



1. 一种分离的多核苷酸，包含核苷酸序列，所述核苷酸序列至少95%相同于选自由下述组成的组的序列：

(a) 编码棘皮动物微管结合蛋白样 4/间变性淋巴瘤激酶 (EML4-ALK)融合多肽的核苷酸序列，所述融合多肽包括 SEQ ID NO: 1 或 SEQ ID NO: 18 的氨基酸序列；

(b) 编码 EML4-ALK 融合多肽的核苷酸序列，所述核苷酸序列包括 SEQ ID NO: 2 或 SEQ ID NO: 19 的核苷酸序列；

(c) 编码 EML4-ALK 融合多肽的核苷酸序列，所述融合多肽包括 EML-4 的 N 端氨基酸序列 (SEQ ID NO: 3 的残基 1-233 或 SEQ ID NO: 3 的残基) 以及 ALK 的激酶结构域 (SEQ ID NO: 5 的残基 1116-1383)；

(d) 包括 EML-4 的 N 端核苷酸序列 (SEQ ID NO: 4 的核苷酸 1-700 或 SEQ ID NO: 4 的核苷酸 1-1486) 以及 ALK 的激酶结构域核苷酸序列 (SEQ ID NO: 6 的核苷酸 3348-4149) 的核苷酸序列；

(e) 包括围绕 EML4-ALK 融合多核苷酸的融合接头 (SEQ ID NO: 2 的核苷酸 700-701 或 SEQ ID NO: 19 的核苷酸 1486-1487) 的至少 6 个邻接核苷酸的核苷酸序列；

(f) 编码包括围绕 EML4-ALK 融合多肽的融合接头 (SEQ ID NO: 1 的残基 233-234 或 SEQ ID NO: 18 的残基 495-496) 的至少 6 个邻接氨基酸的多肽的核苷酸序列； 以及

(g) 互补于 (a) - (f) 的任何核苷酸序列的核苷酸序列。

2. 一种分离的多核苷酸,其在严格杂交条件下杂交于根据权利要求1所述的多核苷酸,其中,所述进行杂交的分离的多核苷酸在严格杂交条件下并不杂交于具有仅由 A 残基或仅由 T 残基构成的核苷酸序列的多核苷酸。
3. 根据权利要求2所述的分离的多核苷酸,其中,所述多核苷酸进一步包括可检测标记。
4. 一种用于产生重组载体的方法,包括将根据权利要求1所述的分离的核酸分子插入到载体中。
5. 一种通过根据权利要求4所述的方法产生的重组载体。
6. 一种用于制备重组宿主细胞的方法,包括将根据权利要求5所述的重组载体引入到宿主细胞中。
7. 一种通过根据权利要求6所述的方法产生的重组宿主细胞。
8. 一种用于产生重组 EML4-ALK 融合多肽或截短的活性 ALK 多肽的方法,所述方法包括在适合于表达所述融合多肽的条件下培养根据权利要求7所述的重组宿主细胞以及回收所述多肽。
9. 一种分离的多肽,包括氨基酸序列,所述氨基酸序列至少 95% 相同于选自由下述组成的组的序列:
  - (a) 编码包括 SEQ ID NO: 1 或 SEQ ID NO: 18 的氨基酸序列的 EML4-ALK 融合多肽的氨基酸序列;
  - (b) 编码包括 EML-4 的 N 端氨基酸序列(SEQ ID NO: 3 的残基 1-233 或 SEQ ID NO: 3 的残基 1-495)和 ALK 的激酶结

构域(SEQ ID NO: 5 的残基 1116-1383)的 EML4-ALK 融合多肽的氨基酸序列; 以及

(c)编码包括围绕 EML4-ALK 融合多肽的融合接头(SEQ ID NO: 1 的残基 233-234 或 SEQ ID NO: 18 的残基 495-496)的至少 6 个邻接氨基酸的多肽的氨基酸序列。

10. 一种利用根据权利要求 5 所述的重组载体或根据权利要求 7 所述的重组宿主细胞产生的重组 EML4-ALK 融合多肽或截短的活性 ALK 多肽。
11. 一种分离的试剂, 其特异性地结合于或检测根据权利要求 9 所述的 EML4-ALK 融合多肽, 但并不结合于或检测野生型 EML-4 或野生型 ALK。
12. 根据权利要求 11 所述的分离的试剂, 其中, 所述试剂是抗体或重同位素标记 (AQUA) 肽。
13. 根据权利要求 11 所述的分离的试剂, 其中, 所述试剂是聚合酶链反应 (PCR) 探针或荧光原位杂交 (FISH) 探针。
14. 根据权利要求 12 所述的重同位素标记 (AQUA) 肽, 其中, 所述肽包括野生型 ALK 中的 EML4-ALK 融合多肽或截短点的融合接头的氨基酸序列。
15. 一种用于在来自哺乳动物癌症的生物样品中检测突变体 ALK 多核苷酸和/或其编码的突变体 ALK 多肽的存在的方法, 所述方法包括以下步骤:
  - (a) 从哺乳动物癌症获得生物样品; 以及
  - (b) 利用至少一种试剂, 所述试剂检测融合多核苷酸、或其编码的融合多肽, 包含具有部分二级蛋白质的部分 ALK,

- 以确定在所述生物样品中是否存在 ALK 突变体多核苷酸和/或其编码的突变体 ALK 多肽。
16. 根据权利要求 15 所述的方法，其中，所述癌症是实体瘤肉瘤或癌。
  17. 根据权利要求 16 所述的方法，其中，所述癌是肺癌。
  18. 根据权利要求 17 所述的方法，其中，所述肺癌是非小细胞肺癌(NSCLC)。
  19. 根据权利要求 15 所述的方法，其中，所述突变体 ALK 多肽是融合多肽，所述融合多肽包括具有部分所述二级蛋白质的 ALK(SEQ ID NO: 5)的残基 1116-1383。
  20. 根据权利要求 15 或 16 所述的方法，其中，所述二级蛋白质选自 EML-4(SEQ ID NO: 3)和 TRK-融合基因(TFG)蛋白质(SEQ ID NO: 22)组成的组。
  21. 根据权利要求 20 所述的方法，其中，所述融合多肽包括 EML-4(SEQ ID NO: 3)的残基 1-233 或残基 1-495 或 TFG(SEQ ID NO: 22)的残基 1-138。
  22. 根据权利要求 15 所述的方法，其中，所述融合多核苷酸包括 EML4-ALK 融合多核苷酸(SEQ ID NOs: 2 或 19)或 TFG-ALK 融合多核苷酸(SEQ ID NO: 21)。
  23. 根据权利要求 15 所述的方法，其中，所述融合多肽包括 EML4-ALK 融合多肽(SEQ ID NO: 1 或 18)或 TFG-ALK 融合多肽(SEQ ID NO: 20)。

24. 根据权利要求 15 所述的方法，其中，所述融合多核苷酸是根据权利要求 1 所述的融合多核苷酸。
25. 根据权利要求 15 所述的方法，其中，所述融合多肽是根据权利要求 9 所述的融合多肽。
26. 根据权利要求 15 所述的方法，其中，所述试剂包括根据权利要求 1 所述的多核苷酸和/或根据权利要求 11 所述的至少一种试剂。
27. 根据权利要求 15 所述的方法，其中，所述试剂包括分离的试剂，所述分离的试剂特异性地结合于或检测 TFG-ALK 融合多肽(SEQ ID NO: 20)或 TFG-ALK 融合多核苷酸(SEQ ID NO: 21)，但并不结合于或检测野生型 TFG 或野生型 ALK。
28. 根据权利要求 27 所述的方法，其中，所述试剂是抗体或重同位素标记 (AQUA) 肽。
29. 根据权利要求 27 所述的方法，其中，所述试剂是聚合酶链反应 (PCR) 探针或荧光原位杂交 (FISH) 探针。
30. 根据权利要求 28 所述的方法，其中，所述重同位素标记 (AQUA) 肽包括野生型 ALK 中的 TFG-ALK 融合多肽或截短点的融合接头的氨基酸序列。
31. 根据权利要求 15 所述的方法，其中，所述方法以流式细胞术 (FC)、免疫组织化学 (IHC)、或免疫荧光 (IF) 测定形式来实施。
32. 根据权利要求 15 所述的方法，其中，所述方法以荧光原位杂交 (FISH) 或聚合酶链反应 (PCR) 测定形式来实施。

33. 根据权利要求 15 所述的方法, 其中, 检测所述 ALK 融合多肽的活性。
34. 一种用于确定一种化合物是否抑制以 ALK 融合多肽的表达为特征的哺乳动物实体瘤的进展的方法, 所述方法包括确定所述化合物是否在所述癌症中抑制所述 ALK 融合多肽的表达和/或活性的步骤。
35. 根据权利要求 34 所述的方法, 其中, 所述 ALK 融合多肽包括 ALK(SEQ ID NO: 5)的残基 1116-1383 和一部分所述二级蛋白质。
36. 根据权利要求 34 所述的方法, 其中, 所述二级蛋白质选自由 EML-4(SEQ ID NO: 3)和 TRK-融合基因(TFG)蛋白质(SEQ ID NO: 22)组成的组。
37. 根据权利要求 36 所述的方法, 其中, 所述融合多肽包括 EML-4(SEQ ID NO: 3)的残基 1-233 或残基 1-495 或 TFG(SEQ ID NO: 22)的残基 1-138。
38. 根据权利要求 34 所述的方法, 其中, 利用至少一种检测根据权利要求 1 所述的多核苷酸的试剂、和/或至少一种根据权利要求 11 所述的试剂、和/或至少一种检测 TFG-ALK 融合多核苷酸或多肽的试剂, 来确定所述 ALK 融合多肽的表达和/或活性的抑制。
39. 一种用于对表达 EML4-ALK 融合多肽的癌症的进展进行抑制的方法, 所述方法包括抑制所述 EML4-ALK 融合多肽在所述癌症中的表达和/或活性的步骤。

- 
40. 一种用于对表达 TFG-ALK 融合多肽的实体瘤的进展进行抑制的方法, 所述方法包括抑制所述 TFG-ALK 融合多肽在所述癌症中的表达和/或活性的步骤。
  41. 根据权利要求 39 或 40 所述的方法, 其中, 所述癌症或所述实体瘤是肺癌。
  42. 根据权利要求 41 所述的方法, 其中, 所述肺癌是非小细胞肺癌(NSCLC)。
  43. 根据权利要求 39 或 40 所述的方法, 其中, 用包括 WHI-131 和/或 WHI-154、或它们的类似物的组合物来抑制所述 EML4-ALK 融合多肽或所述 TFG-ALK 融合多肽的表达和/或活性。

## 在人实体瘤中的基因缺损和突变体 ALK 激酶

### 相关申请

本申请要求 2006 年 4 月 14 日提交的目前未决的 USSN 60/792,364 的优先权和权益,将其全部披露内容结合于此作为参考。

### 技术领域

本发明通常涉及与癌症有关的蛋白质和基因,以及涉及癌症的检测、诊断和治疗。

### 背景技术

许多癌症以细胞信号通路的破坏为特征,其导致细胞过程的异常控制,或导致不能控制细胞的生长和增殖。这些破坏经常由特定信号蛋白(如激酶)活性的变化所引起。这些癌症包括实体瘤,如非小细胞肺癌(NSCLC)。在美国,NSCLC 是癌症死亡的主要原因,并且占有所有肺癌的约 87%。在美国每年有约 151,000 个 NSCLC 新病例,并且据估计仅在美国每年 120,000 以上的患者将死于该疾病。参见“*Cancer Facts and Figures 2005*,” American Cancer Society。包括三种不同亚型的 NSCLC 经常仅在它已经转移以后被检测到,因此在诊断的两年内死亡率是 75%。

已知,导致具有异常信号活性的激酶融合蛋白的基因缺失和/或易位可以直接导致某些癌症。例如,已直接证明,BCR-ABL 癌蛋白(一种酪氨酸激酶融合蛋白)是人慢性髓细胞性白血病(CML)的病原体。在至少 90-95%的 CML 病例中发现的 BCR-ABL 癌蛋白是通过基因序列从第 9 号染色体上的 c-ABL 蛋白质酪氨酸激酶易位

到第 22 号染色体上的 BCR 序列而产生的，这产生所谓的费城染色体。参见例如，Kurzock *et al.*, *N. Engl. J. Med.* 319: 990-998 (1988)。在急性淋巴细胞白血病和 NSCLC 病例中也观测到易位。

已描述了导致与各种其它癌症有关的突变体或融合蛋白的基因易位和缺失。例如，Falini *et al.*, *Blood* 99(2): 409-426 (2002)，综述了已知发生在血液癌症中的易位，包括在 ALCL 中发现的 NPM-ALK 融合。到目前为止，仅描述了发生在肺癌中的有限数目的基因易位、缺失、以及突变体蛋白质，包括涉及 Notch3 的 t(15;19) 易位。参见 Dang *et al.*, *J. Natl. Can. Instit.* 92(16): 1355-1357 (2000)。在小细胞和非小细胞肺癌中已发现 RNA 结合蛋白-6(EML-4)表达和/或活性方面的缺损。参见 Drabkin *et al.*, *Oncogene* 8(16): 2589-97 (1999)。然而，到目前为止，还没有描述在人 NSCLC 癌症中涉及蛋白激酶的易位或缺失。

已描述了在间变性大细胞淋巴瘤中由于 NPM 与 ALK 的融合所导致的 ALK 激酶表达方面的缺损。参见 Morris *et al.*, 1994; Shiota *et al.*, 1994。已描述了 ALK 与膜突蛋白、非肌球蛋白重链 9(Tort *et al.* 2001)、网格蛋白 (clarthrin) 重链 (Touriol *et al.*, 2000; Bridge *et al.*, 2001)、原肌球蛋白 3(TPM3)(Lamant *et al.*, 1999)、TRK-融合基因 (TGF)(Hernandez *et al.*, *Am. J. Path.* 160(4): 1487-1493 (2002))以及其它基因的融合。尤其是，据报道在非实体淋巴瘤中有 TGF-ALK 融合，但到目前为止，还未在实体瘤中描述这种融合。已描述了 ALK 在癌症中的一般作用。参见 Pulford *et al.*, *J. Cell Physiol.* 199(3): 330-358 (2004)。然而，到目前为止，还没有描述在 EML-4 表达和/或激活方面的缺损。

高度期望鉴定在人癌症中的突变，因为它可以导致开发新疗法（治疗剂），其靶向这样的融合或突变体蛋白质，以及导致新诊断方法，用于鉴定具有这样的基因突变的患者。例如，BCR-ABL 已变成开发治疗白血病的治疗剂（疗法）的靶。最近，Gleevec®（甲磺酸伊马替尼，STI-571），ABL 激酶的一种小分子抑制剂，已批准用于治疗 CML。这种药物是新类抗增殖剂的第一种药物，其用来

对驱动肿瘤细胞生长的信号通路进行干扰。这种药物的开发相对于 CML 和 ALL 的常规疗法(化疗和放射)而言代表一种显著的进步,其中常规疗法受到众所周知的副作用的烦扰并且经常具有有限的效果,因为它们不能特异性地靶向癌(恶性肿瘤)的基本的原因。同样,已描述了用于特异性地检测患者中的 BCR-ABL 融合蛋白的试剂和方法,以便鉴定最可能应答靶向抑制剂如 Gleevec®的患者。

因此,仍然需要鉴定新的基因突变,如易位或缺失,其导致与人癌症,尤其是实体瘤,包括肺癌如 NSCLC 的进展有关的融合或突变体蛋白质,以及需要开发新的试剂和方法,用于研究和检测这样的融合蛋白。这样的融合蛋白的鉴定将尤其期望产生新方法来选择用于靶向治疗的患者,以及用于筛选能抑制这样的突变体/融合蛋白的新药物。

## 发明内容

根据本发明,在人实体瘤非小细胞肺癌(NSCLC)中现已鉴定了发生在人第2号染色体中的新的基因缺失突变,其导致融合蛋白,该融合蛋白结合部分间变性淋巴瘤激酶(ALK)和二级蛋白质。与 ALK 融合有关的二级蛋白质包括棘皮动物微管结合蛋白样(棘皮类微管关联蛋白样, Echinoderm Microtubule-Associated Protein-Like)4(EML-4)和 TRK-融合基因(TFG)。在非小细胞肺癌患者样品中目前已观测到突变体/融合 ALK 激酶。

因此本发明部分地提供了分离的多核苷酸和载体,其编码所披露的突变体/融合 ALK 多肽,用于检测它们的探针和测定方法,分离的突变体/融合 ALK 多肽,重组突变体多肽,以及用于检测突变体 ALK 多核苷酸和多肽的试剂。所披露的这些新的突变体 ALK 激酶和易位/缺失的鉴定使得可以用新方法来确定在生物样品中突变体 ALK 多核苷酸或多肽的存在,本发明还提供了对抑制突变体激酶蛋白的化合物进行筛选的方法,以及对以突变体 ALK 多核苷酸或多肽的表达为特征的癌症的进展进行抑制的方法。下面更详细地描述本发明的方面和实施方式。

## 附图说明

图 1A 示出了 EML-4 基因和 ALK 基因在第 2 号染色体上的位置 (图片 A), 全长 EML-4 和 ALK 蛋白质的结构域位置, 以及 EML4-ALK 融合蛋白 (短变体) 的结构域位置 (图片 B); 融合接头存在于氨基酸 233-234, 并且融合蛋白包括 ALK 的激酶结构域 (但不包括跨膜和胞外结构域)。还示出了 (在图片 B 中) EML4 外显子 6/内含子 6/ALK 外显子 20 融合接头区的 DNA (和蛋白质) 序列 (分别为 SEQ ID NO:7 和 SEQ ID NO:8)。

图 1B 示出了 EML-4 基因和 ALK 基因在第 2 号染色体上的位置 (图片 A), 和全长 EML-4 和 ALK 蛋白质的结构域位置, 以及 EML4-ALK 融合蛋白 (长变体) 的结构域位置 (图片 B); 融合接头存在于氨基酸 495-496, 并且融合蛋白包括 ALK 的激酶结构域 (但不包括跨膜和胞外结构域)。还示出了 (在图片 B 中) EML4 外显子 13/ALK 外显子 20 融合接头区的 DNA (和蛋白质) 序列 (分别为 SEQ ID NO:24 和 SEQ ID NO:25)。

图 1C 示出了 TFG 基因在第 6 号染色体上和 ALK 基因在第 2 号染色体上的位置 (图片 A), 和全长 TFG 和 ALK 蛋白质的结构域位置, 以及 TFG-ALK 融合蛋白的结构域位置 (图片 B); 融合接头存在于氨基酸 138-139, 并且融合蛋白包括 ALK 的激酶结构域 (但不包括跨膜和胞外结构域)。还示出了 (在图片 B 中) TFG 外显子 3/ALK 外显子 20 融合接头区的 DNA (和蛋白质) 序列 (分别为 SEQ ID NO:26 和 SEQ ID NO:27)。

图 2A 是人 EML4-ALK 融合蛋白 (短变体) 的氨基酸序列 (1 字母代码) (SEQ ID NO:1) (顶部图片), 还示出了编码 DNA 序列 (SEQ ID NO:2) (底部图片); EML-4 部分的残基用斜体表示, 而 ALK 的激酶结构域的残基用黑体表示。

图 2B 是人 EML4-ALK 融合蛋白 (长变体) 的氨基酸序列 (1 字母代码) (SEQ ID NO:18) (顶部图片), 还示出了编码 DNA 序列

(SEQ ID NO:19) (底部图片); EML-4 部分的残基用斜体表示, 而 ALK 的激酶结构域的残基用黑体表示。

图 2C 是人 TFG-ALK 融合蛋白的氨基酸序列(1 字母代码)(SEQ ID NO:20)(顶部图片), 还示出了编码 DNA 序列(SEQ ID NO:21)(底部图片); TFG 部分的残基用斜体表示, 而 ALK 的激酶结构域的残基用黑体表示。

图 3A-3B 是人 EML-4 蛋白质的氨基酸序列(1 字母代码)(SEQ ID NO:3)(SwissProt Accession No. 061936), 还示出了编码 DNA 序列(SEQ ID NO:4)(GeneBank Accession No. NM019063); 保留在短变体缺失突变体中的残基被加上下划线, 而保留在长变体中的残基用斜体表示。

图 4A-4B 是人 ALK 激酶的氨基酸序列(1 字母代码)(SEQ ID NO:5)(SwissProt Accession No. Q9UM73), 还示出了编码 DNA 序列(SEQ ID NO:6)(GeneBank Accession No. HSU66559); 保留在缺失突变体中的残基被加上下划线, 而激酶结构域的残基用粗体表示。

图 4C-4D 是人 TFG 蛋白质的氨基酸序列(1 字母代码)(SEQ ID NO:22)(SwissProt Accession No. Q92734), 还示出了编码 DNA 序列(SEQ ID NO:23)(GeneBank Accession No. NM006070); 保留在缺失突变体中的残基被加上下划线。

图 5 是凝胶图 (gels), 其示出了 (A) 借助于 5'RACE 产物并通过 ALK 引物在 2 轮 PCR 以后 ALK 的检测; UAP 表示通用扩增引物, GSP 表示基因特异性引物, (B) 通过 EML-4 形成的融合基因和通过 RT-PCR 形成的 ALK 缺失突变体的检测, (C) 在人 NSCLC 肿瘤样品中通过 5'RACE 对 EML4-ALK 融合基因 (短和长变体) 的检测, 以及 (D) 在人 NSCLC 肿瘤样品中通过 5'RACE 对 TFG-ALK 融合基因的检测。

图 6 示出了通过采用双色（橙色/绿色）断裂探针（包括 ALK 基因断裂点 2p23 相反侧的探针）的 FISH 测定法，对在 H2228 细胞中通过 EML-4 和 ALK 易位所形成的融合基因的检测；探针大小和位置示于上部图片中。

### 具体实施方式

根据本发明，在人实体瘤非小细胞肺癌（NSCLC）中现已鉴定了导致突变体激酶融合蛋白的先前未知的基因缺失和易位，其中突变体激酶融合蛋白结合部分间变性淋巴瘤激酶(ALK)和部分二级蛋白质。与所发现的 ALK 融合有关的二级蛋白质包括棘皮动物微管结合蛋白样 4（EML-4）和 TRK-融合基因（TFG）。

两种披露的发生在第 2 号染色体上的 EML4 和 ALK 基因之间的缺失产生融合蛋白，其结合 EML-4（401 个氨基酸微管结合蛋白）的 N 端与 ALK（1620 个氨基酸膜酪氨酸激酶）的激酶结构域和 C 端。所得到的 EML4-ALK 融合蛋白（其分别为 796 个氨基酸(短变体)和 1059 个氨基酸(长变体)并保留 ALK 激酶活性）预期会驱动人实体瘤亚型（包括 NSCLC）的增殖和存活。

所披露的发生在第 6 号染色体上的 TFG 基因和第 2 号染色体上的 ALK 基因之间的易位产生融合蛋白，其结合 TFG（一种 400 个氨基酸蛋白质）的 N 端与 ALK（一种 1620 个氨基酸膜酪氨酸激酶）的激酶结构域和 C 端。所得到的 TFG-ALK 融合蛋白（其是 701 个氨基酸）先前在非实体人淋巴瘤中已观测到（Hernandez *et al.* (2002), 上文），但先前在实体瘤中没有描述。TFG-ALK 融合蛋白保留 ALK 激酶活性，并且预期会驱动人实体瘤亚型（包括 NSCLC）的增殖和存活。

虽然在 NSCLC 中已描述了几种导致异常融合蛋白的基因易位或缺失，包括涉及 Notch3 的 t(15;19)易位（参见 Dang *et al.*, 上文），但目前披露的 EML4-ALK 缺失突变体和融合蛋白是新的。类似地，TFG-ALK 易位突变体和融合蛋白，虽然在非实体瘤如淋巴瘤中是

已知的，但在实体瘤 NSCLC 中是新的。EML-4 是一种表达在大多数组织中的微管相关蛋白。到目前为止，还没有报道在 EML-4 表达和/或活性方面的缺损。ALK 是一种膜酪氨酸激酶，并且表达在人类的脑、CNS 组织、以及小肠和睾丸中，但不表达在正常淋巴细胞中。在神经系统的正常发育和功能中它起重要的作用 (Iwahara *et al.*, 1997)。

在间变性大细胞淋巴瘤和成神经细胞瘤中已发现在 ALK 表达和/或激活方面的缺损 (参见 Morris *et al.*, 1994, Osajima-Hakomori *et al.*, 2005)。已描述了 ALK 与膜突蛋白、非肌球蛋白重链 9、网格蛋白重链、原肌球蛋白 3(TPM3)、TRK-融合基因(TFG)、以及其它基因的融合。参见 Tort *et al.*; Touriol *et al.*, Hernandez *et al.*, 上文。有趣的是，所披露的 EML-4 与 ALK 的融合 (短变体) 精确地存在于野生型 (wild-type) ALK 中的相同点 (氨基酸 1058)，如先前针对其它 ALK 融合突变体所描述的。

如下文进一步描述的，目前已分离和测序了 EML4-ALK 缺失突变体和表达的融合蛋白，并且产生了用于表达融合蛋白的 cDNA。因此，本发明部分地提供了：编码 EML4-ALK 融合多肽的分离的多核苷酸；杂交于这样的多核苷酸的核酸探针；以及利用这样的多核苷酸来产生重组突变体 ALK 多肽的方法、载体以及宿主细胞。本发明还部分地提供了：分离的多肽，其包含编码 EML4-ALK 融合多肽的氨基酸序列；重组突变体多肽；以及分离的试剂，其特异性地结合于和/或检测 EML4-ALK 融合多肽，但不结合于或检测野生型 EML-4 或野生型 ALK。本发明的这些方面 (其将在下文进一步详细地描述) 将尤其可用于进一步研究由突变体 ALK 激酶表达/活性驱动的癌症机制，用于鉴定实体瘤 (例如，癌，包括肺癌和肉瘤) 以及其它癌症，其以所披露的 ALK 缺失和易位突变和/或融合蛋白、或突变体 ALK 激酶的表达/活性为特征，以及可用于实施本发明的方法，如下文进一步描述的。

新型 ALK 激酶突变体以及基因缺失和易位突变的鉴定对于实体瘤如 NSCLC 的可能的诊断和治疗具有重要意义，其中实体瘤以

这些融合蛋白中的一种或多种为特征。NSCLC，例如，经常仅在它转移以后才检测到，因此在诊断两年内的死亡率为 75%。因此，高度期望能够尽早地鉴定具有基因突变（其可以导致 NSCLC）的患者。

因此，起因于基因缺失的 EML4-ALK 融合蛋白（短和长变体）和起因于基因易位的 TFG-ALK 融合蛋白（其预期会驱动实体瘤，NSCLC，的增殖和存活）的发现使得可以产生重要的新方法准确地鉴定哺乳动物的实体瘤，包括肺癌（如 NSCLC），以及其它癌症，其中表达 ALK 融合蛋白（如 EML4-ALK 或 TFG-ALK）。这些肿瘤很可能对突变体 ALK 蛋白如 WHI-131 或 WHI-154 的激酶活性的抑制剂起反应。尽早鉴定由突变体 ALK 激酶驱动的癌症的能力将大大有助于临床确定哪种疗法（治疗剂）或疗法组合将最适合于特定患者，从而有助于避免这样的靶向其它激酶的抑制剂处方，这些激酶事实上并不是驱动癌症的主要信号分子。

因此，本发明部分地提供了利用本发明的融合特异性和突变体特异性试剂来检测癌症中 ALK 突变体多核苷酸和/或融合多肽的存在的方法。这样的方法可以用来例如鉴定实体瘤，如 NSCLC，其可能对突变体蛋白质的 ALK 激酶活性的抑制剂起反应。本发明还部分地提供了用于确定化合物是否抑制以 EML4-ALK 融合多肽为特征的癌症的进展的方法。另外，本发明提供了一种通过抑制突变体多肽的表达和/或活性来抑制实体瘤进展的方法，其中实体瘤表达 EML4-ALK 融合多肽或 TFG-ALK 融合多肽。这样的方法将在下文进一步详细地描述。

下文将更详细地描述本发明的另外的方面、优点以及实施方式。将本文引用的所有参考文献的全部内部结合于此作为参考。

## 定义

如在本文中所使用的，以下术语具有指定的含义。

“抗体”是指所有类型的免疫球蛋白，包括 IgG、IgM、IgA、IgD 以及 IgE，包括 F<sub>ab</sub> 或其抗原识别片段，包括嵌合、多克隆以及单克隆抗体。如在本文中所使用的，术语“人源化抗体”是指这样的抗体分子，其中氨基酸已被非抗原结合区代替以便更接近地类似人抗体，同时仍然保留最初的结合能力。

术语“生物上活性的”是指具有天然存在分子的结构、调节、或生化功能的蛋白质。同样，“免疫学上活性的”是指天然、重组、或合成 EML4-ALK 或 TFG-ALK 融合多肽、或其任何寡肽诱导适当动物或细胞中的特异性免疫应答以及与特异性抗体结合的能力。

术语“生物样品”以其最广的含义使用，并且指任何怀疑包含 ALK 融合多核苷酸或多肽或其片段（包括 EML4-ALK 和 TFG-ALK 融合多核苷酸和多肽）的生物样品，并且可以包括细胞、分离自细胞的染色体（例如，一定范围的中期染色体）、基因组 DNA（在溶液中或结合于固体载体如用于 DNA 分析）、RNA（在溶液中或结合于固体载体如用于 RNA 分析）、cDNA（在溶液中或结合于固体载体）、来自细胞的提取物、血液、尿、骨髓、或组织等。

就癌症和突变体 ALK 多核苷酸以及多肽而论，“特征为”是指这样的癌症，其中涉及 ALK 的基因缺失或易位和/或表达的融合多肽是存在的，这是与其中这样的基因缺失和/或融合多肽不存在的癌症相比。突变体多肽的存在可以整体上或部分地驱动这样的癌症的生长和存活。

“共有序列”是指这样的核酸序列，其已被重测序以分离（分解）不需要的碱基（uncalled base），或其已利用 XL-PCR™（Perkin Elmer, Norwalk, Conn.）在 5' 和/或 3' 方向加以延伸并测序过，或其已利用 GELVIEW™ Fragment Assembly 系统（GCG, Madison, Wis.）装配自多于一种 Incyte 克隆的重叠序列，或其已被延伸和装配。

“ALK 激酶-抑制治疗剂”是指任何包含一种或多种化合物（化学的或生物的）的组合物，其直接或间接地、并单独地和/或作为融合蛋白的一部分（如 EML4-ALK 融合蛋白和 TFG-ALK 融合蛋白），抑制野生型或截短的 ALK 激酶的表达和/或活性。

“衍生物”是指编码所披露的融合多核苷酸的核酸序列的化学修饰或所编码的多肽本身。作为上述修饰的例证是用烷基、酰基、或氨基基团代替氢。核酸衍生物将编码保留天然分子的基本生物特性的多肽。

就本文披露的多肽、多核苷酸或试剂而论，“可检测标记”是指化学的、生物的、或其它修饰，包括但不限于荧光、质量、残基、染料、放射性同位素、标记、或标志修饰等，借此可以检测感兴趣分子的存在。

就生物样品中的 ALK 融合多肽而论，“表达”或“表达的”是指与其中该融合多肽未显著表达的对照样品相比显著表达。

“重同位素标记肽”（可与 AQUA 肽互换使用）是指包括至少一个重同位素标记的肽，其适合于绝对量化或检测蛋白质，如在 WO/03016861, “Absolute Quantification of Proteins and Modified Forms Thereof by Multistage Mass Spectrometry” (Gygi *et al.*) 中所描述的，并在下文进一步描述。就上述 AQUA 肽而论，术语“特异性地检测”是指肽将仅检测和量化包含 AQUA 肽序列的多肽和蛋白质并且将不会显著地检测并不包含 AQUA 肽序列的多肽和蛋白质。

“分离的”（或“基本上纯化的”）是指离开它们的天然环境并经分离的或隔离的核酸或氨基酸序列。它们优选为至少 60%、更优选 75%、以及最优选 90%或更大地不含它们天然伴随的其它成分。

“模拟物”是指这样的分子，按照 ALK 融合多肽或其部分的结构开发其结构，因此同样能够实施易位相关的蛋白质样分子的一些或所有作用。

如本文所描述的，“突变体 ALK”或“融合”多核苷酸或多肽是指涉及 ALK 和二级蛋白质（例如，EML-4 或 TFG）的融合多核苷酸或多肽。

“多核苷酸”（或“核苷酸序列”）是指寡核苷酸、核苷酸、或多核苷酸、以及其片断或部分，以及指基因组或合成起源 DNA 或 RNA，其可以是单链或双链的，并且表示有义链或反义链。

“多肽”（或“氨基酸序列”）是指寡肽、肽、多肽、或蛋白质序列、以及其片断或部分，并且指天然存在或合成分子。其中本文陈述的“氨基酸序列”是指天然存在的蛋白质分子的氨基酸序列，“氨基酸序列”和类似术语，如“多肽”或“蛋白质”，并不是将氨基酸序列限于与陈述的蛋白质分子有关的完全、天然氨基酸序列。

“EML4-ALK 融合多核苷酸”是指如本文描述的基本上纯化的 EML4-ALK 缺失突变体基因产物或融合多核苷酸（短或长变体）的核酸序列，其获自任何物种，尤其是哺乳动物，包括牛、羊、猪、鼠、马、以及优选人，并且来自任何来源：天然、合成、半合成、或重组。

“EML4-ALK 融合多肽”是指本文描述的基本上纯化的 EML4-ALK 融合多肽（短或长变体）的氨基酸序列，其获自任何物种，尤其是哺乳动物，包括牛、羊、猪、鼠、马、以及优选人，并且来自任何来源：天然、合成、半合成、或重组。

“TFG-ALK 融合多核苷酸”是指如本文描述的基本上纯化的 TFG-ALK 易位突变体基因产物或融合多核苷酸的核酸序列，其获自任何物种，尤其是哺乳动物，包括牛、羊、猪、鼠、马、以及优选人，并且来自任何来源：天然、合成、半合成、或重组。

“TFG-ALK 融合多肽”是指本文描述的基本上纯化的 TFG-ALK 融合多肽的氨基酸序列，其获自任何物种，尤其是哺乳动物，包括

牛、羊、猪、鼠、马、以及优选人，并且来自任何来源：天然、合成、半合成、或重组。

关于抗体和蛋白质或肽的相互作用，术语“特异性地结合于”（或“特异性地结合”或“特异的结合”）是指，该相互作用取决于在蛋白质上特定结构（即，抗原决定簇或表位）的存在；换言之，抗体识别并结合于特定的蛋白结构而不是一般地结合于蛋白质。就抗体结合于序列或抗原决定簇而不是结合于它对其是特异性的序列或抗原决定簇而论，术语“并不结合”是指并不显著地与其反应，这是与抗体结合于对其抗体是特异性的抗原决定簇或序列相比。

就序列或探针杂交条件而论，术语“严格条件”是“严格性”，其发生在约  $T_m - 5^\circ\text{C}$ （低于探针或序列的解链温度( $T_m$ ) $5^\circ\text{C}$ ）至低于  $T_m$  约  $20^\circ\text{C}$  至  $25^\circ\text{C}$  的范围内。典型的严格条件是：在  $42^\circ\text{C}$  下在溶液中过夜温育，其中溶液包含：50%甲酰胺、5X.SSC(750 mM NaCl, 75 mM 柠檬酸三钠)、50 mM 磷酸钠(pH 7.6)、5X Denhardt 溶液、10%硫酸右旋糖酐、以及 20 微克/ml 变性的、切变的鲑鱼精 DNA，接着在约  $65^\circ\text{C}$  的 0.1XSSC 中洗涤过滤器。如本领域技术人员将明了的，可以改变杂交的严格性以便鉴定或检测相同的或有关的多核苷酸序列。

突变体 ALK 多肽的“变体”是指被一个或多个氨基酸改变的氨基酸序列。该变体可以具有“保守”变化，其中取代的氨基酸具有类似的结构或化学性能，例如，用异亮氨酸代替亮氨酸。更例外地，变体可以具有“非保守”变化，例如，用色氨酸代替甘氨酸。类似的较小变化还可以包括氨基酸缺失或插入，或两者。利用本领域众所周知的计算机程序，例如 DNASTAR 软件，可以找到指导，其用来确定可以取代、插入、或缺失哪些氨基酸残基而不消除生物或免疫活性。

#### A. 在人实体瘤中的突变体 ALK 激酶的鉴定

在对来自非小细胞肺癌(NSCLC)细胞系(包括 H2228)的提取物和来自患者的实体瘤进行全面磷酸化肽分布实验过程中意外地鉴

定了本文披露的新的人基因缺失，其发生在第2号染色体上并导致两种结合具有激酶结构域的EML-4的N端和ALK的C端的融合蛋白变体的表达。NSCLC（一种实体瘤）是肺癌的一种亚型。在这些缺失融合中涉及的蛋白质示于图1A-1B的图片A中。

利用最近描述的用于分离和质谱表征来自复杂混合物的修饰肽的技术（参见美国专利公开号20030044848, Rush *et al.*, “Immunoaffinity Isolation of Modified Peptides from Complex Mixtures”）（“IAP”技术），首次阐述了H2228细胞系的磷酸化分布，如下文的实施例1中进一步描述的。应用IAP技术并利用磷酸酪氨酸特异性抗体（CELL SIGNALING TECHNOLOGY, INC., Beverly, MA, 2003/04 Cat. #9411）鉴定了H2228细胞系表达ALK激酶，但蛋白质被明显截短。该筛选鉴定了在细胞系中的许多其它激活的激酶，包括已知在肺癌中被激活的某些激酶。然后通过5' RACE进行的序列5'至ALK的分析鉴定了激酶被融合到EML-4的N端（参见图6）。

随后利用相同的全面磷酸基分布方式（phospho-profiling approach）对来自NSCLC患者的154个肿瘤样品进行的实验不仅证实那些患者群体中存在EML4-ALK（短变体）突变，而且揭示了在其它患者群体中存在第二种EML4-ALK（长变体）以及存在TFG-ALK突变（参见实施例1B和1C）。

通过利用siRNA沉默来抑制细胞可以证实，突变体ALK蛋白质在这些NSCLC肿瘤中正驱动细胞增殖和存活（参见实施例3）。

通过PCR扩增EML4-ALK融合基因（短和长变体）和TFG-ALK融合基因，然后分离并测序（参见实施例3）。如图1A-1B的图片B所示，EML4-ALK缺失结合具有激酶结构域的野生型EML-4（在短变体中的氨基酸1-233，或在长变体中的氨基酸1-495）N端与野生型ALK（氨基酸1057-1620）的C端（还参见SEQ ID NOs: 3和5）。融合接头正好存在于野生型ALK的C端至跨膜结构域（参见图1A-1B）。EML4-ALK融合多肽分别保留EML-4的N端233或495个氨基酸，其包括该蛋白质的卷曲螺旋结构域。所得到的分别包含796个氨基酸（短变体）或1059个氨基酸（长变体）（参见图

1A-1B 和图 2A-2B 的图片 B (SEQ ID NOs: 1 和 18) 的 EML4-ALK 融合蛋白可保留 ALK 的激酶活性。涉及的外显子以及融合接头示于图 1A-1B (图片 B) 中。融合接头包括来自 EML-4 的内含子 6, 其接着外显子 6 (短变体) 或来自 EML-4 (长变体) 的外显子 13。

如图 1C 的图片 B 所示, TFG-ALK 易位结合具有激酶结构域的野生型 TFG (氨基酸 1-138) 的 N 端与野生型 ALK (氨基酸 1057-1620) 的 C 端 (还参见 SEQ ID NOs: 22 和 5; 以及图 1C 的图片 B 和图 4C (SEQ ID NOs: 20 和 1))。融合接头正好存在于野生型 ALK 的 C 端至跨膜结构域 (参见图 1C) 并保留 ALK 的激酶活性。涉及的外显子以及融合接头示于图 1C (图片 B) 中。该融合接头包括来自 TFG 的外显子 3 和来自 ALK 的外显子 20。

FISH 探针用来在一组 400 个石蜡包埋的人 NSCLC 肿瘤样品中检测 EML4-ALK (短变体) 融合蛋白的存在 (参见实施例 6 和 7; 图 6)。在该样本量中这种短变体突变的发生率非常低。然而, 利用 IAP 技术来检查在另一组来自患者的 154 个冷冻人 NSCLC 肿瘤样品中的全面磷酸化分布, 则检测到 EML4-ALK 融合蛋白 (短和长变体两者)、以及 TFG-ALK 融合蛋白的表达具有更高的发生率 (参见实施例 1B)。

## B. 分离的多核苷酸

本发明部分地提供了: 编码 EML4-ALK 融合多肽的分离的多核苷酸; 杂交于上述多核苷酸的核苷酸探针, 以及用于利用上述多核苷酸来产生重组融合多肽的方法、载体、以及宿主细胞。

除非另有说明, 本文中通过测序 DNA 分子所确定的所有核苷酸序列是利用自动化的 DNA 测序仪 (如来自 Applied Biosystems, Inc. 的型号 373) 加以确定的, 并且由本文确定的 DNA 分子编码的多肽的所有氨基酸序列是利用自动化的肽测序仪加以确定的 (参见实施例 2)。如本技术领域已知的, 对于通过这种自动化方式确定的任何 DNA 序列, 本文确定的任何核苷酸序列可以包含一些误差。通

过自动化方法确定的核苷酸序列通常至少约 90%相同于，更通常至少约 95%至至少约 99.9%相同于测序的 DNA 分子的实际核苷酸序列。实际序列可以通过其它方式更准确地加以确定，包括本领域众所周知的人工 DNA 测序方法。如本技术领域同样已知的，和实际序列相比，在确定的核苷酸序列中的单插入或缺失将在核苷酸序列的翻译中引起移码，以致由确定的核苷酸序列编码的预测的氨基酸序列将完全不同于由测得的 DNA 分子实际编码的氨基酸序列，其开始于上述插入或缺失的点。

除非另有说明，本文陈述的每个核苷酸序列表示为脱氧核糖核苷酸（简写为 A、G、C 以及 T）的序列。然而，核酸分子或多核苷酸的“核苷酸序列”对于 DNA 分子或多核苷酸是指脱氧核糖核苷酸序列，而对于 RNA 分子或多核苷酸则是指核糖核苷酸（A、G、C 以及 U）的相应序列，其中在指定的脱氧核糖核苷酸序列中的每个胸苷脱氧核糖核苷酸(T)被核糖核苷酸尿苷(U)取代。例如，对照具有 SEQ ID NO: 2 的序列(利用脱氧核糖核苷酸简写表示)的 RNA 分子可用来指出 RNA 分子具有这样的序列，其中 SEQ ID NO: 2 的每个脱氧核糖核苷酸 A、G 或 C 已被相应的核糖核苷酸 A、G 或 C 取代，并且每个脱氧核糖核苷酸 T 已被核糖核苷酸 U 取代。

在一种实施方式中，本发明提供了一种分离的多核苷酸，该多核苷酸包含核苷酸序列，而该核苷酸序列至少 95%相同于选自由以下序列组成的组的序列：

(a) 编码棘皮动物微管结合蛋白样 4/间变性淋巴瘤激酶 (EML4-ALK) 融合多肽的核苷酸序列，包括 SEQ ID NO: 1 或 SEQ ID NO: 18 的氨基酸序列；

(b) 编码 EML4-ALK 融合多肽的核苷酸序列，所述核苷酸序列包括 SEQ ID NO: 2 或 SEQ ID NO: 19 的核苷酸序列；

(c) 编码 EML4-ALK 融合多肽的核苷酸序列，其中 EML4-ALK 融合多肽包含 EML-4 的 N 端氨基酸序列 (SEQ ID NO: 3 的残基

1-233 或 SEQ ID NO: 3 的残基 1-495 )以及 ALK 的激酶结构域( SEQ ID NO: 5 的残基 1116-1383 );

(d)核苷酸序列,包含 EML-4 的 N 端核苷酸序列( SEQ ID NO: 4 的核苷酸 1-700 或 SEQ ID NO: 4 的核苷酸 1-1486 ) 以及 ALK 的激酶结构域核苷酸序列 ( SEQ ID NO: 6 的核苷酸 3348-4149 );

(e)核苷酸序列,包含至少 6 个邻接核苷酸,其围绕 EML4-ALK 融合多核苷酸的融合接头 ( SEQ ID NO: 2 的核苷酸 700-701 或 SEQ ID NO: 19 的核苷酸 1486-1487 );

(f) 编码多肽的核苷酸序列, 包含至少 6 个邻接氨基酸, 其围绕 EML4-ALK 融合多肽的融合接头( SEQ ID NO: 1 的残基 233-234 或 SEQ ID NO: 18 的残基 495-496 ); 以及

(g) 核苷酸序列, 其互补于 (a) (-f) 的任何核苷酸序列。

利用本文提供的信息,如图 2 中的核苷酸序列( SEQ ID NO: 2 ), 并利用标准克隆和筛选方法,如利用 mRNA 作为原料来克隆 cDNA 的那些方法,可以获得本发明的核酸分子,其编码本发明的突变体 ALK 多肽。作为本发明的例证,在图 2 中描述的 EML4-ALK 融合多核苷酸(短变体)(SEQ ID NO: 2)分离自人 NSCLC 细胞系的基因组 DNA (如在在下面的实施例 2 中进一步描述的)。还可以在其它癌症(包括实体瘤)的基因组 DNA 或 cDNA 文库中鉴定融合基因,其中发生所披露的 EML4-ALK 基因缺失(第 2 号染色体)。

确定的 EML4-ALK 融合基因的核苷酸序列 ( SEQ ID NOs: 2 和 19 ) 分别编码 796 个氨基酸的激酶融合蛋白 (短变体) 和 1059 个氨基酸的激酶融合蛋白 (长变体) (参见图 2A-B(SEQ ID NOs: 1 和 18)和图 1A-B)。EML4-ALK 融合多核苷酸包括野生型 EML-4 的部分核苷酸序列 (参见图 3(SEQ ID NO: 4)), 其编码具有野生型 ALK 的部分核苷酸序列 (参见图 4(SEQ ID NO: 6)) 的蛋白质的 N 端 (氨基酸 1-233(短变体)或氨基酸 1-495(长变体)), 其中上述部分核苷酸序列编码上述蛋白质的激酶结构域和 C 端。参见图 1A-B。激酶结

构域包含在短变体融合蛋白（由短变体融合多核苷酸的核苷酸 874-1704 编码）中的残基 292-568 或在长变体融合蛋白（由长变体融合多核苷酸的核苷酸 1663-2494 编码）中的残基 555-831。参见图 2A-2B。

如所指出的，本发明部分地提供了 EML4-ALK 融合蛋白的成熟形式。根据信号假说，由哺乳动物细胞分泌的蛋白质具有信号或分泌前导序列，其在生长蛋白质链穿过糙面内质网的输出已被引发以后剪切自成熟蛋白质。大多数哺乳动物细胞并且甚至昆虫细胞在相同的特异性下剪切分泌的蛋白质。然而，在某些情况下，分泌蛋白质的剪切并不是完全均匀的，其导致两种或更多种蛋白质的成熟物质。另外，很久已经知道，分泌的蛋白质的剪切特异性最终由完全蛋白质的一级结构确定，即，它是多肽的氨基酸序列所固有的。

具有例如由沉积的 cDNA 克隆编码的氨基酸序列的成熟 EML4-ALK 多肽是指，通过在哺乳动物细胞（例如，3T3 细胞，如下所述）中表达完全可读框所产生的成熟形式的这种融合蛋白，其中完全可读框是由沉积的克隆或其它编码成熟融合多肽的克隆的人 DNA 序列编码的。

如所指出的，本发明的多核苷酸可以具有 RNA 的形式，如 mRNA，或具有 DNA 的形式，包括，例如，cDNA 和基因组 DNA（通过克隆获得或合成产生）。该 DNA 可以是双链或单链的。单链 DNA 或 RNA 可以是编码链，还称作有义链，或它可以是非编码链，还称作反义链。

本发明的分离的多核苷酸是核酸分子、DNA 或 RNA，其已离开它们的天然环境。例如，对于本发明来说，包含在载体中的重组 DNA 分子被认为是分离的。分离的 DNA 分子的另外的实例包括保持在异源宿主细胞中的重组 DNA 分子或在溶液中的经纯化（部分地或基本上）的 DNA 分子。分离的 RNA 分子包括本发明的 DNA 分子的体内或体外 RNA 转录物。根据本发明的分离的核酸分子进一步包括上述合成产生的分子。

本发明的分离的多核苷酸包括图 2A-B 所示的 DNA 分子 (SEQ ID NOs: 2 和 19); 图 1A-B 所示的包含用于成熟 EML4-ALK 融合蛋白的编码序列的 DNA 分子 (SEQ ID NOs: 1 和 18); 以及这样的 DNA 分子, 其包含显著不同于上述序列的序列, 但由于遗传密码的简并, 其仍然编码本发明的 ALK 突变体多肽。遗传密码在本技术领域是众所周知的, 因此本领域技术人员可以常规地产生上述简并变体。

在另一种实施方式中, 本发明提供了一种分离的多核苷酸, 该多核苷酸编码 EML4-ALK 融合多肽, 其包括包含在上述沉积的 cDNA 克隆中的 EML4-ALK 融合核苷酸序列。优选地, 上述核酸分子将编码由沉积的 cDNA 克隆或另一种克隆 (其表达本文描述的全长 EML4-ALK 融合蛋白) 编码的成熟融合多肽。在另一种实施方式中, 本发明提供了一种分离的核苷酸序列, 该核苷酸序列编码 EML4-ALK 融合多肽, 其包含 EML-4 的 N 端氨基酸序列 (SEQ ID NO: 3 的残基 1-233 或 SEQ ID NO: 3 的残基 1-495) 以及 ALK 的激酶结构域 (SEQ ID NO: 5 的残基 1116-1383)。在一种实施方式中, 包含 ALK 的激酶结构域的多肽包括 SEQ ID NO: 5 的残基 1057-1620 (参见图 1, 图片 B)。在另一种实施方式中, 上述 EML-4 的 N 端氨基酸序列和 ALK 的激酶结构域分别由包含 SEQ ID NO: 4 的核苷酸 1-700 或 SEQ ID NO: 4 的核苷酸 1-1486 和 SEQ ID NO: 6 的核苷酸 3171-4860 的核苷酸序列加以编码。

本发明进一步提供了分离的多核苷酸, 该多核苷酸包含这样的核苷酸序列, 其具有互补于本发明的突变体 ALK 多核苷酸之一的序列。上述分离的分子, 尤其是 DNA 分子, 可作为探针, 用于通过与染色体的原位杂交进行基因作图, 以及用于检测 EML4-ALK 融合蛋白在人组织中的表达, 例如, 通过 RNA 印迹分析, 如在下面的 F 部分中进一步描述的。

本发明进一步涉及本文描述的分离的核酸分子的片断。本发明的分离的 EML4-ALK 多核苷酸的片断是指长度至少约 15 个核苷酸的片断, 更优选至少约 20 个核苷酸, 进一步更优选至少约 30 个核

苷酸，并且甚至更优选至少约 40 个核苷酸，如本文所讨论的，其可以用作诊断探针和引物。当然，根据本发明，也可以使用长度约 50-1500 个核苷酸的更大片断，因为这些片断对应于大多数（如果不是所有）沉积的 cDNA 的突变体 ALK 核苷酸序列或如图 2 所示（SEQ ID NO: 2）或其它克隆，其表达融合多核苷酸，如图 2A-B 所示（SEQ ID NOs: 2 或 19）。长度至少 20 个核苷酸的片断，例如，是指这样的片断，其包括 20 个或更多邻接相应核苷酸序列（片断由其衍生）的碱基。本领域技术人员可以用常规方法产生上述 DNA 片断，并且可以例如通过限制性内切核酸酶剪切或声处理剪切可获自沉积的 cDNA 克隆的 DNA 来完成，或根据本文披露的序列加以合成。可替换地，可以直接合成产生上述片断。

本发明的优选的核酸片断或探针包括这样的核酸分子，其编码 EML4-ALK 融合基因产物的融合接头（参见图 1A-B，图片 B）。例如，在某些优选的实施方式中，本发明的分离的多核苷酸包括核苷酸序列/片断，其包含围绕 EML4-ALK 融合多核苷酸的融合接头（SEQ ID NO: 2 的核苷酸 700-701 或 SEQ ID NO: 19 的核苷酸 1486-1487）的至少 6 个邻接核苷酸（参见图 1A-B，图片 B（SEQ ID NOs: 8 和））。在另一种优选的实施方式中，本发明的分离的多核苷酸包括编码多肽的核苷酸序列/片断，该多肽包含围绕 EML4-ALK 融合多肽的融合接头（SEQ ID NO: 1 的残基 233-234 或 SEQ ID NO: 18 的残基 495-496）的至少 6 个邻接氨基酸（参见图 1A-B，底部图片（SEQ ID NOs: 7 和 24））。

在另一个方面，本发明提供了一种分离的多核苷酸，其在严格杂交条件下杂交于如本文所描述的本发明的突变体 ALK 多核苷酸的一部分。“严格杂交条件”是指在 42°C 下在溶液中过夜温育，其中所述溶液包含：50% 甲酰胺、5 X.SSC（750 mM NaCl、75 mM 柠檬酸三钠）、50 mM 磷酸钠（pH 7.6）、5X Denhardt 溶液、10% 硫酸右旋糖酐、以及 20 微克/ml 变性的、切变的鲑鱼精 DNA，接着在约 65°C 下在 0.1X SSC 中洗涤过滤器。

杂交于多核苷酸的“一部分”的多核苷酸是指这样的多核苷酸(DNA或RNA),其杂交于参比多核苷酸的至少约15个核苷酸(nt),更优选至少约20个nt,进一步更优选至少约30个nt,以及甚至更优选约30-70个nt。如在上文所讨论的和在下文更详细描述,这些杂交于多核苷酸的“一部分”的多核苷酸可用作诊断探针和引物。

当然,根据本发明,杂交于更大部分参比多核苷酸(例如,在图2中描述的成熟EML4-ALK融合多核苷酸(SEQ ID NO: 2)),例如,长度为部分50-750个nt,或杂交于整个长度的参比多核苷酸的多核苷酸也可以用作探针,因为多核苷酸对应于大多数(如果不是所有)的沉积的cDNA的核苷酸序列或图2A-B所示的核苷酸序列(SEQ ID NOs: 2或19),或图1A-B(图片B)所示的核苷酸序列(SEQ ID NOs: 7和24)。

“长度至少20个核苷酸”的部分多核苷酸,例如,是指来自参比多核苷酸的核苷酸序列的20个以上邻接核苷酸。如所指出的,这样的部分可以诊断上用作探针(按照常规DNA杂交技术)或用作引物,通过聚合酶链反应(PCR)用于靶序列的扩增,如例如在MOLECULAR CLONING, A LABORATORY MANUAL, 2<sup>nd</sup> Ed., Sambrook, J., Fritsch, E. F. and Maniatis, T., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)中所描述的,将其全部披露内容以引用方式结合于本文。当然,仅杂交于多聚腺苷酸序列(如图2所示的EML4-ALK序列(SEQ ID NO: 2)的3'端多聚(腺苷)段)或仅杂交于T(或U)残基的互补段序列的多核苷酸将不包括在本发明的用来杂交于本发明的部分核酸的多核苷酸中,因为这样的多核苷酸将杂交于任何包含多聚(腺苷)段或其补体的核酸分子(例如,实际上任何双链cDNA克隆)。

如所指出的,本发明的核酸分子,其编码本发明的突变体ALK多肽,可以包括但不限于那些单独地编码成熟多肽的氨基酸序列的核酸分子;用于成熟多肽和另外序列的编码序列,如那些编码前导或分泌序列的编码序列,如前体蛋白序列、或原蛋白序列或前原蛋白序列;成熟多肽的编码序列,具有或没有上述另外的编码序列,

连同另外的、非编码序列，包括例如但不限于内含子和非编码 5' 和 3' 序列，如转录的、非翻译序列，其在转录、mRNA 加工中起作用，包括剪接和聚腺苷酸化信号，例如，mRNA 的核糖体结合和稳定性；另外的编码序列，其编码另外的氨基酸如那些提供另外功能的氨基酸。

因此，编码多肽的序列可以被融合到标记序列，如对肽进行编码的序列，其促进融合多肽的纯化。在本发明的该方面的某些优选实施方式中，标记氨基酸序列是六组氨酸肽，如提供在 pQE 载体中的标记(Qiagen, Inc.)，其中许多可商业上获得。如在 Gents *et al.*, *Proc. Natl. Acad. Sci. USA* 86: 821-824 (1989) 中所描述的，例如，六组氨酸提供融合蛋白的方便纯化。“HA” 标记是另一种可用于纯化的肽，其对应于衍生自流感血凝素蛋白的表位，这已由 Wilson *et al.*, *Cell* 37: 767 (1984) 加以描述。如下面所讨论的，其它这样的融合蛋白包括本身在 N 端或 C 端融合到 Fc 的 EML4-ALK 融合多肽。

本发明进一步涉及本发明的核酸分子的变体，其编码本文披露的 EML4-ALK 融合多肽的部分、类似物或衍生物。变体可以是天然存在的，如天然等位基因变体。“等位基因变体” 是指基因的多种替换形式之一，其中基因占据生物体的染色体上的给定基因座。参见，例如，GENES II, Lewin, B., ed., John Wiley & Sons, New York (1985)。可以利用本领域已知的诱变技术来产生非天然存在的变体。

这样的变体包括那些由核苷酸取代、缺失或加成（添加）所产生的变体。取代、缺失或加成可以涉及一个或多个核苷酸。可以在编码区、非编码区、或两者中改变变体。在编码区中的变化可以产生保守或非保守氨基酸取代、缺失或加成。其中尤其优选的是沉默取代、加成以及缺失，其并不改变本文披露的突变体 ALK 多肽的性能和活性（例如，激酶活性）。在这方面还尤其优选的是保守取代。

本发明的进一步的实施方式包括分离的多核苷酸，其包含至少 90% 相同于，以及更优选至少 95%、96%、97%、98% 或 99% 相同

于本发明的突变体 ALK 多核苷酸的核苷酸序列（例如，编码具有图 2 所示的完全氨基酸序列(SEQ ID NO: 1)的 EML4-ALK 融合多肽的核苷酸序列，或编码 EML-4 的 N 端和 ALK 的激酶结构域的核苷酸序列（参见图 1，图片 B；以及图 3 和 4）；或互补于上述典型序列的核苷酸）。

具有至少，例如，95% “相同” 于参比核苷酸序列（其编码突变体 ALK 多肽）的核苷酸序列的多核苷酸是指，多核苷酸的核苷酸序列相同于参比序列，只是对于参比核苷酸序列（其编码突变体 ALK 多肽）的每 100 个核苷酸，多核苷酸序列可以包括多达 5 个点突变。换言之，为了获得其核苷酸序列至少 95% 相同于参比核苷酸序列的多核苷酸，参比序列中的多达 5% 的核苷酸可以缺失或用另一种核苷酸取代，或参比序列中的总核苷酸的多达 5% 的核苷酸可以被插入参比序列。参比序列的这些突变可以发生在参比核苷酸序列的 5' 或 3' 端位置或这些末端位置之间的任何地方，其单独地散布在参比序列中的核苷酸之间或散布在参比序列内的一个或多个邻接组中。

事实上，可以利用已知计算机程序如 Bestfit 程序（Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, Wis. 53711）来常规地确定是否任何特定的核酸分子至少 90%、95%、96%、97%、98% 或 99% 相同于，例如，图 2A-B 所示的核苷酸序列（SEQ ID NOs: 2 和 19）或相同于上述沉积的 cDNA 克隆的核苷酸序列。Bestfit 采用 Smith 和 Waterman 的局部同源算法，*Advances in Applied Mathematics* 2: 482-489 (1981)，来发现两个序列之间的最好同源节段。当利用 Bestfit 或任何其它序列对比程序来确定特定序列是否，例如，95% 相同于根据本发明的参比 EML4-ALK 融合多核苷酸序列或截短的 ALK 多核苷酸序列时，当然要设置参数，使得相对于参比核苷酸序列的全长来计算同一性的百分比，以及在参比序列中多达核苷酸总数 5% 的同源缺口（间隙）是允许的。

本发明在其范围内包括这样的核酸分子，其至少 90%、95%、96%、97%、98%或 99%相同于图 2 所示的核酸序列 (SEQ ID NO: 2)，或相同于沉积的 cDNA 的核酸序列，而不考虑它们是否编码具有 ALK 激酶活性的多肽。这是因为，甚至在特定核酸分子并不编码具有 ALK 激酶活性的融合多肽的情况下，本领域技术人员将仍然知道如何将核酸分子用作，例如，杂交探针或聚合酶链反应 (PCR) 引物。本发明的并不编码具有激酶的多肽的核酸分子的应用尤其包括 (1) 分离 cDNA 文库中的 EML4-ALK 缺失基因、或截短的 ALK 基因、或其等位基因变体；(2) 原位杂交 (例如，“FISH”) 于中期染色体扩展 (染色体铺展, chromosomal spreads) 以提供 EML4-ALK 缺失基因或截短的 ALK 基因的精确的染色体位置，如在 Verma *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES, Pergamon Press, New York (1988) 中所描述的；以及 RNA 印迹分析，用于检测在特定组织中 EML4-ALK 融合蛋白或截短的 ALK 激酶 mRNA 的表达。

然而，优选的核酸分子具有这样的序列，这些序列至少 95% 相同于本发明的突变体 ALK 多肽或相同于沉积的 cDNA 的核酸序列，其事实上编码具有 ALK 激酶活性的融合多肽。上述活性可以类似于，但不一定相同于本文披露的 EML4-ALK 融合蛋白 (全长蛋白质、成熟蛋白质、或保留激酶活性的蛋白质片段) 的活性，如用特定生物分析法测得的。例如，通过确定 ALK 磷酸化一个或多个酪氨酸 (其包含肽底物，例如，胰岛素受体底物 1 或 2 (IRS1, IRS2)，其是 ALK 激酶的底物) 的能力可以检查 ALK 的激酶活性。

由于遗传密码的简并，本领域技术人员将立即认识到，大量的具有的序列至少 90%、95%、96%、97%、98%、或 99% 相同于沉积的 cDNA 的核酸序列或图 2A-B 所示的核酸序列 (SEQ ID NOs: 2 和 19) 的核酸分子将编码具有 ALK 活性的突变体多肽。事实上，因为这些核苷酸序列的简并变体都编码相同的多肽，所以这对于本领域技术人员是显而易见的，即使没有进行上述比较测定。在本领域将进一步明了，对于这样的并不是简并变体的核酸分子，合理数目的核酸分子将同样编码保留 ALK 激酶活性的多肽。这是因为本

领域技术人员完全明了较少可能或不可能显著影响蛋白质功能的氨基酸取代（例如，用第二种脂族氨基酸代替一种脂族氨基酸）。

例如，关于如何进行表型沉默氨基酸取代的指导提供在 Bowie *et al.*, “Deciphering the Message in Protein Sequences: Tolerance to Amino Acid Substitutions,” *Science* 247: 1306-1310 (1990)中，其描述了两种主要方式来研究氨基酸序列对变化的耐受性。第一种方法依赖进化过程，其中通过自然选择来接收或排斥突变。第二种方法利用基因工程来在克隆的基因的特定位置引入氨基酸变化，以及利用选择或筛选来鉴定保持功能的序列。这些研究已揭示了，蛋白质惊人地耐受氨基酸取代。熟悉这种技术的技术人员还明了，在蛋白质的一定位置，哪些氨基酸变化可能是允许的。例如，大多数隐蔽氨基酸残基需要非极性侧链，而表面侧链的很少特征通常是保守的。其它上述表型沉默取代描述在 Bowie *et al.* (上文)，以及其中引用的参考文献中。

本领域中众所周知的和一般可获得的 DNA 序列分析的方法可以用来实施本发明的任何多核苷酸实施方式。该方法可以采用这样的酶如 DNA 聚合酶 I 的克林诺片断、SEQUENASE® (US Biochemical Corp, Cleveland, Ohio)、Taq 聚合酶 (Perkin Elmer)、耐热 T7 聚合酶 (Amersham, Chicago, IL)、或重组聚合酶和校正外切核酸酶的组合如由 Gibco BRL (Gaithersburg, Md.) 出售的 ELONGASE 扩增系统。优选地，该过程是借助于仪器自动进行，如 Hamilton Micro Lab 2200 (Hamilton, Reno, Nev.)、Peltier Thermal Cycler (PTC200; MJ Research, Watertown, Mass.) 以及 ABI 377 DNA 测序仪 (Perkin Elmer)。

编码本发明的突变体 ALK 多肽的多核苷酸序列可以使用部分核苷酸序列加以延伸并采用本领域已知的各种方法来检测上游序列如启动子和调节元件。例如，可以采用的一种方法，“限制位点” PCR，使用通用引物来挽回（恢复，retrieve）相邻于已知基因座的未知序列 (Sarkar, G, *PCR Methods Applic.* 2: 318-322 (1993))。尤其是，在存在接头序列的引物和对于已知区特异性的引物的情况下首

先扩增基因组 DNA。示例性的引物是那些在本文的实施例 2 中描述的引物。扩增序列然后经受第二轮 PCR，其中使用相同的接头引物和第一种引物内部的另一种特异性引物。借助于适当的 RNA 聚合酶转录每轮 PCR 的产物并利用反转录酶进行测序。

利用基于已知区的趋异性引物，反向 PCR 还可以用来扩增或延伸序列 (Triglia *et al.*, *Nucleic Acids Res.* 16: 8186 (1988))。利用 OLIGO 4.06 引物分析软件 (National Biosciences Inc., Plymouth, Minn.)、或另一种适当的程序，引物可以被设计成长度为 22-30 个核苷酸，具有 50%或更大的 GC 含量，以及在约 68-72°C 的温度下退火成靶序列。该方法使用多种限制酶以在基因的已知区域中产生适当的片断。然后通过分子内连接反应环化该片断并用作 PCR 模板。

可以使用的另一种方法是捕捉 PCR，其涉及 DNA 片断的 PCR 扩增，其中 DNA 片断相邻于人和酵母人工染色体 DNA 中的已知序列 (Lagerstrom *et al.*, *PCR Methods Applic.* 1: 111-119 (1991))。在该方法中，多限制酶消化和连接作用还可以用来在进行 PCR 以前将工程双链序列放入 DNA 分子的未知部分中。可以用来挽回未知序列的另一种方法是描述在 Parker *et al.*, *Nucleic Acids Res.* 19: 3055-3060 (1991)中的方法。另外，可以使用 PCR、套式引物、以及 PROMOTERFINDER®文库来步查 (walk in) 基因组 DNA (Clontech, Palo Alto, Calif.)。这种方法避免需要筛选文库并且可用于发现内含子/外显子接头。

当筛选全长 cDNA 时，优选使用已大小选择的文库以包括更大的 cDNA。并且，随机引物文库是优选的，因为它们将包含更多序列，其包含基因的 5' 区。随机引物文库可以尤其优选用于这样的情况，其中寡 d(T)文库并不产生全长 cDNA。基因组文库可以用于延伸序列到 5' 和 3' 非转录调控区。

可商业上获得的毛细管电泳系统可以用来分析测序或 PCR 产物的大小或证实测序或 PCR 产物的核苷酸序列。尤其是，毛细管测

序可以采用用于电泳分离的可流动聚合物，激光活化的四种不同的荧光染料（对于每种核苷酸用一种），以及通过电荷耦合器件照相机检测发射波长。利用适当的软件（例如，GENOTYPER™和SEQUENCE NAVIGATOR™，Perkin Elmer）可以将输出/光强度转换成电信号并且可以计算机控制从装载样品到计算机分析和电子数据显示的整个过程。毛细管电泳尤其优选用于小片 DNA（可能有有限量存在于特定样品中）的测序。

### C. 载体和宿主细胞

本发明还提供：重组载体，其包含本发明的分离的多核苷酸；宿主细胞，其是借助于重组载体而基因工程的；以及通过重组技术，重组 EML4-ALK 多肽或其片断的生产。

利用众所周知的技术如侵染、转导、转染、转位、电穿孔以及转化，可以将重组构建物引入到宿主细胞中。载体可以是，例如，噬菌体、质粒、病毒或反转录病毒载体。反转录病毒载体可以是可复制的或复制缺陷的。在后一种情况下，病毒繁殖通常将仅发生在互补宿主细胞中。

可以将多核苷酸连接于包含用于在宿主中繁殖的选择性标记的载体。通常，在沉淀物中，如磷酸钙沉淀物，或在与带电荷脂质的复合物中，引入质粒载体。如果载体是病毒，则利用适当的包装细胞系，它可以被体外包装，然后被转导到宿主细胞中。

优选包括相对于感兴趣的多核苷酸的顺式作用调控区的载体。在引入宿主以后，适当的反式作用因子可以由宿主供给，由互补载体供给或由载体本身供给。关于这点，在某些优选的实施方式中，载体提供特异性表达，其可以是可诱导的和/或细胞类型特异性的。在上述载体中特别优选的是那些通过容易控制的环境因素（如温度和营养添加剂）可诱导的载体。

可用于本发明的表达载体包括衍生自染色体的载体、衍生自附加体的载体以及衍生自病毒的载体，例如，衍生自细菌质粒、噬菌

体、酵母附加体、酵母染色体元件、病毒（如杆状病毒、乳多空病毒、痘苗病毒、腺病毒、鸡痘病毒、假狂犬病病毒以及反转录病毒）的载体，以及衍生自其组合的载体，如黏粒和噬菌粒。

包含本发明的 EML4-ALK 多核苷酸的 DNA 插入片段应操作性地连接于适当的启动子，如  $\lambda$  噬菌体 PL 启动子，大肠杆菌 lac、trp 以及 tac 启动子，SV40 早期和晚期启动子以及反转录病毒 LTR 的启动子（仅列举几种）。其它适宜的启动子是本领域技术人员已知的。表达构建物将进一步包括转录起始位点，终止位点，以及在转录区用于翻译的核糖体结合位点。由构建物表达的成熟转录物的编码部分将优选包括在最初起始的翻译和终止密码子（UAA、UGA 或 UAG），其适当地定位在待翻译的多肽的末端。

如所指出的，表达载体将优选包括至少一种选择性标记。这样的标记包括二氢叶酸还原酶或用于真核细胞培养的新霉素抗性基因以及用于大肠杆菌和其它细菌中培养的四环素或氨基青霉素抗性基因。适当宿主的代表性实例包括但不限于细菌细胞，如大肠杆菌、链霉菌以及鼠伤寒沙门氏杆菌细胞；真菌细胞，如酵母细胞；昆虫细胞如果蝇 S2 和夜蛾 Sf9 细胞；动物细胞如 CHO、COS 以及 Bowes 黑色素瘤细胞；以及植物细胞。用于上述宿主细胞的适当的培养基和条件在本领域是已知的。

在载体中，优选供细菌之用的载体包括可获自 Qiagen 的 pQE70、pQE60 和 pQE-9；可获自 Stratagene 的 pBS 载体、Phagescript 载体、Bluescript 载体、pNH8A、pNH16a、pNH18A、pNH46A；以及可获自 Pharmacia 的 ptrc99a、pKK223-3、pKK233-3、pDR540、pRIT5。其中优选的真核载体是可获自 Stratagene 的 pWLNEO、pSV2CAT、pOG44、pXT1 和 pSG；以及可获自 Pharmacia 的 pSVK3、pBPV、pMSG 和 pSVL。其它适宜的载体对于本领域技术人员来说将是显而易见的。

在适合于供本发明之用的已知细菌启动子中包括大肠杆菌 lacI 和 lacZ 启动子、T3 和 T7 启动子、gpt 启动子、 $\lambda$ PR 和 PL 启动子

以及 trp 启动子。适宜的真核启动子包括 CMV 立即早期启动子、HSV 胸苷激酶启动子、早期和晚期 SV40 启动子、反转录病毒 LTR 的启动子，如劳氏肉瘤病毒(RSV)的启动子、以及金属硫蛋白启动子，如小鼠金属硫蛋白-I 启动子。

在酵母（酿酒酵母）中，可以使用多种载体，其包含组成型或诱导型启动子如  $\alpha$  因子、乙醇氧化酶、以及 PGH。关于综述，参见 Ausubel *et al.* (1989) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, N.Y, and Grant *et al.*, *Methods Enzymol.* 153: 516-544 (1997)。

可以通过磷酸钙转染、DEAE-葡聚糖介导转染、阳离子脂质介导转染、电穿孔、转导、侵染或其它方法来实现将构建物引入宿主细胞。这样的方法描述在许多标准实验室手册中，如 Davis *et al.*, BASIC METHODS IN MOLECULAR BIOLOGY (1986)。

可以通过将增强子序列插入到载体中来增加 DNA 的转录，其中 DNA 通过高等真核生物编码本发明的 EML4-ALK 融合多肽。增强子是 DNA 的顺式作用元件，通常约 10 至 300 bp，其用来在给定的宿主细胞类型中增加启动子的转录活性。增强子的实例包括 SV40 增强子，其位于复制起点的后侧 (late side) (在碱基对 100 至 270 处)，巨细胞病毒早期启动子增强子，在复制起点的后侧上的多瘤增强子，以及腺病毒增强子。

为了将翻译蛋白分泌进入内质网的腔、进入周质间隙或进入胞外环境，可以将适当的分泌信号加入到表达多肽中。该信号对于多肽而言可以是内源的或它们可以是异源信号。

可以以修饰形式，如融合蛋白（例如，GST-融合），来表达多肽，并且不仅可以包括分泌信号，而且包括另外的异源功能区。例如，可以将另外的氨基酸（尤其是带电荷的氨基酸）区加入到多肽的 N 端中或在纯化期间或在随后的处理和贮存期间改善在宿主细胞中的稳定性和持久性。同样地，可以将肽部分加入到多肽中以促进纯化。可以在多肽的最终制备以后除去这样的区域。将肽部分加

入到多肽以引起分泌或排泄，从而改善稳定性以及促进纯化（除了别的之外），在本领域是熟悉和常规的技术。优选的融合蛋白包含来自免疫球蛋白的异源区，其可用来溶解蛋白质。

通过众所周知的方法，包括硫酸铵或乙醇沉淀、酸提取、阴离子或阳离子交换层析、磷酸纤维素层析、疏水作用层析、亲和层析、羟基磷灰石层析和凝集素层析，EML4-ALK 融合多肽可以回收和纯化自重组细胞培养物。更优选地，高效液相层析（“HPLC”）用于纯化。本发明的多肽包括天然纯化产物，化学合成步骤的产物，以及通过重组技术由原核生物或真核生物宿主产生的产物，其中上述宿主包括，例如，细菌、酵母、高等植物、昆虫以及哺乳动物细胞。取决于在重组生产过程中采用的宿主，本发明的多肽可以是糖基化的或可以是非糖基化的。此外，本发明的多肽还可以包括最初修饰的甲硫氨酸残基，在某些情况下作为宿主介导过程的结果。

因此，在一种实施方式中，本发明提供了一种通过在适合于表达融合多肽的条件下培养重组宿主细胞并回收（恢复，recover）多肽而用于产生重组 EML4-ALK 融合多肽的方法，如下所述。适合于宿主细胞生长和从上述细胞表达重组多肽的培养条件是本领域技术人员熟知的。参见，例如，CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel FM *et al.*, eds., Volume 2, Chapter 16, Wiley Interscience.

#### D. 分离的多肽

本发明还部分地提供了分离的突变体 ALK 激酶多肽及其片段。在一种实施方式中，本发明提供了一种分离的多肽，该多肽包含一种氨基酸序列，而该氨基酸序列至少 95% 相同于选自由以下组成的组的序列：

(a) 编码包括 SEQ ID NO: 1 或 SEQ ID NO: 18 的氨基酸序列的 EML4-ALK 融合多肽的氨基酸序列；

(b) 编码 EML4-ALK 融合多肽的氨基酸序列, 该 EML4-ALK 融合多肽包含 EML-4 的 N 端氨基酸序列 (SEQ ID NO: 3 的残基 1-233 或 SEQ ID NO: 3 的残基 1-495) 以及 ALK 的激酶结构域 (SEQ ID NO: 5 的残基 1116-1383); 以及

(c) 编码多肽的氨基酸序列, 该多肽包含围绕 EML4-ALK 融合多肽的融合接头 (SEQ ID NO: 1 的残基 233-234 或 SEQ ID NO: 18 的残基 495-496) 的至少 6 个邻接氨基酸。

在一种优选的实施方式中, 提供了本发明的重组突变体 ALK 多肽, 如上所述, 其可以利用重组载体或重组宿主细胞来产生。

本领域中应当明了, 可以变化 EML4-ALK 融合多肽或截短的活性 ALK 激酶多肽的某些氨基酸序列而不会显著影响突变体蛋白质的结构或功能。如果设想在序列中的上述差异, 则应当记住, 在蛋白质上将存在关键区域, 其决定活性 (例如, ALK 的激酶结构域)。通常, 可以取代形成三级结构的残基, 只要使用了执行类似功能的残基。在其它情况下, 如果变化发生在蛋白质的非关键区, 则残基的类型可以是完全不重要的。

因此, 本发明进一步包括 EML4-ALK 融合多肽的变化, 其保留基本的 ALK 激酶活性或其包括 EML-4 或 ALK 蛋白质的其它区, 如下文讨论的蛋白质部分。这样的突变体包括缺失、插入、倒位、重复、以及类型取代 (例如, 用一种亲水残基取代另一种亲水残基, 但通常不用强亲水残基取代强疏水残基)。较小变化或上述 “中性” 氨基酸取代将通常对活性几乎没有影响。

通常看作保守取代的是脂族氨基酸 Ala、Val、Leu 以及 Ile 之间的彼此替代, 羟基残基 Ser 和 Thr 的互换, 酸性残基 Asp 和 Glu 的交换, 酰胺残基 Asn 和 Gln 之间的取代, 碱性残基 Lys 和 Arg 的交换以及芳族残基 Phe、Tyr 之间的替代。本领域技术人员已知的保守氨基酸取代的实例是: 芳族: 苯丙氨酸色氨酸酪氨酸; 疏水的: 亮氨酸异亮氨酸缬氨酸; 极性的: 谷氨酰胺天冬酰胺; 碱性的: 精氨酸赖氨酸组氨酸; 酸性的: 天冬氨酸谷氨酸; 较小的: 丙氨酸丝

氨酸苏氨酸甲硫氨酸甘氨酸。如上面详细说明的，关于哪些氨基酸变化可能是表型沉默的（即，不可能对功能具有显著的有害影响）的进一步指导可以在 *Bowie et al.*, *Science* 247, 上文中找到。

优选以分离的形式提供本发明的多肽，并且优选基本上是纯化的。重组产生的本发明的 EML4-ALK 融合多肽的变体可以通过在 *Smith* 和 *Johnson*, *Gene* 67: 31-40 (1988) 中描述的一步法基本上加以纯化。

本发明的多肽包括：图 2A-B 的 EML4-ALK 融合多肽 (SEQ ID NOs: 1 和 18) (不管包括还是不包括前导序列)；氨基酸序列，该序列编码 EML4-ALK 融合多肽，其包含 EML-4 的 N 端氨基酸序列 (SEQ ID NO: 3 的残基 1-233 或 SEQ ID NO: 3 的残基 1-495) 和 ALK 的激酶结构域 (SEQ ID NO: 5 的残基 1116-1383)；和氨基酸序列，该序列编码多肽，其包含围绕 EML4-ALK 融合多肽的融合接头 (SEQ ID NO: 1 的残基 233-234 或 SEQ ID NO: 18 的残基 495-496) 的至少 6 个邻接氨基酸 (还参见图 1A-B, 底部图片)；以及这样的多肽，其至少 90%，优选至少 95%，以及进一步更优选至少 96%、97%、98% 或 99% 相似于上述多肽。

两种多肽的“%相似性”是指相似性得分，其是通过利用 Bestfit 程序 (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, Wis. 53711) 和用于确定相似性的默认设置来比较两种多肽的氨基酸序列而产生的。Bestfit 使用 Smith 和 Waterman 的局部同源算法 (*Advances in Applied Mathematics* 2: 482-489 (1981)) 来发现两种序列之间相似性最好的区段。

具有的氨基酸序列至少，例如，95% “相同” 于本发明的 EML4-ALK 融合多肽的参比氨基酸序列的多肽是指，多肽的氨基酸序列相同于参比序列，只是对于突变体 ALK 多肽的参比氨基酸序列的每 100 个氨基酸，多肽序列可以包括多达 5 个氨基酸变化。换言之，为了获得具有的氨基酸序列至少 95% 相同于参比氨基酸序列的多肽，在参比序列中的多达 5% 的氨基酸残基可以缺失或用另一

种氨基酸取代，或在参比序列中总氨基酸残基的多达 5%的氨基酸可以被插入到参比序列中。参比序列的这些变化可以发生在参比氨基酸序列的氨基或羧基末端位置或发生在那些末端位置之间的任何地方，其单独地散布在参比序列中的残基之间或散布在参比序列内的一个或多个邻接基团中。

当利用 Bestfit 或任何其它序列对比程序来确定特定序列是否，例如，95%相同于根据本发明的参比序列时，当然要设置参数，使得相对于参比氨基酸序列的全长来计算同一性百分比，以及在同源方面多达 5%的参比序列中氨基酸残基总数的缺口（间隙）是允许的。

本发明的 EML4-ALK 融合多肽可以用作 SDS-PAGE 凝胶或分子筛凝胶过滤柱上的分子量标记，例如，利用本领域技术人员熟知的方法。

如下文进一步详细描述，本发明的多肽还可以用来产生融合多肽特异试剂，如多克隆和单克隆抗体，或截短的多肽特异试剂，其可用于测定来检测突变体 ALK 多肽表达（如下所述），或作为能够增强或抑制突变体 ALK 蛋白质的功能/活性的激动剂和拮抗剂。另外，这样的多肽可以用于酵母双杂交系统以“捕捉”EML4-ALK 融合多肽或截短的 ALK 激酶多肽结合蛋白，其还是根据本发明的候选激动剂和拮抗剂。该酵母双杂交系统描述在 Fields 和 Song, *Nature* 340: 245-246 (1989)中。

在另一个方面，本发明提供了一种肽或多肽，其包含本发明的多肽的含有表位的部分，例如，包含 EML4-ALK 融合多肽的融合接头的表位（参见图 1A-B，底部图片）。这种多肽部分的表位是本发明的多肽的免疫原性或抗原性表位。“免疫原性表位”被定义为一部分蛋白质，当全蛋白是免疫原时其诱导抗体应答。这些免疫原性表位被认为限于分子上的几个位置。另一方面，抗体可以结合的蛋白质分子的区域被定义为“抗原性表位”。蛋白质的免疫原性表位的数目通常少于抗原性表位的数目。参见，例如，Geysen *et al.*,

*Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1983)。下文更详细地描述了本发明的融合多肽特异性抗体的生产。

由含有抗原性表位的肽或多肽引起的抗体可用于检测模拟蛋白质，而不同肽的抗体可以用于跟踪蛋白质前体的不同区的命运，其中蛋白质前体经受翻译后加工。肽和抗肽抗体可以用于各种各样的用于模拟蛋白质的定性或定量测定，例如在竞争测定中，因为研究表明，在免疫沉淀测定中，甚至较短的肽（例如，约9个氨基酸）可以结合和代替更大的肽。参见，例如，Wilson *et al.*, *Cell* 37: 767-778 (1984) at 777。本发明的抗肽抗体还可用于模拟蛋白质的纯化，例如，通过吸附层析，其中利用本领域众所周知的方法。下文更详细地描述免疫测定形式。

重组突变体 ALK 多肽也在本发明的范围内，并且可以利用本发明的多核苷酸来生产，如在上文部分 B 中所描述的。例如，本发明部分地提供了一种通过在适合于表达融合多肽和回收多肽的条件下培养重组宿主细胞(如上所述)来生产重组 EML4-ALK 融合多肽的方法。适合于宿主细胞的生长和从上述细胞表达重组多肽的培养条件是本领域技术人员熟知的。

## E. 突变体特异性试剂

可用于实施所披露方法的突变体 ALK 多肽特异性试剂尤其包括融合多肽特异性抗体和 AQUA 肽(重同位素标记肽)，其对应于、以及适合于检测和量化在生物样品中的 EML4-ALK 融合多肽表达，其中生物样品来自癌症，如哺乳动物的实体肉瘤或肿瘤。同样有用的是截短特异性试剂，如抗体、AQUA 肽、或核酸探针，其适合于检测是否存在本发明的截短的 ALK 激酶多核苷酸或多肽。融合多肽特异性试剂是任何试剂(生物的或化学的)，其能够特异性地结合于、检测和/或量化在生物样品中的已表达的 EML4-ALK 融合多肽的存在/水平。该术语包括但不限于下文讨论的优选的抗体和 AQUA 肽试剂，以及等效试剂是在本发明的范围内。

## 抗体

适合于供实施本发明的方法使用的试剂包括 EML4-ALK 融合多肽特异性抗体和 TFG-ALK 融合多肽特异性抗体。本发明的融合特异性抗体是分离的一个抗体或多个抗体，其特异性地结合本发明的 EML4-ALK 融合多肽（例如，SEQ ID NO: 1）但并不显著地结合野生型 EML-4 或野生型 ALK，或特异性地结合本文描述的 TFG-ALK 融合多肽（例如，SEQ ID NO: 20）但并不显著地结合野生型 TFG 或野生型 ALK。其它适宜的试剂包括表位特异性抗体，其特异性地结合于野生型 ALK 蛋白质序列的胞外结构域中的表位（该结构域并不存在于本文披露的截短的、活性 ALK 激酶中），因此能够检测样品中野生型 ALK 的有无。

人 EML4-ALK 或 TFG-ALK 融合多肽特异性抗体还可以结合于其它哺乳动物物种，例如鼠或兔中的高度同源的和相当的表位肽序列，并且反之亦然。可用于实施本发明的方法的抗体包括（a）单克隆抗体，（b）纯化的多克隆抗体，其特异性地结合于靶多肽（例如，EML4-ALK 融合多肽（参见图 1A-B，底部图片）或 TFG-ALK 融合多肽（参见图 1C，底部图片）的融合接头，（c）如在以上（a）-（b）中所描述的抗体，其结合其它非人物种（例如，小鼠、大鼠）中的相当的和高度同源的表位或磷酸化位点，以及（d）上述（a）-（c）的片断，其结合于由本文披露的示例性抗体结合的抗原（或更优选表位）。

如本文中所使用的，术语“抗体”是指所有类型的免疫球蛋白，包括 IgG、IgM、IgA、IgD 以及 IgE。抗体可以是单克隆或多克隆的并且可以具有任何起源物种，包括（例如）小鼠、大鼠、兔、马或人，或可以是嵌合抗体。参见，例如，M. Walker *et al.*, *Molec. Immunol.* 26: 403-11 (1989); Morrision *et al.*, *Proc. Nat'l. Acad. Sci.* 81: 6851 (1984); Neuberger *et al.*, *Nature* 312: 604 (1984)。抗体可以是按照在美国专利第 4,474,893 号(Reading)或美国专利第 4,816,567 号(Cabilly *et al.*)中所披露的方法制备的重组单克隆抗体。抗体还可以

是按照在美国专利第 4,676,980 号(Segel *et al.*)中所披露的方法制备的化学构建的特异性抗体。

本发明的 EML4-ALK 融合多肽特异性抗体的优选表位位点是肽片断，其基本上包括人 EML4-ALK 融合多肽序列 (SEQ ID NOs: 1 和 18) 的约 11 至 17 个氨基酸，该片断围绕融合接头 (其存在于短变体融合蛋白中的残基 233-234 处和长变体融合蛋白中的残基 495-496 处(参见图 1A-B(底部图片))。应当明了，特异性地结合更短或更长的围绕 EML4-ALK 融合多肽的融合接头的肽/表位的抗体是在本发明的范围内。

类似地，可用于实施所披露方法的 TFG-ALK 融合多肽特异性抗体的优选表位位点是一种肽片断，其基本上包括人 TFG-ALK 融合多肽序列 (SEQ ID NO: 20) 的约 11 至 17 个氨基酸，该片断围绕融合接头 (其存在于残基 137-138 处) (参见图 1C(底部图片))。

本发明并不限于使用抗体，而是包括相当的分子，如蛋白质结合结构域或核酸适体，其以融合蛋白质或截短的蛋白质特异性方式，基本上结合于可用于本发明的方法的 EML4-ALK 或 TFG-ALK 融合多肽特异性抗体所结合的共同表位。参见，例如，Neuberger *et al.*, *Nature* 312: 604 (1984)。在下文中进一步描述的本发明的方法中，可以适当地采用上述相当的非抗体试剂。

可用于实施本发明的方法的多克隆抗体可以按照标准技术加以制备，其中按照已知方法，用围绕所期望的融合蛋白质特异性表位 (例如，本文描述的 ALK 融合蛋白的融合接头) 的抗原免疫适宜的动物 (例如，兔、山羊等)，从动物收集免疫血清，从免疫血清分离多克隆抗体，以及纯化具有期望特异性的多克隆抗体。抗原可以是合成的包含期望表位序列的肽抗原，按照众所周知的技术加以选择 and 构建。参见，例如，ANTIBODIES: A LABORATORY MANUAL, Chapter 5, p. 75-76, Harlow & Lane Eds., Cold Spring Harbor Laboratory (1988); Czernik, *Methods In Enzymology*, 201: 264-283 (1991); Merrifield, *J. Am. Chem. Soc.* 85: 21-49 (1962))。如下文进一步描述的，可以对如本文所述制备的多克隆抗体进行筛选和分离。

单克隆抗体还可以有利地用于本发明的方法，并且可以按照 Kohler 和 Milstein 的众所周知的技术，*Nature* 265: 495-97 (1975); Kohler and Milstein, *Eur. J. Immunol.* 6: 511 (1976); 还参见 CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel *et al.* Eds. (1989)，用杂交瘤细胞系加以生产。如此产生的单克隆抗体是高度特异性的，并且改善本发明所提供的测定方法的选择性和特异性。例如，可以将包含适当抗原（例如，包含 EML4-ALK 融合多肽的融合接头的合成肽）的溶液注射到小鼠中，并在足够时间（与常规技术一致）以后，处死小鼠并获得脾细胞。然后，通常在存在聚乙二醇的情况下，通过用骨髓瘤细胞融合它们来固定脾细胞，以产生杂交瘤细胞。兔融合杂交瘤，例如，可以如在美国专利第 5,675,063 号（K. Knight, 1997 年 10 月 7 日公布）中所描述的来产生。然后在适宜的选择培养基如次黄嘌呤-氨基蝶呤-胸苷（HAT）中生长杂交瘤细胞，并筛选上清液以获得具有期望特异性的单克隆抗体，如下所述。通过常规方法如沉淀、离子交换或亲和层析等，可以从组织培养上清液回收分泌的抗体。

通过本领域技术人员已知的重组技术，还可以在大肠杆菌中产生单克隆 Fab 片段。参见，例如，W. Huse, *Science* 246: 1275-81 (1989); Mullinax *et al.*, *Proc. Nat'l Acad. Sci.* 87: 8095 (1990)。如果一种同种型的单克隆抗体优选用于特定用途，则可以通过利用用来分离类别转换变体的同胞选择技术，从分泌不同同种型的单克隆抗体的亲代杂交瘤，直接、通过从最初融合进行选择、或二次制备来制备特定的同种型（Steplewski, *et al.*, *Proc. Nat'l. Acad. Sci.*, 82: 8653 (1985); Spira *et al.*, *J. Immunol. Methods*, 74: 307 (1984)）。结合单克隆抗体的位点的抗原可以通过 PCR 加以克隆并且单链抗体产生为噬菌体展示重组抗体或在大肠杆菌中的可溶抗体（参见，例如，ANTIBODY ENGINEERING PROTOCOLS, 1995, Humana Press, Sudhir Paul editor）。

更进一步地，美国专利第 5,194,392 号，Geysen (1990)描述了检测或确定单体（氨基酸或其它化合物）的序列的一般方法，该序列是拓扑相当的表位（即，“模拟表位”），其互补于感兴趣抗体的

特定互补位（抗原结合位点）。更一般地，该方法涉及检测或确定单体的序列，该序列是拓扑相当的配体，其互补于感兴趣的特定受体的配体结合位点。类似地，美国专利第 5,480,971 号，Houghten *et al.* (1996)，披露了线性 C<sub>1</sub>-C-烷基全烷基化寡肽和这样的肽的集合和文库，以及利用上述寡肽集合和文库来确定全烷基化寡肽的序列的方法，其中全烷基化寡肽优先结合于感兴趣的受体分子。因此，本发明的含有表位的肽的非肽类似物也可以通过这些方法常规地加以制备。

可以按照标准技术并针对表位和融合蛋白特异性，来筛选可用于本发明的方法的抗体（不管是多克隆或单克隆）。参见，例如，Czernik *et al.*, *Methods in Enzymology*, 201: 264-283 (1991)。例如，可以相对于肽文库并通过 ELISA 来筛选抗体以确保对于所期望的抗原的特异性，以及如果需要的话，确保仅与，例如，本发明的 EML4-ALK 融合多肽而不与野生型 EML-4 或野生型 ALK 的反应性的特异性。还可以通过相对于包含靶蛋白的细胞制剂的蛋白质印迹来试验抗体以证实仅与所期望的靶具有反应性以及确保没有明显结合于涉及 ALK 的其它融合蛋白。融合蛋白特异性抗体的生产、筛选、以及应用是本领域技术人员已知的，并且已有描述。参见，例如，美国专利公开号 20050214301，Wetzel *et al.*, September 29, 2005。

可用于本发明的方法的融合多肽特异性抗体可以呈现与其它融合蛋白中的类似融合表位或与野生型 EML-4、野生型 TFG、以及野生型 ALK（其形成融合接头）中的表位的某些有限的交叉反应性。这并不是意外的，因为大多数抗体呈现一定程度的交叉反应性，并且抗肽抗体将经常与相对于免疫肽具有高同源性或同一性的表位进行交叉反应。参见，例如，Czernik，上文。通过蛋白质印迹连同已知分子量的标记可以容易地表征与其它融合蛋白的交叉反应性。可以检查交叉反应蛋白质的氨基酸序列以鉴定高度同源于或相同于 EML4-ALK 或 TFG-ALK 融合多肽序列并且抗体与其结合的位

点。通过阴性选择并利用肽柱上的抗体纯化可以除去不期望的交叉反应性（例如，选出结合野生型 EML-4 和/或野生型 ALK 的抗体）。

可用于实施本文所披露方法的本发明的 EML4-ALK 融合多肽特异性抗体（以及 TFG-ALK 融合多肽特异性抗体）对于人融合多肽是理想地特异性的，但并不仅限于结合人类本身。本发明包括抗体的生产和应用，其中抗体还结合其它哺乳动物物种（例如，小鼠、大鼠、猴）的保守的和高度同源的或相同的表位。通过与本文披露的人 EML4-ALK 融合多肽序列（SEQ ID NOs: 1 和 18）或本文披露的人 TFG-ALK 融合多肽序列（SEQ ID NO: 20）进行标准序列比较，如利用 BLAST，可以容易地鉴定在其它物种中的高度同源的或相同的序列。

可以通过在特定测定形式中的应用，例如流式细胞术（FC）、免疫组织化学（IHC）、和/或免疫细胞化学（ICC），来进一步表征和证实在本发明的方法中采用的抗体。在下面的部分 F 中进一步描述了在上述方法中 ALK 融合多肽特异性抗体的应用。抗体还可以有利地共轭于荧光染料（例如，Alexa488、PE）、或标记如量子点，连同其它信号转导（磷酸-AKT、磷酸-Erk 1/2）和/或细胞标记物（细胞角蛋白）抗体用于多参数分析，如在下面的部分 F 中进一步描述的。

在实施本发明的方法时，利用这些野生型蛋白质的抗体（磷酸特异性或总抗体）还可以有利地检查在给定生物样品中野生型 EML-4、野生型 TFG、和/或野生型 ALK 的表达和/或活性。例如，总 ALK 和磷酸化-位点特异性抗体可商业上获得（参见 CELL SIGNALING TECHNOLOGY, INC., Beverly MA, 2005/06 Catalogue, #'s 3341, 3342）。如上所述，还可以按照标准方法来产生上述抗体。公开了人 EML-4、TFG、以及 ALK 的氨基酸序列（参见图 3A 和 4A-4C，以及参比的 SwissProt Accession Nos.），如同公开了来自其它物种的这些蛋白质的序列那样。

在生物样品（例如，肿瘤样品）中，野生型 EML-4、TFG、以及野生型 ALK 表达和/或激活的检测，连同 EML4-ALK 和/或

TFG-ALK 融合多肽表达, 可以提供关于是否融合蛋白单独驱动肿瘤、或是否野生型 ALK 也被激活并驱动肿瘤的信息。这样的信息临床上可用于评估靶向融合蛋白或野生型蛋白质、或两者, 或可能最有利于抑制肿瘤的进展, 以及有利于选择适当的疗法或它们的组合。对野生型 ALK 激酶胞外结构域特异的抗体 (其并不存在于本文披露的截短的活性 ALK 激酶中), 可以特别用于确定突变体 ALK 激酶的存在。

应当明了, 在实施上述方法时, 可以使用多于一种的抗体。例如, 可以同时采用一种或多种 EML4-ALK 融合多肽特异性抗体连同对于另一种激酶、受体、或激酶底物 (在其中表达 EML4-ALK 融合多肽的癌症中, 其怀疑被、或潜在被激活) 特异的一种或多种抗体, 以在包含来自上述癌症的细胞的生物样品中检测这样的其它信号分子的活性。

本领域技术人员将明了, 上述本发明的 EML4-ALK 融合多肽以及其含有融合接头表位的片断可以结合有免疫球蛋白(IgG)的部分恒定结构域, 从而产生嵌合多肽。这些融合蛋白可促进纯化并显示出增加的体内半寿期 (半衰期)。例如, 对于由人 CD4-多肽的开始的两个结构域和哺乳动物免疫球蛋白的重链或轻链的恒定区的各种结构域构成的嵌合蛋白, 情况就是如此 (EPA 394,827; Traunecker *et al.*, *Nature* 331: 84-86 (1988))。具有二硫连接的二聚体结构 (由于 IgG 部分) 的融合蛋白还可以比单独的单体 EML4-ALK 融合多肽更有效地结合和中和其它分子 (Fountoulakis *et al.*, *J Biochem* 270: 3958-3964 (1995))。

### 重同位素标记肽 (AQUA 肽)

可用于实施所披露方法的 EML4-ALK 或 TFG-ALK 融合多肽特异性试剂还可以包括重同位素标记肽, 其适合于生物样品中已表达的 ALK 融合多肽或截短的 ALK 激酶多肽的绝对量化。已描述了 AQUA 肽的生产以及 AQUA 肽用于复杂混合物中蛋白质 (AQUA) 的绝对量化。参见 WO/03016861, “Absolute Quantification of Proteins

and Modified Forms Thereof by Multistage Mass Spectrometry,” Gygi *et al.* 以及 Gerber *et al. Proc. Natl. Acad. Sci. U.S.A.* 100: 6940-5 (2003) (将其教导的全部内容以引用方式结合于本文)。

AQUA 方法采用将已知量的至少一种重同位素标记肽标准(其具有可以通过 LC-SRM 层析法检测的独特特征)引入消化的生物样品中,以便通过与肽标准的比较来确定生物样品中具有相同序列和蛋白质修饰的肽的绝对量。简单地说, AQUA 方法具有两个阶段:肽内标准选择和证实以及方法开发;以及利用证实的肽内标准进行实施以检测和量化样品中的靶蛋白。该方法是一种有力的技术,用于检测和量化在复杂的生物混合物中的给定肽/蛋白质,如细胞溶胞产物,并且可以用来,例如,量化由于药物处理而引起的蛋白质磷酸化的变化,或量化在不同生物状态蛋白质水平的差异。

通常,为了开发适宜的内标准,基于其氨基酸序列和用来消化的特定蛋白酶,选择在靶蛋白质序列内的特定肽(或修饰肽)。然后通过固相肽合成产生该肽,使得一个残基被相同的包含稳定同位素( $^{13}\text{C}$ 、 $^{15}\text{N}$ )的残基替代。结果是这样的肽,其化学上相同于通过蛋白水解形成的其天然配对物,但容易通过 MS 并借助于 7-Da 质量移动加以区分。然后通过 LC-MS/MS 评估新合成的 AQUA 内标准肽。通过反向层析、电离效率、以及借助于碰撞诱导解离的片段化,该方法提供了关于肽保留的定性信息。选择用于天然和内标准肽集合的信息和丰富的片断离子,然后作为层析保留的函数快速连续加以具体监测以形成基于肽标准的独特分布的选择反应监测(LC-SRM)方法。

AQUA 策略的第二阶段是其实施以测量来自复杂混合物的蛋白或修饰蛋白的量。通常通过 SDS-PAGE 凝胶电泳来分级分离全细胞溶胞产物,然后切除与蛋白质迁移一致的凝胶区。该过程以后接着是在有 AQUA 肽存在的情况下进行凝胶内蛋白水解以及 LC-SRM 分析(参见 Gerber *et al.* 上文)。将 AQUA 肽掺加(穿刺, spiked)到通过用蛋白酶消化全细胞溶胞产物所获得的复杂肽混合物,然后经受免疫亲和纯化,如上所述。通过消化(例如胰蛋白酶

消化)形成的天然肽的保留时间和片段化模式相同于先前确定的 AQUA 内标准肽;因此,利用 SRM 实验的 LC-MS/MS 分析导致内标准和直接来自极端复杂肽混合物的分析物的高度特异和敏感性的测量。

因为加入 AQUA 肽的绝对量(例如,250 fmol),所以曲线下面的面积比可以用来确定在最初细胞溶胞产物中蛋白质或蛋白质的磷酸化形式的精确的表达水平。此外,在凝胶内消化过程中当形成天然肽时,内标准是存在的,使得从凝胶片的肽提取效率、样品处理(包括真空离心)过程中的绝对损失、以及在引入 LC-MS 系统期间的易变性并不影响天然和 AQUA 肽丰度的确定的比率。

开发了一种 AQUA 肽标准,其用于通过 IAP-LC-MS/MS 方法先前在靶蛋白内鉴定的已知序列。如果位点被修饰,则可以开发一种 AQUA 肽,其在位点内掺入特定残基的修饰形式;以及开发第二种 AQUA 肽,其掺入残基的未修饰形式。以这种方式,两种标准可以用来检测和量化在生物样品中位点的修饰和未修饰形式两者。

还可以通过检查蛋白质的基本氨基酸序列和确定由蛋白酶剪切产生的肽的边界来产生肽内标准。可替换地,可以用蛋白酶实际消化蛋白质,然后可以对产生的特定肽片断进行测序。适宜的蛋白酶包括但不限于丝氨酸蛋白酶(例如,胰蛋白酶,一种与膜缔合的丝氨酸蛋白酶)、金属蛋白水解酶(例如,PUMP1)、胰凝乳蛋白酶、组织蛋白酶、胃蛋白酶、嗜热菌蛋白酶、羧肽酶等。

按照一个或多个准则选择靶蛋白内的肽序列,以优化使用该肽作为内标准。优选地,选择肽的大小以将在其它非靶蛋白中重复肽序列的机会降到最低程度。因此,肽优选为至少约6个氨基酸。还优化肽的大小以使电离频率最大化。因此,比约20个氨基酸更长的肽并不是优选的。优选的范围为约7至15个氨基酸。还选择在质谱法过程中不可能是化学活性的肽序列,因此避免包含半胱氨酸、色氨酸、或甲硫氨酸的序列。

可以选择并不包括靶区的修饰区的肽序列,使得肽内标准可以用来确定蛋白质的所有形式的量。可替换地,围绕修饰氨基酸的肽内标准可以期望用来仅检测和量化靶蛋白的修饰形式。可以一起使用用于修饰和未修饰区两者的肽标准,以确定在特定样品中修饰的程度(即,确定修饰形式相当于蛋白质总量的多少部分)。例如,用于已知在特定位点被磷酸化的蛋白质的磷酸化和非磷酸化形式两者的肽标准可以用来量化样品中磷酸化形式的量。

利用一种或多种标记氨基酸来标记肽(即,标记是肽的实际部分),或较少优选地可以在按照标准方法合成以后连接标记。优选地,标记是基于以下考虑事项所选择的质量变化标记:质量应是只有移动片断质量才有的,其中移动片断质量是对低本底的光谱区进行MS分析所产生的;离子质量特征成分是标记部分的一部分,其在MS分析中优选呈现独特的离子质量特征;标记的组分原子的质量总和优选独特地不同于所有可能的氨基酸的片断。因此,通过在所得到的质谱中的离子/质量模式,标记氨基酸和肽容易与未标记的分开。优选地,离子质量特征成分将质量赋予并不匹配任何20种天然氨基酸的残基质量的蛋白质片断。

标记在MS的片段化条件下应是耐用的并且不会经受不利的片段化。在一定范围的条件下,尤其是在变性条件下,标记化学性能应是有效的,并且标记标志优选仍然可溶于所选择的MS缓冲系统中。标记优选并不抑制蛋白质的电离效率并且不是化学活性的。标记可以包含两种或更多种同位素不同的物质的混合物以在每个标记片断位置产生独特的质谱模式。稳定同位素,如 $^2\text{H}$ 、 $^{13}\text{C}$ 、 $^{15}\text{N}$ 、 $^{17}\text{O}$ 、 $^{18}\text{O}$ 、或 $^{34}\text{S}$ ,属于优选的标记。还可以制备掺入不同同位素标记的肽内标准对。重同位素标记可以掺入其中的优选的氨基酸残基包括亮氨酸、脯氨酸、缬氨酸以及苯丙氨酸。

按照它们的质荷( $m/z$ )比,以及优选地还按照它们在色谱柱(例如,HPLC柱)上的保留时间来表征肽内标准。与相同序列的未标记肽共洗脱的内标准被选择为最佳内标准。然后通过任何适当的方式,例如通过利用例如氩或氮作为碰撞气体的碰撞诱导解离(CID)

对肽进行分割，来分析内标准。然后例如通过多级质谱法( $MS^n$ )对片断进行分析，以获得片断离子谱，从而获得肽片段化特征。优选地，肽片断具有显著的  $m/z$  比差异，以使得峰值对应于每个很好分开的片断，以及获得靶肽特有的特征。如果在第一阶段未获得适宜的片断特征，则进行 MS 的另一阶段，直到获得独特的特征。

$MS/MS$  和  $MS^3$  谱中的片断离子对于感兴趣的肽通常是高度特异的，并且，连同 LC 方法，可以高度选择性地检测和量化复杂蛋白质混合物，如细胞溶胞产物，中的靶肽/蛋白质，其中细胞溶胞产物包含数千或上万种蛋白质。可以测定潜在地包含感兴趣的靶蛋白/肽的任何生物样品。优选采用粗制或部分纯化的细胞提取物。通常，样品具有至少 0.01 mg 的蛋白质，通常浓度为 0.1-10 mg/mL，并且可以调节到期望的缓冲浓度和 pH。

然后将对应于待检测/量化的靶蛋白的已知量的标记肽内标准，优选约 10 飞摩尔，加入到生物样品，如细胞溶胞产物中。然后用一种或多种蛋白酶消化掺加样品适当的时间，以允许进行消化。然后进行分离（例如，通过 HPLC、反相 HPLC、毛细管电泳、离子交换层析法等）以从样品中的其它肽分离标记内标准和其相应的靶肽。微毛细管 LC 是一种优选的方法。

然后通过监测 MS 中的选择反应来检查每种分离的肽。这涉及利用通过表征肽内标准所获得的先前知识，然后获得 MS，以连续监测在  $MS/MS$  或  $MS^n$  谱用于感兴趣的肽和内标准两者的特定离子。在洗脱以后，计算肽标准和靶肽峰的曲线下的面积 (AUC)。两个面积的比率提供绝对量化，其可以加以归一化，以获得在分析中使用的细胞数目和蛋白质的分子量，从而提供每个细胞的精确的蛋白质拷贝数。AQUA 方法的进一步的细节描述在 Gygi *et al.* 以及 Gerber *et al.* 上文中。

如上所述，可以期望制备 AQUA 内肽标准（重同位素标记肽）以检测和量化在本发明的突变体 ALK 多肽内的任何独特位点（例如，所披露的 EML4-ALK 融合多肽内的融合接头）。例如，可以制

备 AQUA 磷酸肽，其对应于 EML4-ALK 融合多肽的融合接头序列（参见图 1A-B(底部图片)）或其对应于 EML4、TFG、或 ALK 的截短点。肽标准例如可以产生用于 EML4-ALK 或 TFG-ALK 融合接头并且在 AQUA 方法中这样的标准用来检测和量化生物样品中的融合接头（即，EML4-ALK 融合多肽的存在）。

例如，本发明的示例性的 AQUA 肽包括氨基酸序列 INQVYR（参见图 1，底部图片），其对应于直接地旁侧 EML4-ALK 融合多肽中的融合接头的每个位点的三个氨基酸（参见 SEQ ID NO: 7）。应当明了，还可以构建包含融合接头序列（和另外的残基，其下游或上游）的更大的 AQUA 肽。类似地，可以可替换地构建更小的 AQUA 肽，其包含少于上述序列的所有残基（但仍然包含融合接头本身的点）。上述更大或更短的 AQUA 肽在本发明的范围内，并且可以如上所述进行优选 AQUA 肽的选择和生产（参见 Gygi *et al.*, Gerber *et al.*, 上文）。

### 核酸探针

如在上述部分 B 中详细描述，由本发明提供的融合特异性试剂还包括适合于检测 EML4-ALK 多核苷酸或截短的 ALK 激酶多核苷酸的核酸探针和引物。上述期望的探针其中包括断裂点探针，该断裂点探针对应于产生融合的野生型 EML4 和/或野生型 ALK 基因中的断裂点的两侧。在下面的部分 F 中描述了在测定中上述探针的特定应用，如荧光原位杂交（FISH）或聚合酶链反应（PCR）扩增。可以制备类似的断裂点探针以检测 TFG-ALK 融合多核苷酸的存在（参见图 1C (SEQ ID NO: 21)）。

### **F. 诊断应用和测定形式**

可以以本领域技术人员已知的各种各样的不同测定形式来实施本发明的方法。

## 免疫测定

可用于实施本发明的方法的免疫测定可以是均相免疫测定或多相免疫测定。在均相测定中，免疫反应通常涉及突变体 ALK 激酶多肽-特异性试剂（例如，EML4-ALK 融合多肽-特异性抗体）、标记分析物、以及感兴趣的生物样品。在抗体结合于标记分析物以后，产生自标记的信号被直接或间接地修饰。在均相溶液中进行免疫反应和检测其程度两者。可以采用的免疫化学标记包括自由基、放射性同位素、荧光染料、酶、噬菌体、辅酶等。还可以有利地采用半导体纳米晶体标记、或“量子点”，并且它们的制备和应用已得到很好的描述。一般地参见，K. Barovsky, *Nanotech. Law & Bus.* 1(2): Article 14 (2004)以及其中引用的专利。

在多相测定方式中，试剂通常是生物样品，一种突变体 ALK 激酶多肽-特异性试剂（例如，EML4-ALK 融合特异性抗体），以及用于产生可检测信号的适宜方式。可以使用如下文进一步描述的生物样品。抗体通常被固定在载体上，如珠、平板或载玻片，并且接触在液相中的怀疑包含抗原的样品。然后从液相中分离载体，并采用用于产生这样的信号的方式来检查载体相或液相的可检测信号。该信号与生物样品中分析物的存在有关。用于产生可检测信号的方式包括使用放射性标记、荧光标记、酶标记、量子点等。例如，如果待检测的抗原包含第二结合位点，则在分离步骤以前，结合于该位点的抗体可以共轭于可检测基团并加入到液相反应溶液中。在固体载体上可检测基团的存在表明在试验样品中存在抗原。适宜的免疫测定的实例是放射免疫测定、免疫荧光方法、酶联免疫测定等。

可以用于实施本文披露的方法的免疫测定形式及其变化在本领域是众所周知的。通常参见 E. Maggio, *Enzyme-Immunoassay*, (1980) (CRC Press, Inc., Boca Raton, Fla.); 还参见，例如，美国专利第 4,727,022 号 (Skold *et al.*, “Methods for Modulating Ligand-Receptor Interactions and their Application”); 美国专利第 4,659,678 号 (Forrest *et al.*, “Immunoassay of Antigens”); 美国专利第 4,376,110 号 (David *et al.*, “Immunometric Assays Using

Monoclonal Antibodies”)。适合于形成试剂-抗体复合物的条件是本领域技术人员熟知的。参见同前。EML4-ALK 融合多肽-特异性单克隆抗体可以用于“两位点”或“夹心”测定，其中单杂交瘤细胞系作为标记单克隆抗体和结合单克隆抗体两者的来源。这样的测定描述于美国专利第 4,376,110 号中。可检测试剂的浓度应是足够的，使得和本底相比可以检测 EML4-ALK 或 TFG-ALK 融合多肽的结合。

按照已知技术，如沉淀，可用于实施本文所披露方法的抗体可以共轭于适合于诊断测定的固体载体（例如，珠、平板、载玻片或孔，其形成自诸如胶乳或聚苯乙烯等材料）。按照已知技术，抗体或其它 ALK 融合多肽-结合试剂或截短的 ALK 激酶多肽-结合试剂可以同样共轭于可检测基团如放射性标记（例如， $^{35}\text{S}$ 、 $^{125}\text{I}$ 、 $^{131}\text{I}$ ）、酶标记（例如，辣根过氧化物酶、碱性磷酸酶）、以及荧光标记（例如，荧光素）。

基于细胞的测定，如流式细胞术（FC）、免疫组织化学（IHC）、或免疫荧光（IF），特别期望用于实施本发明的方法，因为这样的测定形式是临床适用的，可以体内检测突变体 ALK 激酶多肽表达，以及避免活性人为变化的风险，其起因于为了获得提取物而操作获自例如肿瘤样品的细胞。因此，在某些优选的实施方式中，本发明的方法是以流式细胞术（FC）、免疫组织化学（IHC）、或免疫荧光（IF）测定形式加以实施。

流式细胞术（FC）可以用来确定，在用目标在于抑制 ALK 激酶活性的药物进行治疗之前、期间、以及以后，突变体 ALK 激酶多肽在哺乳动物肿瘤中的表达。例如，可以通过流式细胞术分析来自骨髓样品的肿瘤细胞的 EML4-ALK 或 TFG-ALK 融合多肽表达和/或激活、以及鉴定癌症细胞类型的标记等（如果希望如此）。可以按照标准方法来实施流式细胞术。参见，例如，Chow *et al.*, *Cytometry (Communications in Clinical Cytometry)* 46: 72-78 (2001)。简单和举例来说，可以采用用于血细胞计数分析的以下程序：在 37°C 下用 2%多聚甲醛固定细胞 10 分钟，接着在冰上的 90%甲醇中进行透化

作用 30 分钟。然后可以用第一 EML4-ALK 或 TFG-ALK 融合多肽-特异性抗体对细胞进行染色，洗涤并用荧光标记第二抗体加以标记。然后用流式细胞仪（例如，Beckman Coulter FC500）并按照所用仪器的具体程序对细胞进行分析。这样的分析将鉴定在肿瘤中表达的 EML4-ALK 或 TFG-ALK 融合多肽的水平。用 ALK-抑制治疗剂治疗肿瘤以后类似的分析将揭示 ALK 融合多肽-表达肿瘤对 ALK 激酶的靶向抑制剂的反应性。

免疫组织化学（IHC）染色还可以用来确定在用目标在于抑制 ALK 激酶活性的药物进行治疗以前、期间、以及以后在哺乳动物癌症（例如，实体瘤样 NSCLC）中突变体 ALK 激酶多肽的表达和/或激活状态。可以按照众所周知的技术进行 IHC。参见，例如，ANTIBODIES: A LABORATORY MANUAL, Chapter 10, Harlow & Lane Eds., Cold Spring Harbor Laboratory (1988)。简单和举例来说，通过用二甲苯接着用乙醇对组织切片进行脱蜡；在水中然后在 PBS 中进行水合；通过在柠檬酸钠缓冲液中加热载玻片来暴露抗原；在过氧化氢中温育切片；在阻断溶液中阻断；在第一抗-EML4-ALK 或抗-TFG-ALK 融合多肽抗体和第二抗体中温育载玻片；以及最后利用 ABC 亲和素/生物素方法并按照制造商的说明进行检测来制备用于免疫组织化学染色的石蜡包埋组织（例如，来自活检的肿瘤组织）。

免疫荧光（IF）测定还可以用来确定在用目标在于抑制 ALK 激酶活性的药物进行治疗以前、期间、以及以后，在哺乳动物癌症中突变体 ALK 激酶多肽的表达和/或激活状态。可以按照众所周知的技术来进行 IF。参见，例如，J.M. Polak and S. Van Noorden (1997) INTRODUCTION TO IMMUNOCYTOCHEMISTRY, 2nd Ed.; ROYAL MICROSCOPY SOCIETY MICROSCOPY HANDBOOK 37, BioScientific/Springer-Verlag。简单和举例来说，患者样品可以被固定得多聚甲醛中，接着固定在甲醇中，用阻断溶液如马血清阻断，用相对于 EML4-ALK 或 TFG-ALK 融合多肽的第一抗体，接着用用荧光染料如 Alexa 488 标记的第二抗体进行温育，然后用落射荧光显微镜进行分析。

在上述测定中采用的抗体可以有利地共轭于荧光染料（例如，Alexa488、PE），或其它标记，如量子点，连同其它信号转导（例如，EGFR、磷酸-AKT、磷酸-Erk 1/2）和/或细胞标记物（例如，细胞角蛋白）抗体用于多参数分析。

用于测量突变体 ALK 激酶多肽的各种各样的其它程序，包括酶联免疫吸附测定（ELISA）、放射免疫测定（RIA）、以及荧光激活细胞分选术（FACS），在本技术领域是已知的，并且提供用于诊断改变或异常水平的 EML4-ALK 或 TFG-ALK 融合多肽表达的基础。通过获自正常哺乳动物受治疗者（优选人）的体液或细胞提取物在适合于复合物形成的条件下结合于相对于 EML4-ALK 或 TFG-ALK 融合多肽的抗体来建立这些融合多肽表达的正常或标准值。可以通过各种方法，但优选通过光度法，来量化标准复合物形成的量。将在受治疗者、对照、以及来自活检组织的疾病样品中表达的 EML4-ALK 或 TFG-ALK 融合多肽的量与标准值进行比较。标准和受治疗者值之间的偏差建立用于诊断疾病的参数。

### 肽和核酸测定

类似地，可以制备用于在生物样品（包括来自肿瘤细胞）中检测/量化已表达的突变体 ALK 激酶多肽的 AQUA 肽并用于标准 AQUA 测定，如在上面的部分 E 中详细描述。因此，在本发明的方法的某些优选实施方式中，ALK 融合多肽-特异性试剂包括重同位素标记的磷酸肽（AQUA 肽），其对应于包含 EML4-ALK 融合多肽或 TFG-ALK 融合多肽的融合接头的肽序列，如在上面的部分 E 中所描述的。

可用于实施本发明的方法的突变体 ALK 多肽-特异性试剂还可以是 mRNA、寡核苷酸或 DNA 探针，其可以直接杂交于、以及检测生物样品中的融合或截短的多肽表达转录物。在上面的部分 B 中详细讨论了这样的探针。简单和举例来说，可以用荧光素标记的 RNA 探针来探测甲醛固定的、石蜡包埋的患者样品，接着用甲酰胺、SSC 以及 PBS 洗涤，然后用荧光显微镜进行分析。同样优选的是

FISH 探针 (包括断裂点探针), 其允许荧光检测基因重排, 如在第 2 号染色体上的 EML4-ALK 缺失突变 (参见实施例 6)。

编码突变体 ALK 激酶多肽的多核苷酸还可以用于诊断目的。可以使用的多核苷酸包括寡核苷酸序列、反义 RNA 和 DNA 分子、以及 PNA。多核苷酸可以用来检测和量化在活检实体瘤组织中的基因表达, 其中 EML4-ALK 或 TFG-ALK 融合多肽或截短的活性 ALK 激酶多肽的表达可以与疾病有关。例如, 诊断测定可以用来区别 EML4-ALK 或 TFG-ALK 融合多肽的缺少、存在、以及过表达, 以及在治疗性干预过程中监测 ALK 融合多肽水平的调节。

在一种优选的实施方式中, 与 PCR 探针的杂交可以用来鉴定对突变体 ALK 多肽进行编码的核酸序列, 其中 PCR 探针能够检测多核苷酸序列, 包括基因组序列, 其编码 ALK 融合多肽或截短的 ALK 激酶多肽、或紧密相关分子。在上面的部分 B 中描述了这样的探针的构建和应用。探针的特异性, 不管是制备自高度特异区, 例如, 在融合接头中的 10 个单一核苷酸, 或较少特异区, 例如, 3' 编码区, 以及杂交或扩增的严格性 (最大、高、中、或低) 将确定探针是否仅鉴定天然存在序列 (其编码突变体 ALK 多肽)、等位基因、或相关序列。

探针还可以用于检测相关序列, 并且应优选包含至少 50% 的来自任何突变体 ALK 多肽编码序列的核苷酸。本发明的杂交探针可以是 DNA 或 RNA 并且衍生自 SEQ ID NO: 2、19、以及 21 的核苷酸序列, 最优选围绕融合接头 (参见图 1A-C, 底部图片以及 SEQ ID NO: 7、24、以及 26), 或衍生自基因组序列, 其包括天然存在的 EML-4、TFG、以及 ALK 多肽的启动子、增强子元件、以及内含子, 如在上面的部分 B 中进一步描述的。

例如, 本发明的 EML4-ALK 融合多核苷酸可以用于 DNA 或 RNA 分析、斑点印迹、或其它膜基技术; 用于 PCR 技术; 或用于纤维素试纸、针、ELISA 或芯片测定, 其利用来自患者活检的流体或组织以检测改变的 ALK 多肽表达。这样的定性或定量方法在本

领域是众所周知的。在一个特定方面，编码本发明的突变体 ALK 多肽的核苷酸序列可以用于这样的测定，其检测各种癌症（包括肺癌）的激活或诱导。突变体 ALK 多核苷酸可以通过标准方法加以标记，并在适合于形成杂交复合物的条件下加入到来自患者的流体或组织样品中。在适当的温育时间以后，洗涤样品，量化信号，并与标准值比较。如果活检或提取样品中的信号量显著不同于可比较的对照样品，则核苷酸序列已杂交于样品中的核苷酸序列，并且样品中编码 EML4-ALK 融合多肽或截短的 ALK 激酶多肽的核苷酸序列的改变的水平存在表明存在有关疾病。这样的测定还可以用来在动物研究中、在临床试验中、或在监测个体患者的治疗中评估特定治疗方案的功效。

为了提供以突变体 ALK 激酶多肽的表达为特征的疾病的诊断基础，建立了表达的正常或标准分布。这可以通过在适合于杂交或扩增的条件下，使获自正常受治疗者（动物或人）的体液或细胞提取物结合于编码 EML4-ALK 或 TFG-ALK 融合多肽的序列或其片段来完成。可以通过比较获自正常受治疗者的值与来自实验的值（其中使用了已知量的基本上纯化的多核苷酸）来量化标准杂交。可以将获自正常样品的标准值与获自有疾病症状的患者样品的值进行比较。标准值与受治疗者值之间的偏差用来确定疾病的存在。

在确定疾病并开始治疗方案以后，可以规则地重复杂交测定以评估患者中表达水平是否开始近似在正常患者中观测到的表达水平。由连续测定获得的结果可以用来显示在几天至数月期间范围的治疗功效。

本发明的突变体 ALK 多核苷酸的另外的诊断应用可以涉及使用聚合酶链反应 (PCR)，一种优选的测定形式，其对于本领域技术人员而言是标准的。参见，例如，MOLECULAR CLONING, A LABORATORY MANUAL, 2nd. edition, Sambrook, J., Fritsch, E. F. and Maniatis, T., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989)。PCR 寡聚体可以化学合成、酶促产生、或制备自重组源。寡聚体将优选由两种核苷酸序列构成，一种具有有义定

向(5'至3')以及另一种具有反义定向(3'至5'),在优化条件下用于鉴定特定基因或条件。相同的两个寡聚体、寡聚体的嵌套集合、或甚至寡聚体的兼并库可以在较少严格的条件下用于检测和/或量化紧密相关的DNA或RNA序列。

还可以用来量化ALK融合多肽或截短的ALK激酶多肽的表达的方法包括放射性标记或生物素化核苷酸、对照核酸的共扩增、以及实验结果被内插到其上的标准曲线(Melby *et al.*, *J. Immunol. Methods*, 159: 235-244 (1993); Duplaa *et al. Anal. Biochem.* 229-236 (1993))。可以通过以ELISA形式进行测定来加速多个样品的量化速度,其中以不同稀度提供感兴趣的寡聚体,以及分光光度或比色反应给出快速量化。

在本发明的另一种实施方式中,本发明的突变体ALK多核苷酸、以及它们近侧和远侧的相邻基因组区,可以用来产生杂交探针,其可用于对天然存在的基因组序列进行作图。利用众所周知的技术,可以将序列作图到特定染色体或到染色体的特异区。这样的技术包括荧光原位杂交(FISH)、FACS、或人工染色体构建,如酵母人工染色体、细菌人工染色体、细菌P1构建或单染色体cDNA文库,如在Price, C. M., *Blood Rev.* 7: 127-134 (1993)以及Trask, B. J., *Trends Genet.* 7: 149-154 (1991)中所综述的。

在一种优选的实施方式中,采用了FISH(如在Verma *et al. HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES*, Pergamon Press, New York, N.Y. (1988)中所描述的)并且FISH可以与其它物理染色体作图技术和基因图谱数据有关。基因图谱的实例可以在1994 Genome Issue of *Science* (265: 1981f)中找到。在物理染色体图谱上编码EML4-ALK或TFG-ALK融合多肽或截短的活性ALK激酶多肽的基因的位置和特定疾病、或特定疾病的素因之间的相关性可以有助于限制与遗传疾病有关的DNA区。本发明的核苷酸序列可以用来检测在正常、载体、或受影响个体之间的基因序列差异。双色断裂点FISH探针,例如,可以用来检测样品中突变体EML-4、TFG、和/或ALK基因的存在或缺少。

染色体制备的原位杂交和物理作图技术如利用确定的染色体标记进行的连锁分析可以用于延伸基因图谱。将基因放置在另一种哺乳动物物种（如小鼠）的染色体上，经常可以揭示相关标记，即使特定人染色体的数目或臂是未知的。通过物理作图，可以将新序列分配给染色体臂、或其部分。这为研究者提供有价值的信息，以便利用定位克隆或其它基因发现技术来寻找疾病基因。在通过基因连锁将疾病或综合征粗略地定位到特定基因组区，例如，AT 到 11q22-23（Gatti *et al.*, *Nature* 336: 577-580 (1988)）以后，任何序列作图到该区域可以表示用于进一步研究的相关或调节基因。本发明的核苷酸序列还可以用来检测在正常、载体、或受影响个体之间起因于易位、倒位等的染色体位置差异。

其它用于核酸检测的适宜方法，如小沟结合共轭寡核苷酸探针（参见，例如，美国专利第 6,951,930 号，“Hybridization-Triggered Fluorescent Detection of Nucleic Acids”）对于本领域技术人员来说已知的。

### 生物样品

可用于实施本发明的方法的生物样品可以获自任何哺乳动物，其中存在或正发展以 EML4-ALK 或 TFG-ALK 融合多肽的表达为特征的癌症。在一种实施方式中，哺乳动物是人，并且该人可以是用于治疗癌症，例如 NSCLC，的 ALK-抑制治疗剂的候选者。该人候选者可以是目前正用、或考虑用 ALK 激酶抑制剂，如 WHI-131 和/或 WHI-154，进行治疗的患者。在另一种实施方式中，哺乳动物是大动物，如马或奶牛，而在其它实施方式中，哺乳动物是小动物，如狗或猫，已知其均发展癌症，包括肺癌。

任何包含来自哺乳动物癌症的细胞（或细胞提取物）的生物样品适合于本发明的方法。在 EML-ALK 融合多肽的情况下，任何癌症，不管是实体还是非实体，都将是适宜的。在 TFG-ALK 的情况下，实体瘤是在本发明的方法的范围内。例如，生物样品可以包含获自渗出液的细胞，如胸膜渗出液的细胞。已知在许多晚期肺癌（包

括 NSCLC) 患者中形成胸膜渗出液(形成在胸腔中的肺外侧并且包含癌细胞的液体), 并且这样的渗出液的存在是不良结果和短存活时间的预兆。已描述了用于获得胸膜渗出液的标准技术并且在本领域是众所周知的(参见 Sahn, *Clin Chest Med.* 3(2): 443-52 (1982))。利用肿瘤标志、细胞角蛋白标志或如所描述的阴性选择的其它方法, 循环肿瘤细胞还可以获自血清(参见 Ma *et al.*, *Anticancer Res.* 23(1A): 49-62 (2003))。对于白血病患者, 血清和骨髓样品可以是特别优选的。对于涉及实体瘤的癌症, 如肉瘤和癌, 生物样品可以包含获自肿瘤活检的细胞, 其可以按照标准临床技术来获得。例如, 在各种各样的癌症(包括成神经细胞瘤和神经外胚层癌症)中已观测到 ALK 的异常表达。参见, 例如, Pulford *et al.*, 上文。然而, 仅在淋巴瘤中描述了 TFG-ALK 易位突变体, 而在实体瘤中先前未观测到。

生物样品可以包含来自癌症的细胞(或细胞提取物), 其中 ALK 融合多肽被表达和/或激活, 但野生型 ALK 激酶则没有。可替换地, 样品可以包含来自癌症的细胞, 其中突变体 ALK 多肽和野生型 ALK 激酶两者均被表达和/或激活, 或其中野生型 ALK 激酶和/或 EML-4 和/或 TFG 被表达和/或活性的, 但突变体 ALK 多肽则没有。

按照标准技术, 上述生物样品的细胞提取物可以制备成粗制物, 或经过部分(或完全)纯化, 并用于本发明的方法。可替换地, 可以以优选的测定形式采用包含全细胞的生物样品, 如免疫组织化学(IHC)、流式细胞术(FC)、免疫荧光(IF)、以及荧光原位杂交(FISH), 如上文进一步描述的。这样的全细胞测定是有利的, 因为它们将肿瘤细胞样品的操作降到最低程度, 因此降低了改变细胞的体内信号/激活状态和/或引入人工信号的风险。全细胞测定还是有利的, 因为它们仅在肿瘤细胞中表征表达和信号, 而不是肿瘤和正常细胞的混合物。

在实施所披露的用于确定是否一种化合物抑制以 EML4-ALK 或 TFG-ALK 融合基因和/或融合多肽为特征的肿瘤的进展的方法时, 还可以有利地采用包含来自哺乳动物骨髓移植模型或异种移植

的细胞的生物样品。优选的异种移植(或移植受体)是小哺乳动物,如小鼠,其躲藏(harbor)表达突变体 ALK 激酶多肽的人肿瘤。躲藏人肿瘤的异种移植在本领域是众所周知的(参见 Kal, *Cancer Treat Res.* 72: 155-69 (1995))并且躲藏人肿瘤的哺乳动物异种移植的生产被很好地描述(参见 Winograd *et al.*, *In Vivo.* 1(1): 1-13 (1987))。类似地,骨髓移植模型的产生和应用被很好地描述(参见,例如, Schwaller, *et al.*, *EMBO J.* 17: 5321-333 (1998); Kelly *et al.*, *Blood* 99: 310-318 (2002))。“以 EML4-ALK 或 TFG-ALK 融合多核苷酸和/或融合多肽为特征的癌症”是指这样的癌症,其中与不存在这样的融合基因和/或融合多肽的癌症相比,存在这样的突变体 ALK 基因和/或表达的多肽。

在评估包含来自哺乳动物癌症肿瘤的细胞的生物样品中的突变体 ALK 多核苷酸存在或多肽表达时,代表细胞的对照样品(其中没有发生上述易位和/或融合蛋白)可以期望用于比较目的。理想地,对照样品包含来自特定癌症(例如, NSCLC)的亚型的细胞,上述亚型代表这样的亚型,其中并没有发生突变(例如, EML4-ALK 缺失突变)和/或并没有表达融合多肽。因此,对照样品与试验生物样品中水平的比较可以确定是否存在突变体 ALK 多核苷酸和/或多肽。可替换地,因为 EML4-ALK 和/或 TFG-ALK 融合多核苷酸和/或多肽可能不存在于大多数癌症中,所以类似地并不表达这样的突变体 ALK 多肽(或躲藏突变体多核苷酸)的任何组织可以用作对照。

以下描述的方法对于以突变体 ALK 多核苷酸和/或多肽为特征的癌症、以及与其有关的治疗决策将具有有价值的诊断应用。例如,生物样品可以获自这样的受治疗者,其先前并没有诊断为具有以 EML4-ALK 缺失突变和/或融合多肽为特征的癌症,也没有经受针对上述癌症的治疗,并且该方法用来诊断上鉴定在这样的受治疗者中的肿瘤为属于肿瘤(例如, NSCLC)的亚型,其中存在和/或表达 EML4-ALK 融合多核苷酸和/或多肽。本发明的方法还可以用来监测在用组合物(包括 ALK 激酶-抑制治疗剂或治疗剂的组合)对

受治疗者进行治疗以后表达突变体 ALK 激酶多肽的癌症的进展或抑制。

可以在初步评估或外科监视步骤以后或以前进行这样的诊断测定。本发明的鉴定方法可以有利地用作诊断来确定患有癌症的患者，如 NSCLC，其由 EML4-ALK 和/或 TFG-ALK 融合蛋白或由截短的 ALK 激酶所驱动，这些患者将很可能对目标在于抑制 ALK 激酶活性的疗法起反应，如 WHI-131 和/或 WHI-154 或它们的类似物。选择上述患者的能力也将用于临床评估未来 ALK-靶向疗法的功效以及用于给予患者的上述药物的未来处方。

### 诊断方法

选择性地鉴定癌症（其中存在 EML4-ALK 和/或 TFG-ALK 融合多核苷酸和/或融合多肽）的能力使得能够实现重要的新方法，其用于精确地鉴定上述肿瘤（用于诊断目的），以及获得信息，当作为单药剂用来治疗癌症时，该信息可用于确定这样的肿瘤是否可能对 ALK-抑制治疗剂组合物起反应，或可能部分或完全地对靶向不同激酶的抑制剂不起反应。

因此，在一种实施方式中，本发明提供了一种用于在来自哺乳动物癌症的生物样品中检测突变体 ALK 多核苷酸和/或其编码的突变体 ALK 多肽的存在的方法，所述方法包括以下步骤：

(a) 从哺乳动物癌症获得生物样品；以及

(b) 利用至少一种试剂，其检测融合多核苷酸、或其编码的融合多肽（包含部分具有部分二级蛋白质的 ALK），以确定在所述生物样品中是否存在 ALK 突变体多核苷酸和/或其编码的突变体 ALK 多肽。

在某些优选的实施方式中，癌症是实体瘤肉瘤或癌，而在一种实施方式中，癌是肺癌，如 NSCLC。在另一种优选的实施方式中，突变体 ALK 多肽是一种融合多肽，其包含 ALK（SEQ ID NO: 5）的残基 1116-1383 并具有部分所述二级蛋白质。在另一种优选的实

实施方式中，二级蛋白质选自由 EML-4 (SEQ ID NO: 3) 和 TRK-融合基因 (TFG) 蛋白质 (SEQ ID NO: 22) 组成的组。在又一种优选的实施方式中，融合多肽包含 EML-4 (SEQ ID NO: 3) 的残基 1-233 或残基 1-495 或 TFG (SEQ ID NO: 22) 的残基 1-138。

在其它优选的实施方式中，融合多核苷酸包括 EML4-ALK 融合多核苷酸 (SEQ ID NO: 2 或 19) 或 TFG-ALK 融合多核苷酸 (SEQ ID NO: 21)，而在又一种实施方式中，融合多肽包括 EML4-ALK 融合多肽 (SEQ ID NO: 1 或 18) 或 TFG-ALK 融合多肽 (SEQ ID NO: 20)。在又一种优选的实施方式中，融合多核苷酸是上述的一种 EML4-ALK 融合多核苷酸或多肽。

在更优选的实施方式中，该方法采用一种试剂，其包含 EML4-ALK 融合多核苷酸和/或至少一种 EML4-ALK 融合多肽-特异性试剂 (抗体或 AQUA 肽)，如上所述。在某些优选的实施方式中，该试剂包括一种分离的试剂，其特异性地结合于或检测 TFG-ALK 融合多肽 (SEQ ID NO: 20) 或 TFG-ALK 融合多核苷酸 (SEQ ID NO: 21)，但并不结合于或检测野生型 TFG 或野生型 ALK。在其它优选的实施方式中，试剂是聚合酶链反应 (PCR) 探针或荧光原位杂交 (FISH) 探针。某些优选的实施方式采用重同位素标记 (AQUA) 肽，其包含野生型 ALK 内的 TFG-ALK 融合多肽或截短点的融合接头的氨基酸序列。

在某些优选的实施方式中，以如上所述的流式细胞术 (FC)、免疫组织化学 (IHC)、或免疫荧光 (IF) 测定形式来实施本发明的诊断方法。在另一种优选的实施方式中，检测了 EML4-ALK 或 TFG-ALK 融合多肽的活性。在其它优选的实施方式中，以如上所述的荧光原位杂交 (FISH) 或聚合酶链反应 (PCR) 测定形式来实施本发明的诊断方法。

本发明进一步提供了一种用于确定化合物是否抑制以 EML4-ALK 或 TFG-ALK 融合多核苷酸和/或融合多肽为特征的癌症的进展的方法，所述方法包括以下步骤：确定所述化合物是否抑

制所述 EML4-ALK 或 TFG-ALK 融合多肽在所述癌症中的表达和/或活性。在一种优选的实施方式中,利用至少一种试剂来确定 ALK 融合多肽的表达和/或活性的抑制,其中所述试剂检测本发明的 EML4-ALK 融合多核苷酸或多肽和/或本文披露的 TFG-ALK 融合多核苷酸或多肽。在下面的部分 G 中更详细地讨论了适合于抑制 ALK 激酶活性的化合物。

在上面的部分 B 和 D 中,进一步详细地描述了可用于实施本发明的方法的突变体 ALK 多核苷酸探针和多肽-特异性试剂。在一种优选的实施方式中,ALK 融合多肽-特异性试剂包含一种融合多肽-特异性抗体。在另一种优选的实施方式中,融合多肽-特异性试剂包含一种重同位素标记磷酸肽(AQUA 肽),其对应于 ALK 融合多肽的融合接头(参见图 1A-C(底部图片))。在又一种优选的实施方式中,融合多核苷酸-特异性试剂包含 FISH 探针,其对应于 ALK 融合基因和/或野生型 EML4、TFG、或 ALK 基因的断裂点的融合接头。

上述本发明的方法还可以可选地包括以下步骤:确定所述生物样品中其它激酶,如野生型 ALK 和 EGFR,或其它下游信号分子的表达或激活水平。描出 ALK 融合多肽表达/激活和其它激酶的表达/激活的轮廓以及在给定生物样品中的途径可以提供以下有价值的信息:哪些激酶和途径正驱动疾病,以及哪种治疗方案因此可能是最有益的。

### 化合物筛选

本文描述的新 EML4-ALK 融合多肽的发现还使得可以开发新化合物,其抑制这种突变体 ALK 蛋白质的活性,尤其是其 ALK 激酶活性。因此,本发明还部分地提供了一种用来确定一种化合物是否抑制以 EML4-ALK 融合多核苷酸和/或融合多肽为特征的癌症的进展的方法,所述方法包括以下步骤:确定所述化合物是否抑制所述 EML4-ALK 融合多肽在所述癌症中的表达和/或活性。在一种优选的实施方式中,EML4-ALK 融合多肽的表达和/或活性的抑制是

利用至少一种试剂加以确定，其中所述试剂检测本发明的融合多核苷酸和/或融合多肽。上面已描述了本发明的优选试剂。在下面的部分 G 中更详细地描述了适合于抑制 ALK 激酶活性的化合物。

化合物可以是，例如，激酶抑制剂，如小分子或抗体抑制剂。它可以是相对于多种不同的激酶具有活性的泛激酶抑制剂，或激酶特异性抑制剂。在下面的部分 G 中进一步详细地讨论了 ALK 激酶-抑制化合物。可以在用抑制剂进行治疗之前和以后获得患者生物样品，然后利用上述方法分析抑制剂对 ALK 激酶活性的生物效应，包括下游底物蛋白的磷酸化。这样的药效测定可以用于确定药物的生物活性剂量，其可以优选于最大耐受量。这样的信息还可用于提出通过证明药物作用机制而认可的药物。在下面的部分 G 中进一步描述了具有上述期望的抑制特性的化合物的鉴定。

## G. 癌症的治疗性抑制

根据本发明，现已表明，其中表达 EML4-ALK 融合蛋白的哺乳动物实体瘤癌症 (NSCLC) 的进展可以通过抑制上述癌症中 ALK 激酶的活性加以体内抑制。类似地，如本文所描述的，其中表达 TFG-ALK 融合蛋白的哺乳动物实体瘤癌症的活性可以类似地通过抑制上述癌症中的 ALK 激酶活性加以体内抑制。通过用 ALK 激酶-抑制治疗剂，如小分子激酶抑制剂如 WHI-131 或 WHI-154 接触癌症 (例如，肿瘤) 可以抑制以突变体 ALK 多肽的表达为特征的癌症中的 ALK 活性。如在下面的实施例 2 中进一步描述的，利用示例性类型的 ALK-抑制治疗剂，siRNA，通过抑制融合激酶可以实现例如表达 ALK 融合蛋白的肿瘤的生长抑制。因此，本发明部分地提供了一种通过抑制突变体 ALK 激酶在癌症中的表达和/或活性而用于抑制表达 EML4-ALK 融合多肽的癌症或表达 TFG-ALK 融合多肽的实体瘤的进展的方法。

ALK 激酶-抑制治疗剂可以是任何组合物，该组合物包含至少一种化合物 (生物的或化学的)，其直接或间接地抑制 ALK 激酶体内的表达和/或活性，并且包括下面描述的示例性类型的化合物。这

样的化合物包括这样的治疗剂 (therapeutics), 其直接作用于 ALK 激酶本身、或改进 ALK 活性的蛋白质或分子, 或其通过抑制 ALK 的表达而间接起作用。这样的组合物还包括仅包含单一 ALK 激酶抑制化合物的组合物, 以及包含多种治疗剂 (包括那些相对于其它 RTK 的治疗剂) 的组合物, 其还可以包括非特异性治疗药剂如化疗药物或通用转录抑制剂。

### 小分子抑制剂

在某些优选的实施方式中, 可用于实施本发明的方法的 ALK-抑制治疗剂是靶向小分子抑制剂, 如 WHI-131 和 WHI-154、或它们的类似物。WHI-131 和 WHI-154 是 ALK 的喹唑啉类小分子靶向抑制剂, 并且已描述了它们的性能。参见 Marzec *et al.*, *Lab. Invest.* 85(12): 1544-54 (2005)。已经表明, 这些化合物可以在淋巴瘤细胞中诱导凋亡和抑制增殖。激酶的其他小分子靶向抑制剂在本领域是众所周知的。例如, Gleevec® (STI-571, Imatinib), 其特异性地结合于并阻断 BCR-ABL 融合激酶 (以及其它激酶) 的 ATP-结合位点从而防止该酶的磷酸化和激活, 可商业上获得并且其性能是众所周知的。参见, 例如, Dewar *et al.*, *Blood* 105(8): 3127-32 (2005)。Novartis, Inc. 和 Cephalon, Inc. 目前正开发 ALK 的其他小分子抑制剂。

小分子靶向抑制剂是一类分子, 其通常通过特异性地、并且经常不可逆地结合于酶的催化部位、和/或结合于酶内的 ATP-结合槽或其它结合位点 (其防止酶采取其活性所必须的构象) 来抑制它们的靶酶的活性。可以利用 ALK 激酶三维结构的 X 射线晶体学模拟或计算机模拟来合理地设计小分子抑制剂, 或可以通过高通量筛选抑制 ALK 的化合物文库来发现小分子抑制剂。这样的方法在本领域是众所周知的, 并且已加以描述。例如, 如上所述, 通过检查这样的化合物抑制 ALK 活性、但不抑制其它激酶活性的能力 (在一组激酶中), 和/或通过检查在包含肺癌细胞的生物样品中 ALK 活性的抑制, 可以证实 ALK 抑制的特异性。下面将进一步描述这样的筛选方法。

## 抗体抑制剂

可用于本发明的方法的 ALK 激酶-抑制治疗剂还可以是靶向抗体，其特异性地结合于 ALK 活性所需要的关键催化或结合位点或结构域，以及通过阻止配体、底物或次级分子接近 ALK，和/或防止酶采取其活性所必须的构象，来抑制激酶。已很好地描述了人源化靶特异性抗体的生产、筛选、以及治疗应用。参见，Merluzzi *et al.*, *Adv Clin Path.* 4(2): 77-85 (2000)。可获得用于高通量产生和筛选人源化靶特异性抑制抗体的商用技术和系统，如 Morphosys, Inc. 的 Human Combinatorial Antibody Library(HuCAL®)。

已描述了各种抗受体激酶靶向抗体的生产以及它们在抑制靶向受体的活性中的应用。参见，例如，美国专利公开号 20040202655, “Antibodies to IGF-I Receptor for the Treatment of Cancers,” October 14, 2004, Morton *et al.*; 美国专利公开号 20040086503, “Human anti-Epidermal Growth Factor Receptor Single-Chain Antibodies,” April 15, 2004, Raisch *et al.*; 美国专利公开号 20040033543, “Treatment of Renal Carcinoma Using Antibodies Against the EGFr,” February 19, 2004, Schwab *et al.*。用于产生和应用受体酪氨酸激酶活性-抑制抗体的准化方法在本领域中是已知的。参见，例如，欧洲专利号 EP1423428, “Antibodies that Block Receptor Tyrosine Kinase Activation, Methods of Screening for and Uses Thereof,” June 2, 2004, Borges *et al.*。

噬菌体展示方式还可以用来产生 ALK-特异性抗体抑制剂，并且用于噬菌体文库构建和重组抗体选择的程序提供在众所周知的参考教科书 CURRENT PROTOCOLS IN IMMUNOLOGY, Colligan *et al.* (Eds.), John Wiley & Sons, Inc. (1992-2000), Chapter 17, Section 17.1 中。还参见美国专利第 6,319,690 号, 2001 年 11 月 20 日, Little *et al.*; 美国专利第 6,300,064 号, 2001 年 10 月 9 日, Knappik *et al.*; 美国专利第 5,840,479 号, 1998 年 11 月 24 日, Little *et al.*; 美国专利公开号 20030219839, 2003 年 11 月 27 日, Bowdish *et al.*。

可以产生展示在噬菌体表面上的抗体片断的文库(参见,例如,美国专利 6,300,064, 2001 年 10 月 9 日, Knappik *et al.*) 并筛选用于结合于受体蛋白质酪氨酸激酶(如 ALK)的可溶的二聚体形式。结合于 RTK 的可溶二聚体形式并用于筛选的抗体片断被鉴定为候选分子,用于阻断细胞中靶 RTK 的组成型激活。参见欧洲专利第 EP1423428 号, *Borges et al.*, 上文。

在如上所述的抗体文库筛选中鉴定的结合靶向抗体的 ALK 然后可以进一步筛选它们在体外激酶测定和体内(在细胞系和/或肿瘤中)激酶测定中阻断 ALK 的活性的能力。如上所述,例如,通过检查这样的抗体治疗剂抑制 ALK 激酶活性、但并不抑制其它激酶活性的能力(在一组激酶中),和/或通过检查在包含癌细胞的生物样品中 ALK 活性的抑制,可以证实 ALK 抑制。上文进一步描述了筛选这样的用于 ALK 激酶抑制的化合物的方法。

### 间接抑制剂

可用于实施所披露方法的 ALK-抑制化合物还可以是这样的化合物,其通过抑制蛋白质或分子而不是 ALK 激酶本身的活性来间接地抑制 ALK 活性。这样的抑制治疗剂可以是靶向抑制剂,其调节关键调节激酶的活性,其中关键调节激酶磷酸化或去磷酸化(因此活化或去活化)ALK 本身,或干扰配体的结合。如同其它受体酪氨酸激酶一样,通过连接物蛋白和下游激酶(包括 STAT5 和 AKT)的网络,ALK 调节下游信号。因此,通过靶向这些相互作用或下游蛋白质,可以抑制通过 ALK 活性的细胞生长的诱导和存活。目前正在开发的可以以这种方式使用的药物包括 Wartmanin。

通过利用抑制为 ALK 采取其活性构象所必须的激活分子的结合的化合物还可以间接地抑制 ALK 激酶活性。类似地,例如,已描述了负调节 PDGF 受体酪氨酸激酶的抗 PDGF 抗体的生产和应用。参见美国专利公开号 20030219839, “Anti-PDGF Antibodies and Methods for Producing Engineered Antibodies,” Bowdish *et al.*。

利用 ALK 三维结构的 X 射线晶体学模拟或计算机模拟可以合理地设计 ALK 活性的间接抑制剂，或可以通过高通量筛选用于抑制关键上游调节酶和/或必要的结合分子的化合物文库来发现 ALK 活性的间接抑制剂，其导致 ALK 激酶活性的抑制。这样的方式在本领域是众所周知的，并且已有描述。如上所述，例如，通过检查上述化合物抑制 ALK 活性、但不抑制其它激酶活性的能力（在一组激酶中），和/或通过检查在包含癌细胞（例如，NSCLC 细胞）的生物样品中 ALK 活性的抑制，可以证实通过上述治疗剂的 ALK 抑制。下面将进一步描述用于鉴定化合物的方法，其中化合物抑制以 EML4-ALK 或 TFG-ALK 融合多核苷酸和/或融合多肽为特征的癌症。

### 反义和/或转录抑制剂

ALK 抑制治疗剂还可以包括反义和/或转录抑制化合物，其通过阻断编码 ALK 的基因和/或 EML4-ALK 或 TFG-ALK 融合基因或截短的 ALK 基因的转录来抑制 ALK 激酶活性。例如，已描述了通过用于治疗癌症的反义疗法来抑制各种受体激酶，其包括 VEGFR、EGFR、IGFR、以及 FGFR。参见，例如，美国专利第 6,734,017 号；第 6,710,174 号；第 6,617,162 号；第 6,340,674 号；第 5,783,683 号；第 5,610,288 号。

按照已知技术，可以将反义寡核苷酸设计、构建，并用作相对于靶基因的治疗药剂。参见，例如，Cohen, J., *Trends in Pharmacol. Sci.* 10(11): 435-437 (1989); Marcus-Sekura, *Anal. Biochem.* 172: 289-295 (1988); Weintraub, H., *Sci. AM.* pp. 40-46 (1990); Van Der Krol *et al.*, *BioTechniques* 6(10): 958-976 (1988); Skorski *et al.*, *Proc. Natl. Acad. Sci. USA* (1994) 91: 4504-4508。最近已描述了利用 EGFR 的反义 RNA 抑制剂来抑制人癌的体内生长。参见美国专利公开号 20040047847, “Inhibition of Human Squamous Cell Carcinoma Growth in vivo by Epidermal Growth Factor Receptor Antisense RNA Transcribed from a Pol III Promoter,” March 11, 2004, He *et al.*。类似地，按照上述方法可以制备 ALK-抑制治疗剂，其包含至少一种相

对于哺乳动物 ALK 基因的反义寡核苷酸(参见图 4(SEQ ID NO: 6))或 EML4-ALK 或 TFG-ALK 融合多核苷酸或截短的 ALK 多核苷酸(参见图 2A-C(SEQ ID NO: 2、19、以及 21))。如下文进一步描述的,可以制备和给予包含 ALK-抑制反义化合物的药物组合物。

### 小分子干涉 RNA

小分子干涉 RNA 分子 (siRNA) 组合物 (其通过 RNA 干涉过程抑制 ALK 的翻译,因而抑制其活性),也可以期望用于本发明的方法中。已很好地描述了 RNA 干涉,以及通过引入外源性小双链 RNA 分子 (其包含互补于编码靶蛋白的 mRNA 的序列) 来选择性沉默靶蛋白表达。参见,例如,美国专利公开号 20040038921, “Composition and Method for inhibiting Expression of a Target Gene,” February 26, 2004, Kreutzer *et al.*; 美国专利公开号 20020086356, “RNA Sequence-Specific Mediators of RNA Interference,” June 12, 2003, Tuschl *et al.*; 美国专利公开号 20040229266, “RNA Interference Mediating Small RNA Molecules,” November 18, 2004, Tuschl *et al.*。

例如,如目前示出的 (参见实施例 2), 在表达融合蛋白的人 NSCLC 细胞系中 siRNA-介导的 EML4-ALK 融合蛋白表达的沉默选择性地抑制在那些细胞中的疾病的进展,但在并不表达突变体 ALK 蛋白质的对照细胞中则不是如此。

已表明,双链 RNA 分子(dsRNA)以高度保守的调节机制,称作 RNA 干涉(RNAi), 阻断基因表达。简单地说, RNase III Dicer 将 dsRNA 处理成大约 22 个核苷酸的小分子干涉 RNA(siRNA), 其作为引导序列, 以通过 RNA 诱导的沉默复合 RISC 来诱导靶特异 mRNA 剪切 (参见 Hammond *et al.*, *Nature* (2000) 404: 293-296)。RNAi 涉及催化型反应, 从而通过更长 dsRNA 的连续剪切来产生新的 siRNA。因此, 不同于反义, RNAi 以非化学计量方式降解靶 RNA。当给予细胞或生物体时, 外源性 dsRNA 显示出可以引导通过 RNAi 的内源性信使 RNA(mRNA)的序列特异降解。

现在可商业上获得各种各样的靶特异性 siRNA 产品, 包括用于它们表达和应用在哺乳动物细胞中的载体和系统。参见, 例如, Promega, Inc. (promega.com); Dharmacon, Inc. (dharmacon.com)。可获得关于用于 RNAi 的 dsRNA 的设计、构建、以及应用的详细技术手册。参见, 例如, Dharmacon's "RNAi Technical Reference & Application Guide"; Promega's "RNAi: A Guide to Gene Silencing"。ALK-抑制 siRNA 产品也可以商业上获得, 并且可以适当地用于本发明的方法。参见, 例如, Dharmacon, Inc., Lafayette, CO (Cat Nos. M-003162-03, MU-003162-03, D-003162-07 至 -10 (siGENOME™ SMARTselection and SMARTpool® siRNAs)。

最近已确定, 在哺乳动物中介导 RNAi 方面最有效的是长度少于 49 个核苷酸、并且优选 19-25 个核苷酸的小 dsRNA, 其包含至少一个基本上相同于部分靶 mRNA 序列的序列, 以及该 dsRNA 最佳地在一端具有至少一个 1-4 个核苷酸的突出端。参见美国专利公开号 20040038921, Kreutzer *et al.*, 上文; 美国专利公开号 20040229266, Tuschl *et al.*, 上文。在上述出版物中详细地描述了这样的 dsRNA 的构建, 以及在药物制剂中它们用来沉默靶蛋白的体内表达。

如果在哺乳动物中待靶向的基因序列是已知的, 例如, 则可以产生和试验 21-23 nt RNA 在哺乳动物细胞, 如人或其它灵长类细胞中介导 RNAi 的能力。如果需要, 可以以适当的动物模型试验那些显示出介导 RNAi 的 21-23 nt RNA 分子, 以进一步评估它们的体内有效性。已知的靶位点, 例如, 基于对其它核酸分子 (例如核酶或反义) 的研究确定为有效靶位点的靶位点, 或那些已知与疾病或病症有关的靶位点如那些包含突变或缺失的位点, 也可以用来设计靶向那些位点的 siRNA 分子。

可替换地, 例如通过利用计算机折叠算法, 可以合理地设计/预测有效 dsRNA 的序列, 其筛选感兴趣的靶 mRNA 的靶位点。靶序列可以被电子 (电脑模拟, *in silico*) 分析成特定长度的所有片断或亚序列的清单, 例如 23 个核苷酸片断, 其中利用定制 Perl 原本

或商用序列分析程序如 Oligo、MacVector、或 GCG Wisconsin Package。

各种参数可以用来确定哪些位点是靶 RNA 序列内的最合适靶位点。这些参数包括但不限于二级或三级 RNA 结构、靶序列的核苷酸碱基组成、靶序列的各种区之间的同源度、或在 RNA 转录物中靶序列的相对位置。基于这些确定，可以选择 RNA 转录物内的任何数目的靶位点以筛选 siRNA 分子的功效，例如通过利用体外 RNA 剪切测定、细胞培养、或动物模型。参见，例如，美国专利公开号 20030170891, September 11, 2003, McSwiggen J。最近还描述了一种用于鉴定和选择 RNAi 靶位点的算法。参见美国专利公开号 20040236517, “Selection of Target Sites for Antisense Attack of RNA,” November 25, 2004, Drlica *et al.*。

通常使用的基因转移技术包括磷酸钙、DEAE-葡聚糖、电穿孔和微注射以及病毒方法 (Graham *et al.* (1973) *Virology* 52: 456; McCutchan *et al.*, (1968), *J. Natl. Cancer Inst.* 41: 351; Chu *et al.* (1987), *Nucl. Acids Res.* 15: 1311; Fraley *et al.* (1980), *J. Biol. Chem.* 255: 10431; Capecchi (1980), *Cell* 22: 479)。还可以利用阳离子脂质体将 DNA 引入到细胞中 (Feigner *et al.* (1987), *Proc. Natl. Acad. Sci USA* 84: 7413)。商业上可获得的阳离子脂质剂型包括 Tfx 50(Promega)或 Lipofectamin 200(Life Technologies)。可替换地，病毒载体可以用来将 dsRNA 递送到细胞并介导 RNAi。参见美国专利公开号 20040023390, “siRNA-mediated Gene Silencing with Viral Vectors,” Feb. 4, 2004, Davidson *et al.*。

用于哺乳动物细胞中的 RNAi 的转染和载体/表达系统是商业上可获得的，并且已得到很好描述。参见，例如，Dharmacon, Inc., DharmaFECT™系统；Promega, Inc., siSTRIKE™ U6 Hairpin 系统；还参见 Gou *et al.* (2003) *FEBS* 548, 113-118; Sui, G. *et al.* A DNA vector-based RNAi technology to suppress gene expression in mammalian cells (2002) *Proc. Natl. Acad. Sci.* 99, 5515-5520; Yu *et al.* (2002) *Proc. Natl. Acad. Sci.* 99, 6047-6052; Paul, C. *et al.* (2002)

*Nature Biotechnology* 19, 505-508; McManus *et al.* (2002) *RNA* 8, 842-850.

然后通过将包含 dsRNA 的药物制剂给予哺乳动物，可以在哺乳动物中利用制备的 dsRNA 分子来进行 siRNA 干涉。以足以抑制靶基因表达的剂量给予药物组合物。通常可以以小于 5 mg dsRNA/千克体重/天的剂量给予 dsRNA，并且足以抑制或完全抑制靶基因的表达。通常，dsRNA 的适宜剂量将在 0.01 至 2.5 毫克/受体的千克体重/天的范围内，优选在 0.1 至 200 微克/千克体重/天的范围内，更优选在 0.1 至 100 微克/千克体重/天的范围内，甚至更优选在 1.0 至 50 微克/千克体重/天的范围内，以及最优选在 1.0 至 25 微克/千克体重/天的范围内。包含 dsRNA 的药物组合物每天给予一次，或例如利用本领域众所周知的缓释剂型以多个亚剂量给予。如下文进一步描述的，可以按照标准技术来制备和给予这样的药物组合物。

如上所述，通过制备包含治疗有效量的这样的 dsRNA 的药物制剂，并将该制剂给予患有表达 EML4-ALK 或 TFG-ALK 融合蛋白或截短的活性 ALK 激酶的癌症的人受治疗者，例如，通过直接注射到肿瘤，上述 dsRNA 可以用来在癌症中抑制 ALK 表达和活性。最近已描述了利用 siRNA 抑制剂来类似地抑制其它受体酪氨酸激酶，如 VEGFR 和 EGFR。参见美国专利公开号 20040209832, October 21, 2004, McSwiggen *et al.*; 美国专利公开号 20030170891, September 11, 2003, McSwiggen; 美国专利公开号 20040175703, September 9, 2004, Kreutzer *et al.*

### 治疗组合物; 给药

可以通过本领域已知的任何方式，包括但不限于口服或腹膜途径，其包括静脉内、肌肉内、腹膜内、皮下、经皮、气道（气雾剂）、直肠、阴道和局部（包括颊和舌下）给药，将可用于实施本发明的方法的 ALK 激酶-抑制治疗组合物给予哺乳动物。

对于口服给药，将通常以片剂或胶囊剂的形式、作为散剂或颗粒剂、或作为含水溶液或混悬剂，来提供 ALK-抑制治疗剂。用于

口服的片剂可以包括与药用赋形剂如惰性稀释剂混合的活性组分、崩解剂、粘合剂、润滑剂、甜味剂、增香剂、着色剂以及防腐剂。适宜的惰性稀释剂包括碳酸钠和碳酸钙、磷酸钠和磷酸钙、以及乳糖，而玉米淀粉和海藻酸是适宜的崩解剂。粘合剂可以包括淀粉和明胶，而润滑剂，如果存在的话，将通常是硬脂酸镁、硬脂酸或滑石粉。如果需要的话，片剂可以涂布有一种材料如甘油单硬脂酸酯或二硬脂酸甘油酯，以延迟在胃肠道中的吸收。

用于口服的胶囊剂包括硬（明胶）胶囊剂，其中活性组分与固体稀释剂混合，以及软胶囊剂，其中活性组分与水或油混合，如花生油、液状石蜡或橄榄油。对于肌内、腹膜内、皮下以及静脉内应用，本发明的药物组合物将通常以无菌含水溶液或混悬剂提供，其被缓冲到适当 pH 和等渗性。适宜的含水载体包括林格氏液和等渗氯化钠。载体可以仅仅由含水缓冲液构成（“仅仅”是指不存在任何助剂或包胶物质，其可能会影响或介导 ALK-抑制治疗剂的摄取）。这样的物质包括，例如，胶束结构，如脂质体或衣壳，如下所述。含水混悬剂可以包括悬浮剂如纤维素衍生物、海藻酸钠、聚乙烯吡咯烷酮和黄芪树胶，以及湿润剂如卵磷脂。用于含水混悬剂的适宜的防腐剂包括对羟基苯甲酸乙酯和对羟基苯甲酸正丙酯。

ALK 激酶-抑制治疗组合物还可以包括微囊剂以防止治疗剂（例如，dsRNA 化合物）快速从体内消除，如控释剂型，其包括植入物和微囊化递送系统。可以使用生物可降解的、生物相容聚合物，如乙烯乙酸乙烯酯、聚酞类、聚乙醇酸、胶原、聚原酸酯类、以及聚乳酸。用于制备上述剂型的方法对于本领域技术人员来说将是显而易见的。还可以从 Alza Corporation 和 Nova Pharmaceuticals, Inc. 商业上获得该材料。脂质体混悬剂（包括脂质体，其靶向具有相对于病毒抗原的单克隆抗体的感染细胞）也可以用作药用载体。这些混悬剂可以按照本领域技术人员已知的方法加以制备，例如，如在美国专利第 4,522,811 号、PCT 公开 WO 91/06309、以及欧洲专利公开 EP-A-43075 中所描述的。微囊剂可以包含病毒外壳蛋白。该病毒外壳蛋白可以衍生自或伴随病毒，如多瘤病毒，或它可以是部

分或全部人工的。例如，外壳蛋白可以是多瘤病毒的病毒蛋白 1 和/或病毒蛋白 2、或其衍生物。

ALK-抑制组合物还可以包含递送载体（包括脂质体，用于给予受治疗者）、载体和稀释剂以及它们的盐，和/或可以存在于药用剂型中。例如，用于递送核酸分子的方法描述在以下文献中：Akhtar *et al.*, 1992, *Trends Cell Bio.*, 2, 139; DELIVERY STRATEGIES FOR ANTISENSE OLIGONUCLEOTIDE THERAPEUTICS, ed. Akhtar, 1995, Maurer *et al.*, 1999, *Mol. Membr. Biol.*, 16, 129-140; Hofland 和 Huang, 1999, *Handb. Exp. Pharmacol.*, 137, 165-192; 以及 Lee *et al.*, 2000, ACS Symp. Ser., 752, 184-192。。 Beigelman 等人，美国专利第 6,395,713 号和 Sullivan 等人 PCT WO 94/02595 进一步描述了用于递送核酸分子的一般方法。这些程序可以用来递送实际上任何核酸分子。

可以通过本领域技术人员已知的各种方法将 ALK-抑制治疗剂给予哺乳动物肿瘤，上述方法包括但不限于，胶囊化在脂质体中，通过离子导入，或通过加入到其它载体中，如水凝胶、环糊精、生物可降解的纳米胶囊、以及生物粘附微球，或通过蛋白质载体（O'Hare 和 Normand, 国际 PCT 公开号 WO 00/53722）。可替换地，通过直接注射或通过使用灌注泵，局部递送治疗剂/载体组合。可以利用标准针和注射器方法，或通过无针技术如那些在 Conry *et al.*, 1999, *Clin. Cancer Res.*, 5, 2330-2337 和 Barry *et al.*, 国际 PCT 公开号 WO 99/31262 中所描述的无针技术进行组合物的直接注射（不管是皮下、肌肉、或皮内）。

ALK 激酶-抑制治疗剂的药用剂型包括上述化合物的盐，例如，酸加成盐，例如，盐酸、氢溴酸、乙酸、以及苯磺酸的盐。药物组合物或剂型是指具有一定形式的组合物或剂型，其适合于给予（例如，系统给予）细胞或患者，包括例如人。适当的形式部分地取决于应用或进入途径，例如口服、经皮、或通过注射。这样的形式应当不会阻止组合物或剂型到达靶细胞。例如，注入血流的药物组合物应当是可溶的。其它因素是本领域已知的，并且包括诸如毒性和形式等的考虑事项，其阻止组合物或剂型发挥其效应。

导致系统吸收（即，药物系统吸收或蓄积在血流中，接着分布于全身）的给予途径是所期望的并且包括但不限于：静脉内、皮下、腹膜内、吸入、口服、肺内以及肌内。这些给予途径的每一种将ALK-抑制治疗机暴露于可接近的疾病组织或肿瘤。已表明，药物进入循环的速率是分子量或大小的函数。使用包含本发明的化合物的脂质体或其它药物载体可以潜在地将药物定位在，例如，某些组织类型，如网状内皮系统（RES）的组织。可以促进药物缔合于细胞（如，淋巴细胞和巨噬细胞）表面的脂质体剂型也是有用的。通过利用巨噬细胞和淋巴细胞免疫识别异常细胞（如癌细胞）的特异性，这种方式可以提供增强药物递送到靶细胞。

“药用剂型”是指一种组合物或剂型，其便于将本发明的核酸分子有效分布在最适合于它们所期望活性的物理位置。适用于具有本发明的核酸分子的剂型的制剂的非限制性实例包括：P糖蛋白抑制剂（如 Pluronic P85），其可以增强药物进入 CNS（Jolliet-Riant 和 Tillement, 1999, *Fundam. Clin. Pharmacol.*, 13, 16-26）；生物可降解聚合物，如聚(DL-交酯-共羟基乙酸)微球，用于在脑内植入后的缓释递送（Emerich *et al.*, 1999, *Cell Transplant*, 8, 47-58）（Alkermes, Inc. Cambridge, Mass.）；以及负荷纳米微粒，如那些由聚氰基丙烯酸正丁酯构成的负荷纳米微粒，其可以递送药物穿过血脑屏障并可以改变神经元摄取机制（*Prog Neuro-psychopharmacol Biol Psychiatry*, 23, 941-949, 1999）。可用于本发明的方法的ALK-抑制化合物的递送策略的其它非限制性实例包括在 Boado *et al.*, 1998, *J. Pharm. Sci.*, 87, 1308-1315；Tyler *et al.*, 1999, *FEBS Lett.*, 421, 280-284；Pardridge *et al.*, 1995, *PNAS USA.*, 92, 5592-5596；Boado, 1995, *Adv. Drug Delivery Rev.*, 15, 73-107；Aldrian-Herrada *et al.*, 1998, *Nucleic Acids Res.*, 26, 4910-4916；以及 Tyler *et al.*, 1999, *PNAS USA.*, 96, 7053-7058 中所描述的材料。

包含表面修饰脂质体（其含有聚乙二醇脂质(PEG 修饰、或长循环脂质体，或隐形脂质体)）的治疗组合物还可以适当地用于本发明的方法。这些剂型提供一种用于在靶组织中增加药物蓄积的方法。通过单核吞噬细胞系统（MPS 或 RES），这类药物载体耐调理

和消除，从而可以获得更长的血液循环时间并增强组织暴露于胶囊化药物 (Lasic *et al.* *Chem. Rev.* 1995, 95, 2601-2627; Ishiwata *et al.*, *Chem. Pharm. Bull.* 1995, 43, 1005-1011)。已表明，这样的脂质体选择性地蓄积在肿瘤中，大概通过外渗和捕捉在新血管形成的靶组织中 (Lasic *et al.*, *Science* 1995, 267, 1275-1276; Oku *et al.*, 1995, *Biochim. Biophys. Acta*, 1238, 86-90)。长循环脂质体可增强 DNA 和 RNA 的药物动力学和药效学，尤其与已知蓄积在 MPS 的组织中的常规阳离子脂质体相比 (Liu *et al.*, *J. Biol. Chem.* 1995, 42, 24864-24870; Choi *et al.*, 国际 PCT 公开号 WO 96/10391; Ansell *et al.*, 国际 PCT 公开号 WO 96/10390; Holland *et al.*, 国际 PCT 公开号 WO 96/10392)。与阳离子脂质体相比，长循环脂质体还可以在更大程度上保护药物免受核酶降解，其是基于它们避免蓄积在代谢侵袭性 MPS 组织如肝和脾中的能力。

治疗组合物可以包括药物有效量的在药用载体或稀释剂中的所期望的化合物。用于治疗用途的可接受的载体或稀释剂在制药领域是众所周知的，并且描述在例如 REMINGTON'S PHARMACEUTICAL SCIENCES, Mack Publishing Co. (A. R. Gennaro, Ed. 1985) 中。例如，可以提供防腐剂、稳定剂、染料以及增香剂。其包括苯甲酸钠、山梨酸以及对羟基苯甲酸的酯。此外，可以使用抗氧化剂和悬浮剂。

药物有效剂量是防止、抑制疾病状态的发生、或治疗疾病状态 (减轻症状到一定程度、优选所有症状) 所需要的剂量。药物有效剂量取决于疾病类型、所用的组合物、给药途径、待治疗哺乳动物的类型、所考虑的特定哺乳动物的物理特性、同时发生的用药、以及医疗领域技术人员将明了的其它因素。通常，给予 0.1 mg/kg 到 100 mg/kg 体重/天的活性组分之间的量，其取决于带负电荷聚合物的效能。

约 0.1 mg 至约 140 mg/千克体重/天的剂量水平可用于治疗上述病症 (约 0.5 mg 至约 7 g/患者/天)。可以结合于载体材料以产生单剂量形式的活性组分的量随待治疗的宿主和特定的给予方式而变化。剂量单位形式通常包含约 1 mg 至约 500 mg 之间的活性组分。

应当明了，用于任何特定患者的具体剂量水平取决于各种因素，包括所采用特定化合物的活性、年龄、体重、一般健康状况、性别、饮食、给药时间、给药途径、排泄速率、药物组合以及接受治疗的特定疾病的严重性。

为了给予非人动物，还可以将组合物加入到动物饲料或饮用水中。可以方便地配制动物饲料和饮用水组合物，使得动物连同其饮食获得治疗适宜量的组合物。还可以方便地提供组合物作为预混合料，用于加入到饲料或饮用水中。

可用于实施本发明的 ALK-抑制治疗剂可以包含如上所述的单一化合物，或多种化合物的组合，不管是相同类型的抑制剂（即，抗体抑制剂）、或不同类型的抑制剂（即，抗体抑制剂和小分子抑制剂）。这样的化合物组合在抑制融合蛋白-表达癌症的进展时可以增加总治疗效应。例如，治疗组合物可以是小分子抑制剂，如单独的 WHI-131 和/或 WHI-154，或连同靶向 ALK 活性的其它抑制剂和/或其它小分子抑制剂。除一种或多种靶向抑制剂之外，治疗组合物还可以包括一种或多种非特异性化疗药物。最近已表明，这样的组合可以在许多癌症中提供协同肿瘤杀伤效应。这样的组合在抑制 ALK 活性和肿瘤体内生长方面的有效性可以如下所述加以评估。

### 突变体 ALK 激酶-抑制化合物的鉴定

本发明还部分地提供了一种方法，用来确定一种化合物是否抑制以 EML4-ALK 或 TFG-ALK 融合多核苷酸和/或融合多肽为特征的癌症的进展，其中通过确定该化合物是否抑制 EML4-ALK 或 TFG-ALK 融合多肽或截短的 ALK 激酶多肽在癌症中的活性。在某些优选的实施方式中，通过检查包含来自骨髓、血液、胸膜渗出液、或肿瘤的细胞的生物样品来确定 ALK 活性的抑制。在另一种优选的实施方式中，利用本发明的至少一种突变体 ALK 多核苷酸或多肽-特异性试剂来确定 ALK 活性的抑制。

试验化合物可以是如上所述的任何类型的治疗剂或组合物。用于评估化合物功效（体外和体内）的方法被很好建立并且在本领域是已知的。例如，利用其中 ALK 被激活的细胞或细胞提取物可以试验组合物体外抑制 ALK 的能力。一组化合物可以用来试验化合物对于 ALK 的特异性（不同于其它靶，如 EGFR 或 PDGFR）。

可以使用的另一种药物筛选技术可以高通量筛选对于感兴趣的蛋白质具有适当结合亲合力的化合物，如在公开的 PCT 申请 W084/03564 中所描述的。以这种方法，当应用于突变体 ALK 多肽时，大量的不同小试验化合物被合成在固相基板上，如塑料针或某些其它表面。试验化合物与突变体 ALK 多肽、或其片段反应，然后洗涤。然后通过本领域众所周知的方法检测结合的突变体多肽（例如，EML4-ALK 融合多肽）。还可以将纯化的突变体 ALK 多肽直接涂布在平板上，用于上述药物筛选技术。可替换地，非中和抗体可以用来捕捉肽并将它固定在固体载体上。

对于发现是 ALK 体外活性的有效抑制剂的化合物，然后可以检查其体外抑制表达 EML4-ALK 或 TFG-ALK 融合多肽和/或截短的 ALK 激酶多肽的癌症的进展的能力，其中利用，例如，哺乳动物异种移植躲藏人肿瘤，如 NSCLC。以这种方式，可以在最精密地类似患者的生物装置中观测药物的效应。可以通过用磷酸化-特异性抗体进行分析来确定在癌细胞或周围基质细胞中药物改变信号的能力。还可以通过用凋亡特异性标志如剪切的胱冬裂酶 3 和剪切的 PARP 进行分析来观测药物在诱导细胞死亡或抑制细胞增殖中的有效性。类似地，可以采用由突变体 ALK 蛋白质驱动的哺乳动物骨髓移植体（例如，小鼠），其躲藏人白血病。在这种方式中，将已知由突变体 ALK 激酶驱动的骨髓细胞移植到小鼠中。可以监测癌细胞的生长。然后可以用药物治疗小鼠，并外部观测药物治疗对癌症表型或进展的影响。然后处死小鼠并除去移植的骨髓，供通过 IHC 和蛋白质印迹等进行分析。

这样的化合物的毒性和治疗功效可以通过在细胞培养或实验动物中的标准药理学程序来确定，例如，用于确定 LD50（50%群体

的致死剂量)和ED50(在50%群体中的治疗有效剂量)。毒性和治疗效应之间的剂量比率是治疗指数并且它可以表示为比率LD50/ED50。呈现高治疗指数的化合物是优选的。

上文和下文引用的所有参考文献的教导以引用方式结合于本文。以下实施例仅用来进一步说明本发明,而不是限制其范围,其由所附的权利要求所限定。本发明包括本文教导的方法的改进和变化,其对于本领域技术人员来说是显而易见的。

## 实施例 1

### 通过全面磷酸肽分布鉴定实体瘤中的 ALK 激酶活性

#### A. 人 NSCLC 细胞系的分布

利用最近描述的、用于从复杂混合物分离和质谱表征修饰肽的有力技术(“IAP”技术,参见 Rush *et al.*, 上文)检查了在 22 种人 NSCLC 细胞系(包括 H2228)中激酶激活的全面磷酸化分布。利用磷酸酪氨酸特异性抗体(CELL SIGNALING TECHNOLOGY, INC., Beverly, MA, 2003/04 Cat. #9411)实施 IAP 技术,来从 NSCLC 细胞系的提取物中分离并随后表征含磷酸酪氨酸的肽。

具体地说, IAP 方式用来促进鉴定酪氨酸激酶,其对每个 NSCLC 细胞系中的蛋白质磷酸化负责。尤其是,考虑到非典型的或异常的激酶活性。

#### 细胞培养

所有细胞培养试剂购买自 Invitrogen, Inc.。检查了总共 41 种人 NSCLC 细胞系。人 NSCLC 细胞系, H520、H838、H1437、H1563、H1568、H1792、H1944、H2170、H2172、H2228、H2347、A549、H441、H1703、H1373、以及 H358, 获自 American Type Culture Collection, 并培养在 RPMI 1640 培养基中, 该培养基含有 10% FBS 并加以调节以包含 2 mM L-谷氨酰胺、1.5 g/L 碳酸氢钠、4.5 g/L 葡

葡萄糖、10 mM HEPES、1.0 mM 丙酮酸钠、青霉素/链霉素。另外的6种人 NSCLC 细胞系, HCC78、Cal-12T、HCC366、HCC15、HCC44、以及 LOU-NH91, 购买自 DSMZ 并在包含 10% FBS 和青霉素/链霉素的 RPMI 1640 中进行培养。将细胞保持在 37°C 的 5% CO<sub>2</sub> 培养箱中。

为了进行免疫亲和沉淀和免疫印迹实验, 使细胞生长至 80% 汇合, 然后在收获以前在没有 FBS 的 RPMI 培养基中饥饿过夜。

### 磷酸肽免疫沉淀

将 1 亿个细胞溶解在脲裂解缓冲液 (20 mM N-羧乙基哌嗪-N'-2-乙烷磺酸、pH 8.0、9 M 脲、1 mM 钒酸钠、2.5 mM 焦磷酸钠、1 mM  $\beta$ -甘油磷酸) 中。超声处理溶胞产物并通过离心作用清除溶胞产物。用 DTT 还原清除的溶胞产物, 然后用碘乙酰胺烷基化, 如先前所描述的 (参见 Rush, *et al.*, *Nat. Biotechnol.* 23(1): 94-101 (2005))。然后用 20 mM N-羧乙基哌嗪-N'-2-乙烷磺酸稀释样品 4 次以降低脲浓度至 2M, 然后在室温下用胰蛋白酶消化过夜同时温和摇动。

用 Sep-Pak C18 柱粗略纯化消化的肽, 如先前描述的 (参见 Rush *et al.*, 上文)。冷冻并干燥洗脱物并将干燥肽溶解在 1.4 ml 的 MOPS IP 缓冲液 (50 mM MOPS/NaOH pH 7.2、10 mM Na<sub>2</sub>PO<sub>4</sub>、50 mM NaCl) 中并通过离心作用除去不溶物质。在 4°C 下用耦联于蛋白质 G 琼脂糖珠 (Roche) 的 160  $\mu$ g 的磷酸酪氨酸 100 抗体 (细胞信号传导技术) 进行免疫沉淀过夜。然后用 1 ml MOPS IP 缓冲液洗涤该珠 3 次, 并用 1 ml HPLC 级冷 dH<sub>2</sub>O 洗涤两次。用 60  $\mu$ l 0.1% TFA 从珠洗脱磷酸肽, 接着用 40  $\mu$ l 0.1% TFA 进行第二次洗脱, 然后汇集级分。利用 ZipTip 柱 (Millipore) 浓缩洗脱的肽, 并用 LC-MS/MS 进行分析。用 LTQ 离子阱质谱仪 (ThermoFinnigan) 收集质谱。

## 通过 LC-MS/MS 质谱法进行分析

浓缩 IP 洗脱物(100  $\mu$ l)中的肽并利用停止之后再行的提取尖管 (StageTips)(参见 *Rappsilber et al., Anal. Chem., 75(3): 663-70 (2003)*) 与洗脱的抗体分离。用 1  $\mu$ l 的 60% MeCN、0.1% TFA 将肽从微柱洗脱到 7.6  $\mu$ l 的 0.4%乙酸/0.005%七氟丁酸 (HFBA) 中。

对每个磷酸肽样品进行 LC-MS 分析两次。用 C18 反相树脂 (Magic C18AQ, 5  $\mu$ m 颗粒, 200 Å 孔径, Michrom Bioresources, Auburn, CA) 装填熔融硅石微毛细管柱 (125  $\mu$ m  $\times$  18 cm)。用自动进样器 (LC Packings Famos, San Francisco, CA) 将样品(4  $\mu$ L)装到此柱上, 然后通过 0.1%甲酸中的 7 至 30%乙腈的 55 分钟线性梯度将样品洗脱到质谱仪中。利用具有在管线中的流动分流器的二元 HPLC 泵 (Agilent 1100, Palo Alto, CA) 并以大约 1000 nL/min 来递送梯度。用混合线性离子阱-7 Tesla 离子回旋共振傅里叶变换仪 (LTQ-FT, Thermo Finnigan, San Jose, CA) 对洗脱肽离子进行质量分析。

采用了顶七方法, 从而基于在 ICR 细胞中的先前 MS 测量扫描期间获得的测量结果收集在线性离子阱中的 7 个数据相关 MS/MS 扫描, 其中同时操作线性离子阱和傅里叶变换仪。以 375-1800 m/z 进行 MS 扫描, 其中自动增益控制 (AGC) 靶为  $8 \times 10^6$  以及质量分辨率为 105。对于 MS/MS, AGC 是  $8 \times 10^6$ , 动态排斥时间是 25 s, 以及通过电荷态筛选来排斥单电荷离子。

## 数据库分析和赋值

利用 TurboSequest 软件 (v.27, rev.12)(ThermoFinnigan) 和复合正向/反向 IPI 人蛋白质数据库, 将 MS/MS 谱赋予肽序列。搜索参数是: 胰蛋白酶作为蛋白酶; 1.08 Da 前体质量耐受性; 在半胱氨酸上的静态修饰 (+57.02146, 酰胺甲基化); 以及在丝氨酸、苏氨酸和酪氨酸 (+79.96633 Da, 磷酸化), 赖氨酸 (+8.01420,

13C615N2), 精氨酸 (+6.02013, 13C6) 以及甲硫氨酸 (+15.99491, 氧化) 上的动态修饰。靶/引诱物数据库方式用来建立适当的得分-过滤准则以使建立的假-正赋值率小于 1%。除了过电荷相关 XCorr 阈值 ( $z=1$ ,  $XCorr \geq 1.5$ , 对于  $z=2$ ,  $XCorr \geq 2.2$ , 对于  $z=3$ ,  $XCorr \geq 3.3$ ) 以外, 需要赋值包含磷酸酪氨酸, 以具有 -5 至 +25 ppm 的质量精度, 以及包含所有轻的或所有重的赖氨酸/精氨酸残基。

利用定制量化程序, Vista (Bakalarski *et al.*, 原稿在准备中) 进一步评估通过这些准则的赋值以计算峰面积以及最终的每种肽的重和轻形式之间的相对丰度。用在 MS 扫描中低于 15 的信噪鉴定的肽不认为是定量。对于那些仅在条件之一下发现的肽而是使用信噪比。

搜索了于 2004 年 8 月 24 日发布的 NCBI 人数据库, 该数据库包含 27,175 种蛋白质, 其允许氧化型甲硫氨酸 (M+16) 和磷酸化 (Y+80) 作为动态修饰。由至少三位科学家综述了支持赋值序列的最终清单的所有谱 (本文未示出) 以确定它们的可靠性。

上述 IAP 分析鉴定了 2000 种以上的非丰余含磷酸酪氨酸的肽、1,500 种以上的磷酸酪氨酸位点、以及大于 1,000 种的酪氨酸磷酸化蛋白质, 其大多数是新的, 并来自已检查的细胞系 (数据未示出)。在许多细胞系中, 如 EGFR、Her2、Her3、EphA2 以及 Met, 已知与 NSCLC 信号有关的受体酪氨酸激酶观测到被酪氨酸磷酸化。在多种细胞系中观测到高水平的 EGFR 磷酸肽, 上述多种细胞系包括 HCC827 和 H3255, 已知该两种细胞系表达增加水平的 EGFR 的基因活化形式, 这证实了该方法可以鉴定已知在 NSCLC 细胞系中为活性的受体酪氨酸激酶。

三种细胞系表达的受体酪氨酸激酶未在其它 NSCLC 细胞系中观测到。在 HCC78、H2228、以及 H1703 细胞系中分别观测到来自 Ros、ALK、以及 PDGFR $\alpha$  的大量酪氨酸磷酸化肽。选择高度表达 ALK 的 NSCLC 细胞系 H2228, 用于进一步检查。

## B. 人 NSCLC 肿瘤样品的分布

IAP 技术,基本上如在上面的部分 A 中所描述的,随后用来检查来自 NSCLC 患者的一组 154 个人肿瘤样品的全面磷酸分布。组织获自中国第二湘雅医院。

将冷冻组织样品切成小片,均化在裂解缓冲液(20 mM N-羟乙基哌嗪-N'-2-乙烷磺酸 pH 8.0、9 M 脲、1 mM 钒酸钠,补充有 2.5 mM 焦磷酸钠、1 mM β-甘油磷酸,1 ml 裂解缓冲液用于 100 mg 冷冻组织)中,其中使用 polytron 两次,每次 20 秒。然后对匀浆液进行简单超声处理。用 DTT 还原清除的溶胞产物并用碘乙酰胺加以烷基化,如先前所述(参见 Rush *et al.*, *Nat. Biotechnol.* 23(1): 94-101 (2005))。然后用 20 mM N-羟乙基哌嗪-N'-2-乙烷磺酸稀释样品 4 次以将脲浓度降低至 2M,并在室温下用胰蛋白酶消化过夜同时温和摇动。

用 Sep-Pak C18 柱粗略纯化消化的肽,如先前描述的(参见 Rush *et al.*, 上文)。冷冻并干燥洗脱物并将干燥肽溶解在 1.4 ml MOPS IP 缓冲液(50 mM MOPS/NaOH pH 7.2、10 mM Na<sub>2</sub>PO<sub>4</sub>、50 mM NaCl)中并通过离心作用除去不溶物质。在 4°C 下用耦联于蛋白质 G 琼脂糖珠(Roche)的 160 μg 的磷酸酪氨酸 100 抗体(细胞信号传导技术)进行免疫沉淀过夜。然后用 1 ml MOPS IP 缓冲液洗涤该珠 3 次,并用 1 ml HPLC 级冷 dH<sub>2</sub>O 洗涤两次。用 60 μl 0.1% TFA 从珠洗脱磷酸肽,接着用 40 μl 0.1% TFA 进行第二次洗脱,然后汇集级分。利用 ZipTip 柱(Millipore)浓缩洗脱的肽,并用 LC-MS/MS 进行分析。用 LTQ 离子阱质谱仪(ThermoFinnigan)收集质谱。

如上文在部分 A 中所描述的,进行磷酸肽免疫沉淀,接着进行 LC-MS/MS 分光测定分析。基本上如上文在部分 A 中所描述的,进行数据库搜索和序列赋值,但利用于 2004 年 8 月 24 日发布的包含 27,970 种蛋白质的 NCBI 人数据库。

上述 IAP 分析鉴定了 2000 种以上的非冗余含磷酸酪氨酸的肽、1,500 种以上的磷酸酪氨酸位点、以及 1,000 种以上的酪氨酸磷酸化

蛋白质，其来自已检查的人肿瘤样品（数据未示出）。在许多肿瘤中，如 EGFR、Her2、Her3、EphA2 以及 Met，已知与 NSCLC 信号有关的受体酪氨酸激酶再次观测到被酪氨酸磷酸化。在多个肿瘤样品中再次观测到高水平的 EGFR 磷酸肽，这证实了该方法可以鉴定已知在 NSCLC 细胞系中为活性的受体酪氨酸激酶。

在其它 NSCLC 细胞系和肿瘤中未观测到 5 位患者样品表达的受体酪氨酸激酶。在患者 CS010/11、CS045、以及 CS110L 中观测到大量的来自 ALK 的酪氨酸磷酸化肽。这三种肿瘤，其高度表达 ALK，被选择用于进一步检查。

## 实施例 2

### 三种 ALK 融合基因的分离和测序

#### A. 在人 NSCLC 细胞系的测序

鉴于在 NSCLC 细胞系 H2228 中检测到 ALK 激酶的高磷酸化水平，进行了在编码 ALK 的激酶结构域的序列上的 cDNA 末端的 5'快速扩增，以便确定是否存在嵌合 ALK 转录物。

#### 互补 DNA 末端的快速扩增

RNeasy Mini Kit (Qiagen) 用来从 H2228 细胞系提取 RNA。通过使用 DNeasy Tissue Kit (Qiagen) 来提取 DNA。借助于 5'RACE 系统 (Invitrogen) 来快速扩增 cDNA 末端，其中引物 ALK-GSP1 用于 cDNA 合成而 ALK-GSP2 和 ALK-GSP3 用于嵌套式 PCR 反应。

#### 5'RACE

图 5(图片 A)示出了通过 5'RACE 对 EML4-ALK 融合基因(短变体)的检测以及在 2 轮以后 PCR 扩增产物的检测。用 PCR 纯化试剂盒 (Qiagen) 纯化 PCR 产物并利用 ALK-GSP3，一种 ABI 3130 毛细管自动 DNA 测序仪 (Applied Biosystems) 进行测序。所得到

的产物的序列分析揭示了 ALK 的激酶结构域和 C 端被融合到 EML-4 基因 N 端 (参见图 1, 图片 B)。EML4-ALK 融合基因 (短变体) 是框内的并将 EML-4 的开始的 233 个氨基酸融合到 ALK 的最后的 562 个氨基酸 (参见图 1, 图片 B)。EML-4 和 ALK 基因均位于第 2 号染色体上, 因此通过该两个基因座之间的基因缺失而产生融合基因。

使用了以下引物:

ALK-GSP1: 5'-GCAGTAGTTGGGGTTGTAGTC (SEQ ID NO: 9)

ALK-GSP2: 5'-GCGGAGCTTGCTCAGCTTGT (SEQ ID NO: 10)

ALK-GSP3: 5'-TGCAGCTCCTGGTGCTTCC (SEQ ID NO: 11)

### PCR 测定

进行 RT-PCR 分析以证实在融合蛋白中 EML-4 的 N 端是完整的 (参见图 6(图片 B))。cDNA 的第一条链合成自 2.5 mg 的总 RNA, 其中借助于具有寡(dT)<sub>20</sub> 的 SuperScript™ III 第一条链合成系统 (Invitrogen)。然后, 借助于引物对 EML-Atg 和 ALK-GSP3 来扩增 EML4-ALK 融合基因。借助于引物对 EML-4-43 和 ALK-GSP3、EML-4-94 和 ALK-GSP3、以及 EML4-202 和 ALK-GSP3, 来检测交互融合。对于基因组 PCR, 通过使用高保真 Platinum Taq DNA 聚合酶(Invitrogen)并借助于引物对 EML-4-atg 和 ALK-tga 来进行融合基因的扩增。

使用了以下引物:

ALK-GSP3: 5'-TGCAGCTCCTGGTGCTTCC (SEQ ID NO: 12)

EML4-Atg: 5'-CGCAAGATGGACGGTTTGGC (SEQ ID NO: 13)

EML4-43: 5'-TGTTCAAGATCGCCTGTCAGCTCT (SEQ ID NO: 14)

EML4-94: 5'-TGAAATCACTGTGCTAAAGGCGGC (SEQ ID NO: 15)

EML4-202: 5'-AAGCCCTCGAGCAGTTATTCCCAT (SEQ ID NO: 16)

ALK-Tga: 5'-GAATTCCGCCGAGCTCAGGGCCCAG (SEQ ID NO: 17)

值得注意的是，在 EML4-ALK 融合（短变体）中，ALK 部分，以在其它 ALK 融合（如发生在 ALCL 中的 NPM-ALK 融合）中观测到的在 ALK 中精确的相同点，被融合到 EML-4 部分。按照基因组 DNA，进一步对 H2228 细胞系中的 ALK 的激酶结构域进行测序并发现是野生型。因此，在 H2228 中发现的缺失突变并不影响 ALK 激酶结构域。另外，野生型 EML-4 仅在 EML4-ALK 融合蛋白（短变体）中不存在的位点被酪氨酸磷酸化，这提示在融合蛋白中为保守的 N 端卷曲螺旋结构域（参见图 1A）可以起作用，以二聚化和活化 ALK、以及促进与野生型 ALK 的相互作用。

## B. 在人 NSCLC 细胞系中的测序

类似地，鉴于在来自患者 CS010/11、CS045、以及 CS110 的人 NSCLC 肿瘤样品中检测到 ALK 激酶的高磷酸化水平，进行了在编码 ALK 的激酶结构域的序列上的 cDNA 末端的 5'快速扩增，以便确定在这些肿瘤中是否存在嵌合 ALK 转录物。

基本上如上面在部分 A 中所描述的，进行互补 DNA 端的快速扩增和 5'RACE，其中引物 ALK-GSP1 用于 cDNA 合成而 ALK-GSP2 和 ALK-GSP3 用于嵌套式 PCR 反应。

图 5（图片 C）示出了在两个患者样品中通过 5'RACE 对 EML4-ALK 融合基因（短和长变体两者）的检测，在一位患者中 TFG-ALK 融合基因的检测，以及在 2 轮以后 PCR 扩增产物的检测。基本上如在上面的部分 A 中所描述的，对 PCR 产物进行纯化和测序。所得到的产物的序列分析揭示了 ALK 的激酶结构域和 C 端以

两种不同的变体融合到 EML-4 基因 N 端(参见图 1A-1B, 图片 B)。EML4-ALK 融合基因是框内的并将 EML-4 的开始的 233 个氨基酸(短变体)或开始的 495 个氨基酸(长变体)融合到 ALK 的最后的 562 个氨基酸(参见图 1A-1B, 图片 B)。EML-4 和 ALK 基因均位于第 2 号染色体上, 因此通过这两个基因座之间的基因缺失来产生融合基因。在患者 CS045 中融合基因(短变体)的观测结果证实了在人细胞系 H2228 中这种突变体基因的发现。

TFG-ALK 融合基因也是框内的并且将 TFG 的开始的 138 个氨基酸融合到 ALK 的最后的 562 个氨基酸(参见图 1C, 图片 B)。TFG 和 ALK 基因位于不同的染色体上(分别为第 6 号和第 2 号染色体), 因此通过这两个基因座之间的基因易位来产生融合基因。有趣的是, TFG 与 ALK 的融合发生在针对两种 EML4-ALK 变体的融合所观测到的在 ALK 中的精确相同的点, 这表明在实体瘤中在此点 ALK 的截短可以经常发生。

如在上面的部分 A 中所描述的, 使用相同的引物。基本上如在上面的部分 A 中所描述的, 进行 RT-PCR 分析, 以证实 EML-4 和 TFG 的 N 端在融合蛋白中是完整的(参见图 6(图片 B))。用于 EML-4 和 ALK 的引物对是如在上面的部分 A 中所描述的。以下引物对用于 TFG:

TFG-F1: 5'-TTTGTTAATGGCCAGCCAAGACCC-3 (SEQ ID NO: 28)

值得注意的是, 在两种 EML4-ALK 融合变体中, ALK 部分, 以在其它 ALK 融合(如发生在 ALCL 中的 NPM-ALK 融合)中观测到的在 ALK 中精确的相同点, 被融合到 EML-4 部分。另外, 野生型 EML-4 仅在 EML4-ALK 融合蛋白中不存在的位点被酪氨酸磷酸化, 这提示在融合蛋白中为保守的 N 端卷曲螺旋结构域(参见图 1A-1B)可以起作用, 以二聚化和活化 ALK、以及促进与野生型 ALK 的相互作用。同样值得注意的是, TG 部分与 ALK 的融合也发生在 ALK 中的精确相同的点, 并且确实是在人淋巴瘤中已描述了

在此点的 TFG 与 ALK 的融合(参见 Hernandez *et al.* (2002), 上文), 但在人实体瘤, 如 NSCLC, 中先前并没有描述。

### 实施例 3

#### 利用 siRNA 抑制 ALK 融合-表达哺乳动物实体瘤的生长

为了证实 ALK 的截短的/融合形式正驱动在 NSCLC 细胞系 H2228 以及来自患者 CS010/11、CS045、以及 CS110 的 NSCLC 肿瘤样品中的细胞生长和存活, 可以检查 siRNA (相对于 ALK) 抑制这些细胞和肿瘤生长的能力。

ALK SMARTpool siRNA 双链体 (专有靶序列-数据未示出) 可以购买自, 例如, Dharmacon Research, Inc. (Lafayette, CO)。非特异 SMARTpool siRNA 用作对照。借助于电穿孔并用 siRNA 转染细胞。简单地说, 利用方形波电穿孔仪 (BTX Genetronics, San Diego, CA) 对  $2 \times 10^7$  个细胞 (H2228) 发送脉冲一次 (20ms; 275V, K562 20ms; 285V), 在室温下温育 30 分钟, 然后转移到含有 30 ml RPMI-1640/10% FBS 的 T150 烧瓶中。

借助于 CellTiter 96 AQueous One 溶液细胞增殖测定 (Promega) 来确定能生存细胞的数目。通过使用 OriginPro 6.1 软件 (OriginLab) 来计算  $IC_{50}$ 。可以通过 Cleaved-Caspase-3 (Cell Signaling Technology) 的流式细胞分析来确定在 48 小时时凋亡细胞的百分率。

免疫印迹分析揭示, 在 siRNA 转染进入 H2228 细胞或来自患者 CS010/11、CS045、以及 CS110 的肿瘤细胞以后 72 小时, ALK 的表达被特异性地和显著地降低。ALK 的负调节预期会导致细胞生长的强抑制。用 ALK siRNA 治疗也预期会导致这些实体瘤细胞的增加的凋亡。这些结果将进一步表明, H2228 细胞系和患者肿瘤中的突变体/融合 ALK 激酶正驱动这些 NSCLC 细胞的增殖和生长, 以及通过利用 siRNA 来抑制 ALK 激酶表达和活性可以抑制这样的生长和增殖。

## 实施例 4

### 利用 WI-131 和/或 WI-154 来抑制 ALK 融合-表达哺乳动物实体瘤的生长

为了进一步证实突变体 ALK 融合蛋白正驱动来自患者 CS010/11、CS045、以及 CS110 的 NSCLC 细胞系 H2228 和 NSCLC 肿瘤细胞的生长和生存,可以用 ALK 激酶的靶向抑制剂,如 WI-131 和/或 WI-154,来处理细胞。WI-131 和 W-154 是 ALK 激酶的喹唑啉类小分子抑制剂,并且已描述了在 T 细胞淋巴瘤中它们相对于 NPM-ALK 融合蛋白的活性。参见 Marzec et al., 上文。

简单地说,培养 NSCLC 细胞,然后用 CellTiter 96 AQueous One Solution Cell Proliferation Assay (Promega)并按照制造商的建议,进行细胞生长抑制测定。简单地说,将 1000 至 5000 个细胞接种到平底 96 孔平板上并在含有 10% FBS 的完全培养基中生长。在 24 小时以后,细胞培养基改变成含有 10% FBS 的 100  $\mu$ l 完全生长培养基,其包含各种浓度的药物,然后另外温育细胞 72 小时。将每种药物浓度施加于细胞的三个孔。在温育结束时,将 20  $\mu$ l 的 CellTiter 96 AQueous One 溶液加入到每个孔中,并温育平板 1-4 小时。利用 Titan Multiskan Ascent 微量板读数器 (Titertek Instrument) 在 490 nm 处读出吸光度。生长抑制可以表示为读自经治疗细胞与未治疗细胞的吸光度百分率的平均值 $\pm$ SD 值。重复测定至少三次。

这样的分析预期会证实,ALK 融合蛋白 (EML4-ALK(短和长变体)、TFG-ALK) 正驱动人 NSCLC 肿瘤(其中表达这些突变体蛋白质)亚型的生长和存活,以及通过利用靶向抑制剂,如 WI-131 和/或 WI-154,来抑制融合 ALK 激酶的活性可以抑制上述细胞。

## 实施例 5

### ALK 融合蛋白驱动转化哺乳动物细胞系的生长和存活

为了证实一种或多种 ALK 融合蛋白的表达可以将正常细胞转化成癌性表型，可以用上述 cDNA 构建物（实施例 2）来转化 3T# 细胞，其分别表达 EML4-ALK（短和长变体）或 TFG-ALK 融合蛋白。

简单地说，将细胞保持在 RPMI-1640 培养基（Invitrogen）中，其含有 10% 小牛血清(FBS)(Sigma)和 1.0 ng/ml IL-3(R&D Systems)。如前所述进行反转录病毒上清液的生产 and 转导。参见 Schwaller *et al.*, *Embo J.* 17(18): 5321-33 (1998)。用包含 MSCV-Neo/EML4-ALK（或 TFG-ALK）载体的反转录病毒上清液转导 3T3 细胞并选择用于 G418（1 mg/ml）。然后在 PBS 中洗涤细胞三次以后，通过平板转导细胞来评估转化细胞在软琼脂上生长的能力。如果需要，为了获得剂量反应曲线，如上所述（参见实施例 3）用相对于 ALK 的 siRNA 处理细胞，并用 CellTiter 96 AQueous One 溶液细胞增殖测定（Promega）来确定能生存细胞的数目。可以借助于 OriginPro 6.1 软件（OriginLab）来计算 IC<sub>50</sub>。可以通过 Cleaved-Caspase-3 的流式细胞分析并利用对于此靶特异的抗体（Cell Signaling Technology）来确定在 48 小时时的凋亡细胞百分率。这样的分析将表明，EML4-ALK 融合蛋白（短或长变体）或 TFG-ALK 融合蛋白的表达可以转化 3T3 细胞，并且当这些细胞由 ALK 融合蛋白驱动时证实软琼脂上的存活和生长，以及进一步，在转化细胞中抑制 ALK 表达可导致降低的生存力和增加的凋亡。

## 实施例 6

### 利用 FISH 测定来检测 EML4-ALK 融合蛋白在人实体瘤中的表达

如前所述，利用荧光原位杂交（FISH）测定来检测人 NSCLC 肿瘤样品中 EML4-ALK 融合蛋白（短变体）的存在。参见，例如，

Verma *et al.* HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES, Pergamon Press, New York, N.Y. (1988)。检查了 200 个以上的石蜡包埋人 NSCLC 肿瘤样品。

ALK 双色、断裂重排探针获自 Vysis (Vysis, Dowers Grove, IL, USA) 并按照制造商的说明加以使用, 其中具有以下改进。简单地说, 再水化石蜡包埋组织切片并在 0.01M 柠檬酸缓冲液 (pH 6.0) 中经受微波抗原恢复 11 分钟。在 37°C 下, 用蛋白酶 (4mg/ml 胃蛋白酶, 2000-3000U/mg) 消化切片 25 分钟, 脱水并杂交于设置在 37°C 下的 FISH 探针 18 小时。在洗涤以后, 将在 Vectashield 封固剂 (Vector Laboratories, Burlingame, CA) 中的 4',6-二脒基-2-苯基吲哚 (DAPI; mg/ml) 用于核对比染色。

ALK 重排探针包括在野生型序列 (SEQ ID NO: 6) 中在 ALK 基因的断裂点 (在核苷酸 3171 处) 相反侧的两种不同标记的探针。当被杂交时, 天然 ALK 区将呈现为橙色/绿色融合信号, 而在此基因座处的重排 (如在 EML4-ALK 缺失突变体中所发生的) 将导致分开的橙色和绿色信号。参见图 6。

FISH 分析揭示了在所研究样本总体中这种短变体 EML4-ALK 突变的相对低发生率 (229 个样品中有一个)。然而, 鉴于世界范围内 NSCLC 的高发生率 (仅在美国每年有 151,00 个以上的新病例), 预计有显著数量的躲藏这种突变体 ALK 的患者, 这些患者可以获利于 ALK-抑制治疗方案。

## 实施例 7

### 利用 PCR 测定来检测人实体瘤中的 ALK 融合蛋白表达

如先前描述的, 还可以利用基因组或反转录酶 (RT) 聚合酶链反应 (PCR) 来检测人实体瘤样品中一种或多种 ALK 融合蛋白的存在。参见, 例如, Cools *et al.*, *N. Engl. J. Med.* 348: 1201-1214 (2003)。简单和举例来说, 利用标准技术, 实体瘤样品可以获自患有例如 NSCLC 的患者。构建了相对于截短的 ALK 激酶或

EML4-ALK 融合蛋白（短或长变体）或 TFG-ALK 融合蛋白的 PCR 探针。RNeasy Mini Kit(Qiagen)可以用来从肿瘤样品提取 RNA。可以借助于 DNeasy Tissue Kit(Qiagen)来提取 DNA。关于 RT-PCR, cDNA 的第一条链合成自, 例如, 2.5  $\mu$ g 总 RNA, 其中使用, 例如, 具有寡(dT)<sub>20</sub> 的 SuperScript™ III 第一条链合成系统 (Invitrogen)。

然后, 通过使用引物对, 例如, EML4-202 和 ALK-GSP3 来扩增 ALK 融合基因 (参见上面的实施例 2)。对于基因组 PCR, 可以借助于高保真 Platinum Taq DNA 聚合酶 (Invitrogen) 并使用引物对, 例如, EML4-202 和 ALK-GSP3 进行融合基因 (参见上面的实施例 2) 的扩增。这样的分析将鉴定患有以截短的 ALK 激酶 (和/或 EML4-ALK 融合蛋白或 TFG-ALK 融合蛋白) 的表达为特征的实体瘤的患者, 该患者是利用 ALK-抑制治疗剂, 如 WHI-131 和/或 WHI154, 进行治疗的候选者。

在H2228细胞系和患者cs045中的EML4-ALK融合

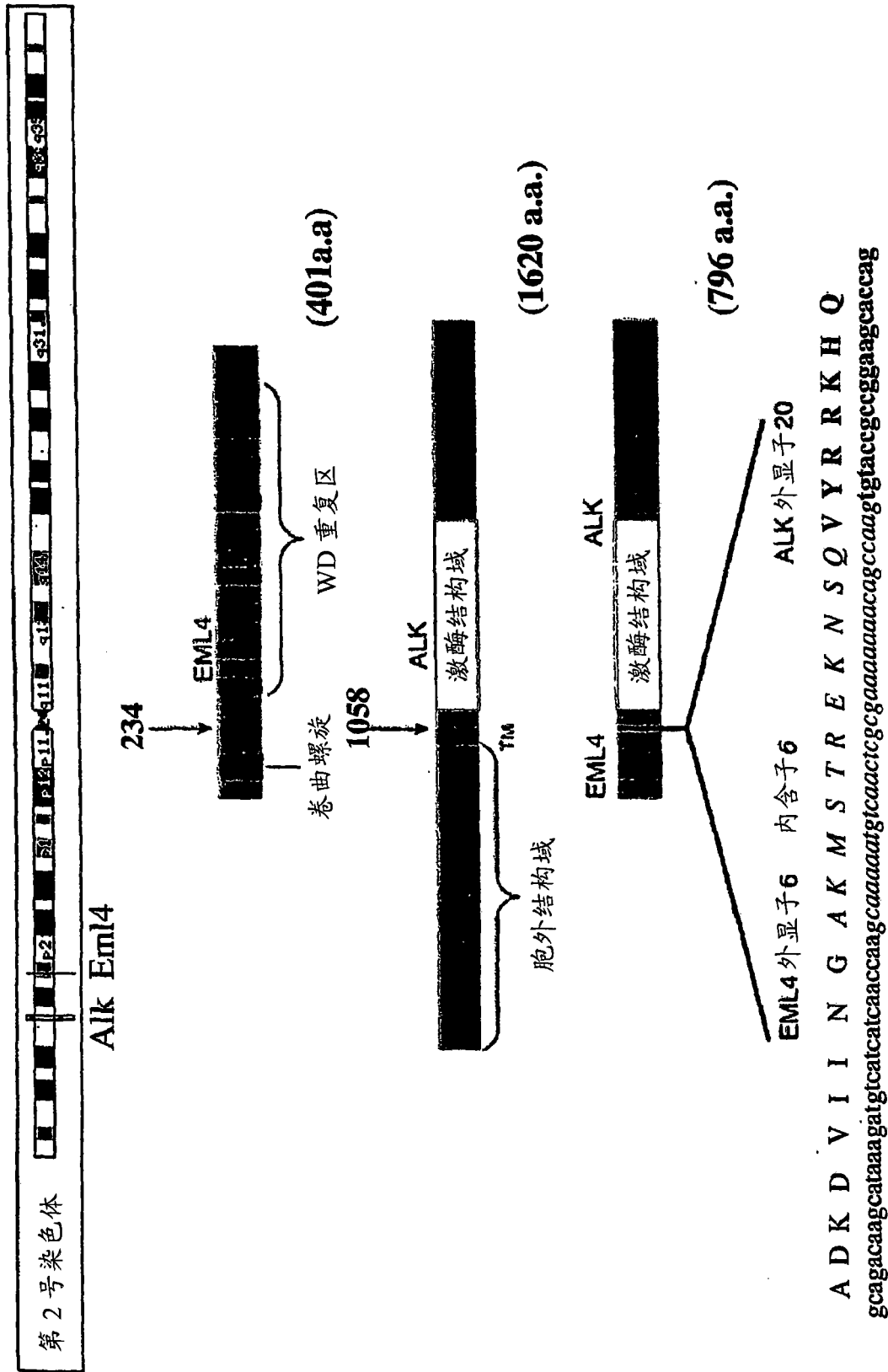
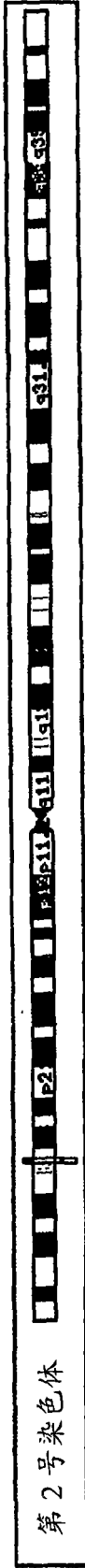


图 1A

在患者 CS010 中的 EML4-ALK 融合



EML4-ALK 融合

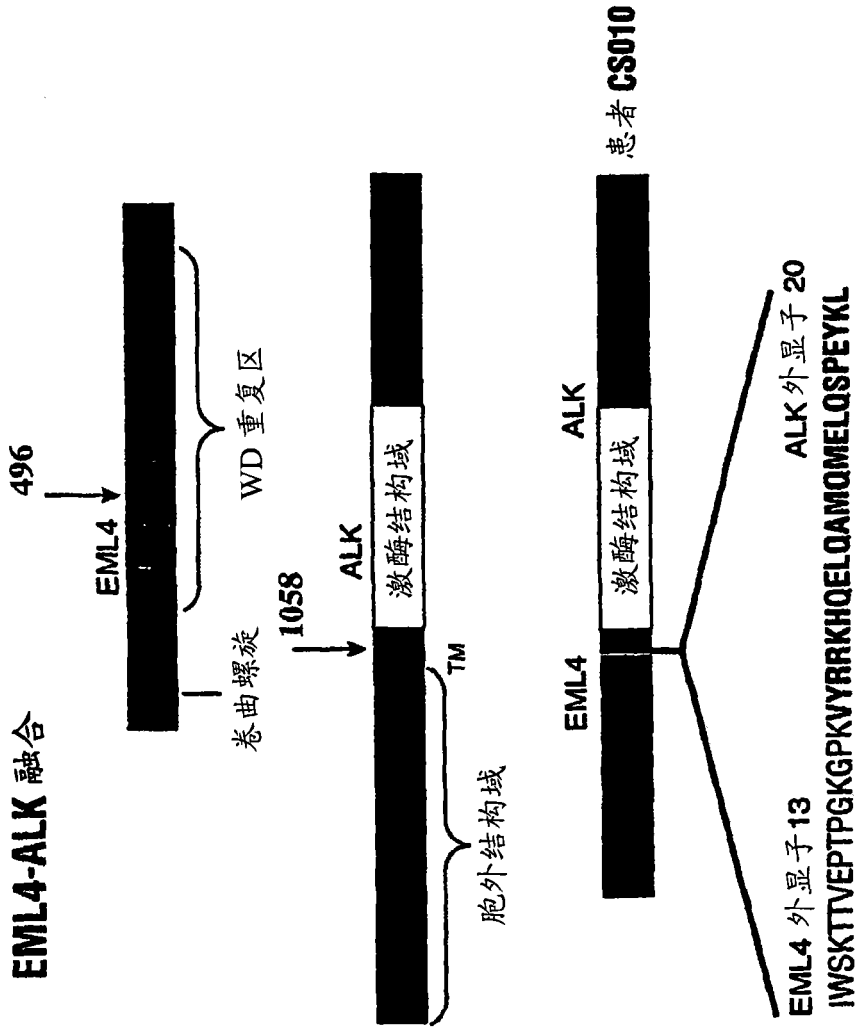


图 1B

在患者 CS110 中的 TFG-ALK 融合

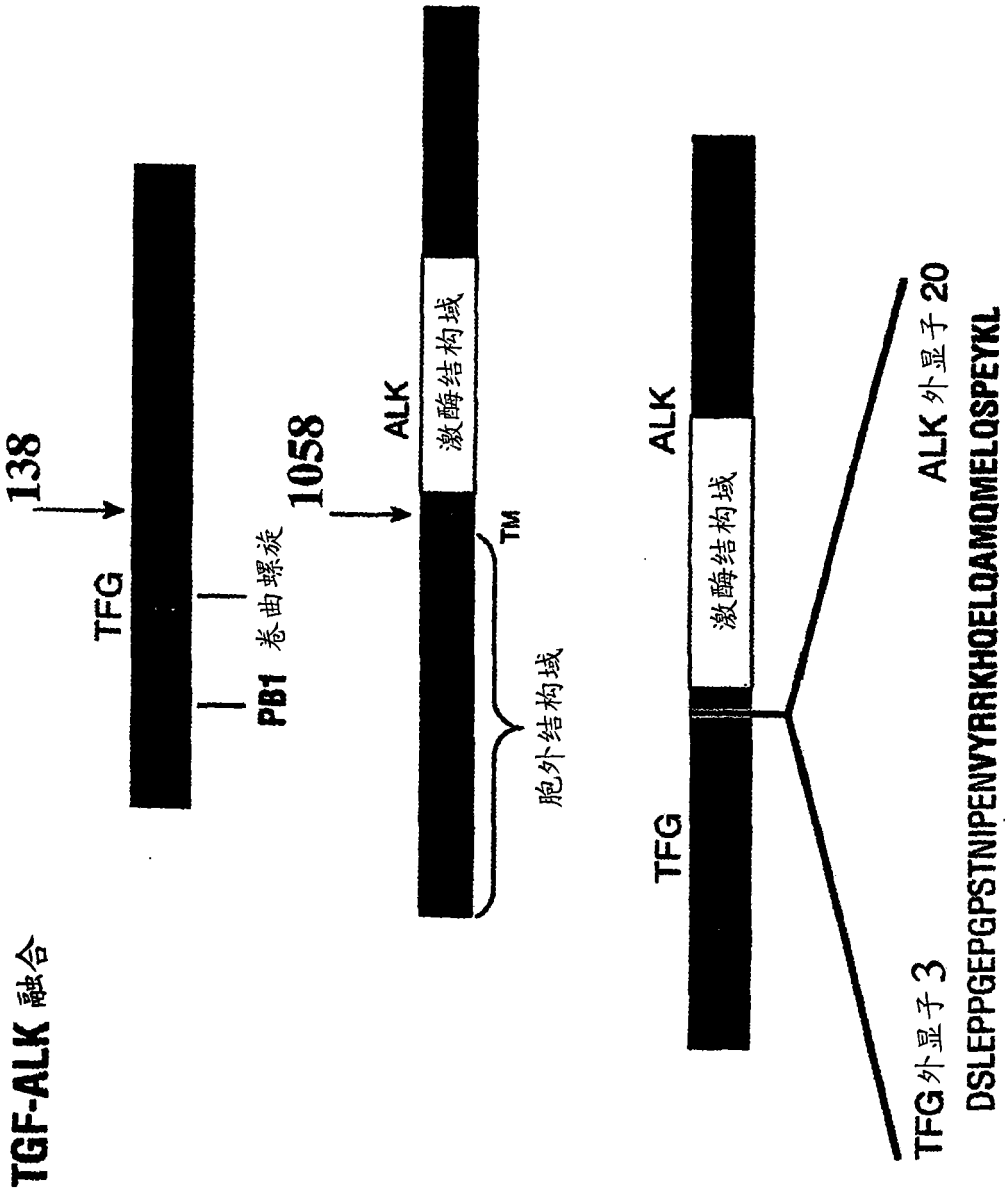


图 1C

MDGFAGSLDDSSISAASTSDVQDRLSALESRVQQQEDEITVLKAALADVLRRLAISEDHVASVKK  
SVSSKGQPSRAVIPMSCITNGSGANRKPSHTSAVSIAGKETLSSAAKSGTEKKKEKPQGQRE  
KKEESHNDQSPQIRASPPQSSQPLQIHRQTPEKSNATPTKSIKRPSAEKSHNSWENSDD  
SRNKLSKIPSTPKLIPKVTKTADKHKDVIINQAKMSTREKNSQVYRRKHQELQAMQMELOQSPEY  
KLSKLRTSTIMTDYNPNYCFAGKTSSISDLKEVPRKNITLIRGLGHGAFGEVYEGQVSGMPN  
DPSPQLQVAVKTLPEVCSEQDELDLFMEALIISKFNHQNIVRCIGVSLQSLPRFILLELM  
AGGDLKSFLRETRPRPSQPSSLAMLDLLHVARDIACGCQYLEENHFHRDIAARNCLL  
TCPGPGRVAKIGDFGMARDIYRASYYRKGGCAMLVVKWMPPEAFMEGIFTSKTDTW  
SFGVLLWEIFSLGYMPYPSKSNQEVLEFVTSGGRMDPPKNCPGPVYRIMTQCWQH  
PEDRPNFAILERIEYCTQDPDVINTALPIEYGPLVEEEEKVPVRPKDPEGVPLLVSQQAKRE  
EERSPAAPPPLTTSSGKAAKKPTAAEVSVRVPRGPAVEGGHVNMFAFSQSNPPSELHKVHGSR  
NKPTSLWNPTYGSWFTEKPTKKNPIAKKEPHDRGNLGLGSCVPPNVATGRLPGASLLEPS  
SLTANMKEVPLFRLRHFCGNVNYGYQQQLPLEAATAPGAGHIEDTILKSKNSMNQPGP

atggacggttccgagcagctctcgatgatagatattctgctgcaagtaacttctgatgttcaagatcgctgtcagctc  
ttgagtcacgagttcagcaacaagaagatgaaatcactgtgctaaaggcgttggctgatgtttgagg  
cgtcttgcaatctctgaagatcatgtgacctcagtgaaaaatcagctcaagtaaggccaaaccaagccctcgag  
cagttattcccatgtcctgtataaccaatggaagtgtgcaaacagaaaaaccaagtcatacagtgctgtctca  
atgacaggaagaaactcttctctgctgctaaaagtgtgacagaaaaaagaaagaaaaaccacaaggacag  
agagaaaaaaaagaggaatctcattctaattgatcaaatgcca caaattcgagcatca ccttctccccagccctct  
tca caa cctctccaaatacacagacaaactccagaaagcaagaatgctactccaccaaaaagcataaaaacgacct  
caccagctgaaaagtacataattcttgggaaaattcagatgatagccgtaataaattgtcgaaaataccttca  
acacccaaattaataccaaaagttacccaaaactgcagacaagcataaagatgtcatcatcaaccaag  
caaaaatgtcaactcgcaaaaaaacagccaag  
tgta cgc cgg aag cacc ag gact gca ag ccat g cag at gg ag ct g cag ag cc ct g ag ta ca ag ct g ag ca ag ct  
ccgacctcgaccatcatgaccgactaca ccccaacta ctgctt gctgg ca ag ac ct cctccatcagtgacc  
tgaaggaggtgcccggaaaaacatcacctcattcggggctctgggcatggcg ccttggggaggtgtatg  
aaggccaggtgtccggaatgcccaacgaccaagccccctgcaagtggctgtgaagacgctgctggaag  
tgtgctctgaacaggacgaactggattcctcatggaagccctgatcatcagcaaatcaaccaccagaac  
attgttgcgtgcattgggggtgagcctgcaatccctgccccggtcatcctgctggagctcatggcgggggg  
agacctcaagtcctcctccgagagaccgccccctgccccgagccagccctcctccctggccatgctggacc  
ttctgcacgtggctcgggacattgctgtggctgtcagtatttgaggaaaaccacttcatccaccgagac  
attgctgcagaaactgctcttgacctgtccaggccctggaagagtggccaagattggagacttcggga  
tgccccgagacatctacagggcgagctactatagaaggaggctgtgccatgctgccagttaagtgga  
tgccccagaggccttcatggaaggaatattcacttctaaaacagacacatggtccttggagtgctgctat  
gggaaatcttcttggatatatgccatacccagcaaaagcaaccaggaagttctggagttgtcacca  
gtggaggccggatggacccaccaagaactgccctgggctgtataccggataatgactcagtgctggc  
aacatcagcctgaagacaggcccaacttgccatcatttggagaggattgaatactgacccaggacc  
ggatgtaatacaaccgcttgcgatagaatatggcacttgggaagaggaagagaaagtgcctgtgaggcca  
aggacctgaggggtcctcctcctggtctctcaacaggcaaaacgggaggaggagcgagcccagctgccccac  
cacctctgctacca cctcctctggcaaggctgcaagaaaacca cagctgcagaggctctgttcgagtcctagaggg  
ccggccgtggaaggggacagtgaaatggcattctcagccaacctcctcgaggtgca caaggtcca cggat  
ccagaaacaagcccacagctgtggaaccaacgtacggctcctggttacagagaaaaccacaaaagaataatc  
ctatagcaagaaggagccaca cga caggggtaacctggggctggagggaagctgta ctgtcccactaacgttgc  
actgggagactccggggcctcactgctcctagagccctctcgtgctgccaataatgaaggaggta cctctgtcag  
gctacgtcactccttgggaatgtcaatta cggctaccagcaacagggttgccttagaagccgctactgcccctgg  
agctggtcatta c gaggata cca tctgaaaagcaagaatagcatgaaccagcctgggccccctga

图 2A

MDGFAGSLDDSSISAASTSDVQDRLSALESRVQQEDEITVLKAALADVLRRRLAISEDHVASVKKSVSQKQ  
 PSPRAVIPMSCITNGSGANRKPSTSAVSIAGKETLSSAAKSGTEKKKEKPOGGQREKKEESHNSNDQSPQIR  
 ASPSPQPSSQPLQIHRQTPESKNATPTKSIKRPSPAEKSHNSWENSDDSRNKLSKIPSTPKLIPKVTKTADK  
 HKDVIINQEGEYIKMFMGRPITMFIPSDVDNYDDIRTELPPEKLEWAYGYRGKDCRANYLLPTGEI  
 VYFIASVVLFNYEERTQRHYLGHTDCVKCLAIHPDKIRIATGQIAGVDKDGRLQPHVVRVWDSVTLSTL  
 QIIGLGTFFERGVGCLDFSKADSGVHLCIIDDNEHMLTVWDWQKKAKGAEIKTNEVVLAVEFHPTDAN  
 TIITCGKSHIFFWTWSGNSLTRKQGIFGKYEKPKFVQCLAFGLNGDVLTDGSDGGVMLIWSKTTVEPTPG  
 KGPKVYRRKHQELQAMQMELOQSPYKLSKLRSTIMTDYNPNYCFAGKTSSISDLKEVPRKNITLI  
 RGLGHGAFGEVYEGQVSGMPNDPSPLQVAVKTLPEVCSEQDELDLMEALIISKFNHQNIVR  
 CIGVSLQSLPRFILLELMAGGDLKSFLRETRRPSQPSSLAMLDLLHVARDIACGCQYLEENH  
 FIHRDIAARNCLLTCPPGPRVAKIGDFGMARDIYRASYRKGKGCAMLVVKWMPPEAFMEGI  
 FTSKTD'TWSFGVLLWEIFSLGYMPYPSKSNQEVLEFVTSGGMRMDPPKNCPPVYRIMTQCW  
 QHQPEDRPNFAILERIEYCTQDPDVINTALPIEYGPLVEEEEKVPVRPKDPEGVPPLLVSQOAKR  
 EERSPAAPPPLPTSSGKAAKKPTAAEISVRVPRGPAVEGGHVNMAFSQSNPPSELHKVHGSRNK  
 P'TSLWNPTYGSWFTEKPTKKNPIAKKEPHDRGNLGLGEGSCTVPPNVATGRLPGASLLEPSSLTA  
 NMKEVPLFRLRHFP CGNVNYGYQQQLPLEAATAPGAGHYEDTILKSKNSMNQPGP

ATGGACGGTTTCGCCGGCAGTCTCGATGATAGTATTTCTGCTGCAAGTACTTCTGATGTTCAAGATCG  
 CCTGTACGCTCTTGAGTCACGAGTTCAGCAACAAGAAGATGAAATCACTGTGCTAAAGGCGGCTTTG  
 GCTGATGTTTTGAGGCGTCTTGCAATCTCTGAAGATCATGTGGCCTCAGTGAAAAATCAGTCTCAAG  
 TAAAGGCCAACCAAGCCCTCGAGCAGTTATTCCCATGTCTGTATAACCAATGGAAGTGGTGCAAAC  
 AGAAAAACAAGTCATACCAGTGTCTCAATTGCAGGAAAAGAACTCTTTCATCTGCTGCTAAAAG  
 TGGTACAGAAAAAAGAAAGAAAAACCACAAGGACAGAGAGAAAAAAGAGGAATCTCATTCTAATG  
 ATCAAAGTCCACAAATTCGAGCATCACCTTCTCCCAGCCCTCTTCAACCTCTCCAAATACACAGA  
 CAAACTCCAGAAAGCAAGAATGCTACTCCCACAAAAGCATAAAACGACCATCACCAGCTGAAAAGT  
 CACATAATTCTTGGGAAAATTCAGATGATAGCCGTAATAAATTGTGCGAAAATACCTTCAACCCCAAAT  
 TAATACCAAAGTTACCAAACCTGCAGACAAGCATAAAGATGTCATCATCAACCAAGAAGGAGAATAT  
 ATTAATAATGTTTATGCGCGGTTCGGCCAATTACCATGTTCAATCCTTCCGATGTTGACAACATGATGAC  
 ATCAGAACCGAACTGCCTCCTGAGAAGCTCAAAGTGGGATGGGCATATGGTTATCGAGGAAAGGACT  
 GTAGAGCTAAGTGTACTTCTTCCGACCGGAAATAGTTTATTTCATTGCATCAGTAGTAGTACTAT  
 TTAATTATGAGGAGAGAACTCAGCGACACTAGCTGGGCCATACAGACTGTGTGAAATGCCTGTCTATA  
 CATCCTGACAAAATTAGGATTGCAACTGGACAGATAGCTGGCGTGGATAAAGATGGAAAGGCGCTCTAC  
 AACCCACGTCAGAGTGTGGGATTCTGTTACTCTATCCACACTGCAGATTATTGGACTTGGCACTTTT  
 GAGCGTGGAGTAGGATGCCTGGATTTTTCAAAGCAGATTCAGGTGTTCAATTATGATTATTGATGA  
 CTCCAATGAGCATATGCTTACTGTATGGGACTGGCAGAAGAAAAGCAAAGGAGCAGAAATAAAGACA  
 ACAAAATGAAGTTGTTTTGGCTGTGGAGTTTCACCCAACAGATGCAAATACCATAATTACATGCGGTAA  
 ATCTCATATTTTTCTTCTGGACCTGGAGCGGCAATTCACATAACAAGAAAACAGGGAATTTTTGGGAAAT  
 ATGAAAAGCCAAAATTTGTGCAGTGTAGCATTCTTGGGGAATGGAGATGTTCTTACTGGGAGACTCA  
 GGTGGAGTCACTGTTATATGGAGCAAACTACTGTAGAGCCACACCTGGGAAAGGACCTAAAAGTGT  
 ACCGCCGGAAGCACCAGGAGCTGCAAGCCATGCAGATGGAGCTGCAGAGCCCTGAGTACAAG  
 CTGAGCAAGCTCCGCACCTCGACCATCATGACCGACTACAACCCCAACTACTGCTTTGCTGGC  
 AAGACCTCCTCATCAGTGACCTGAAGGAGGTGCCGCGGAAAAACATCACCTCATTCCGGG  
 GTCTGGGCCATGGCGCCTTTGGGGAGGTGTATGAAGGCCAGGTGTCCGGAATGCCCAA  
 CGACCCAAGCCCCCTGCAAGTGGCTGTGAAGACGCTGCCTGAAGTGTGCTCTGAACAGG  
 ACGAACTGGATTCCTCATGGAAGCCCTGATCATCAGCAAATTCACCACCAGAACATTG  
 TTCGCTGCATTGGGGTGAACCTGCAATCCCTGCCCGGTTTCATCCTGCTGGAGCTCATG  
 GCGGGGGGAGACCTCAAGTCTTCTCCGAGAGACCCGCCCTCGCCCGAGCCAGCCCTC  
 CTCCCTGGCCATGCTGGACCTTCTGCACGTGGCTCGGGACATTGCCTGTGGCTGTCAGT  
 ATTTGGAGGAAAACCACTTCATCCACCGAGACATTGCTGCCAGAACTGCCTCTTGACCT

图 2B

GTCCAGGCCCTGGAAGAGTGGCCAAGATTGGAGACTTCGGGATGGCCCGAGACATCTAC  
AGGGCGAGCTACTATAGAAAGGGAGGCTGTGCCATGCTGCCAGTTAAGTGGATGCCCC  
AGAGGCCTTCATGGAAGGAATATTCACCTTCTAAAACAGACACATGGTCCTTTGGAGTGC  
TGCTATGGGAAATCTTTTCTCTTGGATATATGCCATACCCAGCAAAAGCAACCAGGAAG  
TTCTGGAGTTTGTCAACCAGTGGAGGCCGGATGGACCCACCCAAGAAGTGCCTGGGCCT  
GTATACCGGATAATGACTCAGTGTCTGGCAACATCAGCCTGAAGACAGGCCCAACTTTGC  
CATCATTTTGGAGAGGATTGAATACTGCACCCAGGACCCGGATGTAATCAACACCGCTTT  
GCCGATAGAATATGGTCCACTTGTGGAAGAGGAAGAGAAAGTGCCTGTGAGGCCCAAGGACC  
CTGAGGGGGTTCCCTCCTCCTGGTCTCTCAACAGGCAAAACGGGAGGAGGAGCGCAGCCCA  
GCTGCCCCACCACCTCTGCCTACCACCTCCTCTGGCAAGGCTGCAAAGAAACCCACAGCTGCA  
GAGATCTCTGTTTCGAGTCCCTAGAGGGCCGGCCGTGGAAGGGGGACACGTGAATATGGCATT  
CTCTCAGTCCAACCCTCCTTCGGAGTTGCACAAGGTCCACGGATCCAGAAACAAGCCCACCAG  
CTTGTGGAACCCAACGTACGGCTCCTGGTTTACAGAGAAACCCACCAAAAAGAATAATCCTAT  
AGCAAAGAAGGAGCCACACGACAGGGGTAACCTGGGGCTGGAGGGAAGCTGTACTGTCCCAC  
CTAACGTTGCAACTGGGAGACTTCCGGGGCCCTCACTGCTCCTAGAGCCCTCTTCGCTGACTG  
CCAATATGAAGGAGGTACCTCTGTTAGGCTACGTCACCTCCCTTGTGGGAATGTCAATTACG  
GCTACCAGCAACAGGGCTTGCCCTTAGAAGCCGCTACTGCCCTGGAGCTGGTCATTACGAGG  
ATACCATTCTGAAAAGCAAGAATAGCATGAACCAGCCTGGGCCCTGA

图 2B (续)

**MNGQLDLSGKLIKAQLGEDIRRIPIHNEDITYDELVLMMQRVFRGKLLSNDEVTIKYKDEDGDLI  
 TIFDSSDLSFAIQCSRLKLTFLVNGQPRPLESSQVKYLRRELIELRNKVNRLDSLEPPGEPGPSTN  
 IPENVYRRKHQELQAMQMELQSPEYKLSKLRSTIMTDYNPNYCFAGKTSSISDLKEVPRKNITLI  
 RGLGHGAFGEVYEGQVSGMPNDPSPLQAVKTLPEVCSEQDELDFLMEALIISKFNHQNIVR  
 CIGVSLQSLPRFILLELMAGGDLKSFLRETRPRPSQPSSLAMLDDLHVARDIACGCQYLEENH  
 FIHRDIAARNCLLTCPPGPRVAKIGDFGMARDIYRASYYRKGGCAMLVVKWMPPEAFMEGI  
 FTSKTDTSWFGVLLWEIFSLGYMPYPSKSNQEVLEFVTSGGRMDPPKNCPGPVYRIMTQCW  
 QHQPEDRPNFAILERIEYCTQDPDVINTALPIEYGPLVEEEEKVPVPRKDPEGVPPLLVSQAKR  
 EERSPAAPPPLPTSSGKAAKKPTAAEISVRVPRGPAVEGGHVNMAFSQSNPPSELHKVHGSRNK  
 PTSLWNPTYGSWFTEKPTKNNPIAKKEPHDRGNLGLGESCCTVPPNVATGRLPGASLLLEPSSLTA  
 NMKEVPLFRLRHFP CGNVNYGYQQQLPLEAATAPGAGHYEDTILKSKNSMNQPGP**

**ATGAACGGACAGTTGGATCTAAGTGGGAAGCTAATCATCAAAGCTCAACTGGGGAGGATATTC  
 GGCGAATTCCTATTATAATGAAGATATTACTTATGATGAATTAGTGCTAATGATGCAACGAGTT  
 TTCAGAGGAAAACCTTCTGAGTAATGATGAAGTAACAATAAAGTATAAAGATGAAGATGGAGATC  
 TTATAACAATTTTGGATAGTTCTGACCTTTCCTTTCGCAATTCAGTGCAGTAGGATACTGAAACTG  
 ACATTATTTGTTAATGGCCAGCCAAGACCCCTTGAATCAAGTCAGGTGAAATATCTCCGTCGAGA  
 ACTGATAGAACTTCGAAATAAAGTGAATCGTTTATTGGATAGCTTGGAAACCCTGGAGAACCA  
 GGACCTTCCACCAATATTCCTGAAAATGTGTACCGCCGGAAGCACCAGGAGCTGCAAGCCATG  
 CAGATGGAGCTGCAGAGCCCTGAGTACAAGCTGAGCAAGCTCCGCACCTCGACCATCATGAC  
 CGACTACAAACCCCACTACTGCTTTGCTGGCAAGACCTCCTCCATCAGTGACCTGAAGGAGGT  
 GCCGCGGAAAAACATCACCCCTCATTCCGGGGTCTGGGCCATGGCGCCTTTGGGGAGGTGT  
 ATGAAGGCCAGGTGTCCGGAATGCCCAACGACCCAAGCCCCCTGCAAGTGGCTGTGAAG  
 ACGCTGCCTGAAGTGTGCTCTGAACAGGACGAACTGGATTTCCTCATGGAAGCCCTGAT  
 CATCAGCAAATTCACCACCAGAACATTGTTTCGCTGCATTGGGGTGAGCCTGCAATCCC  
 TGCCCCGGTTCATCCTGCTGGAGCTCATGGCGGGGGGAGACCTCAAGTCCCTTCTCCGA  
 GAGACCCGCCCTCGCCCGAGCCAGCCCTCCTCCCTGGCCATGCTGGACCTTCTGCACGT  
 GGCTCGGGACATTGCCTGTGGCTGTGAGTATTTGGAGGAAAACCACTTCATCCACCGAG  
 ACATTGCTGCCAGAACTGCCTCTTGACCTGTCCAGGCCCTGGAAGAGTGGCCAAGATT  
 GGAGACTTCGGGATGGCCCGAGACATCTACAGGGCGAGCTACTATAGAAAGGGAGGCT  
 GTGCCATGCTGCCAGTTAAGTGGATGCCCCAGAGGCCCTTCATGGAAGGAATATTCCT  
 TCTAAAACAGACACATGGTCTTTGGAGTGTGCTATGGGAAATCTTTTCTCTTGGATAT  
 ATGCCATACCCAGCAAAAGCAACCAGGAAGTTCTGGAGTTTGTACCAGTGGAGGCGG  
 GATGGACCCACCAAGAACTGCCCTGGGCCTGTATACCGGATAATGACTCAGTGCTGGC  
 AACATCAGCCTGAAGACAGGCCCAACTTTGCCATCATTTTGGAGAGGATTGAATACTGC  
 ACCCAGGACCCGGATGTAATCAACACCGCTTTGCCGATAGAATATGGTCCACTTGTGGAAGA  
 GGAAGAGAAAGTGCCTGTGAGGCCCAAGGACCCCTGAGGGGGTTCCTCCTCCTGGTCTCTCA  
 ACAGGCAAAAACGGGAGGAGGAGCGCAGCCAGCTGCCCAACCACTCTGCCTACCACCTCCT  
 CTGGCAAGGCTGCAAAGAAACCCACAGCTGCAGAGATCTCTGTTTCGAGTCCCTAGAGGGCCG  
 GCCGTGGAAGGGGGACACGTGAATATGGCATTCTCTCAGTCCAACCTCCTTCGGAGTTGCAC  
 AAGGTCCACGGATCCAGAAAACAGCCACCGCTTGTGGAACCCAACGTACGGCTCCTGGTTT  
 ACAGAGAAAACCCACCAAAAAGAATAATCTATAGCAAAGAAGGAGCCACACGACAGGGGTA**

图 2C

ACCTGGGGCTGGAGGGAAGCTGTA  
CTGTCCCACCTAACGTTGCAACTGGGAGACTTCCGGGG  
GCCTCACTGCTCCTAGAGCCCTCTTCGCTGACTGCCAATATGAAGGAGGTACCTCTGTTCAGG  
CTACGTCACTTCCCTTGTGGGAATGTCAATTACGGCTACCAGCAACAGGGCTTGCCCTTAGAA  
GCCGCTACTGCCCCTGGAGCTGGTCATTACGAGGATACCATTCTGAAAAGCAAGAATAGCATG  
AACCAGCCTGGGCCCTGA

图 2C (续)

mdqfagslddsisaastsdvqdrlsaalesrvqqedeitvkaaladvlr  
rlaisedhvasvkksvsskqggspravipmscitngsqanrkpshtsavs  
iaqketlssaaksqtekkkekpggqrekkeeshsndgspqiraspsppqs  
sqplqihrtpesknatptksikrpspaekshnswensddsrnklskips  
tpkllpkvktadkhkdviinqegeyikmfmrgrpitmfi  
 psdvdnyddirtelppekiklewaygyrgkdcranvyl1ptgeivyfias  
 vvvlfnyeertqrhylvghtdcvkclaihpdkiriatgqiagvdkdgrplq  
 phvrwdsvtl1stlqiiglgtfergvglfdskadsgvhlcviddsnehm  
 ltvwdwqkkakgaeikttnevvlavefhptdantiitcgkshiffwtwsg  
 nsltrkqgfigkyekpkfvqclafngdvl1tgdsqgvmliwskttvept  
 pgkqpkgyqiskqikahdgsvftlcqmrngml1tgggkdrkiilwdhdl  
 npereievpdqygtiravaegkadqflvgtsrnfilrgt fndgfgievqg  
 htdelwglathpfdllltcaqdrqvclwnsmehrlwtr1vdepghcad  
 fhpsgtvvaigthsgrwfvldaetrdlvsihdgnelsvmrysidgtfl  
 avgshdnfiylyvvsengrkystrygrctghssyithldwspdnkyimsns  
 gdyeilwrdipngcklirnrscdkdidwttyt1cvlgfvfgvwpegsdgt  
 dinalvrshnrkviavaddfckvhlfgypcskakapshkysahsshvtnv  
 sfthndshlistggkmsi1qwk1vek1slpqnetvadtt1tkapvsste  
 sviqsn1tpppsqplnetaeesrissptllensleqtvepsedhsee  
 eseegsgdlgeplyeepcneiskeqakat1ledqqdpsps

图 3A

atggacggtttcggccgcaatctcgaatgataagatattctgctgcaagtaactctgagttcaagatcgccctg  
tcagctctttgagtcacgattcagcaacaagaagatgaatcactgctaaaggcggctttgctgat  
gtttgagcgctcttcaatctctgaaatcattgtggcctcagtgaaaaatcagctcgaagtaaggcc  
aaccaagccctcagcagttattccatgctctgataaccaatggaagtggtcaaacagaaaacca  
agtcatalaccagtgctgtctcaattcaggaaaaagaaacttctcatctgctgctaaaagtggtacagaa  
aaaaagaagaanaaaccaaggacagagagaaaaaaaggagaaatcattc1aatgatcaaa  
gtccacaaattcagcattcctcctccagccctctcacaacctctcaatacacagacaaaactc  
cagaaagcaagaatgctactcccacaaaagcataaaacgacctaccagctgaaaagtcacat  
aattctgggaaaattcagatgatagcgtaataaattgctgaaaataccttcaaac  
ccaaattaataccaaaagttaccaaaactgcagacaagcataaagatgcatcatcaaccaagaag  
gagaatataaaaatgtttatgctgctggcgaattaccatgctcattcctccgatgtgacaactatg  
atgacatcagaacggaaactgctcctgagaaagctcaaac1ggagtggtgcatatggtatcagagaaa  
ggactgtagagctaatgttaccctcctccgaccgggaaatagttattcattgcatcagtagtact  
atttaattatgaggagagaactcagcagactaccctggccatacagactgtgtgaaatgccttgctat  
acatcctgacaaaattaggattgcaactggacagatagctggcgtggataaagatggaaggcctcta  
caacccacgctcagagtggtggattctgtactctatccacactgcagattatgacttggcactttgag  
cgtaggagtaggatgcttgatttttcaaaaagcagattcaggtgtcattatgtgtattgatgactcfaat  
gagcatatgcttactgtatggactggcagaagaagcaaaaaggagcagaataaaagacaacaaa  
tgaagttgtttggctgtggagttcaccacaacagatgcaaataccataattacatcggtaaatctcata  
ttttcttgacctggagcggcaattcactaacaagaaaacagggaattttgggaaatatgaaaagc  
caaaaattgtgagtttagcattctgggaatggagatgttcttactggagactcaggtggagtcag  
cttatatgagcaaaactactgtagaccacacctggaaaggacctaaaggtgtatatcaaatcag  
caacaaatcaaaagctcatgatgagcagtggttcacacttttcagatgagaatgggatgttactgga  
ggagggaaagacagaaaaataattctgtggatcatgatctgaatcctgaaagagaatagaggttctga  
tcaglatgacacatcagagctgtagcagaaggaaaggcagatcaatttttagtaggcacatcacgaaact  
ttattttacaggaacattaatgatgcttccaataagaagtacagggcatacagatgagctttgggtctt  
gccacacatccttcaagattgtcttgcacatgtctcagacaggcaggtgtcctgtggaactcaatg  
gaacacaggctggaatgaccaggctgtagatgaaccaggacactgtcagattttcatccaagtggca

图 3B

cagtggtggccataggaacgcactcaggcaggtggttgttctggatgcagaaccaagatcagtttct  
 atccacacagacgggaatgaacagctctctgtgatgcgctactcaatagatgtaacctctggctgtagga  
 tctcatgacaacttatttacctctatgtagtctctgaaaatggaagaaaatagcagataggaagggtcac  
 tggacattccagctacatcacacacctgactggtccccagacaacaagtataatgtctaactcgggaga  
 ctatgaaatattgactggacattccaaatggctgcaactaatcaggaaatcagcggattgtaaggacatt  
 gattggacgacalalactgtgtcctaggattcaaglatitgggtctgcccagaaaggatctgatgggaca  
 gataatcaatgcactgggtgcgatcccacaatagaaaggatagctgttggcagatcttggtaaagccac  
 tgtttcagatccctgctccaaaggcaaggctcccagtcacaagtacagtgcccacagcagccatgtcacc  
 aatgtcagtttactcacaatgacagtcacctgatacactgggtgaaaagacatgagcatcattcagtgga  
 aactgtgaaaaatlatcttccctcagaatgagactgtgaggatactacttaaccaaaagccccgtctc  
 tccactgaaagtgcatccaatcctaactcccacaccgctctctcagccctaaatgagacagctgaa  
 gaggaagtagaataagcagttctccacacttctggagaacagcctggaacaaactgtgggccaagtga  
 agaccacagcggagggagagtgaaaggggcagcgggagaccctggtagcctcttatgaagaccat  
 gcaacgagataagcaaggagcagggcaaaagccaccttctggaggaccagcaagacccttcgacctc  
 tctaaccacctggctcagtgcaactcttctcagctgcatgtgatttggataaagttcaggtaaacagg  
 atgggacagtgatggagaatcactgttgattgagatttgggttccatgtgatttcttcaatagcttatttc  
 agtctctcaaalacagccaactaaagttagtttgggtttattgaaaattaacaaacttaatactaggaga  
 agactgaatcattaatgatgtctcaaaactgtgtacctaagggtgtgatgaaatactggaacaaaa  
 cagcagttgcatgatttggaaaacaaaccccttcttatctgaacatgtttctcaggaacaaccagaggta  
 tcacaaacactgtactcatctactggtcagactgtactacttttttttttttctgaaaaagaaaccagaa  
 aaaaatgtactctactgagataccctctcaccctaaatgtgtaatgaaaatitaaataagaaaaactcag  
 ttttggcaagtgaatgggtgttccctcttaaaaaatgcccgtttcttaactaccagtggtatgacagatg  
 ctcttagtctactagagaggtgtcgtcttcttaagtcataatgaggaaacagctccctaaatctgtgtgcaact  
 ctgtttatcttagaactaaagagcattgtgttitaagagcttcaatgtataltaaacctcaatactcag  
 aatgatggattcctcaaggagctcttactagcctaa  
 acattctcaaatgttggatcgaatgaaaggaaaccacatgctttaaactaaactgtaata  
 attacctggcctaattcagctaaagcctcatcataattgttccctcagtaataggagaaatataaatac  
 agtaagttagattatgaattgtgcttgaatattgtgtttgtgtaattatatacagattatagggataa  
 gatactcatcaaatgcaaatcttttttacagaagtggtggtaacagtcacagcagttttttaccacagc  
 atacttaacagacttctgtgtagcagtttttctgggtgggtgctgtaagcttgaagctaatgtgctat  
 cctactctttgggcaatgcatgtattatgattggaaggatlttttttaagttctgttggctagctatggttct  
 agtaccattcctactttaaagagtaattactgacaaatgtatttcttatgtttactttgattataaaaaagt  
 ttgtttgatttttaacttctgctgatttttgalacttctattttttggcaaatcatgttagaaacttggatga  
 gtaagaagctlaagatgcaggcgttacgtgattgtccattccaagtgatcagaactgtcattccctt  
 ctaatatcttccagggataatacaaaacaggtaattcatcatcattggtaatatgaaactccagtgactcc  
 caaggacattacaacattatattcacacgctgtatggaagggtgtgggtgtgtgaaaggggcagtggtg  
 agacactgtgtatcttagalaagaagatgcaccagttgaaaactcagtgtagatcctatgtgtat  
 aggtatctgtatattcttctttttacaactgttaaaaaacctcaaaatagttctctcaaaagagagagatt  
 ccaagcaaccatcttctcagtatgtatgtctgtacatactatcggagcgcgccagtaagatcaggcat  
 atatactgtctgttagcaatgattatcatcatcagatcagcatgtctatctccctgcaagaataactg  
 acatgaacaggcagttctggagaagaagagcatttcttaagtacctggggaalacagctcagtgat  
 cagcagggagllatttgaggacatcagtcaccttgggggtgccatgtacaatgagattataatcatgatac  
 tctcgggtgtagttcaaaagacactactaatacaggaagcgtccagclattaatgtggcaactactg  
 ttaatggcagttaaactgtgataatggttggaaagggtgggttatgaaattgtagatttttagaaaa  
 ctgtgaaatgaaaatgaatccaagtgttcatgtgaagatgtgagccattgctatcatgattcctgtctcatg  
 cgagaaaaattgaaagataaaaaataaaataacaaaatgttccctcttcaaaaaaaaaaaaaaaaaa

图 3B (续)

MGAIGLLWLLPLLLSTAAVGSGMGTGQRAGSPAAGSPLQPREPLSYSRLQ  
 RKSLAVDFVVPSSLFRVYARDLLLPSSSELKAGRPEARGLALDCAPLLR  
 LLGPAPGVSWTAGSPAPAEARTLSRVLKGGSVRKLRRAKQLVLELGEAI  
 LEGCVGPPGEAAVGLLQFNLSLSELSWIRQGEGRRLRIRLMPEKKASEVGR  
 EGRLSAAIRASQPRLLFQIFGTGHSLSLESPTNMPSPPDYFTWNLTWIMK  
 DSFPFLSHRSRYGLECSFDFPCELEYSFPLHDLRNQSWSWRRIPSEEASQ  
 MDLLDGPGAERSKEMPRGSFLLNLSADSKHTILSPWMRSSEHCTLAVS  
 VHRHLQPSGRYIAQLLPHNEAAREILLMPTPGKHGWTVLQGRIGRPNPF  
 RVALEYISSGNRSLSAVDFALKNCSEGTSPGSKMALQSSFTCWNGTVLQ  
 NFEDGFCGWTQGTLSPHTPQWQVRLKDFQDHQDHALLLSTTDVPASE  
 SATVTSATFPAPIKSSPCELMSWLIRGVLRGNVSLVLVENKTGKEQGRM  
 VWHVAAYEGLSLWQWMLPLLDVSDRFWLQMAVWGGQSRRAIVAFDNISI  
 SLDCYLITISGEDKILQNTAPKSRNLFERNPNKELKPGENSPROTPIFDPT  
 VHWLFTTCGASGPHGPTQAQCNNAYQNSNLSVEVGSEGPLKGIQIWKVPA  
 TDTYSISGYGAAGGKGGKNTMMRSHGVSVLGIFNLEKDDMLYILVGQQGE  
 DACPSTNQLIQKVCIGENNVIEEIRVNRSVHEWAGGGGGGGGATYVFKM  
 KDGVPVPLIIAAGGGGRAYAKTDTFHPERLENNSSVLGLNGNSGAAGGG  
 GGWNDNTSLLWAGKSLQEGATGGHSCPQAMKKWGWETRGGFGGGGGC  
 SSGGGGGYIGNAASNNDPEMDGEDGVSFISPLGILYTPALKVMEGHGEVN  
 IKHYLNCSHCEVDECHMDPESHKVICFDHGTVLAEDGVSCIVSPTPEPH  
 LPLSLILSVVTSALVAALVLAFIGIMIVYRRKHQELQAMQELQSPEYKL  
SKLRTSTIMTDYNPNYCFAGKTSSISDLKEVPRKNITLIRGLGHGAFGEV  
YEQVSGMPNDPSPLQVAVKTLPEVCSEQDELDFLMEALIISKFNHQNIV  
RCIGVLSQSLPRFILLELMAGGDLKSFLRETRPRPSQPSLAMLDDLHVA  
RDIACGCQYLEENHFIRHDIARNCLLTCGPGRVAKIGDFGMARDIYRA  
SYRKGGCAMLVVKWMPPEAFMEGIFTSKTDTFWSFGVLLWEIFSLGYMPY  
PSKSNQEVLEFVTSGGRRMDPPKNCPGPVYRIMTQCWQHQPEDRPNFAILL  
ERIEYCTQDPDVINTALPIEYGPLVEEEKVPVRPKDPEGVPPLLVSQQA  
KREEERSPAAPPPLPTTSSGKAACKPTAAEVSVRVPRGPAVEGGHVNMAF  
SQSNPPSELHKVHGRNKPTSLWNPTYGSWFTEKPTKKNNP IAKKEPHDR  
GNLGLEGSCVPPNVATGRLP GASLLEPSSLTANMKEVPLFRLRHFFCG  
NVNYGYQQOGLPLEAATAPGAGHYEDTILKSKNSMNQPGP

图 4A

1 agagacttgc ggcacgcac agtccctctg agatcaggtg gaaggagccg ctgggtacca  
 61 aggactgttc agagcctcct cccatctcgg ggagagcgaa ggggtaggtc gggcccggag  
 121 agcagtgtaa acggcctcct ccggcgggat gggagccatc gggctcctgt ggctgctgcc  
 181 gctgctgctt tccacggcag ctgtgggctc cgggatgggg accggccagc gcgcgggctc  
 241 cccagctgcg gggccgcccg tgcagccccg ggagccactc agctactcgc gcctgcagag  
 301 gaagagtctg gcagttgact tctgtgtgcc ctgctcttc cgtgtctacg cccgggacct  
 361 actgtgccca ccatcctcct cggagctgaa ggctggcagg cccgaggccc gcggctcgtc  
 421 agctctggac tgcgccccg tgcctcaggtt gctggggccc gcgcccgggg tctcctggac  
 481 cgccggttca ccagccccg cagaggcccg gacgtgtcc aggggtgctga agggcggctc  
 541 cgtgcgaag ctccggcgtg ccaagcagtt ggtgtcggag ctgggcaggt aggcgatctt  
 601 ggagggttgc gtcgggcccc ccggggaggc ggctgtgggg ctgctccagt tcaatctcag  
 661 cgagctgttc agttgggtga ttcgccaagg cgaaggcga ctgaggatcc gcctgatgcc  
 721 cgagaagaag gcgtcggag tggcagaga ggaaggctg tccgcgcaa ttcgcccctc  
 781 ccagccccgc cttctcttcc agatctcgg gactggtcat agctccttgg aatccaac  
 841 aaacatgcct tctccttctc ctgattattt tacatggaat ctacactgga taatgaaga  
 901 ctctctcct tctctgtctc atcgcagcgg atatggtctg gactgcagct ttgactccc  
 961 ctgtgagctg gagtattccc ctccactgca tgacctcagg aaccagagct ggtcctggcg  
 1021 ccgcatcccc tccgaggagg cctcccagat ggacttctg gatgggctg gggcagagcg  
 1081 ttctaaggag atgccagag gctcctttct cttctcaac acctcagctg actccaagca  
 1141 caccatcctg agtccgtgga ttaggagcag cagtgcacac tgcacactgg ccgtctcggg  
 1201 gcacaggcac ctgcagccct ctggaaggtg cattgccag ctgctgccc acaacaggc  
 1261 tgcaagagag atcctcctga tgcccactcc aggaagcat ggttgacag tgctccagg  
 1321 aagaatcggg cgtccagaca acccatttgg agtgcccctg gaatacatct ccagtggaaa  
 1381 ccgagcttg tctgcagtg acttcttgc cctgaagaac tgcagtgaag gaacatccc  
 1441 aggctccaag atggcctgc agagctcctt cacttgttgg aatgggacag tctcctcagct

图 4B

1501 tgggcaggcc tgtgacttcc accaggactg tgcccaggga gaagatgaga gccagatgtg  
1561 ccggaaactg cctgtgggtt tttactgcaa ctttgaagat ggcttctgtg gctggacca  
1621 aggcacactg tcacccca caacctcagt gcaggtcagg accctaaagg atgcccgggt  
1681 ccaggaccac caagaccatg ctctattgct cagtaccact gatgtccccg cttctgaaag  
1741 tgctacagtg accagtgcta cgtttcctgc accgatcaag agctctocat gtgagctccg  
1801 aatgtcctgg ctcatctgtg gagtcttgag gggaaacgtg tccttgggtg tagtgagaa  
1861 caaaaccggg aaggagcaag gcaggatggt ctggcatgtc gccgctatg aaggcttgag  
1921 cctgtggcag tggatggtgt tgccctcctc cgatgtgtct gacaggttct ggctgcagat  
1981 ggtcgcagtg tggggacaag gatccagagc catcgtggct tttgacaata tctccatcag  
2041 cctggactgc tactcacca ttacgaggaga ggacaagatc ctgcagaata cagcaccaca  
2101 atcaagaaac ctgtttgaga gaaacccaaa caaggagctg aaaccggggg aaaattcacc  
2161 aagacagacc cccatctttg accctacagt tcattggctg ttcaccacat gtggggccag  
2221 cgggccccat ggccccacc aggcacagt caacaacgcc taccagaact ccgacctgag  
2281 cgtggaggty gggagcagg gcccctgaa aggcattccag atctggaagg tgccagccac  
2341 cgacacctac agcatctcgg gctacggagc tgctggcggg aaaggcggga agaaccacat  
2401 gatgcggctc cacgycgtgt ctgtgctggg catcttcaac ctggagaagg atgacatgct  
2461 gtacatcctg gttggcagc agggagagga cgctgcccc agtacaacc agttaatcca  
2521 gaaagtctgc attggagaga acaatgtgat agaagaaga atccgtgtga acagaagcgt  
2581 ccatgagtg gcaggaggc gaggaggag ggtggagcc acctacgtat ttaagatgaa  
2641 ggatggagtg ccggtgcccc tgatcattgc agccggagg ggtggcagg cctacggggc  
2701 caagacagac acgttccacc cagagagact ggagaataac tcctcggttc tagggctaaa  
2761 cggcaattcc ggagccgag gtggtggagg tggctggaat gataacactt cctgtctctg  
2821 ggccggaaaa tctttgcagg aggtgccc cggagggacat tcctgcccc aggcatgaa  
2881 gaagtggggg tgggagaca gagggggtt cggaggggtt ggagggggc gctcctcagg  
2941 tggaggaggc ggagatata taggcggcaa tgcagcctca aacaatgact ccgaaatgga  
3001 tggggaagat ggggtttcct tcatcagtc actgggcata ctgtacacc cagctttaa  
3061 agtgatggaa ggccacgggg aagtgaatat taagcattat ctaaaactga gtcactgtga  
3121 gtagacgaa tgtcacatg accctgaaag ccacaaggtc atctgcttct gtgaccacgg  
3181 gacggtgctg gctgaggat gcgtctcctg cattgtgtca cccacccc agccacacct  
3241 gccactctg ctgatcctct ctgtggtgac ctctgccc gtggcggcc ttgtcctggc  
3301 tttctccggc atcatgattg tgtaccggcg gaagcaccag gagctgcaag ccatctgag  
3361 ggagctgag agccctgagt acaagctgag caagctccgc acctgacca tcatgaccga  
3421 ctacaacccc aactactgct ttgtgqcaa gacctcctcc atcagtgacc tgaaggagg  
3481 gccgqgaaa aacatcacc tcattcgggg tctggqccat ggcgctttg gggaggtgta  
3541 tgaaggccag gtgtccqaa tgcccacga cccaagccc ctgcaagtgg ctgtgaaagc  
3601 gcatctgaa gtgtgctctg aacaggacga actggattc ctcatggaag cctgtatcat  
3661 cagcaaatc aaccaccaga acattgttcg ctgcatggg gtgagctgc aatccctgcc  
3721 ccggttcata ctgctggagc tcatggcggg gggagacctc aagtecttc tccgagagc  
3781 ccgctctgc ccgagccagc cctcctcctt ggcctgctg gacctctgc acgtgctgc  
3841 ggacattgcc tgtgctctg agtatttggg ggaanaaccac ttcattcacc gacacattgc  
3901 tggcagaaac tgectcttga cctgtccagg ccctggaaga gtggccaaga ttggagactt  
3961 cggctggcc ccgagacatc acagggcag ctactataga aaggagactt gtcgactgc  
4021 gccagttaa tggatgccc cagaggctt catggaagg atattcctt ctanaacaga  
4081 cacatggtcc tttgagtg tctatggga aatctttct cttggtata tgccatacc  
4141 cagcaaaagc aaccaggag ttctggagt tgcaccagt ggagggcggg tggaccacc  
4201 caagaactgc cctggcctg tataccggat aatgactcag tgctggcaac atcagcctga  
4261 agacaggccc aactttgcca tcattttggg gaggattgaa tactgaccc aggaccggg  
4321 tghtaatcaac accgcttgc cgtagaata tggctccact gtggaagagc aagagaaat  
4381 ccctgtgagg cccaaggacc ctgaggggt tcctcctctc ctggctctc aacaggcaaa  
4441 acgggagggg gaggcagcc cagctgccc accacctctg cctaccacct cctctggcaa  
4501 ggctgcaag aaaccacag ctgcagagg ctctgttca gtccctagag ggcggccgt  
4561 ggaaggggga cagtgataa tggcattctc tcagtcaca cctcctcgg agttgcacag  
4621 ggtccacgga tccagaaata agcccaccag cttgtggaac ccaacgtac gctcctggt  
4681 tacagagaaa cccacaaa agaataatcc tatagcaaa agggagccac acgagaggg  
4741 taacctggg ctggaggga gctgtactgt cccacctaac gttgcaactg qgagactcc  
4801 gggggcctca ctgctcctag agccctctc gctgactgc aatatgaaq aggtacctc  
4861 gttcaggcta cgtcactcc cttgtggaa tgtcaattac ggctaccagc aacaggcctt  
4921 gcccttagaa gccgctactg cccctggagc tggctattac gaggatacca ttctgaaaq  
4981 caaqaatagc atgaaccagc ctggggcctg agctcggctg cacacttct tctctctt  
5041 aggatcccta agaccgtga gtagagagag gcaatcaat gctcctca caaacagag  
5101 accaaatgct acgttttgtt ttgtgcaac ctattttgaa gtaccacca aaaagctgta  
5161 tttgaaaat gctttgaaa ggttttgagc atgggttcat cctattctt cgaagaaga  
5221 aatatcata aaaatgagt ataaatacaa ggcagatgt ggttgcataa ggtttttatg  
5281 catgtttgtt gta

图 4B (续)

MNGQLDLSGKLIVKAQLGEDIRRIPIHNEDITYDELVLMMQRVFRGKLLS  
NDEVTIKYKDEDGDLITIFDSSDLSFAIQCSRILKLTFLVNGQPRPLESS  
QVKYLRRELIELRNKVNRLLDLSLEPPGEPGPSTNIPENDTVDGREEKSAS  
DSSGKQSTQVMAASMSAFDPLKNQDEINKNVMSAFGLTDDQVSGPPSAPA  
EDRSGTPDSIASSSSAAHPPGVQPQPPYTGAQTQAGQIEGQMYQQYQQQ  
AGYGAQQPQAPPQQPQQYGIQYSASYSQQTGPQQPQQFQGYGQQPTSQAP  
APAFSGQPQQLPAQPPQQYQASNYPAQTYTAQTSQPTNYTVAPASQPGMA  
PSQPGAYQPRPGFTSLPGSTMTPPPSGPNPYARNRPPFGQGYTQPQPGYR

图 4C

ATGAACGGACAGTTGGATCTAAGTGGGAAGCTAATCATCAAAGCTCAACTTGGGG  
AGGATATTCGGCGAATTCCTATTCATAATGAAGATATTACTTATGATGAATTAGT  
GCTAATGATGCAACGAGTTTTCAGAGGAAAACCTTCTGAGTAATGATGAAGTAACA  
ATAAAGTATAAAGATGAAGATGGAGATCTTATAACAATTTTTGATAGTTCTGACC  
TTTCCTTTGCAATTCAGTGCAGTAGGATACTGAAACTGACATTATTTGTTAATGG  
CCAGCCAAGACCCCTTGAATCAAGTCAGGTGAAATATCTCCGTCGAGAACTGATA  
GAACTTCGAAATAAAGTGAATCGTTTATTTGGATAGCTTGGAAACCACCTGGAGAAC  
CAGGACCTTCCACCAATATTCCTGAAAATGATACTGTGGATGGTAGGGAAGAAAA  
GTCTGCTTCTGATTCTTCTGGAAAACAGTCTACTCAGGTTATGGCAGCAAGTATG  
TCTGCTTTTGATCCTTTAAAAAACCAAGATGAAATCAATAAAAATGTTATGTCAG  
CGTTTGCTTAACAGATGATCAGGTTTCAGGGCCACCCAGTGCTCCTGCAGAAGA  
TCGTTTCAGGAACACCCGACAGCATTGCTTCCTCCTCCTCAGCAGCTCACCCACCA  
GGCGTTCAGCCACAGCAGCCACCATATACAGGAGCTCAGACTCAAGCAGGTCAGA  
TTGAAGGTCAGATGTACCAACAGTACCAGCAACAGGCCGGCTATGGTGCACAGCA  
GCCGCAGGCTCCACCTCAGCAGCCTCAACAGTATGGTATTCAGTATTCAGCAAGC  
TATAGTCAGCAGACTGGACCTCAACAACCTCAGCAGTTCCAGGGATATGGCCAGC  
AACCAACTTCCAGGCACCAGCTCCTGCCTTTTCTGGTCAGCCTCAACAACCTGCC

图 4D

TGCTCAGCCGCCACAGCAGTACCAGGCGAGCAATTATCCTGCACAAACTTACACT  
GCCCAAACCTTCTCAGCCTACTAATTATACTGTGGCTCCTGCCTCTCAACCTGGAA  
TGGCTCCAAGCCAACCTGGGGCCTATCAACCAAGACCAGGTTTTACTTCACTTCC  
TGGAAGTACCATGACCCCTCCTCCAAGTGGGCCTAATCCTTATGCGCGTAACCGT  
CCTCCCTTTGGTCAGGGCTATACCCAACCTGGACCTGGTTATCGATAAGGAGGCT  
CCTCTACACCAATTAATGTAGCTGCTAGCTATTGGCCTCCCAAAGACTCCAGTA  
CTATTTAATTTGTATTGAAGAAGTTCAGAAATTTAAAAGCAGAGCATTTTTTAT  
GATATCATTGTTGGTGTTAATTGAAAGTATAATTTGCTGGAACACAAAGACCAA  
ATGAAAGTTTTTTCCTCCCTGCTTAAAATGTAGCAGCTTCTTAGTTACTTTGGA  
ACACTACTCTTACATGTATAAAGTGATTGACTTGACTTTCTAGCTTCCCTTGTC  
GGAGGATATTTAAAATGCTAGGGTGAGGTTTAGCCATCTTACTTGGCTTTTACTAT  
TAACATGATGTACTAAAGTAGAGCCCTTTGAGAATACAAGATATTATGTATAAAA  
TGTAACACTGATGATAGGTTAATAAAGATGATTGAATCCATTAAAAA

图 4D (续)

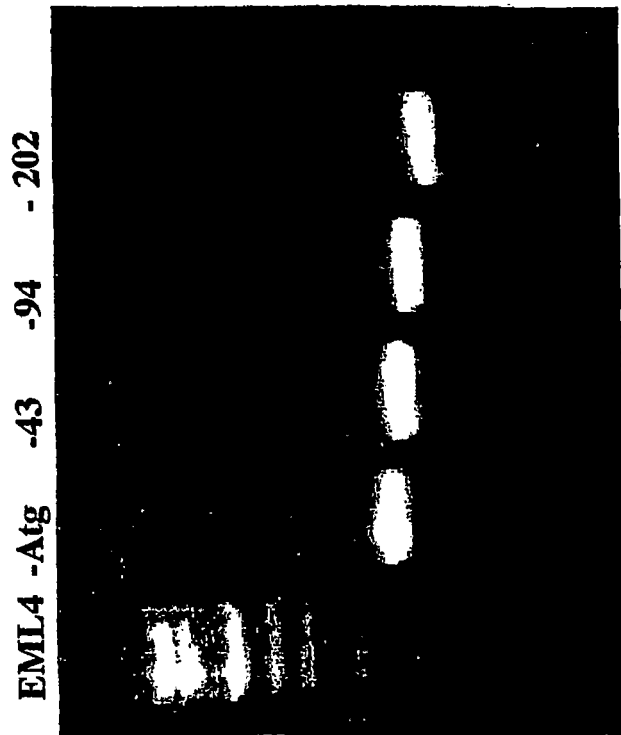


图 5B

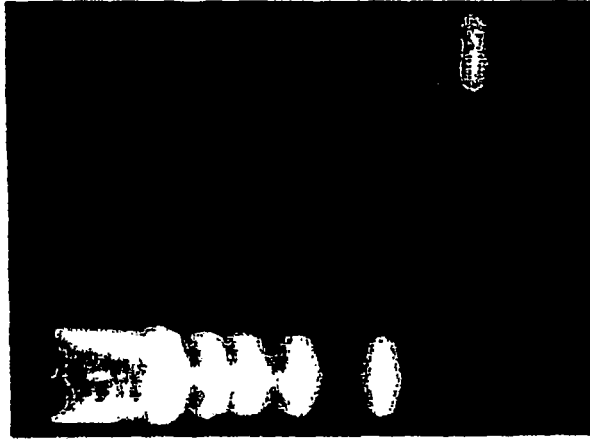


图 5A

← EML4-ALK →

在 CS110 患者中  
的 TFG-ALK 融合

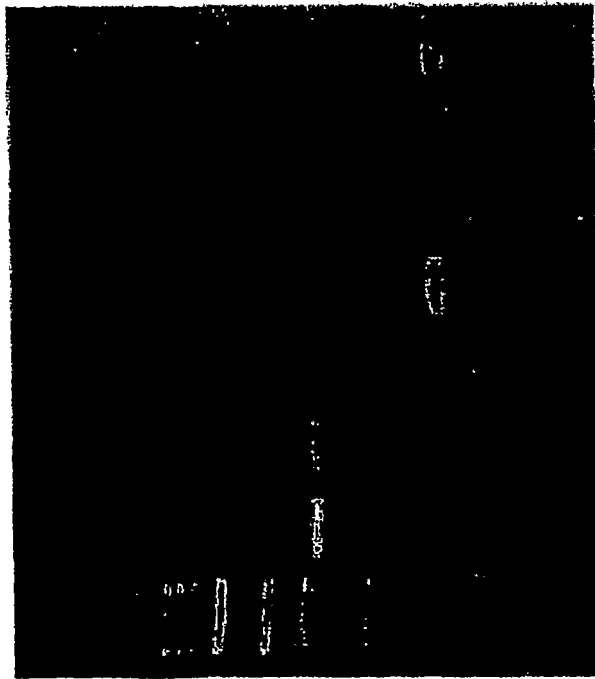
CS010  
CS015\*  
CS110\*



TFG-ALK\*

在 H2228 细胞系和患者  
cs010/011, cs045 中的 EML4-ALK 融合

CS010\*\*  
CS011\*\*  
CS015\*\*  
CS045\*\*  
CS042  
LY18  
H2228\*\*



EML4-ALK\*\*

图 5D

图 5C

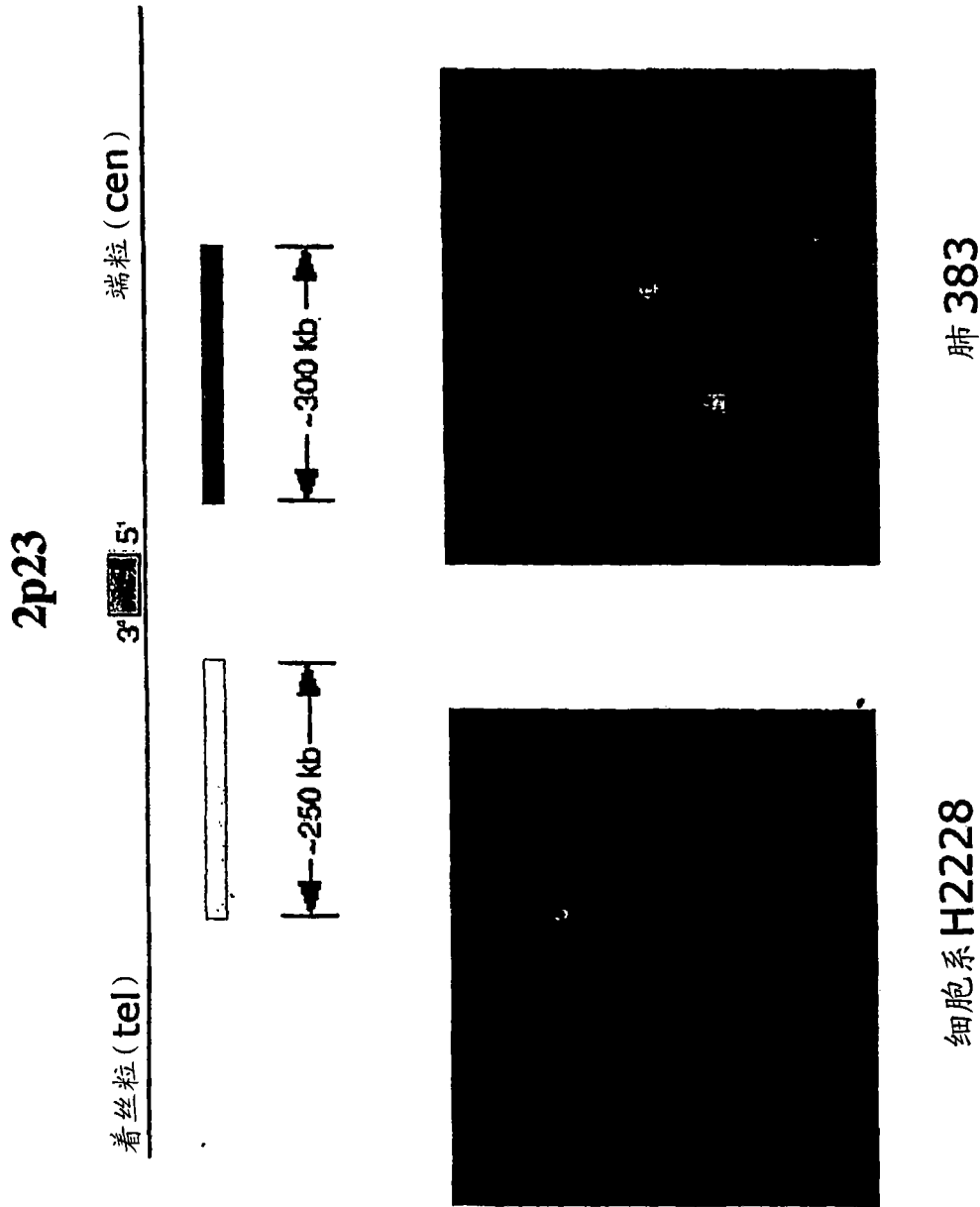


图 6

专利名称(译)	在人实体瘤中的基因缺损和突变体ALK激酶		
公开(公告)号	<a href="#">CN101466721A</a>	公开(公告)日	2009-06-24
申请号	CN200780021431.4	申请日	2007-04-13
[标]申请(专利权)人(译)	细胞信号技术有限公司		
[标]发明人	克拉里萨里科瓦 赫伯特哈克 劳拉沙利文 顾挺磊 安东尼波塞马托 郭爱兰 琼麦克尼尔 俞建		
发明人	克拉里萨·里科瓦 赫伯特·哈克 劳拉·沙利文 顾挺磊 安东尼·波塞马托 郭爱兰 琼·麦克尼尔 俞建		
IPC分类号	C07H21/02 C07K1/00 A61K38/00 C07K16/00 G01N33/53 C12Q1/68		
CPC分类号	C12Q2600/178 G01N33/57423 C12Q2600/156 C07K14/82 C12N9/1205 C07K2319/00 C12Q1/6886 A61K38/00 C12Q2600/136 A61P35/00 C07K19/00 C12N15/62 C12Q1/6841		
代理人(译)	张英		
其他公开文献	CN101466721B		
外部链接	<a href="#">Espacenet</a> <a href="#">SIPO</a>		

摘要(译)

根据本发明，现已在人实体瘤，例如非小细胞肺癌(NSCLC)中鉴定了新的涉及第2号染色体的基因缺失和易位，其导致结合部分间变性淋巴瘤激酶(ALK)激酶与部分二级蛋白质的融合蛋白。二级蛋白质包括棘皮动物微管结合蛋白样4(EML - 4)和TRK - 融合基因(TFG)。保留ALK酪氨酸激酶活性的EML4 - ALK融合蛋白被证实可以驱动以这种突变为特征的NSCLC的增殖和存活。因此，本发明部分地提供了分离的多核苷酸和载体，其编码所披露的突变体ALK激酶多肽，用于检测它的探针，分离的突变体多肽，重组多肽，以及用于检测融合和截短的多肽的试剂。所披露的这种新融合蛋白的鉴定使得可以用新方法来确定在生物样品中这些突变体ALK激酶多肽的存在，本发明还提供了对抑制蛋白质的化合物进行筛选的方法，以及对以突变体多核苷酸或多肽为特征的癌症的进展进行抑制的方法。

