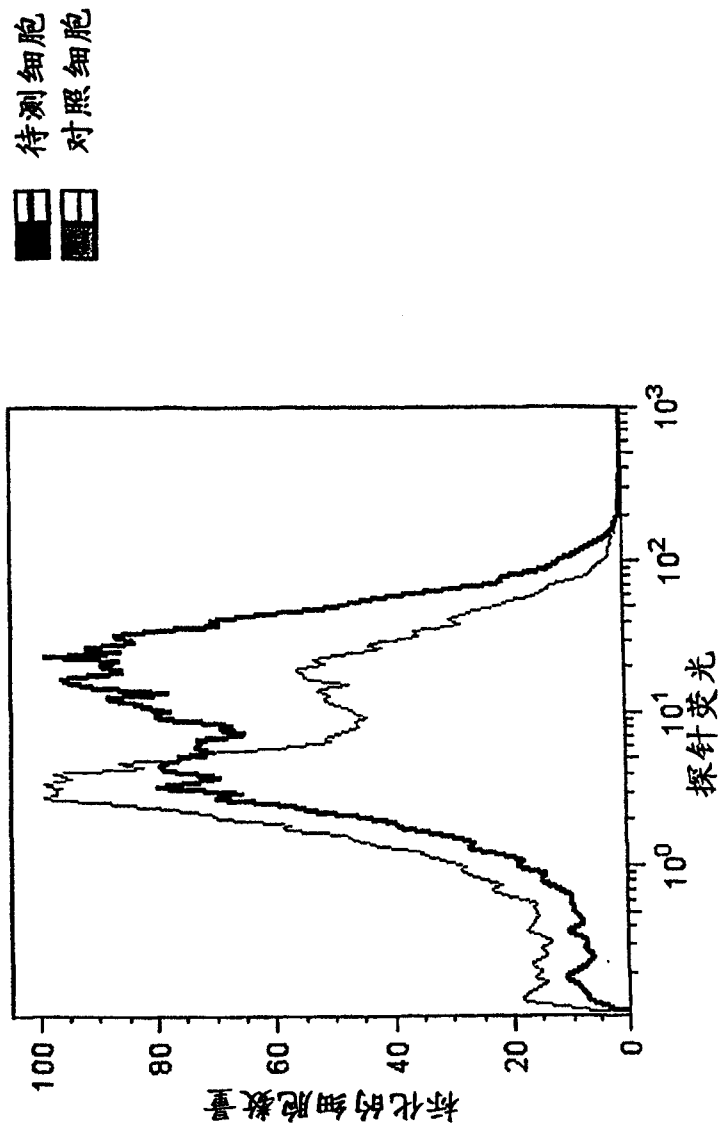


图 48 mcon3

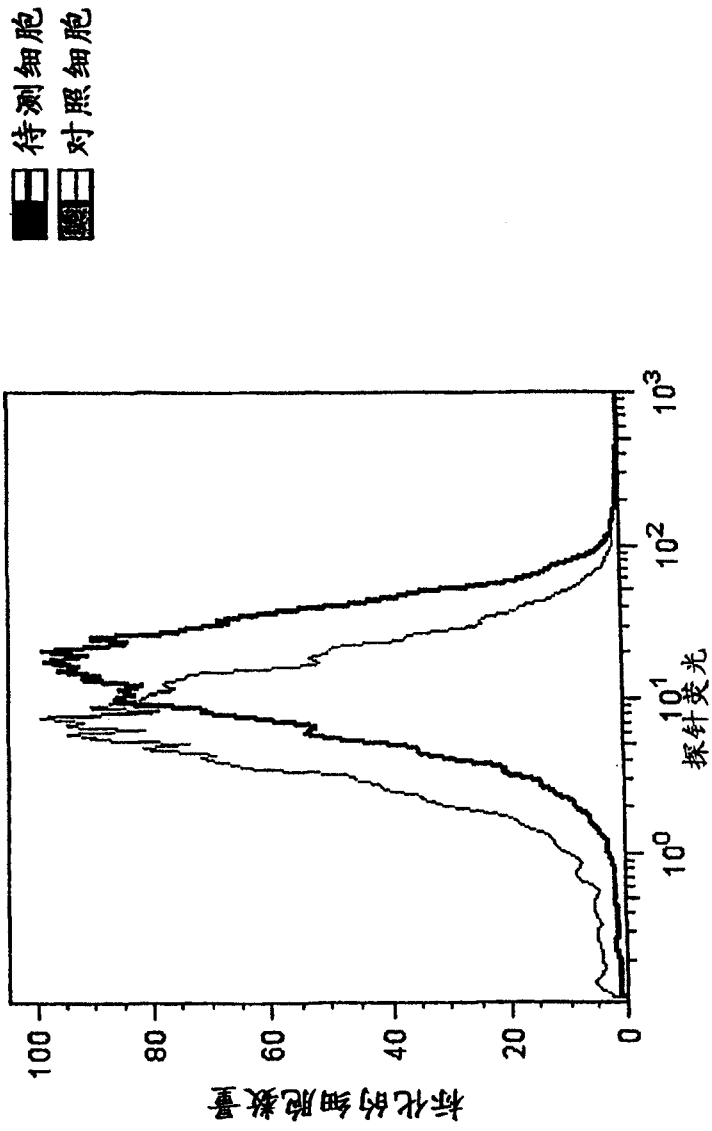


图 49 mcon4

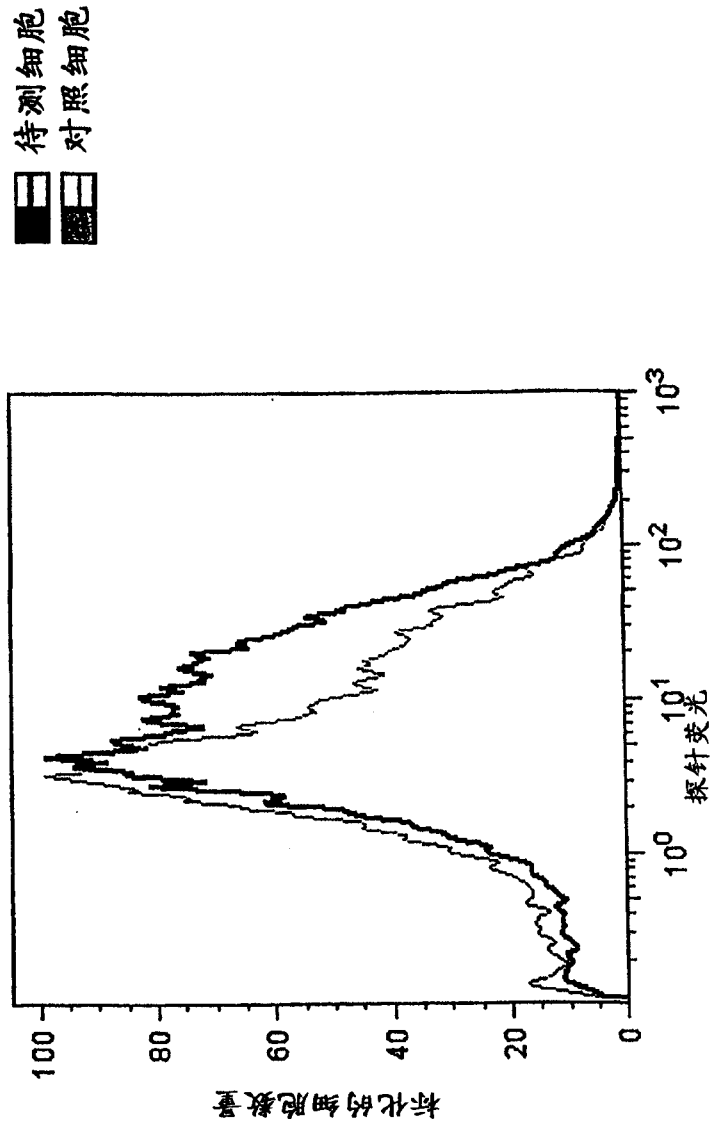


图 50 mcon5

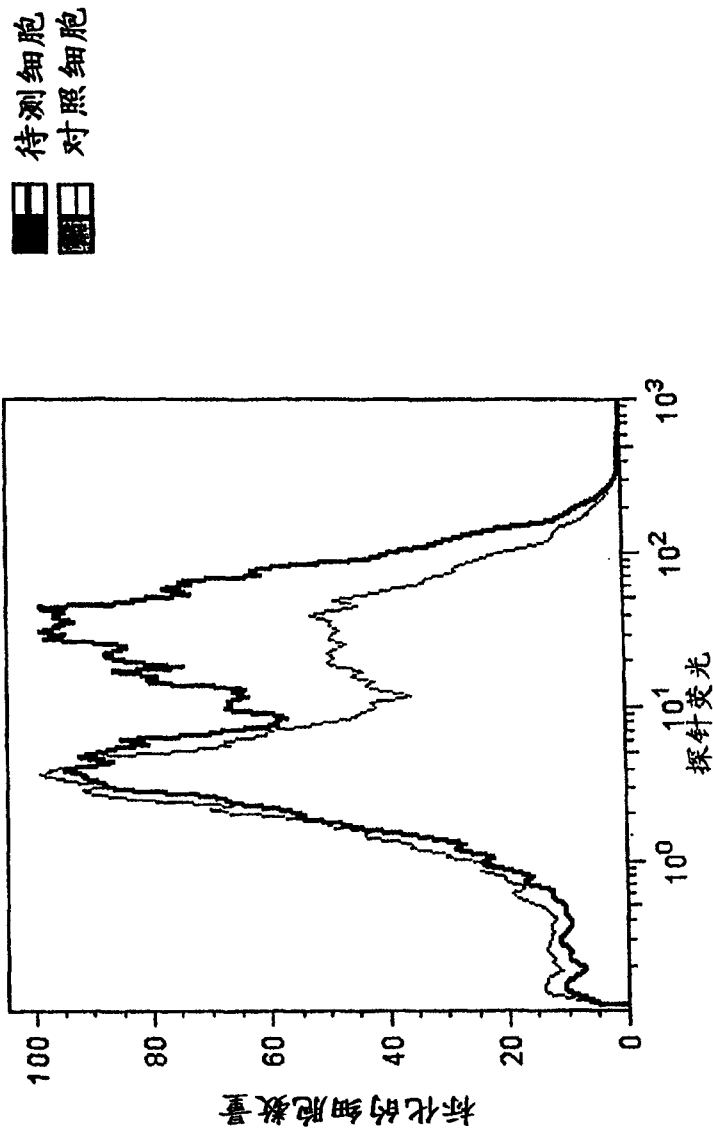


图 51 mcon6

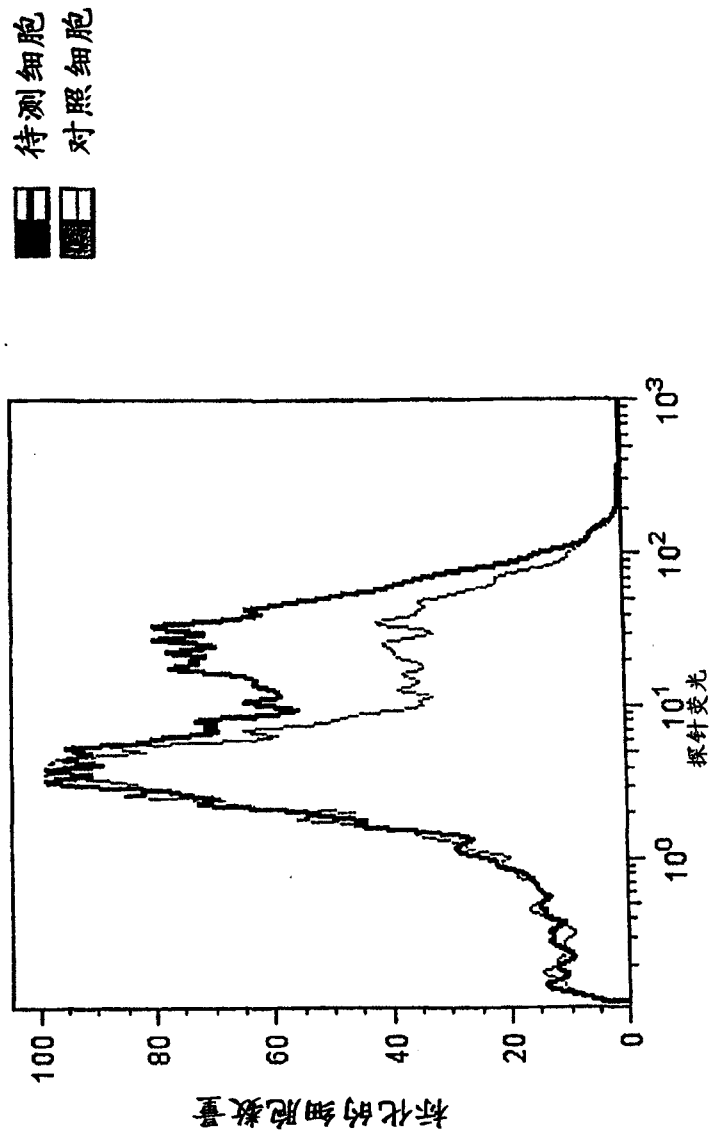


图 52 mcon7

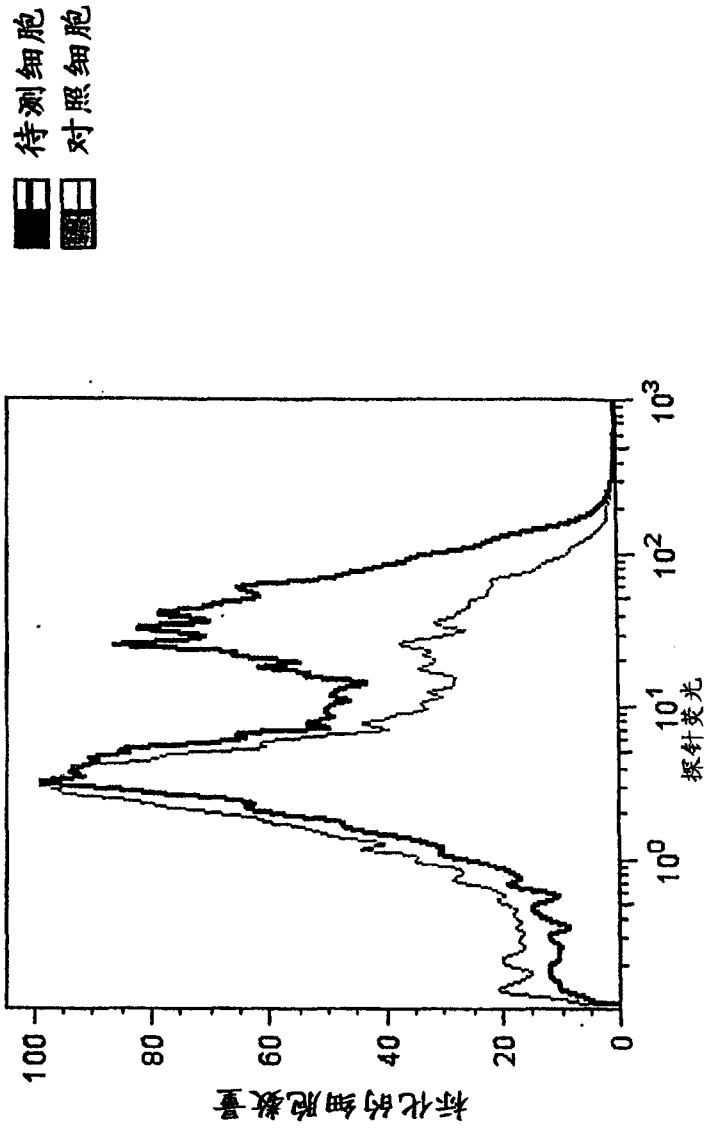


图 53 mcon8

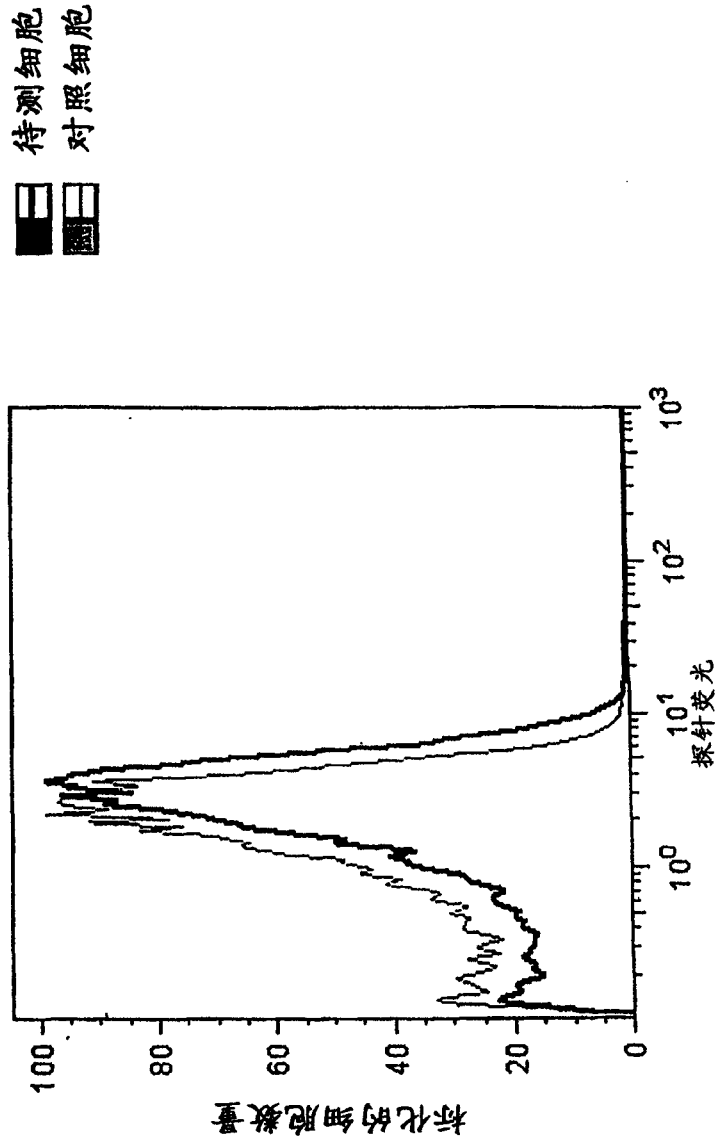


图 54 mcon9

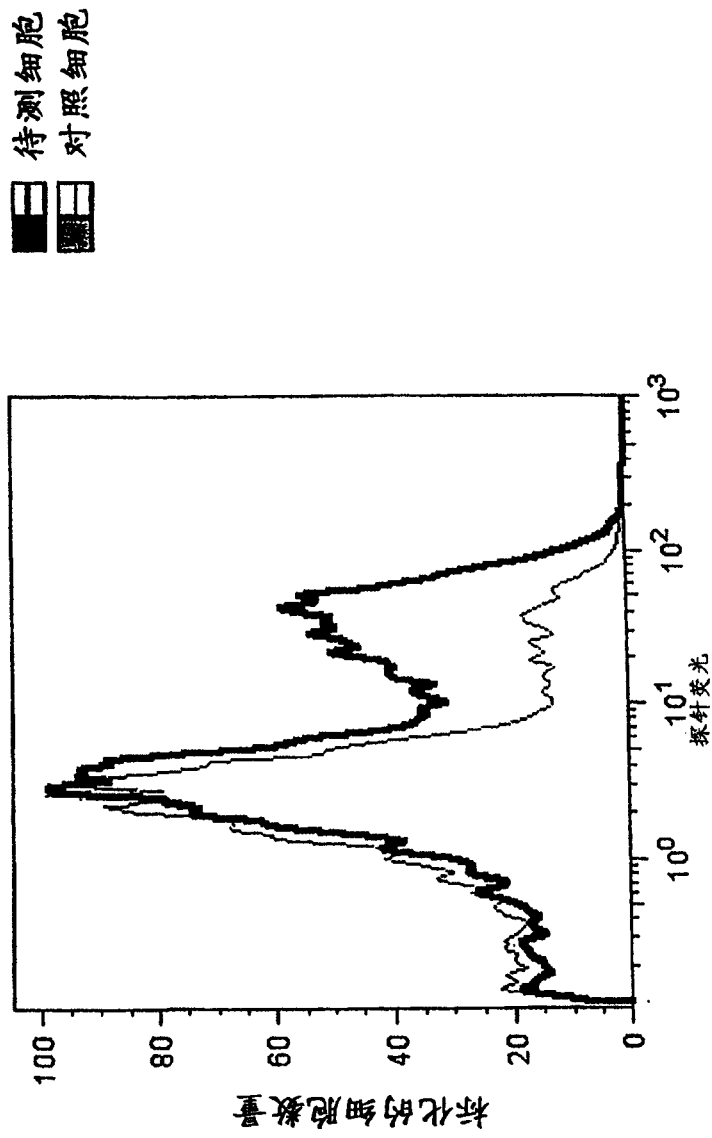


图 55 mcon10

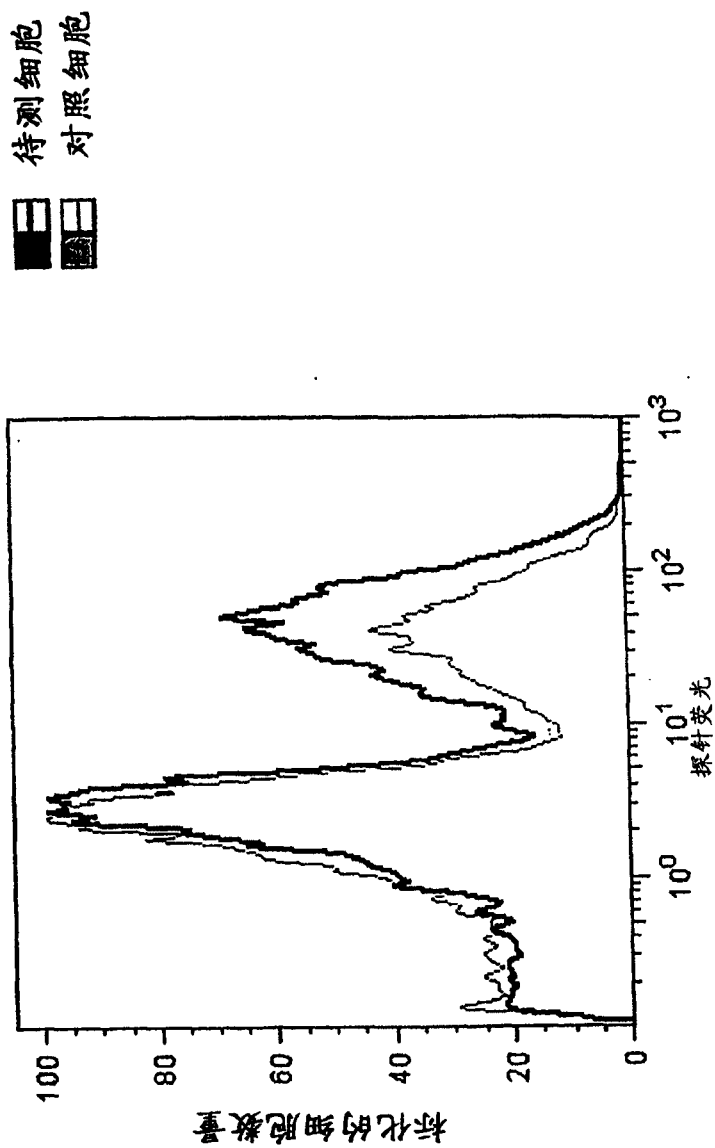


图 56 mcon11

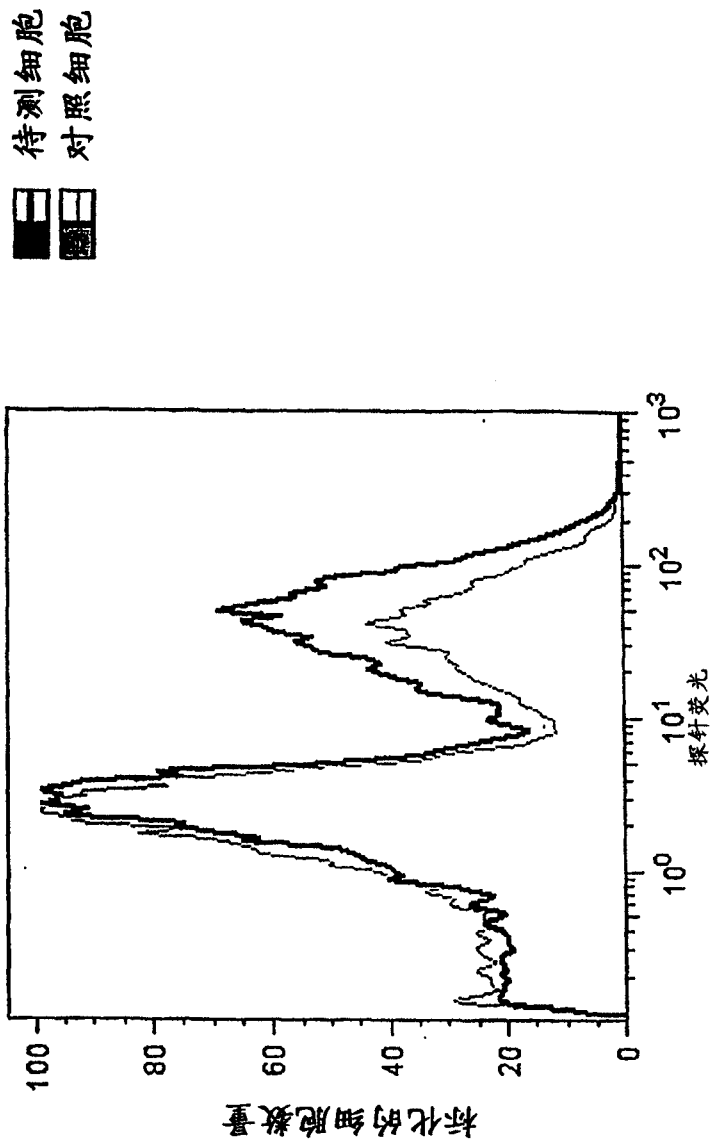


图 56 mcon11

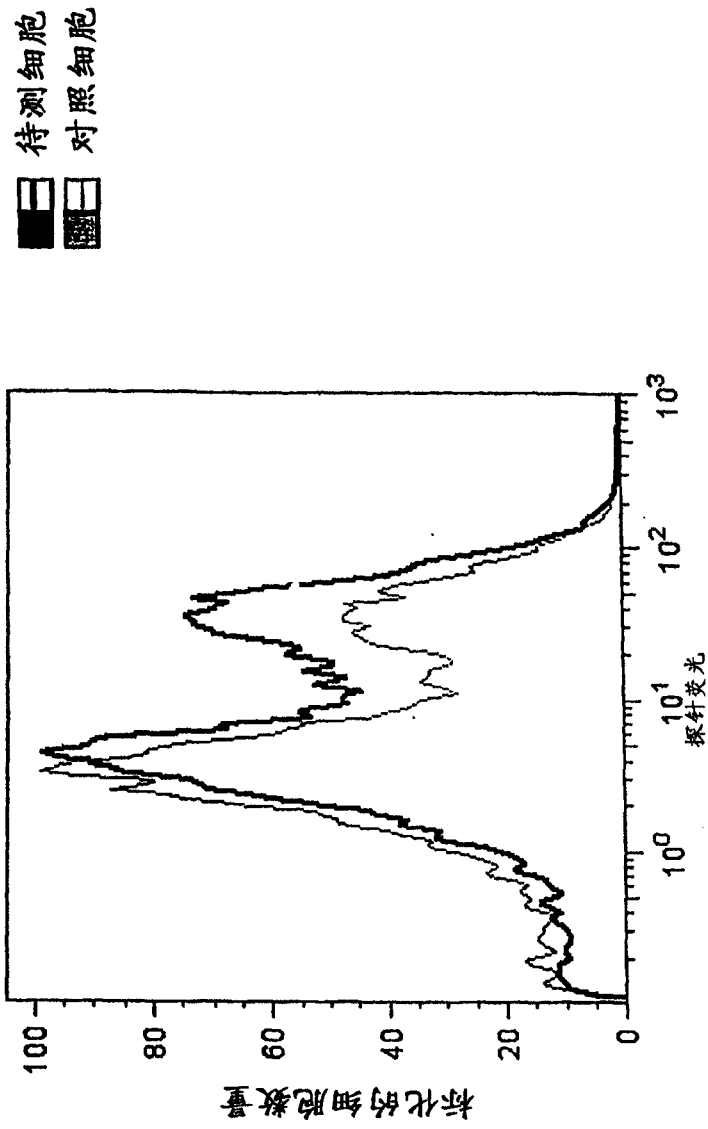


图 57 mcon12

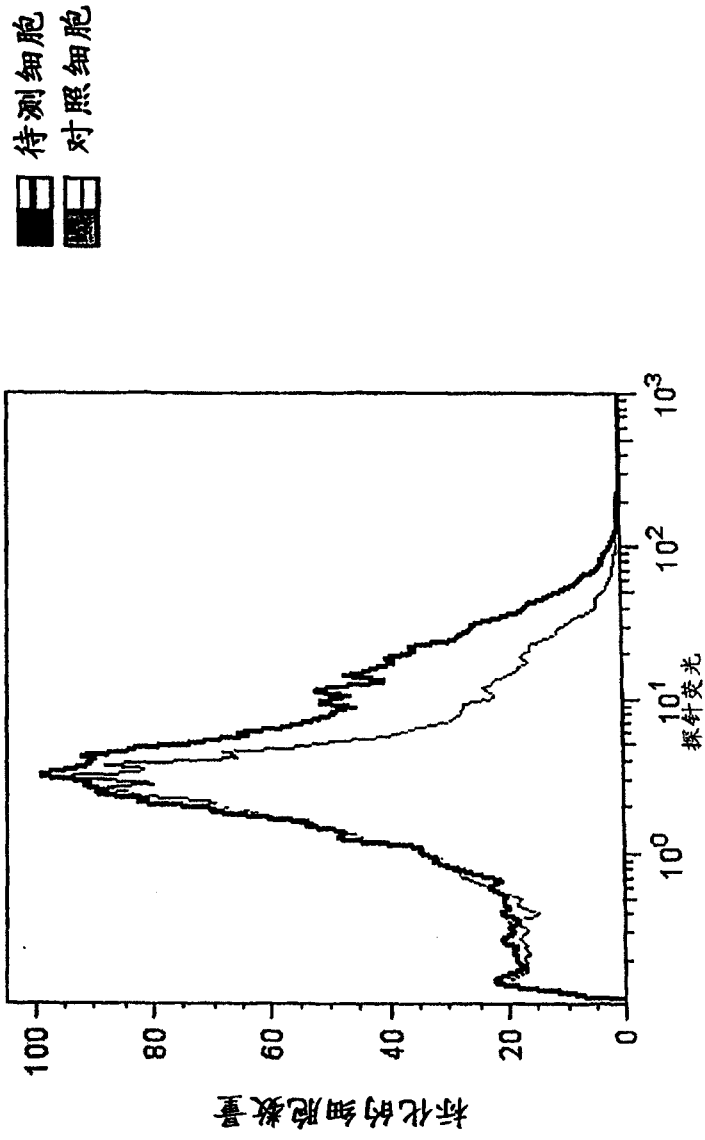


图 58 mcon13

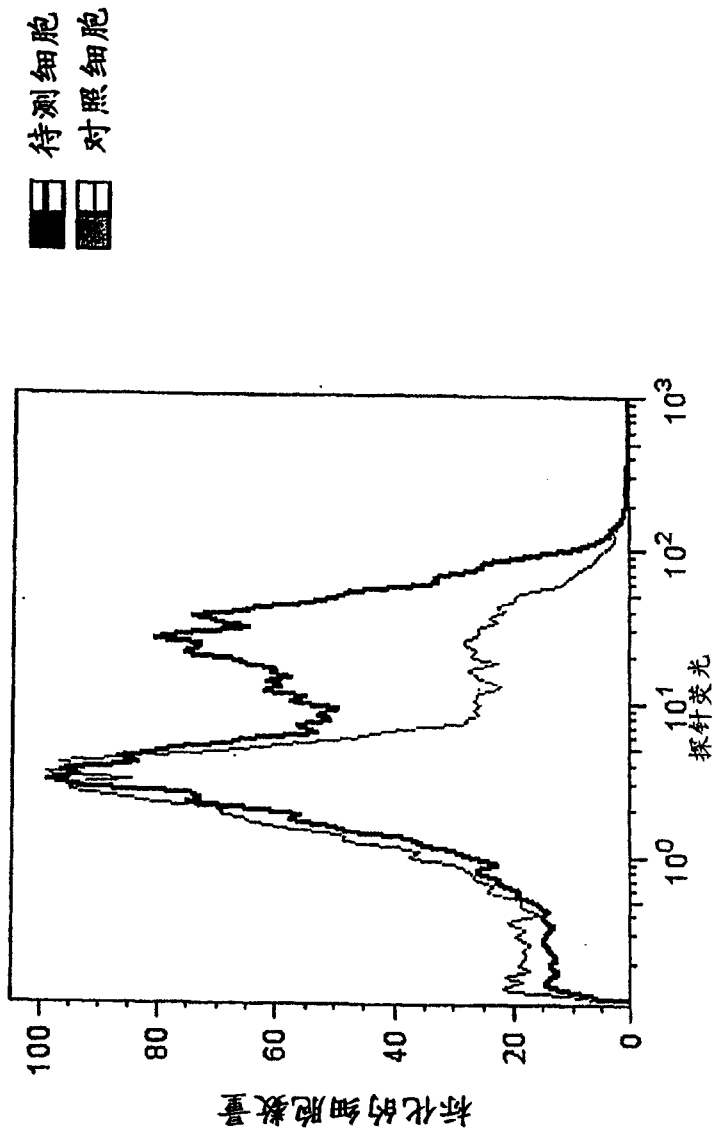


图 59 mcon14

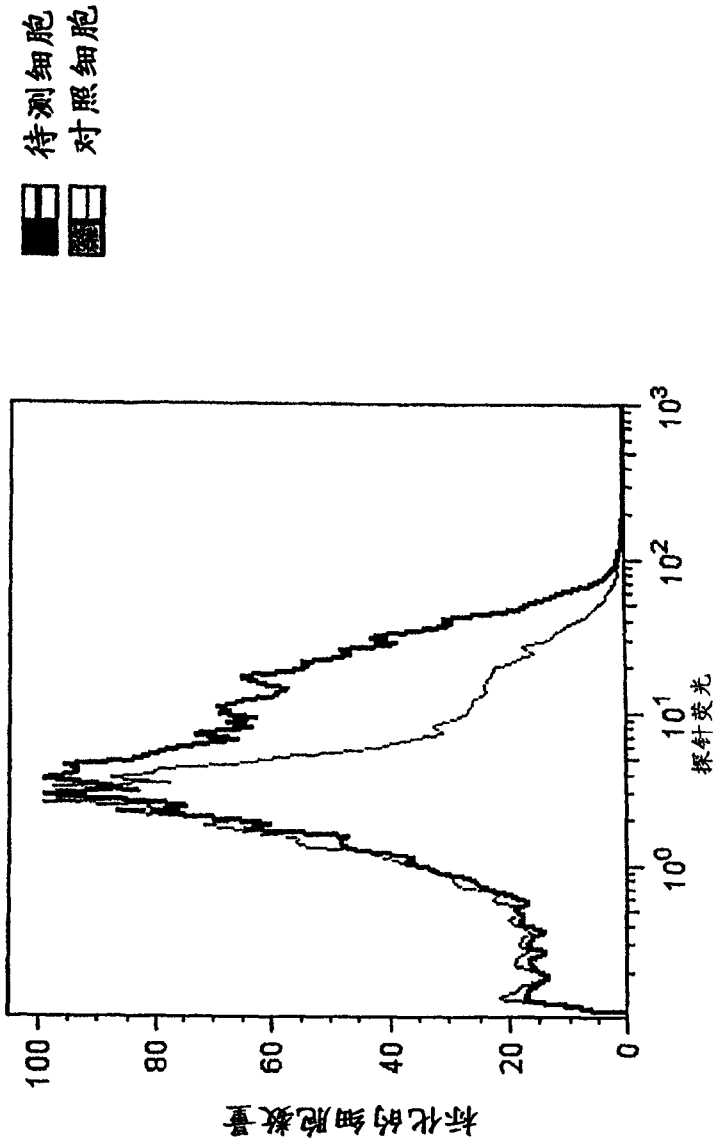


图60 mcon15