



(12) 发明专利

(10) 授权公告号 CN 113436741 B

(45) 授权公告日 2023.02.28

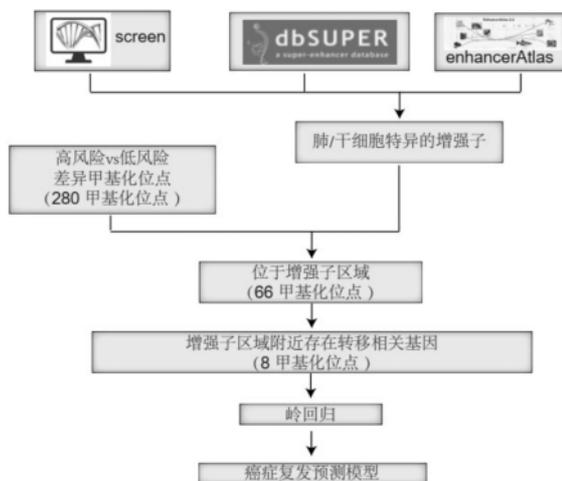
(21) 申请号 202110806001.5	CN 111916154 A, 2020.11.10
(22) 申请日 2021.07.16	CN 109385478 A, 2019.02.26
(65) 同一申请的已公布的文献号 申请公布号 CN 113436741 A	CN 110004226 A, 2019.07.12
(43) 申请公布日 2021.09.24	CN 104046624 A, 2014.09.17
(73) 专利权人 四川大学华西医院 地址 610041 四川省成都市武侯区国学巷 37号	CN 105087758 A, 2015.11.25
(72) 发明人 刘伦旭 邓雨岚 邓森议 陈楠 苏雨桃 夏梁	CN 110807785 A, 2020.02.18
(74) 专利代理机构 成都虹桥专利事务所(普通 合伙) 51124 专利代理师 陈立志	CN 106480210 A, 2017.03.08
(51) Int. Cl. G16H 50/30 (2018.01) G06Q 10/04 (2023.01) C12Q 1/6886 (2018.01)	CN 108410980 A, 2018.08.17
(56) 对比文件 CN 111564177 A, 2020.08.21	CN 108070655 A, 2018.05.25
	CN 108977539 A, 2018.12.11
	CN 110157804 A, 2019.08.23
	CN 105861644 A, 2016.08.17
	CN 108315418 A, 2018.07.24
	CN 105319364 A, 2016.02.10
	CN 108913776 A, 2018.11.30
	CN 104899474 A, 2015.09.09
	CN 105825077 A, 2016.08.03
	CN 109661477 A, 2019.04.19
	CN 107447008 A, 2017.12.08
	(续)
	审查员 李晓晖
	权利要求书2页 说明书7页 附图4页

(54) 发明名称
基于组织特异增强子区域DNA甲基化的肺癌
复发预测方法

(57) 摘要

本发明涉及生物医学领域,公开了一种基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,用以解决目前非小细胞肺癌复发模型对非小细胞肺癌复发预测不合理的问题。本发明首先提取肺癌患者的DNA,并进行DNA甲基化检测,得到组织特异增强子区域的DNA甲基化位点的beta值;其中,DNA甲基化位点包括chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419;然后将DNA甲基化位点的beta值代入预先构建的肺癌复发模型,计算出模型得分,并根据模型得分以及多个得分阈值得到预测结果。本发明适用于早期非小

细胞肺癌的复发预测。



CN 113436741 B

[接上页]

(56) 对比文件

CN 109462996 A,2019.03.12	CN 107058523 A,2017.08.18
CN 106544428 A,2017.03.29	CN 109988708 A,2019.07.09
CN 108939048 A,2018.12.07	US 2012185174 A1,2012.07.19
CN 110191725 A,2019.08.30	US 2016078167 A1,2016.03.17
CN 101880715 A,2010.11.10	EP 3260552 A1,2017.12.27
	US 2018355408 A1,2018.12.13

1. 基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,其特征在于,包括如下步骤:

步骤1:提取肺癌患者的DNA,并进行DNA甲基化检测,得到组织特异增强子区域的DNA甲基化位点的beta值;其中,DNA甲基化位点包括chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419;

步骤2:将以上DNA甲基化位点的beta值代入预先构建的肺癌复发模型,计算出模型得分,并根据模型得分以及多个得分阈值得到预测结果;

所述肺癌复发模型的构建步骤包括:

S1:收集早期的非小细胞肺癌病例,包括临床信息匹配的复发与未复发患者,提取其DNA样本,并进行DNA甲基化检测;

S2:将样本的甲基化位点划分为高风险组和低风险组;

S3:结合多个增强子数据库,确定肺组织特异或干细胞特异的增强子区域,并从样本中筛选出位于增强子区域的甲基化位点;

S4:在步骤S3的基础上,从增强子100kb范围内进一步筛选出癌症相关基因,最终得到组织特异增强子区域的chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419八个DNA甲基化位点;

S5:利用岭回归训练组织特异增强子区域的DNA甲基化信号,获得基于组织特异增强子区域DNA甲基化信号的肺癌复发模型。

2. 如权利要求1所述的基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,其特征在于,所述肺癌复发模型如下:

$$ModelScore = \frac{1}{1 + e^{-\left(2.42 \times \beta_{chr1:170667082} + 2.27 \times \beta_{chr12:85280518} - 0.35 \times \beta_{chr17:76561412} - 3.23 \times \beta_{chr11:9759297} + 0.87 \times \beta_{chr15:91915828} - 2.09 \times \beta_{chr16:1079990} - 3.18 \times \beta_{chr2:226797988} - 2.08 \times \beta_{chr7:150477419} + 4.27\right)}}$$

其中, $\beta_{chr1:170667082}$ 、 $\beta_{chr12:85280518}$ 、 $\beta_{chr17:76561412}$ 、 $\beta_{chr11:9759297}$ 、 $\beta_{chr15:91915828}$ 、 $\beta_{chr16:1079990}$ 、 $\beta_{chr2:226797988}$ 、 $\beta_{chr7:150477419}$ 分别是甲基化位点chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419的beta值,ModelScore是模型的得分。

3. 如权利要求1或2所述的基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,其特征在于,所述根据模型得分以及多个得分阈值得到预测结果,包括:若模型得分 $\geq A1$,则认为患者处于高风险组;若模型得分 $\leq A2$,则认为患者处于低风险组;若模型得分在 $A2 \sim A1$ 之间,则认为患者处于未确定组。

4. 如权利要求3所述的基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,其特征在于, $A1$ 为0.6, $A2$ 为0.4。

5. 如权利要求1所述的基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,其特征在于,步骤S1中,所述早期为8版TNM分期中的IB期和IIA期。

6. 如权利要求1所述的基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,其特征在于,步骤S1中,DNA甲基化检测平台为RRBS、WGBS、450K或850K。

7. 如权利要求1所述的基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,其特

征在于,步骤S2利用单侧Wilcoxon秩和检验计算划分甲基化位点,划分阈值为 $P < 0.01$ 且 $\Delta \beta \geq 0.1$ 。

8.如权利要求1所述的基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,其特征在于,步骤S3中所述多个增强子数据库包括以下数据库:screen、SUPERdb和enhancerAtlas。

基于组织特异增强子区域DNA甲基化的肺癌复发预测方法

技术领域

[0001] 本发明涉及生物学领域,特别涉及基于组织特异增强子区域DNA甲基化的肺癌复发预测方法。

背景技术

[0002] 非小细胞肺癌是世界范围内的头号致死癌症。低剂量CT能够在人群中筛选出早期肺癌患者,对这些患者进行手术切除和术后辅助治疗,可以显著提高其整体生存率,但仍然有超过25%的早期(T2N0M0)患者会经历术后复发,而术后复发是预后较差的主要原因。考虑到肺癌存在广泛的异质性,在早期肺癌患者中精确识别其中的高复发人群是重要的科学问题。最新的美国国家综合癌症网络(National Comprehensive Cancer Network,NCCN)指南推荐6个高风险因素,包括肿瘤分化状态、楔形切除术和胸膜浸润等。然而,这些高风险因素的预测效能不足,亟须新的有效模型用于预测早期肺癌复发。

[0003] DNA甲基化是一种表观遗传学修饰,其信号改变与癌症的发生发展密切相关。Sandoval等人开发了一个基于HumanMethylation450芯片(后简称为450K)的非小细胞肺癌复发模型,用于预测I期非小细胞肺癌肺(non-small-cell lung cancer,NSCLC)患者的复发风险。该模型涉及5个基因,分别为HIST1H4F,PCDHGB6,NPBWR1,ALX1和HOXA9,这5个基因的甲基化程度用10个探针的beta值表示,分别为:cg10723962,cg22723502,cg12260798,cg16104915,cg12600174,cg18507379,cg18617005,cg26205771,cg07770968和cg14996220。对于一个基因对应多个探针的情况,用这些探针beta值的均值代表。在检测样本中,对于每一个基因,若其甲基化的beta值不小于0.4,认为其发生甲基化,否则认为该基因未发生甲基化。若该样本中这五个基因均未发生甲基化,或仅一个基因发生甲基化,认为该样本为低风险样本,否则为高风险样本。该非小细胞肺癌复发模型存在以下问题:

[0004] 1.该模型在外部数据的效能不足。当把模型应用于目前常见癌症数据库(如The Cancer Genome Atlas)的早期肺腺癌样本中,基于该模型不具有区分生存的效能($P=0.28$)。

[0005] 2.该模型的部分探针不能被简化亚硫酸氢盐测序(Reduced Representation Bisulfite Sequencing,RRBS)所覆盖,即难以应用于RRBS检测平台。

[0006] 3.该模型没有对阈值(0.4)附近的分类不清情况进行有效处理,如特定基因在第一个样本中的beta值是0.399,在第二个样本beta值是0.401,二者beta值差异不大(仅0.002),但前者被认为是甲基化,而后者被认为是非甲基化,这显然是不合理的。

发明内容

[0007] 本发明要解决的技术问题是:提供一种基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,用以解决目前非小细胞肺癌复发模型对非小细胞肺癌复发预测不合理的问题。

[0008] 为解决上述问题,本发明采用的技术方案是:基于组织特异增强子区域DNA甲基化

的肺癌复发预测方法,包括如下步骤:

[0009] 步骤1:提取肺癌患者的DNA,并进行DNA甲基化检测,得到组织特异增强子区域的DNA甲基化位点的beta值;其中,DNA甲基化位点包括chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419;

[0010] 步骤2:将以上DNA甲基化位点的beta值代入预先构建的肺癌复发模型,计算出模型得分,并根据模型得分以及多个得分阈值得到预测结果。

[0011] 进一步的,所述肺癌复发模型如下:

$$[0012] \quad ModelScore = \frac{1}{1 + e^{-(2.42 \times \beta_{chr1:170667082} + 2.27 \times \beta_{chr12:85280518} - 0.35 \times \beta_{chr17:76561412} - 3.23 \times \beta_{chr11:9759297} + 0.87 \times \beta_{chr15:91915828} - 2.09 \times \beta_{chr16:1079990} - 3.18 \times \beta_{chr2:226797988} - 2.06 \times \beta_{chr7:150477419} + 4.27)}}$$

[0013] 其中, $\beta_{chr1:170667082}$ 、 $\beta_{chr12:85280518}$ 、 $\beta_{chr17:76561412}$ 、 $\beta_{chr11:9759297}$ 、 $\beta_{chr15:91915828}$ 、 $\beta_{chr16:1079990}$ 、 $\beta_{chr2:226797988}$ 、 $\beta_{chr7:150477419}$ 分别是甲基化位点chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419的beta值,ModelScore是模型的得分。

[0014] 进一步的,所述根据模型得分以及多个得分阈值得到预测结果,包括:若模型得分 $\geq A1$,则被认为患者处于高风险组;若模型得分 $\leq A2$,则被认为患者处于低风险组;若模型得分在 $A2 \sim A1$ 之间,则被认为患者处于未确定组。

[0015] 进一步的,A1为0.6,A2为0.4。即:当模型得分 ≥ 0.6 ,则认为患者处于高风险组;若模型得分 ≤ 0.4 ,则认为患者处于低风险组;若模型得分在0.4~0.6之间,则认为患者处于未确定组。

[0016] 进一步的,所述肺癌复发模型的构建步骤包括:

[0017] S1:收集早期的非小细胞肺癌病例,包括临床信息匹配的复发与未复发患者,提取其DNA样本,并进行DNA甲基化检测;

[0018] S2:将样本的甲基化位点划分为高风险组和低风险组;

[0019] S3:结合多个增强子数据库,确定肺组织特异或干细胞特异的增强子区域,并从样本中筛选出位于增强子区域的甲基化位点;

[0020] S4:在步骤S3的基础上,从增强子100kb范围内进一步筛选出癌症相关基因,最终得到组织特异增强子区域的chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419八个DNA甲基化位点;

[0021] S5:利用岭回归训练组织特异增强子区域的DNA甲基化信号(即DNA甲基化位点的beta值),获得基于组织特异增强子区域DNA甲基化信号的肺癌复发模型。

[0022] 进一步的,步骤S1中,所述早期为8版TNM分期中的IB期和IIA期。

[0023] 进一步的,步骤S2中,DNA甲基化检测平台为RRBS、WGBS、450K或850K。

[0024] 进一步的,步骤S2可利用单侧Wilcoxon秩和检验计算划分甲基化位点,划分阈值为 $P < 0.01$ 且 $\Delta \beta \geq 0.1$ 。

[0025] 进一步的,步骤S3中所述多个增强子数据库可包括以下数据库:screen、SUPERdb和enhancerAtlas

[0026] 本发明的有益效果如下:

[0027] 1. 本发明建立肺癌复发模型时针对早期 (I/II期) 非小细胞肺癌的患者, 而现有大多数模型用所有肺癌样本进行建模, 其模型不一定适用于早期患者。

[0028] 2. 本发明建立肺癌复发模型时的训练数据用的是8版TNM分期中的IB期和IIA期, 样本的均一性更好。

[0029] 3. 本发明用到的肺癌复发模型可应用于多种甲基化检测平台, 比如RRBS、WGBS、450K和850K芯片平台, 而大多数现有模型仅能应用于一种甲基化检测平台。

[0030] 4. 本发明用到的肺癌复发模型基于组织特异增强子区域的DNA甲基化信号, 在多套外部数据中能得到良好的验证, 对于靶基因和整合模型都有较好的预测效能, 而且在单细胞中有效。

[0031] 5. 本发明用到的肺癌复发模型的增强子靶基因也可以在原发灶中预测转移相关的亚克隆, 可以为组织特异增强子的调控机制提供线索。

附图说明

[0032] 图1是本发明的肺癌复发模型训练流程图。

[0033] 图2是肺癌复发模型得分的分布图。

[0034] 图3是训练数据的KM曲线。

[0035] 图4是肺癌复发模型在外部数据GSE39279中的KM曲线。

[0036] 图5是肺癌复发模型在外部数据TCGA中的KM曲线。

具体实施方式

[0037] 为了解决目前非小细胞肺癌复发模型对非小细胞肺癌复发预测不合理的问题, 本发明提供了一种基于DNA甲基化的早期非小细胞肺癌复发预测方法, 主要包括建立非小细胞肺癌复发模型和使用非小细胞肺癌复发模型预测。

[0038] 一、建立非小细胞肺癌复发模型

[0039] S1: 收集早期的非小细胞肺癌病例, 包括临床信息匹配的复发与未复发患者, 提取其DNA样本, 并进行DNA甲基化检测;

[0040] S2: 将样本的甲基化位点划分为高风险组和低风险组;

[0041] S3: 结合多个增强子数据库, 确定肺组织特异或干细胞特异的增强子区域, 并从样本中筛选出位于增强子区域的甲基化位点;

[0042] S4: 在步骤S3的基础上, 从增强子100kb范围内进一步筛选出癌症相关基因, 最终得到组织特异增强子区域的chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419八个DNA甲基化位点的beta值;

[0043] S5: 利用岭回归训练组织特异增强子的DNA甲基化信号, 获得基于组织特异增强子区域DNA甲基化信号的肺癌复发模型, 具体的肺癌复发模型如下:

$$[0044] \quad ModelScore = \frac{1}{1 + e^{-\left(2.42 \times \beta_{chr1:170667082} + 2.27 \times \beta_{chr12:85280518} - 0.35 \times \beta_{chr17:76561412} - 3.23 \times \beta_{chr11:9759297} + 0.87 \times \beta_{chr15:91915828} - 2.09 \times \beta_{chr16:1079990} - 3.18 \times \beta_{chr2:226797988} - 2.08 \times \beta_{chr7:150477419} + 4.27\right)}}$$

[0045] 其中, $\beta_{chr1:170667082}$ 、 $\beta_{chr12:85280518}$ 、 $\beta_{chr17:76561412}$ 、 $\beta_{chr11:9759297}$ 、 $\beta_{chr15:91915828}$ 、 $\beta_{chr16:1079990}$ 、 $\beta_{chr2:226797988}$ 、 $\beta_{chr7:150477419}$ 分别是甲基化位点chr1:170667082、chr12:

85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419的beta值,ModelScore是模型的得分。

[0046] 二、使用非小细胞肺癌复发模型预测

[0047] 非小细胞肺癌复发模型建立之后,就可以利用得到非小细胞肺癌复发模型进行非小细胞肺癌复发预测,具体包括:

[0048] 步骤1:提取肺癌患者的DNA,并进行DNA甲基化检测,得到组织特异增强子区域的DNA甲基化位点的beta值;其中,DNA甲基化位点包括chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419;

[0049] 步骤2:将DNA甲基化位点的beta值代入预先构建的肺癌复发模型,计算出模型得分,若模型得分 $\geq A1$,则被认为患者处于高风险组;若模型得分 $\leq A2$,则被认为患者处于低风险组;若模型得分在 $A2 \sim A1$ 之间,则被认为患者处于未确定组。

[0050] 以下再结合附图和实施例对本发明做进一步说明。

[0051] 实施例一种基于组织特异增强子区域DNA甲基化的肺癌复发预测方法,具体流程如下:

[0052] (1)、如图1所示,收集IB/IIA期的非小细胞肺癌病例,包括临床信息匹配的复发与未复发患者,提取其DNA样本,并进行简化亚硫酸氢盐测序。

[0053] (2)、为了使肺癌复发模型可以同时被应用于450K芯片平台,我们筛选出能够被该芯片探针检测的甲基化位点进行后续模型建立。需要说明的是,理论上只要能够检测甲基化位点/区间的技术,都可以应用该模型,所以除了450K,在其他甲基化检测技术(如RRBS、WGBS和850K)中可以达到相同的目的。

[0054] (3)、我们将样本分为高风险组和低风险组,利用单侧Wilcoxon秩和检验计算显著差异的甲基化位点,阈值为 $P < 0.01$ 且 $\Delta \text{beta} \geq 0.1$ 。

[0055] (4)、为了获得肺组织特异或干细胞特异的增强子区域,我们整合三个数据库的资源,包括screen、SUPERdb和enhancerAtlas,要求差异甲基化位点落于增强子区域内。

[0056] (5)、我们要求增强子100kb范围内的靶基因必须是癌症相关基因,得到chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419八个癌症相关组织特异的增强子区域,用于训练模型。

[0057] (6)、我们利用岭回归训练组织特异增强子的DNA甲基化信号,即R包glmnet的cv.glmnet函数,参数为family="binomial",alpha=1,获得基于组织特异增强子区域DNA甲基化信号的肺癌复发模型。该肺癌复发模型涉及的甲基化位点以及其模型系数如表1所示,模型使用的参考基因组版本是hg38。

[0058] 表1.组织特异增强子甲基化复发模型系数

甲基化位点	系数	Delta Beta	P 值	细胞系	来源	邻近基因
chr1:170667082	2.42	0.20	<0.001	NHLF	dbSUPER	PRRX1
chr12:85280518	2.27	0.12	0.006	H9	screen	ALX1 CYGB, ST6GALNA C2, ST6GALNA C1
chr17:76561412	-0.35	0.12	0.010	Lung female adult 30 year, 53 year	screen	SWAP70
[0059] chr11:9759297	-3.23	-0.18	<0.001	IMR90,A549, NHLF	dbSUPER, enhancerAtlas	SLCO3A1 SOX8
chr15:91915828	0.87	-0.12	0.006	Lung	dbSUPER	RHBDD1, IRS1
chr16:1079990	-2.09	-0.10	0.009	UCSCF-4	screen	GIMAP8
chr2:226797988	-3.18	-0.14	0.008	IMR90,Lung, NHLF	dbSUPER	
chr7:150477419	-2.08	-0.14	0.003	Lung female adult 53 year	screen	

[0060] 根据表1,可得到肺癌复发模型如下:

$$[0061] \text{ModelScore} = \frac{1}{1 + e^{-\left(2.42 \times \beta_{\text{chr1:170667082}} + 2.27 \times \beta_{\text{chr12:85280518}} - 0.35 \times \beta_{\text{chr17:76561412}} - 3.23 \times \beta_{\text{chr11:9759297}} + 0.87 \times \beta_{\text{chr15:91915828}} - 2.09 \times \beta_{\text{chr16:1079990}} - 3.18 \times \beta_{\text{chr2:226797988}} - 2.08 \times \beta_{\text{chr7:150477419}} + 4.27\right)}}$$

[0062] 其中, $\beta_{\text{chr1:170667082}}$ 、 $\beta_{\text{chr12:85280518}}$ 、 $\beta_{\text{chr17:76561412}}$ 、 $\beta_{\text{chr11:9759297}}$ 、 $\beta_{\text{chr15:91915828}}$ 、 $\beta_{\text{chr16:1079990}}$ 、 $\beta_{\text{chr2:226797988}}$ 、 $\beta_{\text{chr7:150477419}}$ 分别是甲基化位点 chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419 的 beta 值, ModelScore 是模型的得分。

[0063] 如图2,因为上述肺癌复发模型的得分呈现双峰分布,所以我们根据模型得分的分布特点设了两个划分阈值,将模型得分划分为3个分组,用于对于复发风险程度,以便医生进行更加准确的进行干预治疗。经过优化,本实施例将模型得分 ≥ 0.6 认定为高得分组,此时认为患者处于高风险组;将模型得分 ≤ 0.4 认定为低得分组,此时认为患者处于低风险组;而模型得分在 0.4~0.6 之间认定为中得分组,此时认为患者处于复发未确定组。

[0064] 经过发明人验证,如果使用以上肺癌复发模型以及其得分分组划分进行肺癌复发预测,将具有以下明显技术效果:

[0065] i. 在训练数据中,高得分组比低得分组具有更高的复发风险,如图3所示。

[0066] 相对于低得分组,高得分组的风险比为 8.87 ($P < 0.001$),且该模型独立于常见的临床因素(包括年龄、性别和TNM分期),如表2所示。

[0067] 表2. 训练数据的cox生存分析

变量		风险比 (95% 置信区间)	P 值
单变量分析			
模型	中得分 vs. 低得分	3.61(1.10-11.88)	0.034
	高得分 vs. 低得分	8.87(3.21-24.56)	<0.001
年龄		0.96(0.92-0.99)	0.019
性别	男 vs. 女	0.98(0.46-2.10)	0.967
吸烟		0.65(0.28-1.54)	0.331
分期	II vs. I	1.06(0.47-2.36)	0.890
胸膜浸润		5.43(1.29-23.00)	0.021
分化程度	差 vs 中/良好	1.87(0.86-4.03)	0.112
术后化疗		0.63(0.25-1.57)	0.322
多因素分析			
模型	中得分 vs. 低得分	4.16(1.02-17.04)	0.047
	高得分 vs. 低得分	8.88(2.90-27.19)	<0.001
年龄		0.99(0.94-1.05)	0.802
性别	男 vs. 女	2.35(0.81-6.87)	0.118
吸烟		0.20(0.05-0.76)	0.018
分期	II vs. I	1.60(0.55-4.66)	0.391
胸膜浸润		5.96(1.28-27.81)	0.023
分化程度	差 vs 中/良好	1.65(0.71-3.85)	0.245
术后化疗		0.48(0.15-1.55)	0.222

[0069] ii. 在外部数据GSE39279中, 高分组比低分组具有更高的复发风险, 如图4所示。

[0070] 相对于低分组, 高分组的风险比为2.31 (P=0.004), 且该模型独立于常见的临床因素 (包括年龄、性别、TNM分期和组织学类型), 如表3所示。

[0071] 表3. 外部数据GSE39279的cox生存分析

变量		风险比 (95% 置信区间)	P 值
单因素分析			
模型	中得分 vs. 低得分	1.16(0.62-2.17)	0.646
	高得分 vs. 低得分	2.31(1.31-4.05)	0.004
年龄		0.98(0.96-1.00)	0.109
性别	男 vs. 女	1.47(0.92-2.35)	0.108
吸烟		1.77(0.244-12.75)	0.573
分期	II vs. I	6.71(3.84-11.71)	<0.001
病理亚型	鳞癌 vs. 腺癌	2.01(1.21-3.34)	0.007
多因素分析			
模型	中得分 vs. 低得分	1.06(0.55-2.02)	0.862
	高得分 vs. 低得分	1.83(1.02-3.30)	0.043
年龄		0.98(0.96-1.00)	0.090
性别	男 vs. 女	1.39(0.83-2.33)	0.214
吸烟		3.01(0.40-22.467)	0.283
分期	II vs. I	5.88(3.35-10.32)	<0.001
病理亚型	鳞癌 vs. 腺癌	1.67(0.98-2.84)	0.060

[0073] iii. 在外部数据TCGA中, 高分组比低分组具有更高的复发风险, 如图5所示。

[0074] 相对于低得分组, 高分组的风险比为2.36 ($P < 0.05$), 且该模型独立于常见的临床因素(包括年龄、性别和TNM分期), 如表4所示。

[0075] 表4. 外部数据TCGA的cox生存分析

变量		风险比 (95% 置信区间)	P 值
单因素分析			
模型	中得分 vs. 低得分	2.02(1.01-4.01)	0.046
	高分 vs. 低得分	2.36(1.21-4.63)	0.012
年龄		1.03(1.00-1.06)	0.057
性别	男 vs. 女	1.12(0.67-1.86)	0.670
吸烟		1.07(0.51-2.25)	0.864
分期	II vs. I	1.62(0.96-2.71)	0.069
术后化疗		1.12(0.61-2.05)	0.709
多因素分析			
模型	中得分 vs. 低得分	2.21(1.09-4.50)	0.029
	高分 vs. 低得分	2.83(1.41-5.66)	0.003
年龄		1.04(1.01-1.07)	0.010
性别	男 vs. 女	1.04(0.62-1.75)	0.873
吸烟		1.35(0.62-2.95)	0.446
分期	II vs. I	1.65(0.94-2.87)	0.079
术后化疗		0.94(0.48-1.83)	0.848

[0077] iv. 此外, 肺癌复发模型中增强子靶基因也可以在原发灶中预测转移相关的亚克隆。

[0078] 我们从EpiMap数据库中获得模型中增强子的靶基因, 利用弹性回归训练靶基因, 获得包括ADM、RHOT2、MRPL28和WEE1靶基因表达模型。我们从公共数据库中获得包含肺癌原发灶、转移淋巴结和脑转移的单细胞测序数据。这4个靶基因的表达值在转移组织中显著高于原发灶, 在原发灶中显著高于正常肺组织。其中, 患者19有匹配的原发灶、淋巴结转移和脑转移。我们将原发灶分成4类, 利用diffusion map刻画其进化轨迹, 发现无论是针对淋巴结转移, 还是脑转移, c2都是转移相关亚克隆。在c2中, 增强子靶基因具有最高的表达量, 说明增强子靶基因可以在原发灶中提示转移相关亚克隆。

[0079] (7)、利用步骤(6)训练得到的肺癌复发模型实现肺癌复发预测: 对于待预测的肺癌患者, 首先进行组织特异增强子区域DNA提取, 检测其chr1:170667082、chr12:85280518、chr17:76561412、chr11:9759297、chr15:91915828、chr16:1079990、chr2:226797988、chr7:150477419八个位点的甲基化值, 然后利用八个位点的甲基化值以及肺癌复发模型系数计算得分, 并根据模型得分得到预测结果, 即得出待预测患者是处于低风险组、中风险组还是高风险组。相对于低得分组, 高分组的患者可以建议密切随访或者提早干预治疗。

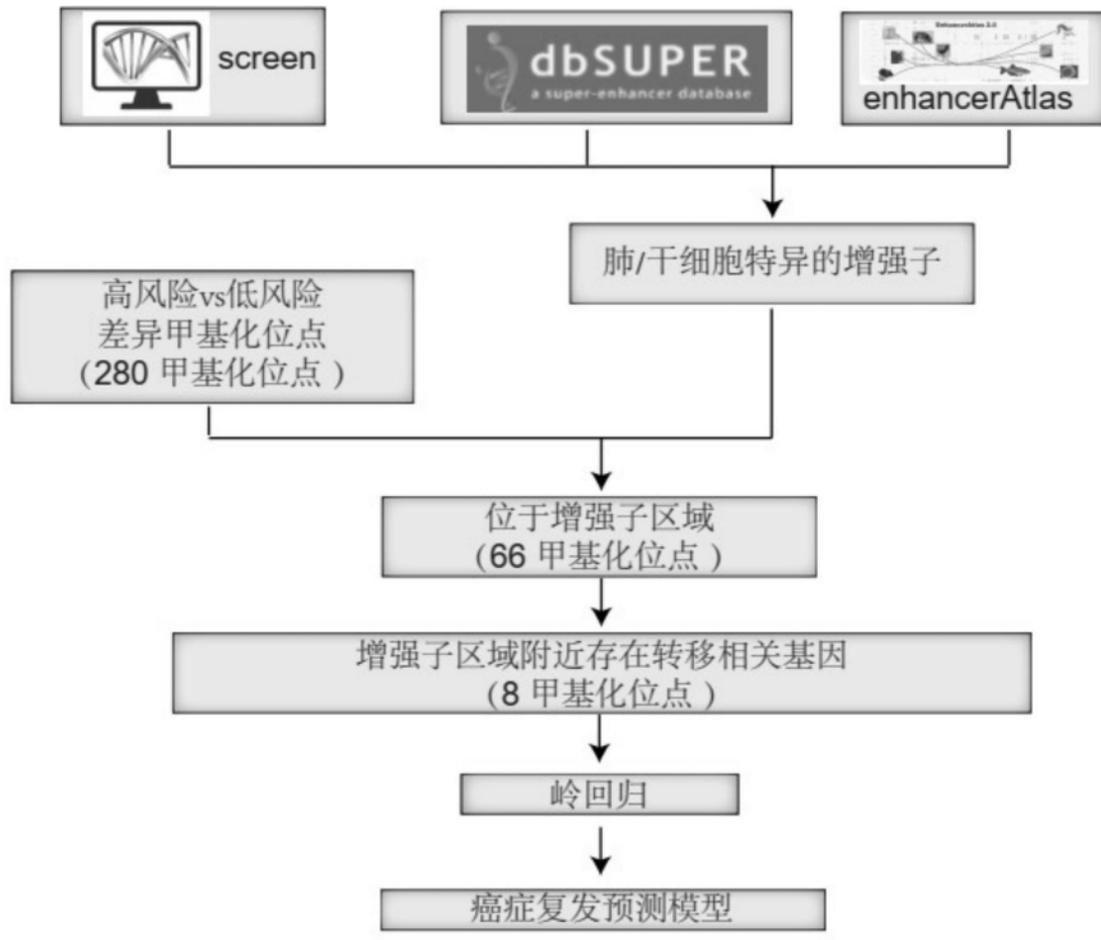


图1

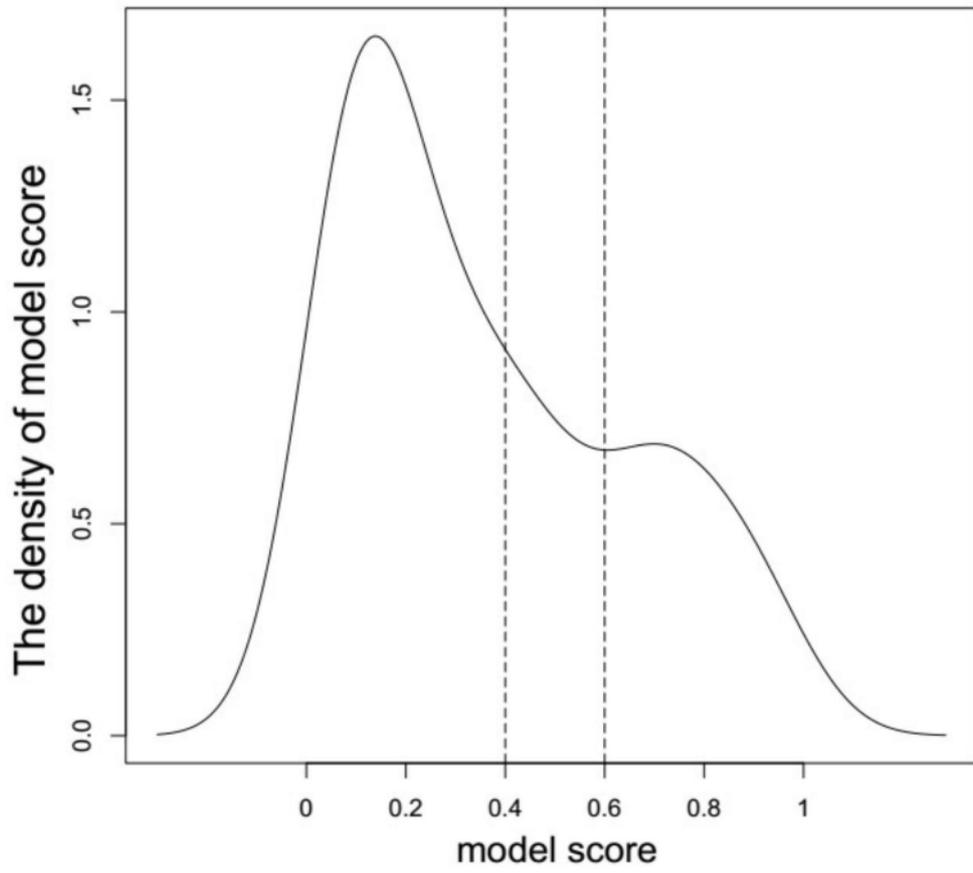


图2

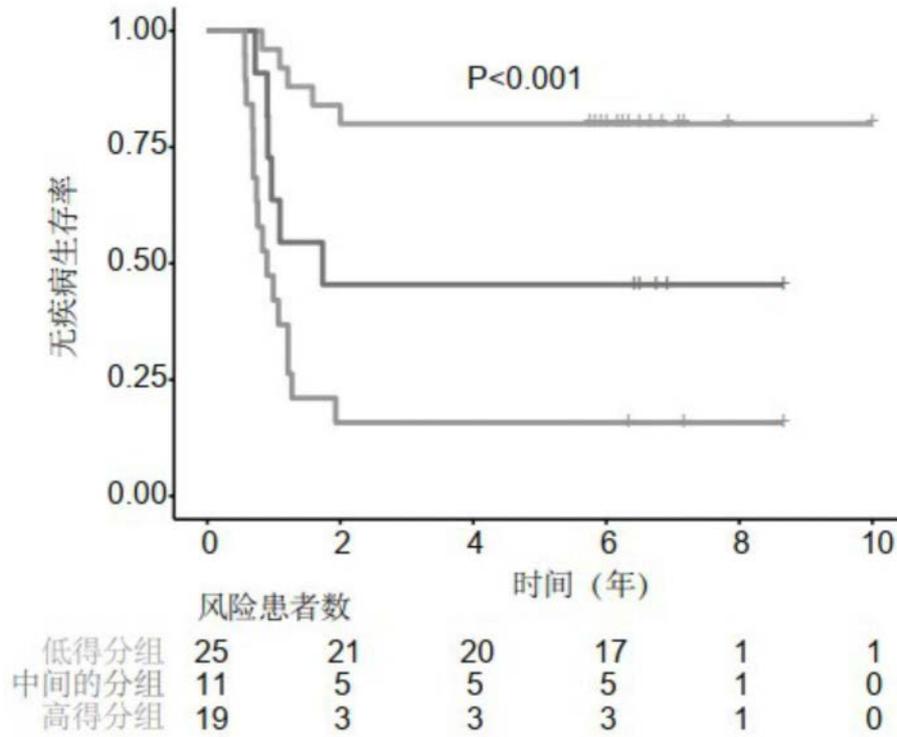


图3

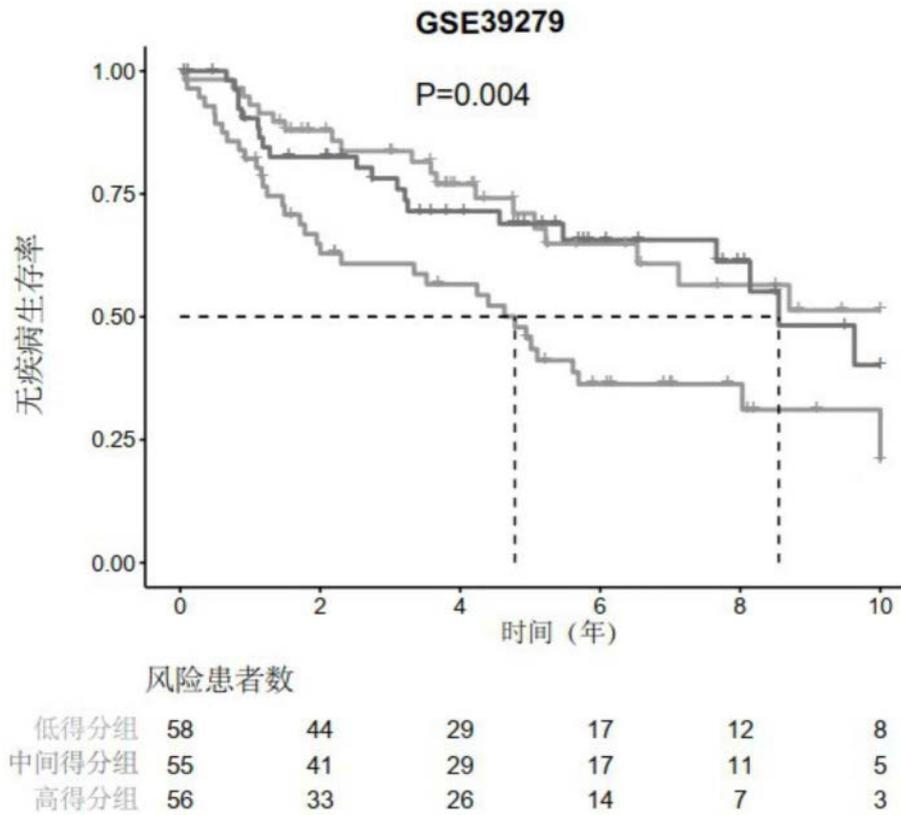


图4

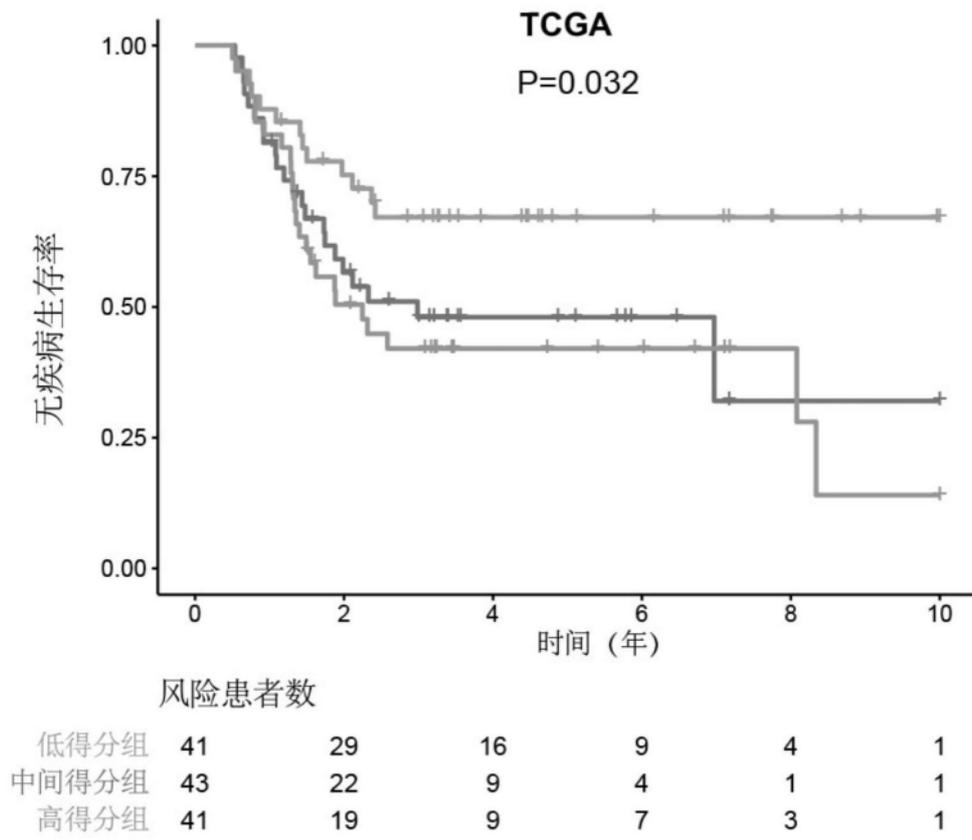


图5