



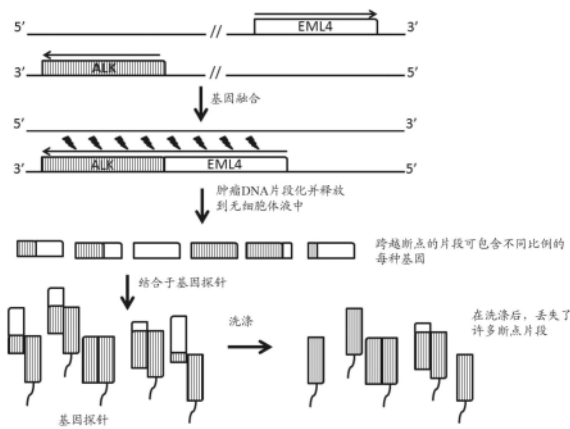
(21) 申请号 201680054945.9  
(22) 申请日 2016.07.21  
(65) 同一申请的已公布的文献号  
申请公布号 CN 108138230 A  
(43) 申请公布日 2018.06.08  
(30) 优先权数据  
62/195,280 2015.07.21 US  
(85) PCT国际申请进入国家阶段日  
2018.03.21  
(86) PCT国际申请的申请数据  
PCT/US2016/043430 2016.07.21  
(87) PCT国际申请的公布数据  
W02017/015513 EN 2017.01.26  
(73) 专利权人 夸登特健康公司  
地址 美国加利福尼亚州  
(72) 发明人 斯特凡尼·安·沃德·莫蒂默  
(74) 专利代理机构 北京安信方达知识产权代理有限公司 11262  
专利代理师 李平 郑霞

(51) Int.Cl.  
C12Q 1/6869 (2018.01)  
C40B 40/06 (2006.01)  
C40B 30/04 (2006.01)  
C12P 19/34 (2006.01)  
(56) 对比文件  
CN 104364391 A, 2015.02.18  
CN 101928767 A, 2010.12.29  
Yun Lei et al..Sequence-specific electrochemical detection of double-strand PCR amplicons of PML/RAR $\alpha$  fusion gene in acute promyelocytic leukemia.《ANALYTICAL AND BIOANALYTICAL CHEMISTRY》.2012,第423-428页.  
Liqing Lin et al..Locked nucleic acids biosensor for detection of BCR/ABL fusion gene using benzoate binuclear copper (II) complex as hybridization indicator.《SENSORS AND ACTUATORS B-CHEMICAL》.2011,第155卷(第1期),第1-7页.

审查员 王震  
权利要求书40页 说明书19页 附图40页

(54) 发明名称  
用于捕获融合基因的锁核酸

(57) 摘要  
本文提供了用于富集样品的包含融合基因的断点的多核苷酸的方法,所述方法包括:a) 将包括多种多核苷酸探针的探针集与多核苷酸的混合物在杂交条件下接触以产生探针捕获的多核苷酸,每种探针被配置为与融合基因特异性杂交,其中探针集包括一种或更多种高亲和力多核苷酸探针(例如,包括一个或更多个锁核酸核苷酸的多核苷酸);和b) 从混合物分离探针捕获的多核苷酸以产生富集包含融合基因的断点片段的多核苷酸的样品。



1. 探针集在制备制剂中的用途,所述制剂基于来自受试者的包含无细胞核酸分子的生物样品中的断点片段的检测为具有或疑似具有癌症的受试者提供诊断或治疗干预,其中所述探针集包含靶向融合基因和非融合基因两者的多核苷酸探针,其中所述多核苷酸探针的子集包含一个或更多个锁核酸核苷酸,其中所述子集中的多核苷酸探针的每种具有(i)与融合基因的序列互补性和(ii)比具有与所述融合基因互补的序列并仅包含未修饰的核苷酸的多核苷酸更大的对所述融合基因的亲和力,其中在使用所述制剂时包括以下步骤:

(a) 提供来自受试者的包含无细胞核酸分子的生物样品;

(b) 将来自所述生物样品的所述无细胞核酸分子与所述探针集在足以产生探针捕获的多核苷酸的杂交条件下接触;

(c) 从混合物分离所述探针捕获的多核苷酸以产生富集包含所述融合基因的断点片段的分离的多核苷酸的样品;

(d) 将所述分离的多核苷酸测序以产生序列;

(e) 基于所述序列检测包含融合基因的断点的多核苷酸;和

(f) 基于断点片段的检测提供诊断或治疗干预。

2. 如权利要求1所述的用途,其中所述多核苷酸探针的每种包含多个锁核酸核苷酸,其中所述锁核酸核苷酸的至少两个间隔不多于30个核苷酸。

3. 如权利要求2所述的用途,其中所述锁核酸核苷酸的所述至少两个间隔不多于15个核苷酸。

4. 如权利要求1所述的用途,其中所述多核苷酸探针的至少一个子集的每种多核苷酸探针的至少50%核苷酸是锁核酸核苷酸。

5. 如权利要求4所述的用途,其中所述多核苷酸探针的至少一个子集的每种多核苷酸探针的至少75%核苷酸是锁核酸核苷酸。

6. 如权利要求1所述的用途,其中所述多核苷酸探针的每种具有比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸高至少1℃的解链温度。

7. 如权利要求6所述的用途,其中所述解链温度高至少10℃。

8. 如权利要求1所述的用途,其中所述多核苷酸探针的每种具有比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸高至少2%的解链温度。

9. 如权利要求8所述的用途,其中所述解链温度高至少10%。

10. 如权利要求1所述的用途,其中所述融合基因为癌症融合基因。

11. 如权利要求1所述的用途,其中所述多核苷酸探针的每种与选自下表的融合基因对的基因

ACBD6/RRP15	KMT2A/ELL
ACSL3_ENST00000357430/ETV1	KMT2A/EP300
ACTB/GLI1	KMT2A/EPS15
AGPAT5/MCPH1	KMT2A/FOXO3
AGTRAP/BRAF	KMT2A/FOXO4
AKAP9_ENST00000356239/BRAF	KMT2A/FRYL
ARFIP1/FHDC1	KMT2A/GAS7
ARID1A/MAST2_ENST00000361297	KMT2A/GMPS
ASPSCR1/TFE3	KMT2A/GPHN
ATG4C/FBXO38	KMT2A/KIAA0284_ENST00000414716
ATIC/ALK	KMT2A/KIAA1524
BBS9/PKD1L1	KMT2A/LASP1
BCR/ABL1	KMT2A/LPP
BCR/JAK2	KMT2A/MAPRE1
BRD3/NUTM1	KMT2A/MLLT1
BRD4_ENST00000263377/NUTM1	KMT2A/MLLT10
C2orf44/ALK	KMT2A/MLLT11
CANT1/ETV4	KMT2A/MLLT3
CARS/ALK	KMT2A/MLLT4_ENST00000392108
CCDC6/RET	KMT2A/MLLT6
CD74_ENST00000009530/NRG1	KMT2A/MYO1F
CD74_ENST00000009530/ROS1	KMT2A/NCKIPSD
CDH11/USP6_ENST00000250066	KMT2A/NRIP3
CDKN2D/WDFY2	KMT2A/PDS5A
CEP89/BRAF	KMT2A/PICALM
CHCHD7/PLAG1	KMT2A/PRRC1

CIC/DUX4L1	KMT2A/SARNP
CIC/FOXO4	KMT2A/SEPT2
CLCN6/BRAF	KMT2A/SEPT5
CLIP1/ROS1	KMT2A/SEPT6
CLTC/ALK	KMT2A/SEPT9_ENST00000427177
CLTC/TFE3	KMT2A/SH3GL1
CNBP/USP6_ENST00000250066	KMT2A/SORBS2
COL1A1/PDGFB	KMT2A/TET1
COL1A1/USP6_ENST00000250066	KMT2A/TOP3A
COL1A2/PLAG1	KMT2A/ZFYVE19
CRTC1/MAML2	KTN1/RET
CRTC3/MAML2	LIFR_ENST00000263409/PLAG1
CTAGE5/SIP1	LMNA/NTRK1_ENST00000392302
CTNNB1/PLAG1	LRIG3/ROS1
DCTN1/ALK	LSM14A/BRAF
DDX5_ENST00000540698/ETV4	MARK4/ERCC2
DNAJB1/PRKACA	MBOAT2/PRKCE
EIF3E/RSP02	MBTD1_ENST00000586178/CXorf67_ENST00000342995
EIF3K/CYP39A1	MEAF6/PHF1
EML4/ALK	MKRN1/BRAF
EPC1/PHF1	MSN/ALK
ERC1/RET	MYB_ENST00000341911/NFIB_ENST00000397581
ERC1/ROS1	MYO5A/ROS1
ERO1L/FERMT2	NAB2/STAT6
ESRP1/RAF1	NACC2/NTRK2
ETV6/ITPR2	NCOA4_ENST00000452682/RET
ETV6/JAK2	NDRG1/ERG
ETV6/NTRK3	NF1/ACCN1



EWSR1/ATF1	NFIA/EHF
EWSR1/CREB1	NFIX_ENST00000360105/MAST1_ENST00000251472
EWSR1/DDIT3	NONO/TFE3
EWSR1/ERG	NOTCH1_ENST00000277541/GABBR2
EWSR1/ETV1	NPM1/ALK
EWSR1/ETV4	NTN1/ACLY
EWSR1/FEV	NUP107/LGR5
EWSR1/FLI1	OMD/USP6_ENST00000250066
EWSR1/NFATC1	PAX3/FOXO1
EWSR1/NFATC2	PAX3/NCOA1
EWSR1/NR4A3	PAX3/NCOA2
EWSR1/PATZ1	PAX5/JAK2
EWSR1/PBX1	PAX7/FOXO1
EWSR1/POU5F1	PAX8/PPARG
EWSR1/SMARCA5	PCM1/JAK2
EWSR1/SP3	PCM1/RET
EWSR1/WT1	PLA2R1/RBMS1
EWSR1/YY1	PLXND1/TMCC1
EWSR1/ZNF384	PPFIBP1/ALK
EWSR1/ZNF444_ENST00000337080	PPFIBP1/ROS1
EZR/ROS1	PRCC/TFE3
FAM131B_ENST00000443739/BRAF	PRKAR1A/RET
FBXL18/RNF216	PTPRK/RSPO3
FCHSD1/BRAF	PWWP2A/ROS1
FGFR1/ZNF703	QKI/NTRK2
FGFR1_ENST00000447712/PLAG1	RAF1/DAZL
FGFR1_ENST00000447712/TACC1	RANBP2/ALK
FGFR3/BAIAP2L1	RBM14/PACS1

FGFR3/TACC3	RGS22/SYCP1
FN1/ALK	RNF130/BRAF
FUS/ATF1	SDC4/ROS1
FUS/CREB3L1	SEC16A_NM_014866.1/NOTCH1_ENST00000277541
FUS/CREB3L2	SEC31A/ALK
FUS/DDIT3	SEC31A/JAK2
FUS/ERG	SEPT8/AFF4
FUS/FEV	SFPQ/TFE3
GATM/BRAF	SLC22A1/CUTA
GMDS/PDE8B	SLC26A6/PRKAR2A
GNAI1/BRAF	SLC34A2/ROS1
GOLGA5/RET	SLC45A3/BRAF
GOPC/ROS1	SLC45A3/ELK4
GPBP1L1/MAST2_ENST00000361297	SLC45A3/ERG
HACL1/RAF1	SLC45A3/ETV1
HAS2/PLAG1	SLC45A3/ETV5_ENST00000306376
HERPUD1/BRAF	SND1/BRAF
HEY1/NCOA2	SQSTM1/ALK
HIP1/ALK	SRGAP3/RAF1
HLA-A/ROS1	SS18/SSX1
HMGA2/ALDH2_ENST00000261733	SS18/SSX2
HMGA2/CCNB1IP1	SS18/SSX4
HMGA2/COX6C	SS18L1/SSX1
HMGA2/EBF1	SSBP2/JAK2
HMGA2/FHIT_ENST00000476844	SSH2/SUZ12
HMGA2/LHFP	STIL/TAL1
HMGA2/LPP	STRN/ALK
HMGA2/NFIB_ENST00000397581	SUSD1/ROD1

HMGA2/RAD51B	TADA2A_ENST00000394395/MAST1_ENST00000251472
HMGA2/WIF1_ENST00000286574	TAF15/NR4A3
HN1/USH1G	TCEA1_ENST00000521604/PLAG1
HNRNPA2B1/ETV1	TCF12/NR4A3
HOOK3/RET	TCF3/PBX1
IL6R/ATP8B2	TECTA/TBCEL
INTS4/GAB2	TFG/ALK
IRF2BP2/CDX1	TFG/NR4A3
JAZF1/PHF1	TFG/NTRK1_ENST00000392302
JAZF1/SUZ12	THRAP3/USP6_ENST00000250066
KIAA1549/BRAF	TMPRSS2/ERG
KIAA1598/ROS1	TMPRSS2/ETV1
KIF5B/ALK	TMPRSS2/ETV4
KIF5B/RET	TMPRSS2/ETV5_ENST00000306376
KLC1/ALK	TP53/NTRK1_ENST00000392302
KLK2/ETV1	TPM3/ALK
KLK2/ETV4	TPM3/NTRK1_ENST00000392302
KMT2A/ABI1	TPM3/ROS1
KMT2A/ABI2	TPM3_ENST00000368530/ROS1
KMT2A/ACTN4	TPM4/ALK
KMT2A/AFF1	TRIM24/RET
KMT2A/AFF3	TRIM27/RET
KMT2A/AFF4	TRIM33_ENST00000358465/RET
KMT2A/ARHGAP26	UBE2L3/KRAS
KMT2A/ARHGEF12	VCL/ALK
KMT2A/BTBD18	VTI1A/TCF7L2
KMT2A/CASC5	YWHAE_ENST00000264335/FAM22A_ENST00000381707
KMT2A/CASP8AP2	YWHAE_ENST00000264335/NUTM2B
KMT2A/CBL	ZC3H7B/BCOR_ENST00000378444
KMT2A/CREBBP	ZCCHC8/ROS1
KMT2A/CT45A2	ZNF700/MAST1_ENST00000251472
KMT2A/DAB2IP	ZSCAN30/BRAF
KMT2A/EEFSEC	

编号	组群 基因	配偶体 基因	编号	组群 基因	配偶体 基因	编号	组群 基因	配偶体 基因
1	ALK	ATIC	44	RET	ERC1	87	FGFR3	TACC3
2	ALK	C2orf44	45	RET	FKBP15	88	FGFR3	TNIP2
3	ALK	C2orf61	46	RET	GOLGA5	89	FGFR3	TPRG1
4	ALK	CAMKMT	47	RET	HOOK3	90	FGFR3	WHSC1
5	ALK	CARS	48	RET	KIF5B	91	ROS1	CCDC6
6	ALK	CENPF	49	RET	KTN1	92	ROS1	CD72
7	ALK	CLTC	50	RET	NCOA4	93	ROS1	CD74
8	ALK	CLIP4	51	RET	PCM1	94	ROS1	CEP85L
9	ALK	COL4A3	52	RET	PRKAR1A	95	ROS1	CLIP1
10	ALK	DCTN1	53	RET	SPECC1L	96	ROS1	CLTC
11	ALK	EIF2AK3	54	RET	SQSTM1	97	ROS1	ERC1
12	ALK	EML4	55	RET	TBL1XR1	98	ROS1	EZR
13	ALK	FANCL	56	RET	TRIM24	99	ROS1	HLA-A
14	ALK	FN1	57	RET	TRIM27	100	ROS1	KDELR2
15	ALK	GTF2IRD1	58	RET	TRIM33	101	ROS1	KIAA1598
16	ALK	HIP1	59	FGFR2	AFF3	102	ROS1	LRIG3

17	ALK	KCNQ5	60	FGFR2	AHCYL1	103	ROS1	MYO5A
18	ALK	KIF5B	61	FGFR2	BICC1	104	ROS1	PPFIBP1
19	ALK	KLC1	62	FGFR2	C10ORF68	105	ROS1	PWWP2A
20	ALK	MPRIP	63	FGFR2	CASP7	106	ROS1	RUNX1
21	ALK	MSN	64	FGFR2	CCAR2	107	ROS1	SDC4
22	ALK	MYH9	65	FGFR2	CCDC6	108	ROS1	SLC34A2
23	ALK	NPM1	66	FGFR2	CD44	109	ROS1	SLC35F1
24	ALK	PPFIBP1	67	FGFR2	CIT	110	ROS1	TFG
25	ALK	PPM1B	68	FGFR2	KIAA1598	111	ROS1	TMEM106B
26	ALK	PPP1CB	69	FGFR2	MGEA5	112	ROS1	TPM3
27	ALK	PRKAR1A	70	FGFR2	NCALD	113	ROS1	TTC28
28	ALK	PTPN3	71	FGFR2	OFD1	114	ROS1	YWHAE
29	ALK	RANBP2	72	FGFR2	PPAPDC1A	115	ROS1	ZCCHC8
30	ALK	RNF213	73	FGFR2	PPHLN1	116	NTRK1	CD74
31	ALK	SEC31A	74	FGFR2	RASAL2	117	NTRK1	IRF2BP2
32	ALK	SOCS5	75	FGFR2	SLC45A3	118	NTRK1	LMNA

33	ALK	SQSTM1	76	FGFR2	TACC2	119	NTRK1	MPRIIP
34	ALK	STK32B	77	FGFR2	TACC3	120	NTRK1	NFASC
35	ALK	STRN	78	FGFR2	TXLNB	121	NTRK1	PEAR1
36	ALK	TFG	79	FGFR2	USP10	122	NTRK1	PLEKHA6
37	ALK	TPM1	80	FGFR2	VCL	123	NTRK1	QKI
38	ALK	TPM3	81	FGFR3	ADD1	124	NTRK1	RABGAP1L
39	ALK	TPM4	82	FGFR3	AES	125	NTRK1	SQSTM1
40	ALK	VCL	83	FGFR3	BAIAP2L1	126	NTRK1	SSBP2
41	RET	AFAP1	84	FGFR3	ELAVL3	127	NTRK1	TFG
42	RET	AKAP13	85	FGFR3	ETV6	128	NTRK1	TP53
43	RET	CCDC6	86	FGFR3	JAKMIP1	129	NTRK1	TPM3

或选自下表的两种或更多种基因之间的融合基因具有序列互补性

ACBD6  
ACCN1  
ACLY  
ACSL3\_ENST00000357430  
ACTB  
AFF4  
AGPAT5  
AGTRAP  
AKAP9\_ENST00000356239  
ALDH2\_ENST00000261733  
ALK  
ARFIP1  
ARID1A  
ASPSCR1  
ATF1  
ATG4C  
ATIC  
ATP8B2  
BAIAP2L1  
BBS9  
BCOR\_ENST00000378444  
BCR  
BRAF  
BRD3  
BRD4\_ENST00000263377  
C2orf44  
CANT1  
CARS  
CCDC6

CCNB1IP1  
CD74\_ENST00000009530  
CDH11  
CDKN2D  
CDX1  
CEP89  
CHCHD7  
CIC  
CLCN6  
CLIP1  
CLTC  
CNBP  
COL1A1  
COL1A2  
COX6C  
CREB1  
CREB3L1  
CREB3L2  
CRTC1  
CRTC3  
CTAGE5  
CTNNB1  
CUTA  
CXorf67\_ENST00000342995  
CYP39A1  
DAZL  
DCTN1  
DDIT3  
DDX5\_ENST00000540698  
DUX4L1  
EBF1  
EHF

EIF3E	FUS	KTN1
EIF3K	GAB2	LGR5
ELK4	GABBR2	LHFP
EML4	GATM	LIFR_ENST00000263409
EPC1	GLI1	LMNA
ERC1	GMDS	LPP
ERCC2	GNAI1	LRIG3
ERG	GOLGA5	LSM14A
ERO1L	GOPC	MAML2
ESRP1	GPBP1L1	MARK4
ETV1	HACL1	MAST1_ENST00000251472
ETV4	HAS2	MAST2_ENST00000361297
ETV5_ENST00000306376	HERPUD1	MBOAT2
ETV6	HEY1	MBTD1_ENST00000586178
EWSR1	HIP1	MCPH1
EZR	HLA-A	MEAF6
FAM131B_ENST00000443739	HMGA2	MKRN1
FAM22A_ENST00000381707	HN1	MSN
FBXL18	HNRNPA2B1	MYB_ENST00000341911
FBXO38	HOOK3	MYO5A
FCHSD1	IL6R	NAB2
FERMT2	INTS4	NACC2
FEV	IRF2BP2	NCOA1
FGFR1	ITPR2	NCOA2
FGFR1_ENST00000447712	JAK2	NCOA4_ENST00000452682
FGFR3	JAZF1	NDRG1
FHDC1	KIAA1549	NF1
FHIT_ENST00000476844	KIAA1598	NFATC1
FLI1	KIF5B	NFATC2
FN1	KLC1	NFIA
FOXO1	KLK2	NFIB_ENST00000397581
FOXO4	KRAS	NFIX_ENST00000360105



NONO		SP3
NOTCH1_ENSTC0000277541	PRKAR2A	SQSTM1
NPM1	PRKCE	SRGAP3
NR4A3	PTPRK	SS18
NRG1	PWWP2A	SS18L1
NTN1	QKI	SSBP2
NTRK1_ENST00000392302	RAD51B	SSH2
NTRK2	RAF1	SSX1
NTRK3	RANBP2	SSX2
NUP107	RBM14	SSX4
NUTM1	RBMS1	STAT6
NUTM2B	RET	STRN
OMD	RGS22	SUSD1
PACS1	RNF130	SUZ12
PATZ1	RNF216	SYCP1
PAX3	ROD1	TACC1
PAX5	ROS1	TACC3
PAX7	RRP15	TADA2A_ENST00000394395
PAX8	RSPO2	TAF15
PBX1	RSPO3	TBCEL
PCM1	SDC4	TCEA1_ENST00000521604
PDE8B	SEC16A_NM_014866.1	TCF12
PDGFB	SEC31A	TCF3
PHF1	SEPT8	TCF7L2
PKD1L1	SFPQ	TECTA
PLA2R1	SIP1	TFE3
PLAG1	SLC22A1	TFG
PLXND1	SLC26A6	THRAP3
POU5F1	SLC34A2	TMCC1
PPARG	SLC45A3	TMPRSS2
PPFIBP1	SMARCA5	TP53
PRCC	SND1	TPM3

TPM3\_ENST00000368530  
TPM4  
TRIM24  
TRIM27  
TRIM33\_ENST00000358465  
UBE2L3  
USH1G  
USP6\_ENST00000250066  
VCL  
VTI1A  
WDFY2  
WIF1\_ENST00000286574  
WT1  
YWHAE\_ENST00000264335  
YY1  
ZC3H7B  
ZCCHC8  
ZNF384  
ZNF444\_ENST00000337080  
ZNF700  
ZNF703  
ZSCAN30

12. 如权利要求1所述的用途,其中所述多核苷酸探针的每种与距所述融合基因的断点不多于500个核苷酸的断点区域具有序列互补性。

13. 如权利要求1所述的用途,其中所述多核苷酸探针的每种与跨越所述融合基因中的断点的序列具有序列互补性。

14. 如权利要求1所述的用途,其中所述多核苷酸探针的每种具有小于500个核苷酸的长度。

15. 如权利要求1所述的用途,其中所述多核苷酸探针的每种具有在20和200个核苷酸之间的长度。

16. 如权利要求1所述的用途,其中所述多核苷酸探针的每种具有在80和160个核苷酸之间的长度。

17. 如权利要求1所述的用途,其中所述断点片段的每个具有在140个核苷酸和180个核苷酸之间的长度。

18. 如权利要求1所述的用途,其中所述多核苷酸探针偶联于固体支持物。

19. 如权利要求1所述的用途,其中所述探针集包括一种或更多种天然多核苷酸探针。

20. 如权利要求1所述的用途,其中所述多核苷酸探针包括与所述融合基因中包括的核

酸序列的断点区域杂交的至少一种多核苷酸探针、和与所述融合基因中包括的核酸序列的非断点区域杂交的至少一种天然多核苷酸探针。

21. 如权利要求1所述的用途,其中所述多核苷酸探针的每种提供对所述融合基因中包括的核酸序列的断点区域的至少50%覆盖。

22. 如权利要求1所述的用途,其中(d)包括向所述分离的多核苷酸附加包含具有不同的条形码序列的条形码的标签,以生成加标签的亲本多核苷酸。

23. 如权利要求22所述的用途,还包括扩增所述加标签的亲本多核苷酸以产生加标签的子代多核苷酸。

24. 如权利要求23所述的用途,还包括(i)将所述加标签的子代多核苷酸测序以产生序列读段,其中每个序列读段包括条形码序列和来源于所述分离的多核苷酸的指定的一种的序列,和(ii)至少基于所述条形码序列将所述序列读段分组为家族。

25. 如权利要求24所述的用途,还包括将被分组在每一个家族以内的所述序列读段进行比较以确定每一个家族的共有序列,其中每一个共有序列对应于所述加标签的亲本多核苷酸中的独特多核苷酸。

26. 一种用于捕获融合基因的断点片段的方法,所述方法用于非诊断目的,所述方法包括:

(a) 在足以进行以下的条件下将提供的生物样品与多核苷酸探针接触,所述多核苷酸探针包含一个或更多个锁核酸核苷酸:

i. 允许所述多核苷酸探针与所述断点片段之间的杂交以在混合物中提供探针捕获的多核苷酸,该多核苷酸探针与所述断点片段具有序列互补性并对所述融合基因具有比具有所述序列互补性且仅包含未修饰的核苷酸的多核苷酸更大的亲和力;和

ii. 允许从所述混合物富集或分离所述探针捕获的多核苷酸,其中所述多核苷酸探针与所述断点片段具有序列互补性,

(b) 洗脱所述探针捕获的多核苷酸,以从所述探针分离所述捕获的多核苷酸;以及

(c) 直接测序洗脱的多核苷酸或使用洗脱的多核苷酸用于产生测序文库,

其中所述生物样品包含或疑似包含无细胞核酸分子,该无细胞核酸分子包括所述融合基因的断点片段。

27. 如权利要求26所述的方法,其中所述多核苷酸探针包含多个锁核酸核苷酸,其中所述锁核酸核苷酸的至少两个间隔不多于30个核苷酸。

28. 如权利要求27所述的方法,其中所述锁核酸核苷酸的至少两个间隔不多于15个核苷酸。

29. 如权利要求26所述的方法,其中所述多核苷酸探针的至少50%的核苷酸是锁核酸核苷酸。

30. 如权利要求29所述的方法,其中所述多核苷酸探针的至少75%的核苷酸是锁核酸核苷酸。

31. 如权利要求26所述的方法,其中所述多核苷酸探针具有比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸高至少1℃的解链温度。

32. 如权利要求31所述的方法,其中所述解链温度高至少10℃。

33. 如权利要求26所述的方法,其中所述多核苷酸探针具有比具有与所述融合基因互

补的序列且仅包含未修饰的核苷酸的多核苷酸高至少2%的解链温度。

34. 如权利要求33所述的方法,其中所述解链温度高至少10%。

35. 如权利要求26所述的方法,其中所述融合基因为癌症融合基因。

36. 如权利要求26所述的方法,其中所述多核苷酸探针与选自下表的融合基因对的基因

ACBD6/RRP15	KMT2A/ELL
ACSL3_ENST00000357430/ETV1	KMT2A/EP300
ACTB/GLI1	KMT2A/EPS15
AGPAT5/MCPH1	KMT2A/FOXO3
AGTRAP/BRAF	KMT2A/FOXO4
AKAP9_ENST00000356239/BRAF	KMT2A/FRYL
ARFIP1/FHDC1	KMT2A/GAS7
ARID1A/MAST2_ENST00000361297	KMT2A/GMPS
ASPSCR1/TFE3	KMT2A/GPHN
ATG4C/FBXO38	KMT2A/KIAA0284_ENST00000414716
ATIC/ALK	KMT2A/KIAA1524
BBS9/PKD1L1	KMT2A/LASP1
BCR/ABL1	KMT2A/LPP
BCR/JAK2	KMT2A/MAPRE1
BRD3/NUTM1	KMT2A/MLLT1
BRD4_ENST00000263377/NUTM1	KMT2A/MLLT10
C2orf44/ALK	KMT2A/MLLT11
CANT1/ETV4	KMT2A/MLLT3
CARS/ALK	KMT2A/MLLT4_ENST00000392108
CCDC6/RET	KMT2A/MLLT6
CD74_ENST00000009530/NRG1	KMT2A/MYO1F
CD74_ENST00000009530/ROS1	KMT2A/NCKIPSD
CDH11/USP6_ENST00000250066	KMT2A/NRIP3
CDKN2D/WDFY2	KMT2A/PDS5A
CEP89/BRAF	KMT2A/PICALM
CHCHD7/PLAG1	KMT2A/PRRC1

CIC/DUX4L1	KMT2A/SARNP
CIC/FOXO4	KMT2A/SEPT2
CLCN6/BRAF	KMT2A/SEPT5
CLIP1/ROS1	KMT2A/SEPT6
CLTC/ALK	KMT2A/SEPT9_ENST00000427177
CLTC/TFE3	KMT2A/SH3GL1
CNBP/USP6_ENST00000250066	KMT2A/SORBS2
COL1A1/PDGFB	KMT2A/TET1
COL1A1/USP6_ENST00000250066	KMT2A/TOP3A
COL1A2/PLAG1	KMT2A/ZFYVE19
CRTC1/MAML2	KTN1/RET
CRTC3/MAML2	LIFR_ENST00000263409/PLAG1
CTAGE5/SIP1	LMNA/NTRK1_ENST00000392302
CTNNB1/PLAG1	LRIG3/ROS1
DCTN1/ALK	LSM14A/BRAF
DDX5_ENST00000540698/ETV4	MARK4/ERCC2
DNAJB1/PRKACA	MBOAT2/PRKCE
EIF3E/RSP02	MBTD1_ENST00000586178/CXorf67_ENST00000342995
EIF3K/CYP39A1	MEAF6/PHF1
EML4/ALK	MKRN1/BRAF
EPC1/PHF1	MSN/ALK
ERC1/RET	MYB_ENST00000341911/NFIB_ENST00000397581
ERC1/ROS1	MYO5A/ROS1
ERO1L/FERMT2	NAB2/STAT6
ESRP1/RAF1	NACC2/NTRK2
ETV6/ITPR2	NCOA4_ENST00000452682/RET
ETV6/JAK2	NDRG1/ERG
ETV6/NTRK3	NF1/ACCN1

EWSR1/ATF1	NFIA/EHF
EWSR1/CREB1	NFIX_ENST00000360105/MAST1_ENST00000251472
EWSR1/DDIT3	NONO/TFE3
EWSR1/ERG	NOTCH1_ENST00000277541/GABBR2
EWSR1/ETV1	NPM1/ALK
EWSR1/ETV4	NTN1/ACLY
EWSR1/FEV	NUP107/LGR5
EWSR1/FLI1	OMD/USP6_ENST00000250066
EWSR1/NFATC1	PAX3/FOXO1
EWSR1/NFATC2	PAX3/NCOA1
EWSR1/NR4A3	PAX3/NCOA2
EWSR1/PATZ1	PAX5/JAK2
EWSR1/PBX1	PAX7/FOXO1
EWSR1/POU5F1	PAX8/PPARG
EWSR1/SMARCA5	PCM1/JAK2
EWSR1/SP3	PCM1/RET
EWSR1/WT1	PLA2R1/RBMS1
EWSR1/YY1	PLXND1/TMCC1
EWSR1/ZNF384	PPFIBP1/ALK
EWSR1/ZNF444_ENST00000337080	PPFIBP1/ROS1
EZR/ROS1	PRCC/TFE3
FAM131B_ENST00000443739/BRAF	PRKAR1A/RET
FBXL18/RNF216	PTPRK/RSPO3
FCHSD1/BRAF	PWWP2A/ROS1
FGFR1/ZNF703	QKI/NTRK2
FGFR1_ENST00000447712/PLAG1	RAF1/DAZL
FGFR1_ENST00000447712/TACC1	RANBP2/ALK
FGFR3/BAIAP2L1	RBM14/PACS1

FGFR3/TACC3	RGS22/SYCP1
FN1/ALK	RNF130/BRAF
FUS/ATF1	SDC4/ROS1
FUS/CREB3L1	SEC16A_NM_014866.1/NOTCH1_ENST00000277541
FUS/CREB3L2	SEC31A/ALK
FUS/DDIT3	SEC31A/JAK2
FUS/ERG	SEPT8/AFF4
FUS/FEV	SFPQ/TFE3
GATM/BRAF	SLC22A1/CUTA
GMDS/PDE8B	SLC26A6/PRKAR2A
GNAI1/BRAF	SLC34A2/ROS1
GOLGA5/RET	SLC45A3/BRAF
GOPC/ROS1	SLC45A3/ELK4
GPBP1L1/MAST2_ENST00000361297	SLC45A3/ERG
HACL1/RAF1	SLC45A3/ETV1
HAS2/PLAG1	SLC45A3/ETV5_ENST00000306376
HERPUD1/BRAF	SND1/BRAF
HEY1/NCOA2	SQSTM1/ALK
HIP1/ALK	SRGAP3/RAF1
HLA-A/ROS1	SS18/SSX1
HMGA2/ALDH2_ENST00000261733	SS18/SSX2
HMGA2/CCNB1IP1	SS18/SSX4
HMGA2/COX6C	SS18L1/SSX1
HMGA2/EBF1	SSBP2/JAK2
HMGA2/FHIT_ENST00000476844	SSH2/SUZ12
HMGA2/LHFP	STIL/TAL1
HMGA2/LPP	STRN/ALK
HMGA2/NFIB_ENST00000397581	SUSD1/ROD1

HMGA2/RAD51B	TADA2A_ENST00000394395/MAST1_ENST00000251472
HMGA2/WIF1_ENST00000286574	TAF15/NR4A3
HN1/USH1G	TCEA1_ENST00000521604/PLAG1
HNRNPA2B1/ETV1	TCF12/NR4A3
HOOK3/RET	TCF3/PBX1
IL6R/ATP8B2	TECTA/TBCEL
INTS4/GAB2	TFG/ALK
IRF2BP2/CDX1	TFG/NR4A3
JAZF1/PHF1	TFG/NTRK1_ENST00000392302
JAZF1/SUZ12	THRAP3/USP6_ENST00000250066
KIAA1549/BRAF	TMPRSS2/ERG
KIAA1598/ROS1	TMPRSS2/ETV1
KIF5B/ALK	TMPRSS2/ETV4
KIF5B/RET	TMPRSS2/ETV5_ENST00000306376
KLC1/ALK	TP53/NTRK1_ENST00000392302
KLK2/ETV1	TPM3/ALK
KLK2/ETV4	TPM3/NTRK1_ENST00000392302
KMT2A/ABI1	TPM3/ROS1
KMT2A/ABI2	TPM3_ENST00000368530/ROS1
KMT2A/ACTN4	TPM4/ALK
KMT2A/AFF1	TRIM24/RET
KMT2A/AFF3	TRIM27/RET
KMT2A/AFF4	TRIM33_ENST00000358465/RET
KMT2A/ARHGAP26	UBE2L3/KRAS
KMT2A/ARHGEF12	VCL/ALK
KMT2A/BTBD18	VTI1A/TCF7L2
KMT2A/CASC5	YWHAE_ENST00000264335/FAM22A_ENST00000381707
KMT2A/CASP8AP2	YWHAE_ENST00000264335/NUTM2B
KMT2A/CBL	ZC3H7B/BCOR_ENST00000378444
KMT2A/CREBBP	ZCCHC8/ROS1
KMT2A/CT45A2	ZNF700/MAST1_ENST00000251472
KMT2A/DAB2IP	ZSCAN30/BRAF
KMT2A/EEFSEC	



编号	组群 基因	配偶体 基因	编号	组群 基因	配偶体 基因	编号	组群 基因	配偶体 基因
1	ALK	ATIC	44	RET	ERC1	87	FGFR3	TACC3
2	ALK	C2orf44	45	RET	FKBP15	88	FGFR3	TNIP2
3	ALK	C2orf61	46	RET	GOLGA5	89	FGFR3	TPRG1
4	ALK	CAMKMT	47	RET	HOOK3	90	FGFR3	WHSC1
5	ALK	CARS	48	RET	KIF5B	91	ROS1	CCDC6
6	ALK	CENPF	49	RET	KTN1	92	ROS1	CD72
7	ALK	CLTC	50	RET	NCOA4	93	ROS1	CD74
8	ALK	CLIP4	51	RET	PCM1	94	ROS1	CEP85L
9	ALK	COL4A3	52	RET	PRKAR1A	95	ROS1	CLIP1
10	ALK	DCTN1	53	RET	SPECC1L	96	ROS1	CLTC
11	ALK	EIF2AK3	54	RET	SQSTM1	97	ROS1	ERC1
12	ALK	EML4	55	RET	TBL1XR1	98	ROS1	EZR
13	ALK	FANCL	56	RET	TRIM24	99	ROS1	HLA-A
14	ALK	FN1	57	RET	TRIM27	100	ROS1	KDELR2
15	ALK	GTF2IRD1	58	RET	TRIM33	101	ROS1	KIAA1598
16	ALK	HIP1	59	FGFR2	AFF3	102	ROS1	LRIG3

17	ALK	KCNQ5	60	FGFR2	AHCYL1	103	ROS1	MYO5A
18	ALK	KIF5B	61	FGFR2	BICC1	104	ROS1	PPFIBP1
19	ALK	KLC1	62	FGFR2	C10ORF68	105	ROS1	PWWP2A
20	ALK	MPRIIP	63	FGFR2	CASP7	106	ROS1	RUNX1
21	ALK	MSN	64	FGFR2	CCAR2	107	ROS1	SDC4
22	ALK	MYH9	65	FGFR2	CCDC6	108	ROS1	SLC34A2
23	ALK	NPM1	66	FGFR2	CD44	109	ROS1	SLC35F1
24	ALK	PPFIBP1	67	FGFR2	CIT	110	ROS1	TFG
25	ALK	PPM1B	68	FGFR2	KIAA1598	111	ROS1	TMEM106B
26	ALK	PPP1CB	69	FGFR2	MGEA5	112	ROS1	TPM3
27	ALK	PRKAR1A	70	FGFR2	NCALD	113	ROS1	TTC28
28	ALK	PTPN3	71	FGFR2	OFD1	114	ROS1	YWHAE
29	ALK	RANBP2	72	FGFR2	PPAPDC1A	115	ROS1	ZCCHC8
30	ALK	RNF213	73	FGFR2	PPHLN1	116	NTRK1	CD74
31	ALK	SEC31A	74	FGFR2	RASAL2	117	NTRK1	IRF2BP2
32	ALK	SOCS5	75	FGFR2	SLC45A3	118	NTRK1	LMNA

33	ALK	SQSTM1	76	FGFR2	TACC2	119	NTRK1	MPRIIP
34	ALK	STK32B	77	FGFR2	TACC3	120	NTRK1	NFASC
35	ALK	STRN	78	FGFR2	TXLNB	121	NTRK1	PEAR1
36	ALK	TFG	79	FGFR2	USP10	122	NTRK1	PLEKHA6
37	ALK	TPM1	80	FGFR2	VCL	123	NTRK1	QKI
38	ALK	TPM3	81	FGFR3	ADD1	124	NTRK1	RABGAP1L
39	ALK	TPM4	82	FGFR3	AES	125	NTRK1	SQSTM1
40	ALK	VCL	83	FGFR3	BAIAP2L1	126	NTRK1	SSBP2
41	RET	AFAP1	84	FGFR3	ELAVL3	127	NTRK1	TFG
42	RET	AKAP13	85	FGFR3	ETV6	128	NTRK1	TP53
43	RET	CCDC6	86	FGFR3	JAKMIP1	129	NTRK1	TPM3

或选自下表的两种或更多种基因之间的融合基因具有序列互补性

ACBD6	CCNB1IP1
ACCN1	CD74_ENST00000009530
ACLY	CDH11
ACSL3_ENST00000357430	CDKN2D
ACTB	CDX1
AFF4	CEP89
AGPAT5	CHCHD7
AGTRAP	CIC
AKAP9_ENST00000356239	CLCN6
ALDH2_ENST00000261733	CLIP1
ALK	CLTC
ARFIP1	CNBP
ARID1A	COL1A1
ASPSCR1	COL1A2
ATF1	COX6C
ATG4C	CREB1
ATIC	CREB3L1
ATP8B2	CREB3L2
BAIAP2L1	CRTC1
BBS9	CRTC3
BCOR_ENST00000378444	CTAGE5
BCR	CTNNB1
BRAF	CUTA
BRD3	CXorf67_ENST00000342995
BRD4_ENST00000263377	CYP39A1
C2orf44	DAZL
CANT1	DCTN1
CARS	DDIT3
CCDC6	DDX5_ENST00000540698
	DUX4L1
	EBF1
	EHF

EIF3E	FUS	KTN1
EIF3K	GAB2	LGR5
ELK4	GABBR2	LHFP
EML4	GATM	LIFR_ENST00000263409
EPC1	GLI1	LMNA
ERC1	GMDS	LPP
ERCC2	GNAI1	LRIG3
ERG	GOLGA5	LSM14A
ERO1L	GOPC	MAML2
ESRP1	GPBP1L1	MARK4
ETV1	HACL1	MAST1_ENST00000251472
ETV4	HAS2	MAST2_ENST00000361297
ETV5_ENST00000306376	HERPUD1	MBOAT2
ETV6	HEY1	MBTD1_ENST00000586178
EWSR1	HIP1	MCPH1
EZR	HLA-A	MEAF6
FAM131B_ENST00000443739	HMGA2	MKRN1
FAM22A_ENST00000381707	HN1	MSN
FBXL18	HNRNPA2B1	MYB_ENST00000341911
FBXO38	HOOK3	MYO5A
FCHSD1	IL6R	NAB2
FERMT2	INTS4	NACC2
FEV	IRF2BP2	NCOA1
FGFR1	ITPR2	NCOA2
FGFR1_ENST00000447712	JAK2	NCOA4_ENST00000452682
FGFR3	JAZF1	NDRG1
FHDC1	KIAA1549	NF1
FHIT_ENST00000476844	KIAA1598	NFATC1
FLI1	KIF5B	NFATC2
FN1	KLC1	NFIA
FOXO1	KLK2	NFIB_ENST00000397581
FOXO4	KRAS	NFIX_ENST00000360105

NONO		SP3
NOTCH1_ENSTC0000277541	PRKAR2A	SQSTM1
NPM1	PRKCE	SRGAP3
NR4A3	PTPRK	SS18
NRG1	PWWP2A	SS18L1
NTN1	QKI	SSBP2
NTRK1_ENST00000392302	RAD51B	SSH2
NTRK2	RAF1	SSX1
NTRK3	RANBP2	SSX2
NUP107	RBM14	SSX4
NUTM1	RBMS1	STAT6
NUTM2B	RET	STRN
OMD	RGS22	SUSD1
PACS1	RNF130	SUZ12
PATZ1	RNF216	SYCP1
PAX3	ROD1	TACC1
PAX5	ROS1	TACC3
PAX7	RRP15	TADA2A_ENST00000394395
PAX8	RSPO2	TAF15
PBX1	RSPO3	TBCEL
PCM1	SDC4	TCEA1_ENST00000521604
PDE8B	SEC16A_NM_014866.1	TCF12
PDGFB	SEC31A	TCF3
PHF1	SEPT8	TCF7L2
PKD1L1	SFPQ	TECTA
PLA2R1	SIP1	TFE3
PLAG1	SLC22A1	TFG
PLXND1	SLC26A6	THRAP3
POU5F1	SLC34A2	TMCC1
PPARG	SLC45A3	TMPRSS2
PPFIBP1	SMARCA5	TP53
PRCC	SND1	TPM3

TPM3\_ENST00000368530  
TPM4  
TRIM24  
TRIM27  
TRIM33\_ENST00000358465  
UBE2L3  
USH1G  
USP6\_ENST00000250066  
VCL  
VTI1A  
WDFY2  
WIF1\_ENST00000286574  
WT1  
YWHAE\_ENST00000264335  
YY1  
ZC3H7B  
ZCCHC8  
ZNF384  
ZNF444\_ENST00000337080  
ZNF700  
ZNF703  
ZSCAN30

37. 如权利要求26所述的方法,其中所述多核苷酸探针与距所述融合基因的断点不多于500个核苷酸的断点区域具有序列互补性。

38. 如权利要求26所述的方法,其中所述多核苷酸探针与跨越所述融合基因中的断点的序列具有序列互补性。

39. 如权利要求26所述的方法,其中所述多核苷酸探针具有小于500个核苷酸的长度。

40. 如权利要求26所述的方法,其中所述多核苷酸探针具有在20和200个核苷酸之间的长度。

41. 如权利要求26所述的方法,其中所述多核苷酸探针具有在80和160个核苷酸之间的长度。

42. 如权利要求26所述的方法,其中所述断点片段的每个具有在140个核苷酸和180个核苷酸之间的长度。

43. 如权利要求26所述的方法,其中所述多核苷酸探针偶联于固体支持物。

44. 如权利要求26所述的方法,其中所述多核苷酸探针提供对所述融合基因中包括的核酸序列的断点区域的至少50%覆盖。

45. 一种用于富集样品的包含融合基因的断点的多核苷酸的方法,所述方法用于非诊

断目的,所述方法包括:

(a) 将探针集与包括无细胞DNA或片段化的基因组DNA多核苷酸的混合物在杂交条件下接触以产生探针捕获的多核苷酸,所述探针集包含多种多核苷酸探针,所述多种多核苷酸探针中的每种探针具有(i)与融合基因的序列互补性和(ii)比具有与所述融合基因互补的序列并仅包含未修饰的核苷酸的多核苷酸更大的对所述融合基因的亲和力;和

(b) 从所述混合物分离所述探针捕获的多核苷酸以产生富集包含所述融合基因的断点片段的多核苷酸的样品。

46. 如权利要求45所述的方法,其中高亲和力多核苷酸包含一个或更多个锁核酸核苷酸。

47. 如权利要求45所述的方法,还包括从所述探针分离捕获的多核苷酸。

48. 如权利要求45所述的方法,还包括将所分离的多核苷酸测序。

49. 如权利要求45所述的方法,其中所述多种多核苷酸探针的每种包含一个或更多个锁核酸核苷酸。

50. 如权利要求49所述的方法,其中所述多种多核苷酸探针的每种包含多个锁核酸核苷酸,其中所述锁核酸核苷酸的至少两个间隔不多于30个核苷酸。

51. 如权利要求50所述的方法,其中所述锁核酸核苷酸的所述至少两个间隔不多于15个核苷酸。

52. 如权利要求45所述的方法,其中所述多种多核苷酸探针的至少一个子集的每种多核苷酸探针的至少50%核苷酸是锁核酸核苷酸。

53. 如权利要求52所述的方法,其中所述多种多核苷酸探针的至少一个子集的每种多核苷酸探针的至少75%核苷酸是锁核酸核苷酸。

54. 如权利要求45所述的方法,其中所述多种多核苷酸探针的每种具有比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸高至少1℃的解链温度。

55. 如权利要求54所述的方法,其中所述解链温度高至少10℃。

56. 如权利要求45所述的方法,其中所述多种多核苷酸探针的每种具有比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸高至少2%的解链温度。

57. 如权利要求56所述的方法,其中所述解链温度高至少10%。

58. 如权利要求45所述的方法,其中所述融合基因为癌症融合基因。

59. 如权利要求45所述的方法,其中所述多种多核苷酸探针的每种与选自下表的融合基因对的基因



ACBD6/RRP15	KMT2A/ELL
ACSL3_ENST00000357430/ETV1	KMT2A/EP300
ACTB/GLI1	KMT2A/EP315
AGPAT5/MCPH1	KMT2A/FOXO3
AGTRAP/BRAF	KMT2A/FOXO4
AKAP9_ENST00000356239/BRAF	KMT2A/FRYL
ARFIP1/FHDC1	KMT2A/GAS7
ARID1A/MAST2_ENST00000361297	KMT2A/GMPS
ASPSCR1/TFE3	KMT2A/GPHN
ATG4C/FBXO38	KMT2A/KIAA0284_ENST00000414716
ATIC/ALK	KMT2A/KIAA1524
BBS9/PKD1L1	KMT2A/LASP1
BCR/ABL1	KMT2A/LPP
BCR/JAK2	KMT2A/MAPRE1
BRD3/NUTM1	KMT2A/MLLT1
BRD4_ENST00000263377/NUTM1	KMT2A/MLLT10
C2orf44/ALK	KMT2A/MLLT11
CANT1/ETV4	KMT2A/MLLT3
CARS/ALK	KMT2A/MLLT4_ENST00000392108
CCDC6/RET	KMT2A/MLLT6
CD74_ENST00000009530/NRG1	KMT2A/MYO1F
CD74_ENST00000009530/ROS1	KMT2A/NCKIPSD
CDH11/USP6_ENST00000250066	KMT2A/NRIP3
CDKN2D/WDFY2	KMT2A/PDS5A
CEP89/BRAF	KMT2A/PICALM
CHCHD7/PLAG1	KMT2A/PRRC1

CIC/DUX4L1	KMT2A/SARNP
CIC/FOXO4	KMT2A/SEPT2
CLCN6/BRAF	KMT2A/SEPT5
CLIP1/ROS1	KMT2A/SEPT6
CLTC/ALK	KMT2A/SEPT9_ENST00000427177
CLTC/TFE3	KMT2A/SH3GL1
CNBP/USP6_ENST00000250066	KMT2A/SORBS2
COL1A1/PDGFB	KMT2A/TET1
COL1A1/USP6_ENST00000250066	KMT2A/TOP3A
COL1A2/PLAG1	KMT2A/ZFYVE19
CRTC1/MAML2	KTN1/RET
CRTC3/MAML2	LIFR_ENST00000263409/PLAG1
CTAGE5/SIP1	LMNA/NTRK1_ENST00000392302
CTNNB1/PLAG1	LRIG3/ROS1
DCTN1/ALK	LSM14A/BRAF
DDX5_ENST00000540698/ETV4	MARK4/ERCC2
DNAJB1/PRKACA	MBOAT2/PRKCE
EIF3E/RSP02	MBTD1_ENST00000586178/CXorf67_ENST00000342995
EIF3K/CYP39A1	MEAF6/PHF1
EML4/ALK	MKRN1/BRAF
EPC1/PHF1	MSN/ALK
ERC1/RET	MYB_ENST00000341911/NFIB_ENST00000397581
ERC1/ROS1	MYO5A/ROS1
ERO1L/FERMT2	NAB2/STAT6
ESRP1/RAF1	NACC2/NTRK2
ETV6/ITPR2	NCOA4_ENST00000452682/RET
ETV6/JAK2	NDRG1/ERG
ETV6/NTRK3	NF1/ACCN1

EWSR1/ATF1	NFIA/EHF
EWSR1/CREB1	NFIX_ENST00000360105/MAST1_ENST00000251472
EWSR1/DDIT3	NONO/TFE3
EWSR1/ERG	NOTCH1_ENST00000277541/GABBR2
EWSR1/ETV1	NPM1/ALK
EWSR1/ETV4	NTN1/ACLY
EWSR1/FEV	NUP107/LGR5
EWSR1/FLI1	OMD/USP6_ENST00000250066
EWSR1/NFATC1	PAX3/FOXO1
EWSR1/NFATC2	PAX3/NCOA1
EWSR1/NR4A3	PAX3/NCOA2
EWSR1/PATZ1	PAX5/JAK2
EWSR1/PBX1	PAX7/FOXO1
EWSR1/POU5F1	PAX8/PPARG
EWSR1/SMARCA5	PCM1/JAK2
EWSR1/SP3	PCM1/RET
EWSR1/WT1	PLA2R1/RBMS1
EWSR1/YY1	PLXND1/TMCC1
EWSR1/ZNF384	PPFIBP1/ALK
EWSR1/ZNF444_ENST00000337080	PPFIBP1/ROS1
EZR/ROS1	PRCC/TFE3
FAM131B_ENST00000443739/BRAF	PRKAR1A/RET
FBXL18/RNF216	PTPRK/RSP03
FCHSD1/BRAF	PWWP2A/ROS1
FGFR1/ZNF703	QKI/NTRK2
FGFR1_ENST00000447712/PLAG1	RAF1/DAZL
FGFR1_ENST00000447712/TACC1	RANBP2/ALK
FGFR3/BAIAP2L1	RBM14/PACS1

FGFR3/TACC3	RGS22/SYCP1
FN1/ALK	RNF130/BRAF
FUS/ATF1	SDC4/ROS1
FUS/CREB3L1	SEC16A_NM_014866.1/NOTCH1_ENST00000277541
FUS/CREB3L2	SEC31A/ALK
FUS/DDIT3	SEC31A/JAK2
FUS/ERG	SEPT8/AFF4
FUS/FEV	SFPQ/TFE3
GATM/BRAF	SLC22A1/CUTA
GMDS/PDE8B	SLC26A6/PRKAR2A
GNAI1/BRAF	SLC34A2/ROS1
GOLGA5/RET	SLC45A3/BRAF
GOPC/ROS1	SLC45A3/ELK4
GPBP1L1/MAST2_ENST00000361297	SLC45A3/ERG
HACL1/RAF1	SLC45A3/ETV1
HAS2/PLAG1	SLC45A3/ETV5_ENST00000306376
HERPUD1/BRAF	SND1/BRAF
HEY1/NCOA2	SQSTM1/ALK
HIP1/ALK	SRGAP3/RAF1
HLA-A/ROS1	SS18/SSX1
HMGA2/ALDH2_ENST00000261733	SS18/SSX2
HMGA2/CCNB1IP1	SS18/SSX4
HMGA2/COX6C	SS18L1/SSX1
HMGA2/EBF1	SSBP2/JAK2
HMGA2/FHIT_ENST00000476844	SSH2/SUZ12
HMGA2/LHFP	STIL/TAL1
HMGA2/LPP	STRN/ALK
HMGA2/NFIB_ENST00000397581	SUSD1/ROD1

HMGA2/RAD51B	TADA2A_ENST00000394395/MAST1_ENST00000251472
HMGA2/WIF1_ENST00000286574	TAF15/NR4A3
HN1/USH1G	TCEA1_ENST00000521604/PLAG1
HNRNPA2B1/ETV1	TCF12/NR4A3
HOOK3/RET	TCF3/PBX1
IL6R/ATP8B2	TECTA/TBCEL
INTS4/GAB2	TFG/ALK
IRF2BP2/CDX1	TFG/NR4A3
JAZF1/PHF1	TFG/NTRK1_ENST00000392302
JAZF1/SUZ12	THRAP3/USP6_ENST00000250066
KIAA1549/BRAF	TMPRSS2/ERG
KIAA1598/ROS1	TMPRSS2/ETV1
KIF5B/ALK	TMPRSS2/ETV4
KIF5B/RET	TMPRSS2/ETV5_ENST00000306376
KLC1/ALK	TP53/NTRK1_ENST00000392302
KLK2/ETV1	TPM3/ALK
KLK2/ETV4	TPM3/NTRK1_ENST00000392302
KMT2A/ABI1	TPM3/ROS1
KMT2A/ABI2	TPM3_ENST00000368530/ROS1
KMT2A/ACTN4	TPM4/ALK
KMT2A/AFF1	TRIM24/RET
KMT2A/AFF3	TRIM27/RET
KMT2A/AFF4	TRIM33_ENST00000358465/RET
KMT2A/ARHGAP26	UBE2L3/KRAS
KMT2A/ARHGEF12	VCL/ALK
KMT2A/BTBD18	VTI1A/TCF7L2
KMT2A/CASC5	YWHAE_ENST00000264335/FAM22A_ENST00000381707
KMT2A/CASP8AP2	YWHAE_ENST00000264335/NUTM2B
KMT2A/CBL	ZC3H7B/BCOR_ENST00000378444
KMT2A/CREBBP	ZCCHC8/ROS1
KMT2A/CT45A2	ZNF700/MAST1_ENST00000251472
KMT2A/DAB2IP	ZSCAN30/BRAF
KMT2A/EEFSEC	

编号	组群 基因	配偶体 基因	编号	组群 基因	配偶体 基因	编号	组群 基因	配偶体 基因
1	ALK	ATIC	44	RET	ERC1	87	FGFR3	TACC3
2	ALK	C2orf44	45	RET	FKBP15	88	FGFR3	TNIP2
3	ALK	C2orf61	46	RET	GOLGA5	89	FGFR3	TPRG1
4	ALK	CAMKMT	47	RET	HOOK3	90	FGFR3	WHSC1
5	ALK	CARS	48	RET	KIF5B	91	ROS1	CCDC6
6	ALK	CENPF	49	RET	KTN1	92	ROS1	CD72
7	ALK	CLTC	50	RET	NCOA4	93	ROS1	CD74
8	ALK	CLIP4	51	RET	PCM1	94	ROS1	CEP85L
9	ALK	COL4A3	52	RET	PRKAR1A	95	ROS1	CLIP1
10	ALK	DCTN1	53	RET	SPECC1L	96	ROS1	CLTC
11	ALK	EIF2AK3	54	RET	SQSTM1	97	ROS1	ERC1
12	ALK	EML4	55	RET	TBL1XR1	98	ROS1	EZR
13	ALK	FANCL	56	RET	TRIM24	99	ROS1	HLA-A
14	ALK	FN1	57	RET	TRIM27	100	ROS1	KDEL2
15	ALK	GTF2IRD1	58	RET	TRIM33	101	ROS1	KIAA1598
16	ALK	HIP1	59	FGFR2	AFF3	102	ROS1	LRIG3

17	ALK	KCNQ5	60	FGFR2	AHCYL1	103	ROS1	MYO5A
18	ALK	KIF5B	61	FGFR2	BICC1	104	ROS1	PPFIBP1
19	ALK	KLC1	62	FGFR2	C10ORF68	105	ROS1	PWWP2A
20	ALK	MPRIP	63	FGFR2	CASP7	106	ROS1	RUNX1
21	ALK	MSN	64	FGFR2	CCAR2	107	ROS1	SDC4
22	ALK	MYH9	65	FGFR2	CCDC6	108	ROS1	SLC34A2
23	ALK	NPM1	66	FGFR2	CD44	109	ROS1	SLC35F1
24	ALK	PPFIBP1	67	FGFR2	CIT	110	ROS1	TFG
25	ALK	PPM1B	68	FGFR2	KIAA1598	111	ROS1	TMEM106B
26	ALK	PPP1CB	69	FGFR2	MGEA5	112	ROS1	TPM3
27	ALK	PRKAR1A	70	FGFR2	NCALD	113	ROS1	TTC28
28	ALK	PTPN3	71	FGFR2	OFD1	114	ROS1	YWHAE
29	ALK	RANBP2	72	FGFR2	PPAPDC1A	115	ROS1	ZCCHC8
30	ALK	RNF213	73	FGFR2	PPHLN1	116	NTRK1	CD74
31	ALK	SEC31A	74	FGFR2	RASAL2	117	NTRK1	IRF2BP2
32	ALK	SOCS5	75	FGFR2	SLC45A3	118	NTRK1	LMNA

33	ALK	SQSTM1	76	FGFR2	TACC2	119	NTRK1	MPRIIP
34	ALK	STK32B	77	FGFR2	TACC3	120	NTRK1	NFASC
35	ALK	STRN	78	FGFR2	TXLNB	121	NTRK1	PEAR1
36	ALK	TFG	79	FGFR2	USP10	122	NTRK1	PLEKHA6
37	ALK	TPM1	80	FGFR2	VCL	123	NTRK1	QKI
38	ALK	TPM3	81	FGFR3	ADD1	124	NTRK1	RABGAP1L
39	ALK	TPM4	82	FGFR3	AES	125	NTRK1	SQSTM1
40	ALK	VCL	83	FGFR3	BAIAP2L1	126	NTRK1	SSBP2
41	RET	AFAP1	84	FGFR3	ELAVL3	127	NTRK1	TFG
42	RET	AKAP13	85	FGFR3	ETV6	128	NTRK1	TP53
43	RET	CCDC6	86	FGFR3	JAKMIP1	129	NTRK1	TPM3

或选自下表的两种或更多种基因之间的融合基因具有序列互补性



ACBD6	CCNB1IP1
ACCN1	CD74_ENST00000009530
ACLY	CDH11
ACSL3_ENST00000357430	CDKN2D
ACTB	CDX1
AFF4	CEP89
AGPAT5	CHCHD7
AGTRAP	CIC
AKAP9_ENST00000356239	CLCN6
ALDH2_ENST00000261733	CLIP1
ALK	CLTC
ARFIP1	CNBP
ARID1A	COL1A1
ASPSCR1	COL1A2
ATF1	COX6C
ATG4C	CREB1
ATIC	CREB3L1
ATP8B2	CREB3L2
BAIAP2L1	CRTC1
BBS9	CRTC3
BCOR_ENST00000378444	CTAGE5
BCR	CTNNB1
BRAF	CUTA
BRD3	CXorf67_ENST00000342995
BRD4_ENST00000263377	CYP39A1
C2orf44	DAZL
CANT1	DCTN1
CARS	DDIT3
CCDC6	DDX5_ENST00000540698
	DUX4L1
	EBF1
	EHF

EIF3E	FUS	KTN1
EIF3K	GAB2	LGR5
ELK4	GABBR2	LHFP
EML4	GATM	LIFR_ENST00000263409
EPC1	GLI1	LMNA
ERC1	GMDS	LPP
ERCC2	GNAI1	LRIG3
ERG	GOLGA5	LSM14A
ERO1L	GOPC	MAML2
ESRP1	GPBP1L1	MARK4
ETV1	HACL1	MAST1_ENST00000251472
ETV4	HAS2	MAST2_ENST00000361297
ETV5_ENST00000306376	HERPUD1	MBOAT2
ETV6	HEY1	MBTD1_ENST00000586178
EWSR1	HIP1	MCPH1
EZR	HLA-A	MEAF6
FAM131B_ENST00000443739	HMGA2	MKRN1
FAM22A_ENST00000381707	HN1	MSN
FBXL18	HNRNPA2B1	MYB_ENST00000341911
FBXO38	HOOK3	MYO5A
FCHSD1	IL6R	NAB2
FERMT2	INTS4	NACC2
FEV	IRF2BP2	NCOA1
FGFR1	ITPR2	NCOA2
FGFR1_ENST00000447712	JAK2	NCOA4_ENST00000452682
FGFR3	JAZF1	NDRG1
FHDC1	KIAA1549	NF1
FHIT_ENST00000476844	KIAA1598	NFATC1
FLI1	KIF5B	NFATC2
FN1	KLC1	NFIA
FOXO1	KLK2	NFIB_ENST00000397581
FOXO4	KRAS	NFIX_ENST00000360105

NONO		SP3
NOTCH1_ENSTC0000277541	PRKAR2A	SQSTM1
NPM1	PRKCE	SRGAP3
NR4A3	PTPRK	SS18
NRG1	PWWP2A	SS18L1
NTN1	QKI	SSBP2
NTRK1_ENST00000392302	RAD51B	SSH2
NTRK2	RAF1	SSX1
NTRK3	RANBP2	SSX2
NUP107	RBM14	SSX4
NUTM1	RBMS1	STAT6
NUTM2B	RET	STRN
OMD	RGS22	SUSD1
PACS1	RNF130	SUZ12
PATZ1	RNF216	SYCP1
PAX3	ROD1	TACC1
PAX5	ROS1	TACC3
PAX7	RRP15	TADA2A_ENST00000394395
PAX8	RSPO2	TAF15
PBX1	RSPO3	TBCEL
PCM1	SDC4	TCEA1_ENST00000521604
PDE8B	SEC16A_NM_014866.1	TCF12
PDGFB	SEC31A	TCF3
PHF1	SEPT8	TCF7L2
PKD1L1	SFPQ	TECTA
PLA2R1	SIP1	TFE3
PLAG1	SLC22A1	TFG
PLXND1	SLC26A6	THRAP3
POU5F1	SLC34A2	TMCC1
PPARG	SLC45A3	TMPRSS2
PPFIBP1	SMARCA5	TP53
PRCC	SND1	TPM3

TPM3\_ENST00000368530  
TPM4  
TRIM24  
TRIM27  
TRIM33\_ENST00000358465  
UBE2L3  
USH1G  
USP6\_ENST00000250066  
VCL  
VTI1A  
WDFY2  
WIF1\_ENST00000286574  
WT1  
YWHAE\_ENST00000264335  
YY1  
ZC3H7B  
ZCCHC8  
ZNF384  
ZNF444\_ENST00000337080  
ZNF700  
ZNF703  
ZSCAN30

60. 如权利要求45所述的方法,其中所述多种多核苷酸探针的每种与距所述融合基因的断点不多于500个核苷酸的断点区域具有序列互补性。

61. 如权利要求45所述的方法,其中所述多种多核苷酸探针的每种与跨越所述融合基因中的断点的序列具有序列互补性。

62. 如权利要求45所述的方法,其中所述多种多核苷酸探针的每种具有小于500个核苷酸的长度。

63. 如权利要求45所述的方法,其中所述多种多核苷酸探针的每种具有在20和200个核苷酸之间的长度。

64. 如权利要求45所述的方法,其中所述多种多核苷酸探针的每种具有在80和160个核苷酸之间的长度。

65. 如权利要求45所述的方法,其中所述断点片段的每个具有在140个核苷酸和180个核苷酸之间的长度。

66. 如权利要求45所述的方法,其中所述多种多核苷酸探针偶联于固体支持物。

67. 如权利要求45所述的方法,其中所述探针集包括一种或更多种天然多核苷酸探针。

68. 如权利要求45所述的方法,其中所述多种多核苷酸探针包括与所述融合基因中包

括的核酸序列的断点区域杂交的至少一种多核苷酸探针、和与所述融合基因中包括的核酸序列的非断点区域杂交的至少一种天然多核苷酸探针。

69. 如权利要求45所述的方法, 其中所述多种多核苷酸探针的每种提供对所述融合基因中包括的核酸序列的断点区域的至少50%覆盖。

## 用于捕获融合基因的锁核酸

[0001] 交叉引用

[0002] 本申请要求2015年7月21日提交的美国临时专利申请第62/195,280号的优先权，其通过引用全文并入本文。

[0003] 背景

[0004] 基因融合事件是使得基因组中至少两个基因的以前分开的部分在一起的染色体重排。基因融合事件可产生癌症融合基因，其中两个或更多个基因的异常并置(juxtaposition)可以编码融合蛋白，或者一个基因的调节元件可驱动癌基因的异常表达。检测这种癌症融合基因可能是困难的。断点片段不太可能与不包含断点的片段在相同程度上与探针杂交。因此，富集断点片段的杂交方法可能缺乏效力。

[0005] 融合基因是癌细胞中发现的一种体细胞突变形式。检测这种融合基因的能力可用于癌症的诊断和监测。

[0006] 已知癌症中发现的融合基因包括，例如，以下：结肠癌中的APIP/SLC1A2、胰腺癌中的ATG7/RAF1、星形细胞瘤中的BCL6/RAF1、慢性骨髓性白血病中的BCR-ABL、中线癌(midline carcinomas)中的BRD4-NUT、血管肉瘤中的CEP85L/ROS1、乳腺癌中的CLTC/VMP1、肺癌中的ELM4-ALK、黑素瘤中的EWSR1/CREM、T细胞急性淋巴母细胞白血病中的FAM133B/CDK6、低级星形细胞瘤中的KIAA1549-BRAF(在7q34)、粘液表皮样癌中的MECT1-MAML2、滤泡性甲状腺癌中的PAX8-PPARG、乳头状甲状腺癌中的RET-NTRK1、乳腺癌中的SEC16A-NOTCH1、低级星形细胞瘤中的SRGAP3-RAF1(在3p25)、肾癌中的TFE3-TFEB。

[0007] 断点可以发生在参与基因融合的基因的许多不同位置。这种断点可以在基因的某些部分聚簇。

[0008] 检测基因融合的一种方法是通过FISH(荧光原位杂交)。另一种是通过脱氧核糖核酸(DNA)测序。

[0009] 概述

[0010] 在本文认识到对富集断点片段以检测和表征癌症融合基因的方法的需求。

[0011] 本公开内容提供了检测融合基因的方法，其可用于检测疾病诸如癌症。本文提供了用于富集断点片段的方法，诸如以检测和表征融合基因，所述融合基因可以与疾病诸如癌症相关。

[0012] 在一个方面，本公开内容提供了用于为具有或疑似具有癌症的受试者提供诊断或治疗干预的方法，所述方法包括(a)提供来自受试者的包含无细胞核酸分子的生物样品；(b)将来自所述生物样品的无细胞核酸分子与探针集在足以产生探针捕获的多核苷酸的杂交条件下接触，该探针集包含多种多核苷酸探针，其中所述多种多核苷酸探针的每种具有(i)与融合基因的序列互补性和(ii)比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸更大的对所述融合基因的亲和力；(c)从混合物分离所述探针捕获的多核苷酸以产生富集包含所述融合基因的断点片段的分离的多核苷酸的样品；(d)将所分离的多核苷酸测序以产生序列；(e)基于所述序列检测包含融合基因的断点的多核苷酸；和(f)基于断点片段的检测提供诊断或治疗干预。

[0013] 在一些实施方案中,所述多种多核苷酸探针的每种包含一个或更多个锁核酸(LNA)核苷酸。在一些实施方案中,所述多种多核苷酸探针的每种包含多个LNA核苷酸,其中所述LNA核苷酸的至少两个间隔不多于30个核苷酸。在一些实施方案中,所述LNA核苷酸的至少两个间隔不多于15个核苷酸。

[0014] 在一些实施方案中,所述多种多核苷酸探针的至少一个子集的每种多核苷酸探针的至少50%核苷酸是锁核酸(LNA)核苷酸。在一些实施方案中,所述多种多核苷酸探针的至少一个子集的每种多核苷酸探针的至少75%核苷酸是锁核酸(LNA)核苷酸。

[0015] 在一些实施方案中,所述多种多核苷酸探针的每种具有比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸高至少约1℃的解链温度。在一些实施方案中,解链温度高至少约10℃。

[0016] 在一些实施方案中,所述多种多核苷酸探针的每种具有比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸高至少约2%的解链温度。在一些实施方案中,解链温度高至少约10%。

[0017] 在一些实施方案中,融合基因为癌症融合基因。在一些实施方案中,所述多种多核苷酸探针的每种与图2A-2B的融合基因对的基因或选自图3的两种或更多种基因之间的融合基因具有序列互补性。在一些实施方案中,所述多种多核苷酸探针的每种与距融合基因的断点不多于500个核苷酸的断点区域具有序列互补性。在一些实施方案中,所述多种多核苷酸探针的每种与跨越融合基因中的断点的序列(a sequence across a breakpoint in the fusion gene)具有序列互补性。

[0018] 在一些实施方案中,所述多种多核苷酸探针的每种具有小于约500个核苷酸的长度。在一些实施方案中,所述多种多核苷酸探针的每种具有在约20和约200个核苷酸之间的长度。在一些实施方案中,所述多种多核苷酸探针的每种具有在约80和约160个核苷酸之间的长度。

[0019] 在一些实施方案中,所述断点片段的每个具有在约140个核苷酸和180个核苷酸之间的长度。

[0020] 在一些实施方案中,所述多种多核苷酸探针偶联于固体支持物。在一些实施方案中,所述探针集包括一种或更多种天然多核苷酸探针。在一些实施方案中,所述多种多核苷酸探针包括与融合基因中包括的核酸序列的断点区域杂交的至少一种多核苷酸探针、和与融合基因中包括的核酸序列的非断点区域杂交的至少一种天然多核苷酸探针。

[0021] 在一些实施方案中,所述多种多核苷酸探针的每种提供对融合基因中包括的核酸序列的断点区域的至少50%覆盖。

[0022] 在一些实施方案中,(d)包括向分离的多核苷酸附加包含具有不同的条形码序列的条形码的标签,以生成加标签的亲本多核苷酸。在一些实施方案中,方法还包括扩增加标签的亲本多核苷酸,以产生加标签的子代多核苷酸。

[0023] 在一些实施方案中,方法还包括(i)将加标签的子代多核苷酸测序以产生序列读段,其中每个序列读段包括条形码序列和来源于所述分离的多核苷酸的指定的一种的序列,和(ii)至少基于条形码序列将序列读段分组为家族。

[0024] 在一些实施方案中,方法还包括将被分组在每一个家族内的序列读段进行比较以确定每一个家族的共有序列,其中每一个共有序列对应于加标签的亲本多核苷酸中的独特

多核苷酸。

[0025] 在另一个方面,本公开内容提供了用于捕获融合基因的断点片段的方法,包括(a)提供包含或疑似包含包括融合基因的断点片段的无细胞核酸分子的生物样品;和(b)在足以进行以下的条件下将生物样品与多核苷酸探针接触:(i)允许多核苷酸探针与断点片段之间的杂交以在混合物中提供探针捕获的多核苷酸,该多核苷酸探针与断点片段具有序列互补性并具有比具有与融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸更大的对融合基因的亲和力;和(ii)从混合物富集或分离探针捕获的多核苷酸,其中多核苷酸探针与断点片段具有序列互补性。

[0026] 在一些实施方案中,多核苷酸探针包含一个或更多个锁核酸(LNA)核苷酸。在一些实施方案中,多核苷酸探针包含多个LNA核苷酸,其中LNA核苷酸的至少两个间隔不多于30个核苷酸。在一些实施方案中,LNA核苷酸的至少两个间隔不多于15个核苷酸。

[0027] 本公开内容的另一个方面提供了包括多种多核苷酸探针的探针集,其中多核苷酸探针的每种具有(i)与作为无细胞核酸分子的一部分的融合基因的序列互补性和(ii)比具有与融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸更大的对融合基因的亲和力。

[0028] 在一些实施方案中,所述多种多核苷酸探针的每种包含一个或更多个锁核酸核苷酸。在一些实施方案中,所述探针集还包括一种或更多种天然多核苷酸探针。在一些实施方案中,所述多种多核苷酸探针的每种包括与融合基因中包括的核酸序列的断点区域杂交的至少一种多核苷酸探针、和与融合基因中包括的核酸序列的非断点区域杂交的至少一种天然多核苷酸探针。

[0029] 在一些实施方案中,所述多种多核苷酸探针的每种提供对融合基因中包括的核酸序列的断点区域的至少50%覆盖。

[0030] 在一些实施方案中,多种多核苷酸探针与融合基因中的不同基因的一种或两次的部分杂交。

[0031] 在一些实施方案中,探针集还包括固体支持物,其中多种多核苷酸探针偶联于固体支持物。

[0032] 在一些实施方案中,所述多种多核苷酸探针的每种具有比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸高至少约1℃的解链温度。在一些实施方案中,解链温度高至少约10℃。

[0033] 在一些实施方案中,所述多种多核苷酸探针的每种具有比具有与所述融合基因互补的序列且仅包含未修饰的核苷酸的多核苷酸高至少约2%的解链温度。在一些实施方案中,解链温度高至少约10%。

[0034] 在一些实施方案中,融合基因为癌症融合基因。

[0035] 在一些实施方案中,所述多种多核苷酸探针的每种与图2A-2B的融合基因对的基因或选自图3的两种或更多种基因之间的融合基因具有序列互补性。

[0036] 在另一方面,本文公开了高亲和力多核苷酸,包含被配置为与无细胞核酸分子中的融合基因相关的核酸序列特异性杂交的序列。

[0037] 在另一方面,本文公开了高亲和力多核苷酸,其被配置为与融合基因特异性杂交。在一个实施方案中,所述高亲和力多核苷酸包含一个或更多个锁核酸核苷酸。在另一个实



施方案中,高亲和力多核苷酸具有比仅包含天然核苷酸的具有相同序列的多核苷酸高至少以下任一种的解链温度:1℃、2℃、3℃、4℃、5℃、10℃、15℃或20℃。在另一个实施方案中,高亲和力多核苷酸具有比仅包含天然核苷酸的具有相同序列的多核苷酸高至少以下任一种的解链温度:2%、4%、6%、8%或10%。在另一个实施方案中,高亲和力多核苷酸被配置为与癌症融合基因特异性杂交。在另一个实施方案中,高亲和力多核苷酸被配置为与图2A-2B的融合基因对的基因或选自图3的1、2、3、4、5、6、7、8、9、10或更多个基因的至少任一个之间的融合基因特异性杂交。在另一个实施方案中,高亲和力多核苷酸被配置为在距融合基因的断点不多于500个核苷酸的断点区域内杂交。在另一个实施方案中,高亲和力多核苷酸被配置为跨越融合基因中的断点杂交。在另一个实施方案中,高亲和力多核苷酸具有小于约500个核苷酸、在约20个和约200个核苷酸之间、或约80个和约160个核苷酸之间的长度。在另一个实施方案中,高亲和力多核苷酸包含多个锁核酸(LNA)核苷酸,其中LNA核苷酸的至少两个间隔不多于30、20、15、10或5个核苷酸。在另一个实施方案中,多核苷酸中的核苷酸的100%、或90%、75%、50%、20%、10%、或5%或1%的至少任一种是锁核酸核苷酸。在另一个实施方案中,高亲和力多核苷酸具有与融合基因的核苷酸序列完全或大体上互补的核苷酸序列。

[0038] 在另一方面,本公开内容提供了高亲和力多核苷酸探针,其包含被配置为与融合基因特异性杂交的高亲和力多核苷酸。在一个实施方案中,所述高亲和力多核苷酸包含一个或更多个锁核酸核苷酸。在另一个实施方案中,探针包括选自可检测标记、结合部分或固体支持物的官能性。在另一个实施方案中,探针被配置为与融合基因的断点片段杂交。在另一个实施方案中,断点片段具有在约140个核苷酸和约180个核苷酸之间的长度。在另一个实施方案中,片段是无细胞脱氧核糖核酸(DNA)或基因组DNA。在另一个实施方案中,高亲和力多核苷酸结合于固体支持物。

[0039] 在另一方面,本公开内容提供了包含多种多核苷酸探针的探针集,每种探针被配置为与融合基因特异性杂交,其中所述集包括一种或更多种高亲和力多核苷酸探针。在一个实施方案中,所述高亲和力多核苷酸包含一个或更多个锁核酸核苷酸。在另一个实施方案中,所述集包括一种或更多种天然多核苷酸探针。在另一个实施方案中,探针集包括与融合基因中包含的基因的断点区域特异性杂交的至少一种高亲和力多核苷酸探针、和与融合基因中包括的基因的非断点区域杂交的至少一种天然多核苷酸探针。在另一个实施方案中,探针集中的一种或更多种高亲和力多核苷酸探针提供对融合基因中包含的基因的断点区域的至少50%(例如,至少0.5X至5X)覆盖。在另一个实施方案中,探针与融合基因中的不同基因的一种或两种的部分杂交。在另一个实施方案中,探针集被配置为寡核苷酸芯片。在另一个实施方案中,靶序列被高亲和力多核苷酸探针和标准亲和力多核苷酸探针二者靶向。

[0040] 在另一方面,本公开内容提供了包括多个探针集的试剂盒,其中每个探针集与不同基因特异性杂交且探针集的至少一个是本公开内容的探针集。在一个实施方案中,高亲和力多核苷酸包含一个或更多个锁核酸核苷酸。

[0041] 在另一方面,本公开内容提供了用于捕获融合基因的断点片段的方法,包括将断点片段与高亲和力多核苷酸探针在严格杂交条件下接触并允许杂交,其中多核苷酸探针结合于固体支持物且其中多核苷酸探针具有与断点片段的核苷酸序列大体上或完全互补的

核苷酸序列。在一个实施方案中,高亲和力多核苷酸包含一个或更多个锁核酸核苷酸。

[0042] 在另一方面,本公开内容提供了用于富集样品的包含融合基因的断点的多核苷酸的方法,包括:a)将权利要求20的探针集与多核苷酸的混合物在杂交条件下接触以产生探针捕获的多核苷酸;和b)从混合物分离探针捕获的多核苷酸以产生富集包含融合基因的断点片段的多核苷酸的样品。在一个实施方案中,高亲和力多核苷酸包含一个或更多个锁核酸核苷酸。在另一个实施方案中,多核苷酸包括无细胞DNA或片段化的基因组DNA。在另一个实施方案中,方法还包括从探针分离捕获的多核苷酸。在另一个实施方案中,方法还包括将分离的多核苷酸测序。

[0043] 在另一方面,本公开内容提供了诊断受试者中的癌症的方法,包括:a)提供来自受试者的包含多核苷酸的样品;b)将来自样品的无细胞DNA(cfDNA)与权利要求20的探针集在杂交条件下接触以产生探针捕获的多核苷酸;c)从混合物分离探针捕获的多核苷酸以产生富集包含融合基因的断点片段的多核苷酸的样品;d)将分离的多核苷酸测序以产生序列;e)基于序列检测包含融合基因的断点的多核苷酸;和f)基于断点片段的检测诊断癌症。在一个实施方案中,高亲和力多核苷酸包含一个或更多个锁核酸核苷酸。

[0044] 本公开内容的另一个方面提供了一种非暂时性计算机可读介质(a non-transitory computer readable medium),该非暂时性计算机可读介质包含机器可执行代码,该机器可执行代码在被一个或更多个计算机处理器执行时实施本文以上或别处的方法的任一种。

[0045] 本公开内容的另一个方面提供了一种系统,该系统包含一个或更多个计算机处理器及与其耦合的非暂时性计算机可读介质。非暂时性计算机可读介质包含机器可执行代码,该机器可执行代码在被一个或更多个计算机处理器执行时实施本文以上或别处的方法的任一种。

[0046] 从以下详述的描述,本公开内容的另外的方面和优势对本领域技术人员而言将变得明显,详细描述中仅示出和描述了本公开内容的说明性实施方案。如将会意识到的,本公开内容能够具有其他和不同的实施方案,并且其若干细节能够在多种明显的方面进行修改,所有这些都不偏离本公开内容。相应地,附图和描述被认为是本质上是说明性的而不是限制性的。

[0047] 通过引用并入

[0048] 本说明书中提及的所有出版物、专利和专利申请通过引用并入本文,其程度如同每一个单独出版物、专利或专利申请被具体和单独地指明通过引用并入的相同程度。

[0049] 附图简述

[0050] 本发明的新颖的特征在所附权利要求书中被具体陈述。本发明的特征和优势的更好的理解将通过参考陈述其中利用本发明的原理的说明性实施方案的以下详细说明和附图(在本文也为“图(Figure)”和“图(FIG.)”)来获得,在附图中:

[0051] 图1描绘来源于融合基因的断点片段和标准探针捕获程序期间此类片段的丢失;

[0052] 图2A提供癌症融合基因对的列表;图2B提供癌症融合基因对的另一个列表;

[0053] 图3提供癌症融合基因中检测到的基因的列表;

[0054] 图4A-4U提供癌症融合基因对的示例性断点;

[0055] 图5A-B显示探针和/或多核苷酸的不同覆盖深度和平铺(tiling);

[0056] 图6A-6D显示高亲和力探针序列子集和标准亲和力探针序列子集的不同示例性混合物;

[0057] 图7显示64基因组群 (64gene panel), 包括参与基因重排的四四种基因ALK、NKRT1、RET和ROS1;

[0058] 图8显示ALK基因的可被靶向用于更深的覆盖的八个基因组区域;和

[0059] 图9显示计算机控制系统,其被编程或以其他方式被配置为实施本文提供的方法。

[0060] 详细说明

[0061] 尽管本文已经显示和描述了本发明的多种实施方案,但对于本领域技术人员将明显的是,此类实施方案仅通过示例的方式提供。在不偏离本发明的情况下,本领域技术人员可以想到许多变化、改变和替换。应当理解,可以采用本文描述的本发明的实施方案的多种替代选择。

[0062] I. 定义

[0063] 本文所用的“高亲和力多核苷酸”是指包含至少一种化学修饰的多核苷酸,所述至少一种化学修饰在杂交反应中为多核苷酸提供比未如此修饰的相同序列多核苷酸更高的解链温度。在实施方案中,更高解链温度可以是高1°C、2°C、3°C、4°C、5°C、10°C、15°C或20°C的至少任一种。多核苷酸可包括一个或多个核苷酸类似物,即LNA核苷酸。

[0064] 本文所用的“锁核酸”(“LNA”) (有时称为“不可及的RNA”)是指包含至少一个锁核酸(LNA)核苷酸的高亲和力多核苷酸。

[0065] 本文所用的“锁核酸核苷酸”(“LNA核苷酸”)是指修饰的RNA核苷酸,所述修饰的RNA核苷酸在杂交期间为多核苷酸提供与仅通过具有天然核糖核苷酸代替所述修饰的RNA核苷酸而不同于LNA的多核苷酸相比更大的热力学稳定性。在某些实施方案中,修饰的RNA核苷酸的核糖部分被连接2' 氧和4' 碳的额外的桥修饰。

[0066] LNA核苷酸可包括增加LNA与其互补物之间的双链体的热力学稳定性的、在RNA的2' O与4' C之间的任何类型的额外的桥。在一些情况中,BNA,2' 氧和4' 碳被亚甲基基团桥联。在一些情况中,2' -O,4' -C-亚乙基-桥联的核酸(ENA),2' 氧和4' 碳被亚乙基基团桥联。BNA的其他实例可包括但不限于,2',4' -BNA<sup>NC</sup>[NH]、2',4' -BNA<sup>NC</sup>[NMe]和2',4' -BNA<sup>NC</sup>[NBn]。

[0067] “桥联的核酸”(“BNA”)是指2' -O,4' -C-亚甲基-修饰的核酸。

[0068] 其他2' O-修饰的核苷酸诸如2' O-Me也表现更大的稳定性。

[0069] 本文所用的“融合基因”是指染色体重排(倒位、缺失、易位)产生的基因,所述染色体重排使得基因组中至少两个不同基因的此前分离的部分在一起。

[0070] 本文所用的“癌症融合基因”是指癌细胞中的体细胞突变产生的融合基因。

[0071] 本文所用的“断点”是指融合基因中的核苷酸位置,在此位置两个不同基因的部分融合。

[0072] 本文所用的“断点区域”是指可参与基因融合的基因的区域,在该区域可存在断点。

[0073] 本文所用的融合基因的“断点片段”是指包括来自构成融合基因的两个不同基因的序列的融合基因的片段。

[0074] 本文所用的“探针”是指包含官能性的多核苷酸。官能性可以是可检测标记(荧光)、结合部分(生物素)、或固体支持物(磁力可吸引的颗粒或芯片)。

[0075] 本文所用的“天然多核苷酸”或“天然寡核苷酸”是指其中探针中的所有核苷酸是天然核苷酸的多核苷酸或寡核苷酸。

[0076] “互补性”是指核酸与另一核酸序列通过传统Watson-Crick类型或其他非传统类型形成氢键的能力。互补性百分比指示核酸分子中可与第二核酸序列形成氢键 (Watson-Crick碱基配对) 的残基的百分比 (10中的5、6、7、8、9、10分别为50%、60%、70%、80%、90%和100%互补)。“完全互补”是指核酸序列中所有的连续残基将与第二核酸序列中相同数目的连续残基氢键键合。

[0077] 本文所用的“大体上互补”是指跨越8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50或更多个核苷酸的区域60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%的至少任一种的互补性程度,或是指在严格条件下杂交的两个核酸。序列同一性,诸如为了评估互补性百分比的目的,可通过任何合适的比对算法测量,所述比对算法包括但不限于Needleman-Wunsch算法 (参见例如在万维网网页: [ebi.ac.uk/Tools/psa/emboss\\_needle/nucleotide.html](http://ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html)可得的EMBOSS Needle aligner, 任选地以默认设置)、BLAST算法 (参见例如在[blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)可得的BLAST比对工具, 任选地以默认设置)、或Smith-Waterman算法 (参见例如在万维网网页: [ebi.ac.uk/Tools/psa/emboss\\_water/nucleotide.html](http://ebi.ac.uk/Tools/psa/emboss_water/nucleotide.html)可得的EMBOSS Water aligner, 任选地以默认设置)。最佳比对可使用选择的算法的任何合适的参数 (包括默认参数) 评估。

[0078] “杂交”是指其中一个或更多个多核苷酸反应形成经由核苷酸残基的碱基之间的氢键被稳定化的复合体的反应。氢键键合可通过Watson Crick碱基配对、Hoogstein结合、或以符合碱基互补性的任何其他序列特异性方式发生。复合体可包括形成双链体结构的两条链、形成多链复合体的三条或更多条链、单个自杂交链、或这些的任何组合。杂交反应可构成更广泛的过程诸如PCR的起始、或内切核酸酶对多核苷酸的酶促裂解中的一个步骤。与第一序列互补的第二序列被称为第一序列的“互补物”。应用于多核苷酸的术语“可杂交的”是指杂交反应中多核苷酸形成经由核苷酸残基的碱基之间的氢键键合被稳定化的复合体的能力。

[0079] “特异性杂交于 (specifically hybridize to)”或“与……特异性杂交 (hybridizing specifically to)”或“特异性杂交 (specific hybridization)”是指在50%甲酰胺、5x SSC和1%SDS、在42℃孵育或5x SSC和1%SDS、在65℃孵育、和在0.2x SSC和0.1%SDS中在65℃洗涤的条件下,两个多核苷酸之间的稳定双链体的形成。

[0080] 术语“严格杂交条件”指在其下多核苷酸将优先与其靶子序列杂交,并且在较小程度上与其它序列杂交或根本不与其他序列杂交的条件。在核酸杂交实验的上下文中的“严格杂交”是序列依赖性的,并在不同环境参数下是不同的。核酸杂交的广泛指南见于Tijssen(1993)Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes第I部分第2章“Overview of principles of hybridization and the strategy of nucleic acid probe assays”,Elsevier,New York。

[0081] 通常,高度严格的杂交和洗涤条件被选择为比在定义的离子强度和pH下特定序列的热解链温度( $T_m$ )低约5℃。 $T_m$ 为50%的靶序列与完全匹配的探针杂交的温度 (在定义的离子强度和pH下)。非常严格的条件被选择为等于特定探针的 $T_m$ 。

[0082] 严格杂交条件包括包含水、缓冲剂(在pH 6-9或pH 7-8的磷酸盐、tris、SSPE或SSC缓冲液)、盐(钠或钾)、和变性剂(SDS、甲酰胺或吐温)的缓冲液和37℃-70℃、60℃-65℃的温度。

[0083] 用于在DNA印迹或RNA印迹中滤器上具有多于100个互补残基的互补核酸的杂交的严格杂交条件的实例为在42℃, 50%福尔马林与1mg肝素, 杂交被进行过夜。高度严格的洗涤条件的实例为在72℃, 0.15M NaCl, 持续约15分钟。严格洗涤条件的实例为在65℃, 0.2X SSC洗涤, 持续15分钟(参见, Sambrook等人对SSC缓冲液的描述)。通常, 高严格洗涤之前进行低严格洗涤以去除背景探针信号。对于, 例如, 多于100个核苷酸的双链体的示例性中度严格洗涤为在45℃, 1x SSC, 持续15分钟。对于, 例如, 多于100个核苷酸的双链体的示例性低度严格洗涤为在40℃, 4-6x SSC, 持续15分钟。通常, 比起在特定杂交测定中对于不相关探针观察到的2倍(或更高)的信噪比指示检测到特异性杂交。

[0084] II. 概要

[0085] 本文提供了用于检测包括一种或更多种融合基因的多核苷酸的组合物和方法。多核苷酸可以是脱氧核糖核酸(DNA)。本文提供的组合物和方法可以高灵敏度检测异质多核苷酸样品诸如无细胞DNA(“cfDNA”)中的融合基因。

[0086] 来自包括癌细胞的细胞的DNA可以无细胞DNA的形式脱落到血液中。无细胞DNA具有约160个核苷酸的平均长度。由于片段化不在预先指定的点发生, 对于任何基因组基因座, 样品中可见跨越该基因座平铺(tile)的片段。

[0087] 在癌症中, 某些基因共同参与与其他基因的基因融合。例如, 在癌症中, EML4和ALK基因共同经历与彼此的基因融合。参与融合的每个基因的断点可在每个基因中的断点区域(“热点”)出现。当包含这些融合基因的细胞死亡时, 其DNA以cfDNA的形式脱落到血液中。如图1所示, 片段中映射到断点的位置可在片段中的任何地方、接近5'末端、在中间、或接近3'末端出现。因此, cfDNA多核苷酸可具有来自参与融合的任一种基因的非常短或非常长的核苷酸序列。

[0088] 某些DNA测序方法使用序列捕获来富集感兴趣的序列。序列捕获通常包括使用与感兴趣的序列杂交的寡核苷酸探针。探针集策略可包括跨越感兴趣的区域平铺探针。此类探针可以是约120个碱基长。探针集可具有约2x的深度。序列捕获的有效性部分地取决于靶分子中与探针的序列互补(或几乎互补)的序列的长度。

[0089] 然而, 在融合基因的情况中, 映射到断点的多核苷酸可包含来自靶基因的比对于杂交和捕获最佳的序列短的序列。例如, 映射到包括ALK-EML4融合的融合体的cfDNA片段可具有, 例如, ALK基因的150个核苷酸的序列、100个核苷酸的序列、50个核苷酸的序列、25个核苷酸的序列或10个核苷酸的序列。在这一情况中, 如果多核苷酸具有较短的ALK序列, 则捕获该多核苷酸的概率比捕获具有与ALK探针完全互补的序列的多核苷酸的概率低。当序列捕获是多路的、靶向来自许多不同基因的序列时, 问题更尖锐。

[0090] 本文提供了用于捕获映射到融合基因中的断点的多核苷酸片段的材料和方法。此类多核苷酸使用高亲和力多核苷酸探针诸如锁核酸捕获。此类探针具有比天然核苷酸制成的相同序列的探针更高的解链温度。因此, 其产生来自相同样品的捕获产物的更高收率。

[0091] 此类探针可被包括在靶向融合基因和非融合基因二者的探针集中。以这种方式, 与使用仅从天然核苷酸制成的探针捕获的群体相比, 捕获的多核苷酸富集包括融合基因的

那些。

[0092] 示例性探针集可包含,例如,LNA探针的子集。LNA探针可被配置为跨越参与融合基因的基因的断点区域平铺。

[0093] LNA探针中的每一个核苷酸可以是LNA核苷酸。可选地,核苷酸的一部分可以是LNA核苷酸。在某些实施方案中,LNA核苷酸可间隔预定数目的核苷酸。

[0094] 本发明提供了可用于富集包含核酸片段的样品的包含基因融合事件的那些核酸片段的高亲和力多核苷酸。这些高亲和力多核苷酸可包含LNA核苷酸。用LNA核苷酸取代标准核苷酸可使高亲和力多核苷酸的解链温度升高,从而增加高亲和力多核苷酸与包含融合基因的核酸片段之间的双链体的稳定性。

[0095] 基因融合可以与健康细胞向赘生物(肿瘤或腺瘤)的发育相关,且在一些情况中有助于健康细胞向赘生物的发育。检测这些基因融合事件可提供用于检测和/或监测患者中的赘生物的存在的有效方法。然而,断点片段将具有比包括仅来自一种基因的序列的具有类似长度的核酸片段少的来源于断点侧翼的任一种基因的序列。因此,断点片段通常仅能够结合于减少部分的基因探针或基因特异性寡核苷酸。如果杂交和洗涤条件已经对全长或几乎全长的结合优化,包含断点的核酸片段可能以不充足的亲和力杂交并被丢失(参见图1)。而且,在包含来自已经经历和尚未经历基因融合事件的细胞的核酸片段的异质样品中,来自尚未经历基因融合事件的那些细胞的核酸片段可更稳定地结合于基因探针或基因特异性寡核苷酸并竞争性地抑制包含断点的核酸片段的杂交。

[0096] 肿瘤来源的核酸可在无细胞体液中发现。可对来自此类无细胞体液的肿瘤来源的核酸测定包含融合基因的核酸片段以检测赘生物。无细胞体液可包含少量的肿瘤来源的核酸,且肿瘤来源的核酸可与来源于健康组织的核酸混合。本公开内容还提供了用于从来源于无细胞体液的核酸富集包含融合基因的核酸片段的方法。

[0097] III. 测试样品

[0098] A. 受试者类型

[0099] 样品从受试者,例如处于发展癌症的风险的患者收集。受试者可以是不具有已知的癌症风险因素的患者。受试者可以是其仅有的癌症风险因素是年龄和/或性别的患者。在一些情况中,受试者可具有已知的癌症风险因素,例如抽烟或癌症家族史。在一些情况中,受试者可以是具有癌症症状的患者。

[0100] 其他受试者可以是具有此前已通过结肠镜检查或成像检测到的赘生物的患者。可对来源于具有此前检测的赘生物的患者样品测定包含断点的核酸片段以建议治疗的疗程或疗法。可对来源于具有赘生物的患者样品测定包含断点的核酸片段以确定其正在接受的治疗或疗法的效力。

[0101] 其他受试者可以是具有此前检测的赘生物的患者,但其中赘生物不再是可检测的(处于缓解的患者或不具有疾病证据的患者)。可对来源于其中赘生物不再是可检测的患者的样品测定包含断点的核酸片段以检测赘生物的复发或再度出现。

[0102] 其他受试者可以是具有癌症家族史的女性,其中造成家族癌症的遗传缺陷已知是融合基因或怀疑是融合基因。在一些情况中,具有癌症家族史的女性可以是妊娠的并希望确定其所怀的胎儿是否具有该融合基因。在一些情况中,可对来自此类受试者的包含胎儿核酸的样品测定基因融合事件。

[0103] B. 样品类型

[0104] 样品可以从多种来源提取的核酸。核酸可以是但不限于, 基因组DNA、RNA、线粒体DNA、胎儿DNA和miRNA。

[0105] 样品可以从多种包含无细胞核酸的体液来提取, 所述体液包括但不限于血液、血清、血浆、玻璃质、痰液、尿液、泪液、汗液、唾液、精液、粘膜分泌物、粘液、脊髓液、羊水、淋巴液等。体液的收集可使用多种技术实现。在一些情况下, 收集可以包括使用注射器从受试者抽吸体液。在其他情况下, 收集可以包括移液或直接收集流体到收集容器中。

[0106] 在收集体液之后, 可以使用多种技术分离和提取无细胞核酸。在一些情况下, 可以使用商购可得的试剂盒诸如Qiagen Qiaamp® Circulating Nucleic Acid试剂盒方案分离、提取和制备无细胞核酸。在其他实例中, 可以使用Qiagen Qubit™ dsDNA HS测定试剂盒方案、Agilent™ DNA 1000试剂盒或TruSeq™ Sequencing Library Preparation; Low-Throughput (LT) 方案以定量核酸。无细胞核酸可以是胎儿来源的(经由从妊娠受试者取得的流体), 或可以源自受试者自身的组织。无细胞核酸可来源于赘生物(例如肿瘤或腺瘤)。

[0107] 通常, 通过分隔步骤(partitioning step)从体液中提取和分离无细胞核酸, 在该分隔步骤中, 如在溶液中发现的无细胞核酸与细胞和体液的其他不可溶性组分分开。分隔可以包括但不限于技术诸如离心或过滤。在其他情况下, 细胞并非首先与无细胞核酸分隔, 而是经裂解。在一个实例中, 完整细胞的基因组DNA通过选择性沉淀来分隔。无细胞核酸(包括DNA)可以保持可溶并可以与不可溶性基因组DNA分开并被提取。通常, 在添加不同试剂盒特定的缓冲液和其他洗涤步骤之后, 可以使用异丙醇沉淀来沉淀核酸。可以使用进一步的清洁步骤诸如基于二氧化硅的柱以去除污染物或盐。可以针对特定应用优化一般步骤。例如, 可以在整个反应中添加非特异性批量(bulk)载体核酸以优化该程序的某些方面诸如收率。

[0108] 无细胞核酸可以是至多500个核苷酸的长度、至多400个核苷酸的长度、至多300个核苷酸的长度、至多250个核苷酸的长度、至多225个核苷酸的长度、至多200个核苷酸的长度、至多190个核苷酸的长度、至多180个核苷酸的长度、至多170个核苷酸的长度、至多160个核苷酸的长度、至多150个核苷酸的长度、至多140个核苷酸的长度、至多130个核苷酸的长度、至多120个核苷酸的长度、至多110个核苷酸的长度、或至多100个核苷酸的长度。

[0109] 无细胞核酸可以是至少500个核苷酸的长度、至少400个核苷酸的长度、至少300个核苷酸的长度、至少250个核苷酸的长度、至少225个核苷酸的长度、至少200个核苷酸的长度、至少190个核苷酸的长度、至少180个核苷酸的长度、至少170个核苷酸的长度、至少160个核苷酸的长度、至少150个核苷酸的长度、至少140个核苷酸的长度、至少130个核苷酸的长度、至少120个核苷酸的长度、至少110个核苷酸的长度、或至少100个核苷酸的长度。尤其地, 无细胞核酸可以是在140和180个核苷酸之间的长度。

[0110] 样品可来自受试者的组织提取。样品可以是肿瘤活组织检查。肿瘤活组织检查可包含肿瘤和健康组织的混合物。肿瘤活组织检查可被甲醛固定和石蜡包埋。肿瘤可以是活组织检查的至少0.1%、活组织检查的至少0.2%、活组织检查的至少0.5%、活组织检查的至少0.7%、活组织检查的至少1%、活组织检查的至少2%、活组织检查的至少3%、活组织检查的至少4%、活组织检查的至少5%、活组织检查的至少10%、活组织检查的至少15%、活组织检查的至少20%、活组织检查的至少25%、或活组织检查的至少30%。样品可

以是来自健康组织的活组织检查。

[0111] 从组织提取的核酸可以是至多10kb的长度、至多7kb的长度、至多5kb的长度、至多4kb的长度、至多3kb的长度、至多2kb的长度、至多1kb的长度、至多500个核苷酸的长度、至多400个核苷酸的长度、至多300个核苷酸的长度、至多250个核苷酸的长度、至多225个核苷酸的长度、至多200个核苷酸的长度、至多190个核苷酸的长度、至多180个核苷酸的长度、至多170个核苷酸的长度、至多160个核苷酸的长度、至多150个核苷酸的长度、至多140个核苷酸的长度、至多130个核苷酸的长度、至多120个核苷酸的长度、至多110个核苷酸的长度、或至多100个核苷酸的长度。

[0112] 从组织提取的核酸可以是至少5kb的长度、至少4kb的长度、至少3kb的长度、至少2kb的长度、至少1kb的长度、至少500个核苷酸的长度、至少400个核苷酸的长度、至少300个核苷酸的长度、至少250个核苷酸的长度、至少225个核苷酸的长度、至少200个核苷酸的长度、至少190个核苷酸的长度、至少180个核苷酸的长度、至少170个核苷酸的长度、至少160个核苷酸的长度、至少150个核苷酸的长度、至少140个核苷酸的长度、至少130个核苷酸的长度、至少120个核苷酸的长度、至少110个核苷酸的长度、或至少100个核苷酸的长度。

[0113] 在一些情况中,核酸可在提取过程中被修剪并包括在100和400个核苷酸之间的长度的片段。在一些情况中,核酸可在提取后被修剪并可包括在100和400个核苷酸之间的长度的核苷酸。

[0114] 无细胞和组织来源的核酸的分离和纯化可以使用多种方法来完成,所述方法包括但不限于使用由公司诸如Sigma Aldrich、Life Technologies、Promega、Affymetrix、IBI等提供的商业试剂盒和方案。试剂盒和方案还可以是非商购可得的。

[0115] IV. 遗传分析

[0116] 遗传分析包括核苷酸序列变体、拷贝数变异和融合基因的检测。遗传变体可通过测序确定。测序方法可以是大规模平行测序,即,同时(或以快速相继)测序至少100,000、1百万、1千万、1亿、或10亿个多核苷酸分子的任一个。测序方法可以包括但不限于:高通量测序、焦磷酸测序、边合成边测序、单分子测序、纳米孔测序、半导体测序、边连接边测序、边杂交边测序、RNA-Seq(Illumina)、数字基因表达(Helicos)、新一代测序、单分子边合成边测序(SMSS)(Helicos)、大规模并行测序、克隆单分子阵列(Solexa)、鸟枪法测序、Maxam-Gilbert或Sanger测序、引物步移、使用PacBio, SOLiD, Ion Torrent的测序、基于纳米孔的平台或其他测序方法。

[0117] 测序可通过进行序列捕获来变得更加高效,序列捕获即如本文描述的富集样品的感兴趣的靶序列、癌症融合基因的序列和癌症融合基因断点。序列捕获可使用与感兴趣的靶杂交的固定的探针进行。序列捕获可使用附接于官能团、生物素的探针进行,其允许与特定序列杂交的探针通过沉淀(pulldown)从样品富集。在一些情况中,在与官能化的探针杂交之前,来自文库片段的具体序列诸如衔接子序列可被与片段退火互补的、未官能化的多核苷酸序列掩蔽,以减少非特异性或脱靶结合。

[0118] 在一些情况中,无细胞核酸片段或组织来源的核酸片段是产生测序文库的输入。在一些情况中,在制备测序文库之前富集片段的特定序列。富集的片段化的核酸可被附接于适合用于在本文公开的任何测序平台上使用的任何测序衔接子。例如,序列衔接子可包括流通池序列、样品条形码、或二者。在另一个实例中,序列衔接子可以是发夹形的衔接子



和/或包括样品条形码。而且,所得的片段可被扩增和测序。在一些情况中,衔接子不包括测序引物区域。在一些情况中,在测序之前富集测序文库的特定序列。

[0119] 无细胞核酸可包括与种系核酸混合的少量的肿瘤核酸。在一些情况中,肿瘤活组织检查可包括与健康组织混合的少量肿瘤组织,且从此类样品提取而不经富集的核酸可包括与种系核酸混合的少量的肿瘤核酸。提高检测肿瘤核酸,以及特别是遗传序列变体和拷贝数变异的灵敏度和特异性的测序方法在本发明的方法中可以有用的。此类方法描述于例如WO 2014/039556、WO 2014/149134和WO 2015/100427,其每一个通过引用全文并入本文。这些方法不仅可以多达或大于0.1%的灵敏度检测分子,而且可区分这些信号与现在的测序方法中典型的噪音。来自基于血液的样品的无细胞核酸的灵敏度和特异性的提高可使用多种方法实现。一种方法包括样品中核酸分子的高效加标签,将样品中多核苷酸的50%、75%或90%的至少任一种加标签。这提高了样品中的低丰度靶分子将被加标签和随后被测序的可能性,并显著提高检测靶分子的灵敏度。

[0120] 另一种方法包括分子追踪,其鉴定从原始亲本分子冗余地生成的序列读段,并在亲本分子中的每个基因座或位置分配碱基的最可能的身份。这通过降低由扩增和测序错误生成的噪音显著提高检测的特异性,降低假阳性的频率。

[0121] 本公开内容的方法可用于在浓度小于5%、1%、0.5%、0.1%、0.05%或0.01%的非独特地加标签的初始起始遗传物质(罕见核酸)中以至少99%、99.9%、99.99%、99.999%、99.9999%或99.99999%的特异性检测遗传变异。随后可追踪加标签的多核苷酸的序列读段从而以不多于2%、1%、0.1%或0.01%的错误率生成多核苷酸的共有序列。

[0122] V. 基因融合事件和断点区域

[0123] 基因融合事件是使得基因组中至少两个不同基因的以前分开的部分在一起、产生融合基因的染色体重排(倒位、缺失和易位)。融合基因可与赘生物的形成有关和/或导致赘生物的形成。融合基因可以是癌症融合基因。癌症融合基因可以是癌症中存在的体细胞突变导致的融合基因。可形成癌症融合基因的基因对的非限制性实例见于图2A和2B。包括于融合基因中的基因的非限制性实例在图3中列出。

[0124] 图8示出了可被靶向用于更深的覆盖的ALK基因的基因组区域的非限制性实例。图8中的基因组区域可对应于ALK基因的不同变体。此类深度覆盖可通过在测序和用分子条形码坍缩(collapsing)后获得的独特分子的数目来定量,例如,对于典型的变体约2000-3000个分子对比对于图8的基因组区域约4000个分子。一系列的数千个独特分子可对应于大于1000x、2000x、3000x、4000x、5000x或10,000x测序深度。

[0125] 通常,融合基因可产生两个基因的异常并置,其可编码融合蛋白(BCR-ABL1),或者一个基因的调节元件可驱动癌基因(TMPRSS2-ERG)的异常表达。尽管癌症融合基因的复发性质,每个融合基因的断点的确切位置可变化。断点区域是指可参与基因融合的基因的区域,在该区域可存在断点。在一些情况中,断点区域在断点的至多500个核苷酸以内。在一些情况中,断点区域在断点的至多200个核苷酸以内,在断点的至多500个核苷酸以内,在断点的至多750个核苷酸以内,在断点的至多1千碱基(kb)以内,在断点的至多5kb以内,在断点的至多10kb以内,在断点的至多20kb以内,在断点的至多30kb以内,在断点的至多40kb以内,在断点的至多50kb以内,或在断点的至多100kb以内。

[0126] 对于给定基因对的示例性、非限制性断点提供于来自癌症体细胞突变目录

(Catalogue of Somatic Mutations in Cancer) (COSMIC; 参见Forbes等人, *Nucleic Acids Research* (2014) 43:D805-D811) 的图4A-4U。对于每个基因对, 具体的突变ID在第一列提供, 其指示特定类别的检测的或从文献推论的融合构建体。例如, 图4A提供29类检测的或从文献推论的融合构建体。对于每种突变, 5' 和3' 融合配偶体 (5' 和3' 是相对于每种基因的转录物的方向) 各自提供基因名称、最后观察到的外显子、相对于转录物推论的断点、是否以及是否存在插入的序列。对于每个突变ID, 还提供了观察到具有该突变的独特样品的数目和包括具有该特定突变的两种基因的基因融合的百分比。

[0127] 例如, 图4A中的第一行指示, 突变COSF463是EML4-ALK融合, 其中EML4基因被融合在ALK基因的上游。在这一实例中, 最后观察到的EML4外显子是外显子13, 且推论的断点是在对应于EML4基因转录物的位置1751的基因组位置。EML4基因已被融合, 使得融合交界处的第一ALK外显子是外显子20, 且推论的断点位置是对应于ALK基因转录物的位置4080的基因组位置。在5' 配偶体或3' 配偶体基因中不存在另外的插入的序列。COSF463融合基因已经在170种独特样品, 或在COSMIC数据库中包括的所有EML4-ALK融合基因的25%中被检测到。在一些情形中, 诸如COSF488 (图4A, 第5行), 推论的断点包括 '+', 随后是数字, 表示由第一个数字指示的转录物位置下游 (在内含子或UTR中) 的该数量的碱基的基因组位置。如果数字是在括号中, 则位置是近似的。在一些情形中, 诸如COSF488 (图4A, 第5行), 推论的断点包括 '-', 随后是数字, 表示由第一数字指示的转录物位置上游 (在内含子或UTR中) 的该数量的碱基的基因组位置。如果数字是在括号中, 则位置是近似的。'?' 指示精确的断点是未知的。例如, 在COSF488中, 断点是对应于EML4基因转录物的位置2318的基因组位置下游的654碱基, 其已经融合到对应于ALK基因转录物的位置4080的基因组位置上游的位置172碱基。

#### [0128] VI. 高亲和力多核苷酸

[0129] 在一些情况中, 高亲和力多核苷酸可以是至少约450个核苷酸的长度、至少约425个核苷酸的长度、至少约400个核苷酸的长度、至少约375个核苷酸的长度、至少约350个核苷酸的长度、至少约325个核苷酸的长度、至少约300个核苷酸的长度、至少约275个核苷酸的长度、至少约250个核苷酸的长度、至少约225个核苷酸的长度、至少约200个核苷酸的长度、至少约180个核苷酸的长度、至少约160个核苷酸的长度、至少约140个核苷酸的长度、至少约120个核苷酸的长度、至少约100个核苷酸的长度、至少约80个核苷酸的长度、至少约60个核苷酸的长度、至少约40个核苷酸、或至少约20个核苷酸的长度。

[0130] 此外, 在一些情况中, 高亲和力多核苷酸可以是至多约500个核苷酸的长度、至多约450个核苷酸的长度、至多约425个核苷酸的长度、至多约400个核苷酸的长度、至多约375个核苷酸的长度、至多约350个核苷酸的长度、至多约325个核苷酸的长度、至多约300个核苷酸的长度、至多约275个核苷酸的长度、至多约250个核苷酸的长度、至多约225个核苷酸的长度、至多约200个核苷酸的长度、至多约180个核苷酸的长度、至多约160个核苷酸的长度、至多约140个核苷酸的长度、至多约120个核苷酸的长度、至多约100个核苷酸的长度、至多约80个核苷酸的长度、至多约60个核苷酸的长度、至多约40个核苷酸、或至多约20个核苷酸的长度。

[0131] 尤其地, 在一些情况中, 高亲和力多核苷酸可以是在约20和约200个核苷酸之间的长度。此外, 在一些情况中, 高亲和力多核苷酸可以是在约80和约160个核苷酸之间的长度。

[0132] 在某些实施方案中,本发明的高亲和力多核苷酸具有与融合基因的靶序列完全互补或大体上互补的至少10个、至少25个、至少50个、至少100个或至少150个核苷酸的序列。

[0133] 高亲和力多核苷酸可包含一个或更多个LNA核苷酸。在一些情况中,高亲和力多核苷酸内的核苷酸的100%是LNA核苷酸。在一些情况中,高亲和力多核苷酸内的核苷酸的至少90%、至少70%、至少50%、至少20%、至少10%、至少5%或至少1%是LNA核苷酸。在一些情况中,高亲和力多核苷酸内的核苷酸的至多90%、至多70%、至多50%、至多20%、至多10%、至多5%或至多1%是LNA核苷酸。

[0134] 如果高亲和力多核苷酸包含多于一个LNA核苷酸,在一些情况中LNA核苷酸可间隔不多于30个核苷酸、间隔不多于20个核苷酸、间隔不多于15个核苷酸、间隔不多于10个核苷酸、或间隔不多于5个核苷酸。在高亲和力多核苷酸包含多于一个LNA核苷酸的其他情况中,LNA核苷酸可间隔至少30个核苷酸、间隔至少20个核苷酸、间隔至少15个核苷酸、间隔至少10个核苷酸、或间隔至少5个核苷酸。

[0135] 对于代替高亲和力多核苷酸中的天然核苷酸插入的每个LNA核苷酸,在严格条件下,高亲和力多核苷酸与仅包含天然核苷酸的其互补序列的双链体的解链温度可升高至少1℃、至少2℃、至少3℃、至少4℃、至少5℃、至少6℃、至少7℃、至少8℃、至少9℃或至少10℃。具体地,对于代替天然核苷酸插入的每个LNA核苷酸,解链温度可升高在约2℃和约8℃之间。

[0136] 在一些情况中,高亲和力多核苷酸(包括一个或更多个LNA核苷酸)的解链温度可以比具有与高亲和力多核苷酸相同的序列、仅包含天然核苷酸的多核苷酸的解链温度高至少0.5%、高至少1%、高至少2%、高至少3%、高至少4%、高至少5%、高至少10%、高至少15%、高至少20%、高至少25%、高至少30%、高至少35%、高至少40%、高至少45%、高至少50%、高至少55%、高至少60%、高至少65%、高至少70%、高至少75%、高至少80%、高至少85%、高至少90%、高至少95%或高至少100%。

[0137] 在一种配置中,结合的探针可使用结合配偶体的组合亲和纯化。在一个实例中,探针可包含结合配偶体诸如生物素。在亲和纯化步骤中,结合配偶体然后可用作另外的结合配偶体诸如链霉抗生物素的诱饵。在一些情况中,结合的探针可从未结合的探针亲和纯化。在其他情况中,包含结合配偶体和结合的探针的样品多核苷酸链可从未结合的探针亲和纯化。

[0138] 通常,用于捕获结合的探针的任何化学方法可以是合适的。在一些情况中,捕获可通过包括生物素和链霉抗生物素或链霉抗生物素衍生物的方法实现。例如,本公开内容的一个实施方案提供了捕获融合基因的测序文库片段,其中将针对参与融合基因的基因的探针、针对断点区域的探针、和/或针对断点的探针退火至测序文库的解链的链并亲和纯化而与其它测序文库片段分开。

[0139] 磁力的可吸引的颗粒诸如珠可用于分离。任何合适的珠分离技术可与本公开内容的方法一起使用。在一些情况中,珠可用于分离,因为感兴趣的分子可被吸引到珠,且珠可被洗涤以去除未被吸引到珠的溶液组分,允许富集、纯化和/或分离。珠可基于特性诸如大小、密度或介电特性、离子特性和磁特性与溶液中的其他组分分离。在优选的实施方案中,颗粒是磁力的可吸引的。磁力的可吸引的颗粒可利用磁场被引入、混合、去除、和释放到溶液中。利用磁力的可吸引的颗粒的方法也可被自动化。磁力的可吸引的颗粒由包括NEB、

Dynal、Micromod、Turbobeads和Spherotech的许多供应商供应。颗粒可使用官能化化学反应官能化以提供具有结合多核苷酸所需的结合基团的表面。

[0140] 在一些情况中,探针和/或高亲和力多核苷酸被配置为与癌症融合基因杂交。例如,探针和/或高亲和力多核苷酸可与融合基因来源于的任一基因的部分互补。在一些情况中,癌症融合基因可以是选自图2A-2B中呈现的列表的一种或更多种基因。

[0141] 在一些情况中,探针和/或高亲和力多核苷酸可被配置为与断点区域杂交。例如,在一些情况中,探针和/或高亲和力多核苷酸可与断点区域的部分杂交(探针和/或高亲和力多核苷酸可与断点的500个核苷酸以内的序列杂交)。此外,在一些情况中,探针和/或高亲和力多核苷酸可被配置为跨越融合基因中的断点杂交(参见图6C)。例如,探针和/或多核苷酸可与断点任一侧上的序列的部分杂交(参见图6D)。

[0142] VII. 探针和/或多核苷酸的集

[0143] 在一些情况中,提供了探针和/或多核苷酸的集。在一些情况中,集中的所有探针和/或多核苷酸包括LNA核苷酸。在一些情况中,集中的探针和/或多核苷酸的一个子集仅包含天然核苷酸,在以下称为“标准亲和力子集”,且第二个子集包含一个或多个LNA核苷酸,在以下称为“高亲和力子集”。

[0144] 在一个实施方案中,探针集包括针对融合基因的断点区域中的核苷酸序列的一种或更多种探针。

[0145] 探针和/或多核苷酸可用多种覆盖深度提供。例如,在一些情况中,覆盖深度可以是至少0.5x,其中探针或多核苷酸的集靶向区域中平均一半的碱基(参见图5A)。

[0146] 在一些情况中,覆盖深度可以是至少1x,其中探针和/或多核苷酸被设计为使得区域中每个碱基平均仅被一个探针和/或多核苷酸序列靶向。在一些情况中,覆盖深度可以是至少2x,其中探针和/或多核苷酸被设计为使得区域中每个碱基平均被两个探针和/或多核苷酸序列靶向。在一些情况中,探针或多核苷酸的集的覆盖深度可以是至少3x、至少4x或至少5x。在一些情况中,探针和/或多核苷酸可被平铺,其中探针和/或多核苷酸的集被设计为使得连续的靶区域被探针和/或多核苷酸序列覆盖(参见图5B)。

[0147] 在一些情况中,使用探针和/或多核苷酸的标准亲和力子集来富集一些感兴趣的核酸片段,和使用探针和/或多核苷酸的高亲和力子集来富集同一样品中的其他核酸片段,可以是优选的。例如,在一些情况中,探针和/或多核苷酸的标准亲和力子集可靶向外显子组、癌基因或肿瘤抑制基因,且探针和/或多核苷酸的高亲和力子集可靶向融合基因诸如癌症融合基因(例如,图3中所列的基因)。在另一个实例中,在一些情况中,标准亲和力子集以第一覆盖深度靶向参与基因融合的一个或多个基因的连续或非连续部分,包括断点区域,且高亲和力子集以第二覆盖深度靶向断点区域(参见图6A)。在一些情况中,标准亲和力子集以第一覆盖深度靶向每个基因的连续或非连续部分,排除断点区域,且高亲和力子集以第二覆盖深度靶向断点区域(参见图6B)。在一些情况中,标准亲和力子集以第一覆盖深度靶向每个基因的连续或非连续部分,且高亲和力子集以第二覆盖深度靶向断点(参见图6C)。在一些情况中,标准亲和力子集以第一覆盖深度靶向每个基因的连续或非连续部分,且高亲和力子集以第二覆盖深度靶向断点任一侧上的序列,而非断点自身(参见图6D)。

[0148] 在一些情况中,探针和/或多核苷酸的集被配置为靶向多于一个基因,以富集可参与基因融合的一组基因(参见例如图7)。此外,在一些情况中,探针和/或多核苷酸的集被配

置为靶向多于一个基因和其断点或断点区域。

[0149] 在一些情况中,探针和/或多核苷酸的集被配置为靶向特定融合基因。例如,探针和/或多核苷酸可被设计为靶向参与基因融合的一个或二个基因。在一些情况中,探针和/或多核苷酸的集包括靶向单个基因和/或其断点或断点区域的探针和/或多核苷酸。

[0150] 在一些情况中,标准亲和力探针和/或多核苷酸与高亲和力探针和/或多核苷酸混合。在一些情况中,标准亲和力探针和/或多核苷酸与高亲和力探针和/或多核苷酸是分离的并依序使用。此外,在一些情况中,将样品首先与标准亲和力探针接触,且然后将未捕获的核酸片段与高亲和力探针接触。

[0151] 在一些情况中,高亲和力探针集可包括掺杂高亲和力多核苷酸的标准亲和力多核苷酸。在此类探针集中,靶序列可被标准亲和力多核苷酸和高亲和力多核苷酸二者靶向用于杂交。在此类掺杂的集中,高亲和力多核苷酸可仅靶向在断点区域的序列。

[0152] VIII. 试剂盒

[0153] 本公开内容提供了用于富集样品的断点片段的试剂盒。试剂盒可包括本文公开的探针和/或多核苷酸的任一种。在一些情况中,试剂盒可包括多个探针集,其中每个探针集与不同的基因杂交且至少一个探针集被配置为与融合基因杂交并包括一个或更多个高亲和力多核苷酸和/或探针。

[0154] IX. 使用方法

[0155] 本公开内容提供了使用本文公开的探针和/或多核苷酸的任一种用于富集断点片段的方法。此类方法可包括将与融合基因杂交的探针集与多核苷酸的混合物接触以产生探针捕获的多核苷酸,所述探针集中一个或更多个探针和/或多核苷酸是高亲和力多核苷酸和/或探针。然后探针捕获的多核苷酸可被分离以产生富集包含融合基因的断点片段的多核苷酸的样品。在一些情况中,多核苷酸为无细胞DNA。在一些情况中,多核苷酸为片段化的基因组DNA。在一些情况中,探针捕获的多核苷酸被洗脱以从探针分离捕获的多核苷酸。在一些情况中,洗脱的多核苷酸被直接测序或用于产生测序文库。

[0156] 提供了检测融合基因的方法。在方法中,提供了包括至少一种高亲和力多核苷酸的至少一个探针集,其针对参与基因融合的基因。探针集可包括标准亲和力多核苷酸探针和高亲和力多核苷酸探针二者。在一些实施方案中,探针集包括多个探针子集,每个子集针对不同的感兴趣的基因的序列,这些基因的一种或更多种参与癌症中的基因融合,且在一些实例中,这些基因的至少一种不参与基因融合。

[0157] 探针集可在严格杂交条件下与包含DNA诸如cfDNA的样品混合,且DNA可被允许与探针杂交。由于探针集包括高亲和力多核苷酸探针,捕获包括融合基因断点的DNA片段的概率升高。捕获的DNA可从探针分离并测序。可分析序列以检测具有跨越断点的序列的DNA片段,诸如包括来自通常不融合的两个不同基因的序列的DNA片段。融合基因的存在可与疾病诸如癌症相关联。因此,这一方法可用在疾病诸如癌症的诊断中。

[0158] 计算机控制系统

[0159] 本公开内容提供了被编程为实现本公开内容的方法的计算机控制系统。图9显示被编程或以其他方式被配置为检测融合基因和诊断和/或提供疾病诸如癌症的治疗干预的计算机系统901。

[0160] 计算机系统901包括中央处理单元(CPU,本文中也称为“处理器”和“计算机处理

器”) 905,其可以是单核或多核处理器或用于并行处理的多于一个处理器。计算机系统901还包括存储器或存储器位置910(例如,随机存取存储器、只读存储器、闪存存储器)、电子存储单元915(例如,硬盘)、用于与一个或更多个其他系统进行通信的通信接口920(例如,网络适配器)和外围设备925,诸如高速缓冲存储器、其他存储器、数据存储和/或电子显示适配器。存储器910、存储单元915、接口920和外围设备925与CPU905通过通信总线(实线),诸如主板(motherboard)通信。存储单元915可以是用于存储数据的数据存储单元(或数据存储库)。计算机系统901可以借助于通信接口920被可操作地耦合至计算机网络(“网络”)930。网络930可以是因特网(Internet)、互联网(internet)和/或外联网、或与因特网通信的内联网和/或外联网。在一些情况下,网络930为电信和/或数据网络。网络930可以包括一个或更多个计算机服务器,这可以支持分布式计算,诸如云计算。在一些情况下,借助于计算机系统901,网络930可以实现对等网络(peer-to-peer network),其可以使耦合至计算机系统901的设备能够作为客户端或服务器运行。

[0161] CPU 905可以执行一系列的机器可读指令,该机器可读指令可以以程序或软件来体现。指令可以被存储于存储器位置,诸如存储器910中。指令可以被导向CPU 905,其可以随后编程或以其他方式配置CPU 905,以实现本公开内容的方法。由CPU 905进行的操作的实例可以包括读取、解码、执行和写回。

[0162] CPU 905可以是电路诸如集成电路的一部分。系统901的一个或更多个其他组件可以被包含在该电路中。在一些情况下,电路为专用集成电路(ASIC)。

[0163] 存储单元915可以存储文件,诸如驱动程序、库和保存的程序。存储单元915可以存储用户数据,例如,用户偏好和用户程序。在一些情况下,计算机系统901可以包括一个或更多个另外的数据存储单元,该数据存储单元在计算机系统901的外部,诸如位于通过内联网或因特网而与计算机系统901通信的远程服务器上。

[0164] 计算机系统901可以与一个或更多个远程计算机系统通过网络930进行通信。例如,计算机系统901可以与用户(例如,健康管理提供者)的远程计算机系统通信。远程计算机系统的实例包括个人计算机(例如便携式PC)、板型或平板PC(例如**Apple®** iPad、**Samsung®** Galaxy Tab)、电话、智能电话(例如**Apple®** iPhone、Android支持的设备、**Blackberry®**)或个人数字助理。用户可以经由网络930访问计算机系统901。

[0165] 如本文描述的方法可以通过机器(例如,计算机处理器)可执行代码的方式实现,该机器可执行代码被存储在计算机系统901的电子存储位置,诸如,例如存储器910或电子存储单元915上。机器可执行代码或机器可读代码可以以软件的形式提供。在使用期间,代码可以由处理器905执行。在一些情况下,代码可以从存储单元915检索并存储在存储器910上,以用于由处理器905随时访问。在一些情况下,可以排除电子存储单元915,而将机器可执行指令存储于存储器910中。

[0166] 代码可以被预编译并配置为用于与具有适用于执行该代码的处理器一起使用,或者可以在运行时间期间被编译。代码可以以编程语言的形式提供,该编程语言可被选择以便使得代码能够以预编译的或按编译原样(as-compiled)的方式被执行。

[0167] 本文所提供的系统和方法的各方面,诸如计算机系统901,可以以编程来体现。技术的多个方面可以被认为是通常呈一种机器可读介质进行或体现的机器(或处理器)可执

行代码和/或相关数据的形式“产品”或“制品 (articles of manufacture)”。机器可执行代码可以被存储于电子存储单元诸如存储器 (例如, 只读存储器、随机存取存储器、闪存存储器) 或硬盘上。“存储”型介质可以包括计算机、处理器等的任何或所有有形存储器, 或其相关模块, 诸如多种半导体存储器、磁带驱动器、磁盘驱动器等, 其可以在任何时间为软件编程提供非暂时性存储。该软件的所有或部分有时可以通过因特网或多种其他电信网络进行通信。例如, 此类通信可以使得将软件从一个计算机或处理器加载到另一个计算机或处理器, 例如, 从管理服务器或主机加载到应用服务器的计算机平台。因此, 能够携带软件元件的另一类型的介质包括诸如在本地设备之间的物理接口、通过有线和光纤陆线网络以及在多种空中链路 (air-links) 上使用的光波、电波和电磁波。携带此类波的物理元件, 诸如有线或无线链路、光链路等, 也可以被认为是携带软件的介质。如本文使用的, 除非被限制为非暂时性的、有形的“储存”介质, 否则术语诸如计算机或机器“可读介质”指参与将指令提供至处理器用于执行的任何介质。

[0168] 因此, 机器可读介质, 诸如计算机可执行代码, 可以采取多种形式, 包括但不限于有形存储介质、载波介质或物理传输介质。非易失性存储介质包括, 例如光盘或磁盘, 诸如在任何计算机等中的任何存储设备, 诸如可用于实现如附图中示出的数据库等。易失性存储介质包括动态存储器, 诸如此类计算机平台的主存储器。有形的传输介质包括同轴电缆; 铜线和光纤, 包括构成计算机系统内的总线的导线。载波传输介质可以采取电信号或电磁信号或者声波或光波的形式, 诸如在射频 (RF) 和红外 (IR) 数据通信期间生成的那些。因此, 计算机可读介质的常见形式包括, 例如: 软盘 (floppy disk)、软性磁盘 (flexible disk)、硬盘、磁带、任何其他磁介质、CD-ROM、DVD或DVD-ROM、任何其他光学介质、穿孔卡片纸带、具有打孔模式的任何其他物理存储介质、RAM、ROM、PROM和EPROM、FLASH-EPROM、任何其他存储器芯片或盒、传输数据或指令的载波、传输此类载波的缆线或链路, 或者计算机可以从其读取编程代码和/或数据的任何其他介质。这些计算机可读介质的形式中的许多形式可以参与向处理器传送一个或更多个指令的一个或更多个序列以用于执行。

[0169] 计算机系统901可以包括电子显示器935或与之通信, 所述电子显示器935包括用户界面 (UI) 940, 用于提供报告的输出, 所述报告可以包括受试者的诊断或受试者的治疗干预。UI的实例包括但不限于图形用户界面 (GUI) 和基于网络的用户界面。

[0170] 本公开内容的方法和系统可以通过一个或更多个算法来实现。算法可以在由中央处理器905执行后通过软件来实现。算法可以, 例如, 便于融合基因的富集、测序和/或检测。

## 实施例

[0171] 实施例1: 癌症基因和癌症融合基因的富集和测序

[0172] 循环无细胞DNA从癌症患者的血浆使用QIAamp循环核酸试剂盒 (Qiagen) 按照制造商的方案分离, 除了用AmpureXP珠 (Beckman Coulter) 进行双侧SPRI以去除 >500bp 的片段和保留所有较低分子量的片段。然后末端修复所得的 ~160-bp cfDNA片段 (5ng至30ng) 并连接到具有分子条形码标签和下游下一代测序 (HiSeq2500, Illumina) 所需的序列的衔接子。使用与连接的衔接子序列互补的引物经10个循环扩增连接的cfDNA。

[0173] 为了富集包括融合基因的感兴趣的区域, 将所得的cfDNA文库在95°C变性和然后在65°C首先与封闭所加入的序列的寡核苷酸杂交, 并然后与120-nt生物素化的RNA寡核苷

酸(Agilent Technologies)杂交并还与120-nt生物素化的RNA/LNA或DNA/LNA寡核苷酸(Exiqon)在严格杂交缓冲液中杂交16小时。使用链霉抗生物素珠(Invitrogen)捕获杂交反应,洗涤以去除非靶向的cfDNA片段,并使用氢氧化钠洗脱。将所得的富集的文库扩增另外12个循环并在HiSeq2500(Illumina)上测序。

[0174] 实施例2序列捕获

[0175] 无细胞DNA从癌症患者分离。

[0176] 提供探针集,所述探针集被配置为捕获具有68个靶基因的序列的多核苷酸,所述68个靶基因包括参与基因重排的四个基因。探针集包括子集,每个子集针对组群中的68个基因中的一个。针对不参与基因重排的基因的每个子集是标准亲和力和子集(仅包括非高亲和力多核苷酸,仅具有天然核苷酸的多核苷酸)。针对参与基因重排的基因的每个子集是高亲和力子集(包括至少一个高亲和力多核苷酸)。集具有跨越外显子的2X平铺。在高亲和力子集中,高亲和力多核苷酸仅针对基因的断点区域。高亲和力子集掺杂有高亲和力多核苷酸,所以高亲和力多核苷酸和标准亲和力多核苷酸二者针对断点区域中的序列。

[0177] 将无细胞DNA和探针集在严格杂交条件下合并并孵育过夜。从混合物分离具有结合的cfDNA的探针集。从探针分离结合的多核苷酸并测序。鉴定了包括跨越断点的序列的多核苷酸。

[0178] 虽然本文已经显示和描述了本发明的优选实施方案,但对于本领域技术人员将明显的是,此类实施方案仅通过示例的方式提供。并非意图将本发明限制于本说明书中提供的具体实例。虽然已参考以上提及的说明书描述了本发明,但本文实施方案的描述和说明并不意图以限制性的意义来解释。在不偏离本发明的情况下,本领域技术人员现在将想到许多变化、改变和替换。此外,应当理解,本发明的所有方面并不限于本文阐述的取决于多种条件和变量的具体描写、配置或相对比例。应当理解,在实践本发明时可以采用本文描述的本发明的实施方案的各种替代选择。因此可以预期,本发明还应涵盖任何此类的替代选择、修改、变化或等同物。以下权利要求意图界定本发明的范围,并且从而涵盖在这些权利要求范围内的方法和结构及其等同物。



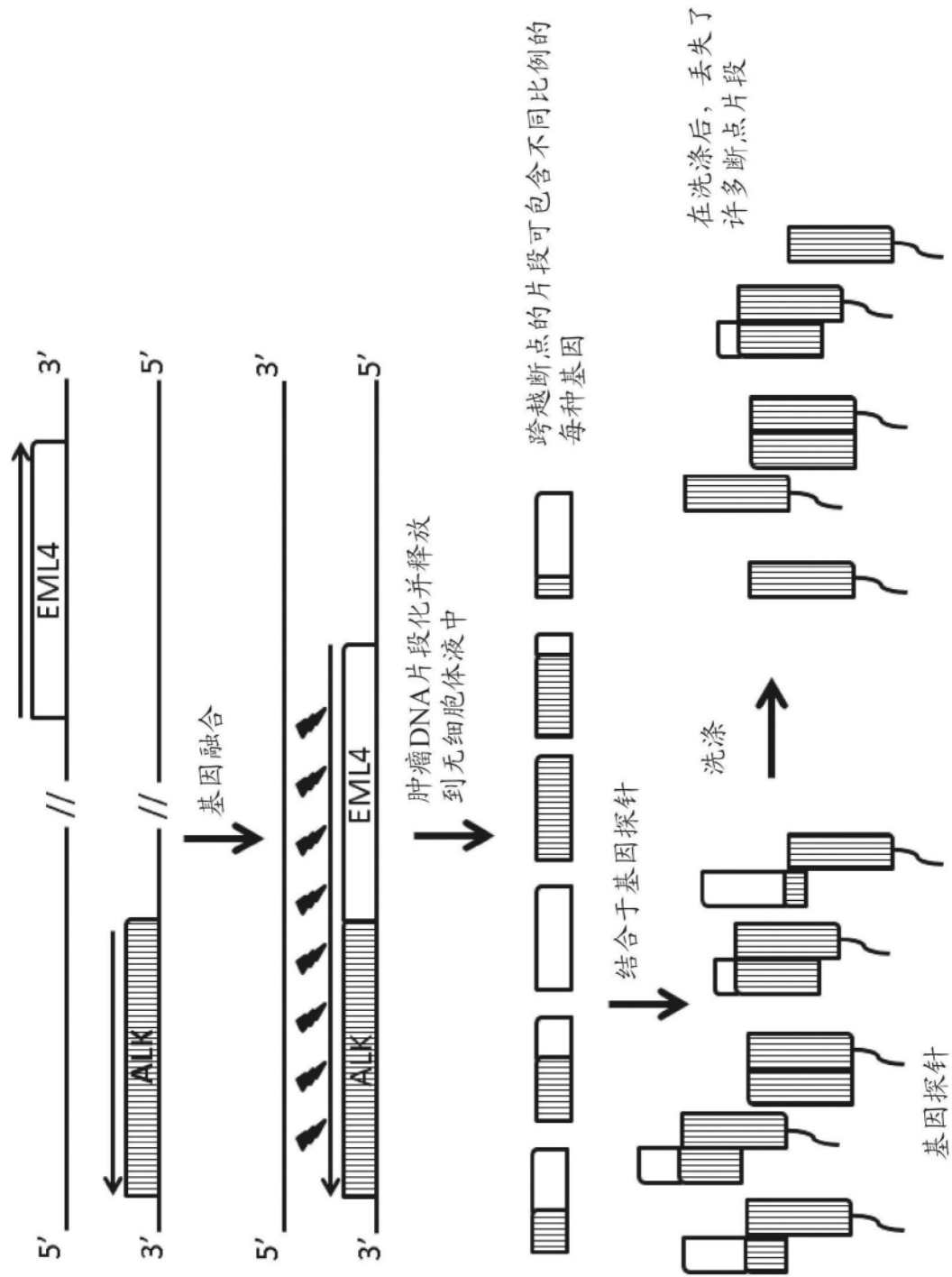


图1

ACBD6/RRP15	KMT2A/ELL
ACSL3_ENST00000357430/ETV1	KMT2A/EP300
ACTB/GLI1	KMT2A/EP515
AGPAT5/MCPH1	KMT2A/FOXO3
AGTRAP/BRAF	KMT2A/FOXO4
AKAP9_ENST00000356239/BRAF	KMT2A/FRYL
ARFIP1/FHDC1	KMT2A/GAS7
ARID1A/MAST2_ENST00000361297	KMT2A/GMPS
ASPSCR1/TFE3	KMT2A/GPHN
ATG4C/FBXO38	KMT2A/KIAA0284 ENST00000414716
ATIC/ALK	KMT2A/KIAA1524
BBS9/PKD1L1	KMT2A/LASP1
BCR/ABL1	KMT2A/LPP
BCR/JAK2	KMT2A/MAPRE1
BRD3/NUTM1	KMT2A/MLLT1
BRD4_ENST00000263377/NUTM1	KMT2A/MLLT10
C2orf44/ALK	KMT2A/MLLT11
CANT1/ETV4	KMT2A/MLLT3
CARS/ALK	KMT2A/MLLT4 ENST00000392108
CCDC6/RET	KMT2A/MLLT6
CD74_ENST0000009530/NRG1	KMT2A/MYO1F
CD74_ENST0000009530/ROS1	KMT2A/NCKIPSD
CDH11/USP6_ENST00000250066	KMT2A/NRIP3
CDKN2D/WDFY2	KMT2A/PDS5A
CEP89/BRAF	KMT2A/PICALM
CHCHD7/PLAG1	KMT2A/PRRC1

图2A

CIC/DUX4L1	KMT2A/SARNP
CIC/FOXO4	KMT2A/SEPT2
CLCN6/BRAF	KMT2A/SEPT5
CLIP1/ROS1	KMT2A/SEPT6
CLTC/ALK	KMT2A/SEPT9 ENST00000427177
CLTC/TFE3	KMT2A/SH3GL1
CNBP/USP6 ENST00000250066	KMT2A/SORBS2
COL1A1/PDGFB	KMT2A/TET1
COL1A1/USP6 ENST00000250066	KMT2A/TOP3A
COL1A2/PLAG1	KMT2A/ZFYVE19
CRTC1/MAML2	KTN1/RET
CRTC3/MAML2	LIFR ENST00000263409/PLAG1
CTAGE5/SIP1	LMNA/NTRK1 ENST00000392302
CTNNB1/PLAG1	LRIG3/ROS1
DCTN1/ALK	LSM14A/BRAF
DDX5 ENST00000540698/ETV4	MARK4/ERCC2
DNAJB1/PRKACA	MBOAT2/PRKCE
EIF3E/RSPO2	MBTD1 ENST00000586178/CXorf67 ENST00000342995
EIF3K/CYP39A1	MEAF6/PHF1
EML4/ALK	MKRN1/BRAF
EPC1/PHF1	MSN/ALK
ERC1/RET	MYB ENST00000341911/NFIB ENST00000397581
ERC1/ROS1	MYO5A/ROS1
ERO1L/FERMT2	NAB2/STAT6
ESRP1/RAF1	NACC2/NTRK2
ETV6/ITPR2	NCOA4 ENST00000452682/RET
ETV6/JAK2	NDRG1/ERG
ETV6/NTRK3	NF1/ACCN1

图2A (续)

EWSR1/ATF1	NFIA/EHF
EWSR1/CREB1	NFIX_ENST00000360105/MAST1_ENST00000251472
EWSR1/DDIT3	NONO/TFE3
EWSR1/ERG	NOTCH1_ENST00000277541/GABBR2
EWSR1/ETV1	NPM1/ALK
EWSR1/ETV4	NTN1/ACLY
EWSR1/FEV	NUP107/LGR5
EWSR1/FLI1	OMD/USP6_ENST00000250066
EWSR1/NFATC1	PAX3/FOXO1
EWSR1/NFATC2	PAX3/NCOA1
EWSR1/NR4A3	PAX3/NCOA2
EWSR1/PATZ1	PAX5/JAK2
EWSR1/PBX1	PAX7/FOXO1
EWSR1/POU5F1	PAX8/PPARG
EWSR1/SMARCA5	PCM1/JAK2
EWSR1/SP3	PCM1/RET
EWSR1/WT1	PLA2R1/RBMS1
EWSR1/YY1	PLXND1/TMCC1
EWSR1/ZNF384	PPFIBP1/ALK
EWSR1/ZNF444_ENST00000337080	PPFIBP1/ROS1
EZR/ROS1	PRCC/TFE3
FAM131B_ENST00000443739/BRAF	PRKAR1A/RET
FBXL18/RNF216	PTPRK/RSP03
FCHSD1/BRAF	PWWP2A/ROS1
FGFR1/ZNF703	QKI/NTRK2
FGFR1_ENST00000447712/PLAG1	RAF1/DAZL
FGFR1_ENST00000447712/TACC1	RANBP2/ALK
FGFR3/BAIAP2L1	RBM14/PACS1

图2A (续)

FGFR3/TACC3	RGS22/SYCP1
FN1/ALK	RNF130/BRAF
FUS/ATF1	SDC4/ROS1
FUS/CREB3L1	SEC16A_NM_014866.1/NOTCH1_ENST000000277541
FUS/CREB3L2	SEC31A/ALK
FUS/DDIT3	SEC31A/JAK2
FUS/ERG	SEPT8/AFF4
FUS/FEV	SFPQ/TFE3
GATM/BRAF	SLC22A1/CUTA
GMDS/PDE8B	SLC26A6/PRKAR2A
GNAI1/BRAF	SLC34A2/ROS1
GOLGA5/RET	SLC45A3/BRAF
GOPC/ROS1	SLC45A3/ELK4
GPBP1L1/MAST2_ENST00000361297	SLC45A3/ERG
HACL1/RAF1	SLC45A3/ETV1
HAS2/PLAG1	SLC45A3/ETV5_ENST00000306376
HERPUD1/BRAF	SND1/BRAF
HEY1/NCOA2	SQSTM1/ALK
HIP1/ALK	SRGAP3/RAF1
HLA-A/ROS1	SS18/SSX1
HMGA2/ALDH2_ENST00000261733	SS18/SSX2
HMGA2/CCNB1IP1	SS18/SSX4
HMGA2/COX6C	SS18L1/SSX1
HMGA2/EBF1	SSBP2/JAK2
HMGA2/FHIT_ENST00000476844	SSH2/SUZ12
HMGA2/LHFP	STIL/TAL1
HMGA2/LPP	STRN/ALK
HMGA2/NFIB_ENST00000397581	SUSD1/ROD1

图2A(续)

HMGA2/RAD51B	TADA2A_ENST00000394395/MAST1_ENST000000251472
HMGA2/WIF1_ENST00000286574	TAF15/NR4A3
HN1/USH1G	TCEA1_ENST00000521604/PLAG1
HNRNPA2B1/ETV1	TCF12/NR4A3
HOOK3/RET	TCF3/PBX1
IL6R/ATP8B2	TECTA/TBCEL
INTS4/GAB2	TFG/ALK
IRF2BP2/CDX1	TFG/NR4A3
JAZF1/PHF1	TFG/NTRK1_ENST00000392302
JAZF1/SUZ12	THRAP3/USP6_ENST00000250066
KIAA1549/BRAF	TMPRSS2/ERG
KIAA1598/ROSI	TMPRSS2/ETV1
KIF5B/ALK	TMPRSS2/ETV4
KIF5B/RET	TMPRSS2/ETV5_ENST00000306376
KLC1/ALK	TP53/NTRK1_ENST00000392302
KLK2/ETV1	TPM3/ALK
KLK2/ETV4	TPM3/NTRK1_ENST00000392302
KMT2A/ABI1	TPM3/ROSI
KMT2A/ABI2	TPM3_ENST00000368530/ROSI
KMT2A/ACTN4	TPM4/ALK
KMT2A/AFF1	TRIM24/RET
KMT2A/AFF3	TRIM27/RET
KMT2A/AFF4	TRIM33_ENST00000358465/RET
KMT2A/ARHGAP26	UBE2L3/KRAS
KMT2A/ARHGEF12	VCL/ALK
KMT2A/BTBD18	VTI1A/TCF7L2
KMT2A/CASC5	YWHAE_ENST00000264335/FAM22A_ENST000000381707
KMT2A/CASP8AP2	YWHAE_ENST00000264335/NUTM2B

图2A (续)

KMT2A/CBL	ZC3H7B/BCOR_ENST00000378444
KMT2A/CREBBP	ZCCHC8/ROS1
KMT2A/CT45A2	ZNF700/MAST1_ENST00000251472
KMT2A/DAB2IP	ZSCAN30/BRAF
KMT2A/EEFSEC	

图2A (续)

融合基因对

编号	红群 基因	配偶体 基因	编号	红群 基因	配偶体 基因	编号	红群 基因	配偶体 基因
1	ALK	ATIC	44	RET	ERC1	87	FGFR3	TACC3
2	ALK	C2orf44	45	RET	FKBP15	88	FGFR3	TNIP2
3	ALK	C2orf61	46	RET	GOLGA5	89	FGFR3	TPRG1
4	ALK	CAMKMT	47	RET	HOOK3	90	FGFR3	WHSC1
5	ALK	CARS	48	RET	KIF5B	91	ROS1	CCDC6
6	ALK	CENPF	49	RET	KTN1	92	ROS1	CD72
7	ALK	CLTC	50	RET	NCOA4	93	ROS1	CD74
8	ALK	CLIP4	51	RET	PCM1	94	ROS1	CEP85L
9	ALK	COL4A3	52	RET	PRKAR1A	95	ROS1	CLIP1
10	ALK	DCTN1	53	RET	SPECC1L	96	ROS1	CLTC
11	ALK	EIF2AK3	54	RET	SQSTM1	97	ROS1	ERC1
12	ALK	EML4	55	RET	TBL1XR1	98	ROS1	EZR
13	ALK	FANCL	56	RET	TRIM24	99	ROS1	HLA-A
14	ALK	FN1	57	RET	TRIM27	100	ROS1	KDELR2
15	ALK	GTF2IRD1	58	RET	TRIM33	101	ROS1	KIAA1598
16	ALK	HIP1	59	FGFR2	AFF3	102	ROS1	LRIG3

图2B



17	ALK	KCNQ5	60	FGFR2	AHCYL1	103	ROS1	MYO5A
18	ALK	KIF5B	61	FGFR2	BICC1	104	ROS1	PPFIBP1
19	ALK	KLC1	62	FGFR2	C10ORF68	105	ROS1	PWWP2A
20	ALK	MPRIIP	63	FGFR2	CASP7	106	ROS1	RUNX1
21	ALK	MSN	64	FGFR2	CCAR2	107	ROS1	SDC4
22	ALK	MYH9	65	FGFR2	CCDC6	108	ROS1	SLC34A2
23	ALK	NPM1	66	FGFR2	CD44	109	ROS1	SLC35F1
24	ALK	PPFIBP1	67	FGFR2	CIT	110	ROS1	TFG
25	ALK	PPM1B	68	FGFR2	KIAA1598	111	ROS1	TMEM106B
26	ALK	PPP1CB	69	FGFR2	MGEA5	112	ROS1	TPM3
27	ALK	PRKAR1A	70	FGFR2	NCALD	113	ROS1	TTC28
28	ALK	PTPN3	71	FGFR2	OFD1	114	ROS1	YWHAE
29	ALK	RANBP2	72	FGFR2	PPAPDC1A	115	ROS1	ZCCHC8
30	ALK	RNF213	73	FGFR2	PPHLN1	116	NTRK1	CD74
31	ALK	SEC31A	74	FGFR2	RASAL2	117	NTRK1	IRF2BP2
32	ALK	SOC5	75	FGFR2	SLC45A3	118	NTRK1	LMNA

图2B (续)

33	ALK	SQSTM1	76	FGFR2	TACC2	119	NTRK1	MPRI
34	ALK	STK32B	77	FGFR2	TACC3	120	NTRK1	NFASC
35	ALK	STRN	78	FGFR2	TXLNB	121	NTRK1	PEAR1
36	ALK	TFG	79	FGFR2	USP10	122	NTRK1	PLEKHA6
37	ALK	TPM1	80	FGFR2	VCL	123	NTRK1	QKI
38	ALK	TPM3	81	FGFR3	ADD1	124	NTRK1	RABGAP1L
39	ALK	TPM4	82	FGFR3	AES	125	NTRK1	SQSTM1
40	ALK	VCL	83	FGFR3	BAIAP2L1	126	NTRK1	SSBP2
41	RET	AFAP1	84	FGFR3	ELAVL3	127	NTRK1	TFG
42	RET	AKAP13	85	FGFR3	ETV6	128	NTRK1	TP53
43	RET	CCDC6	86	FGFR3	JAKMIP1	129	NTRK1	TPM3

图2B(续)

ACBD6  
 ACCN1  
 ACLY  
 ACSL3\_ENST00000357430  
 ACTB  
 AFF4  
 AGPAT5  
 AGTRAP  
 AKAP9\_ENST00000356239  
 ALDH2\_ENST00000261733  
 ALK  
 ARFIP1  
 ARID1A  
 ASPSCR1  
 ATF1  
 ATG4C  
 ATIC  
 ATP8B2  
 BAIAP2L1  
 BBS9  
 BCOR\_ENST00000378444  
 BCR  
 BRAF  
 BRD3  
 BRD4\_ENST00000263377  
 C2orf44  
 CANT1  
 CARS  
 CCDC6

图3

CCNB1IP1	CD74_ENST00000009530
CDH11	
CDKN2D	
CDX1	
CEP89	
CHCHD7	
CIC	
CLCN6	
CLIP1	
CLTC	
CNBP	
COL1A1	
COL1A2	
COX6C	
CREB1	
CREB3L1	
CREB3L2	
CRTC1	
CRTC3	
CTAGE5	
CTNNB1	
CUTA	
CXorf67_ENST00000342995	
CYP39A1	
DAZL	
DCTN1	
DDIT3	
DDX5_ENST00000540698	
DUX4L1	
EBF1	
EHF	

图3 (续)

EIF3E	
EIF3K	
ELK4	
EML4	
EPC1	
ERC1	
ERCC2	
ERG	
ERO1L	
ESRP1	
ETV1	
ETV4	
ETV5_ENST00000306376	
ETV6	
EWSR1	
EZR	
FAM131B_ENST000000443739	
FAM22A_ENST000000381707	
FBXL18	
FBXO38	
FCHSD1	
FERMT2	
FEV	
FGFR1	
FGFR1_ENST000000447712	
FGFR3	
FHDC1	
FHIT_ENST000000476844	
FLI1	
FN1	
FOXO1	
FOXO4	

图3 (续)

FUS	GAB2	GABBR2	GATM	GLI1	GMD5	GNAI1	GOLGA5	GOPC	GPBP1L1	HACL1	HAS2	HERPUD1	HEY1	HIP1	HLA-A	HMGA2	HN1	HNRNPAB2B1	HOOK3	IL6R	INTS4	IRF2BP2	ITPR2	IJAK2	IJAZF1	KIAA1549	KIAA1598	KIF5B	KLC1	KLK2	KRAS
-----	------	--------	------	------	------	-------	--------	------	---------	-------	------	---------	------	------	-------	-------	-----	------------	-------	------	-------	---------	-------	-------	--------	----------	----------	-------	------	------	------

图3 (续)

KTN1	
LGR5	
LHFP	
LIFR_ENST00000263409	
LMNA	
LPP	
LRIG3	
LSM14A	
MAML2	
MARK4	
MAST1_ENST00000251472	
MAST2_ENST00000361297	
MBOAT2	
MBTD1_ENST00000586178	
MCPH1	
MEAF6	
MKRN1	
MSN	
MYB_ENST00000341911	
MYO5A	
NAB2	
NACC2	
NCOA1	
NCOA2	
NCOA4_ENST00000452682	
NDRG1	
NF1	
NFATC1	
NFATC2	
NFIA	
NFIB_ENST00000397581	
NFIX_ENST00000360105	

图3 (续)

NONO  
 NOTCH1\_ENST00000277541  
 NPM1  
 NR4A3  
 NRG1  
 NTN1  
 NTRK1\_ENST00000392302  
 NTRK2  
 NTRK3  
 NUP107  
 NUTM1  
 NUTM2B  
 OMD  
 PACS1  
 PATZ1  
 PAX3  
 PAX5  
 PAX7  
 PAX8  
 PBX1  
 PCM1  
 PDE8B  
 PDGFB  
 PHF1  
 PKD1L1  
 PLA2R1  
 PLAG1  
 PLXND1  
 POU5F1  
 PPARG  
 PPFIBP1  
 PRCC

图3 (续)

PRKAR1A  
 PRKAR2A  
 PRKCE  
 PTPRK  
 PWWP2A  
 QKI  
 RAD51B  
 RAF1  
 RANBP2  
 RBM14  
 RBMS1  
 RET  
 RGS22  
 RNF130  
 RNF216  
 ROD1  
 ROS1  
 RRP15  
 RSPO2  
 RSPO3  
 SDC4  
 SEC16A\_NM\_014866.1  
 SEC31A  
 SEPT8  
 SFPQ  
 SIP1  
 SLC22A1  
 SLC26A6  
 SLC34A2  
 SLC45A3  
 SMARCA5  
 SND1

图3 (续)

SP3  
 SQSTM1  
 SRGAP3  
 SS18  
 SS18L1  
 SSBP2  
 SSH2  
 SSX1  
 SSX2  
 SSX4  
 STAT6  
 STRN  
 SUSD1  
 SUZ12  
 SYCP1  
 TACC1  
 TACC3  
 TADA2A\_ENST00000394395  
 TAF15  
 TBCEL  
 TCEA1\_ENST00000521604  
 TCF12  
 TCF3  
 TCF7L2  
 TECTA  
 TFE3  
 TFG  
 THRAP3  
 TMCC1  
 TMPRSS2  
 TP53  
 TPM3

图3 (续)

TPM3\_ENST00000368530  
 TPM4  
 TRIM24  
 TRIM27  
 TRIM33\_ENST00000358465  
 UBE2L3  
 USH1G  
 USP6\_ENST00000250066  
 VCL  
 VTI1A  
 WDFY2  
 WIF1\_ENST00000286574  
 WT1  
 YWHAE\_ENST00000264335  
 YY1  
 ZC3H7B  
 ZCCHC8  
 ZNF384  
 ZNF444\_ENST00000337080  
 ZNF700  
 ZNF703  
 ZSCAN30

图3 (续)

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF463	EML4	13	1751	-	ALK	20	4080	-	170	25%
COSF474	EML4	6	929+220	-	ALK	20	4080	-	93	13.68%
COSF465	EML4	20	2504	-	ALK	20	4080	-	48	7.06%
COSF734	EML4	6	929	-	ALK	20	4080	-	31	4.56%
COSF488	EML4	18	2318+654	y	ALK	20	4080-172	-	4	.59%
COSF1545	EML4	6	929	-	ALK	20	4080-18	-	2	.29%
COSF1376	EML4	18	2318	-	ALK	20	4080	-	2	.29%
COSF480	EML4	2	470	-	ALK	20	4080-117	-	2	.29%
COSF475	EML4	15	2029	-	ALK	20	4080	-	2	.29%
COSF733	EML4	17	2229	y	ALK	20	4080	-	1	.15%
COSF1063	EML4	13	1751	-	ALK	20	4080-69	-	1	.15%
COSF1065	EML4	14	1903	-	ALK	20	4080	-	1	.15%
COSF1128	EML4	14	1903	-	ALK	20	4080	-	1	.15%
COSF1297	EML4	6	929	-	ALK	19	3975	-	1	.15%
COSF1368	EML4	17	2229+2522	-	ALK	20	4126	-	1	.15%
COSF1540	EML4	13	1751	-	ALK	20	4080-90	-	1	.15%
COSF1542	EML4	14	1903	-	ALK	20	4080-123	-	1	.15%
COSF1543	EML4	2	470	-	ALK	20	4080	-	1	.15%
COSF731	EML4	20	2504	-	ALK	20	4080-18	-	1	.15%
COSF493	EML4	6	929+805	y	ALK	20	4080-115	-	1	.15%
COSF491	EML4	14	1903	y	ALK	20	4129	-	1	.15%
COSF490	EML4	20	2504+182	-	ALK	20	4080-67	-	1	.15%

图4A

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF489	EML4	13	1751+447	-	ALK	20	4080-161	-	1	.15%
COSF476	EML4	6	929+(7320)	-	ALK	20	4080	-	1	.15%
COSF464	EML4	20	2504+545	-	ALK	20	4080-232	-	1	.15%
COSF462	EML4	13	1751+(3600)	-	ALK	20	4080-297	-	1	.15%
COSF410	EML4	13	1751+1485	-	ALK	20	4080-1254	-	1	.15%
COSF414	EML4	13	1751+2575	-	ALK	20	4080-203	-	1	.15%
COSF473	EML4	?	?	-	ALK	?	?	-	307	45%
总的突变									680	100%

图4A (续)

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF1272	CCDC6	1	535	-	RET	12	2369	-	583	98.65%
COSF1516	CCDC6	2	685	-	RET	12	2369	-	3	.51%
COSF1532	CCDC6	1	535+1111	-	RET	12	2369-807	-	1	.17%
COSF1533	CCDC6	1	535+1054	-	RET	12	2369-?	-	1	.17%
COSF1518	CCDC6	8	1462+848	-	RET	11	2237	-	1	.17%
COSF1480	CCDC6	?	?	-	RET	?	?	-	2	0%
总的突变									591	100%

图4B

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF125	TMPRSS2	1(utr)	79	-	ERG	4	312	-	476	20.82%
COSF128	TMPRSS2	2	150	-	ERG	4	312	-	38	1.66%
COSF123	TMPRSS2	1(utr)	79	-	ERG	2(utr)	124	-	27	1.18%
COSF126	TMPRSS2	1(utr)	79	-	ERG	5	530	-	21	.92%
COSF127	TMPRSS2	2	150	-	ERG	2(utr)	124	-	7	.31%
COSF736	TMPRSS2	2	150	-	ERG	3	226	-	4	.17%
COSF130	TMPRSS2	3	373	-	ERG	4	312	-	3	.13%
COSF135	TMPRSS2	3	373	-	ERG	2(utr)	124	-	3	.13%
COSF124	TMPRSS2	1(utr)	79	-	ERG	3	226	-	3	.13%
COSF116	TMPRSS2	5	580	-	ERG	4	312	-	3	.13%
COSF129	TMPRSS2	2	150	-	ERG	5	530	-	3	.13%
COSF139	TMPRSS2	1(utr)	79+?	-	ERG	4	312	-	1	.04%
COSF138	TMPRSS2	1(utr)	79	-	ERG	4	312	-	1	.04%
COSF118	TMPRSS2	4	460	-	ERG	4	312	-	1	.04%
COSF121	TMPRSS2	?	?	-	ERG	?	?	-	1695	74%
总的突变									2286	100%

图4C



突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到 的外显子	推论的 断点	插入的 序列	基因名称	第一个观察到 的外显子	推论的 断点	插入的 序列		
COSF1492	NCOA4_ENST00000452682	8	1014	-	RET	12	2369	-	351	97.23%
COSF1341	NCOA4_ENST00000452682	7	870+196	-	RET	12	2369-1570	-	3	.83%
COSF1531	NCOA4_ENST00000452682	8	1014+21	-	RET	12	2369-1591	-	1	.28%
COSF1499	NCOA4_ENST00000452682	8	996	-	RET	12	2369-?	-	1	.28%
COSF1501	NCOA4_ENST00000452682	8	1014+1186	-	RET	11	2273	-	1	.28%
COSF1502	RET	11	2271	-	NCOA4_ENST00000452682	9	1015-490	-	1	.28%
COSF1495	RET	11	2368+?	-	NCOA4_ENST00000452682	8	871-?	-	1	.28%
COSF1493	NCOA4_ENST00000452682	7	870+42	-	RET	12	2369-1673	-	1	.28%
COSF1497	RET	11	2368+?	-	NCOA4_ENST00000452682	9	1015	-	1	.28%
总的突变									361	100%

图4D

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF1233	KIF5B	15	2183	-	RET	12	2327	-	36	59.02%
COSF1231	KIF5B	16	2372+476	-	RET	12	2327-436	-	7	11.48%
COSF1610	KIF5B	16	2372	-	RET	12	2327	-	3	4.92%
COSF1254	KIF5B	22	2897	-	RET	12	2327	-	3	4.92%
COSF1235	KIF5B	23	3002	-	RET	12	2327	-	2	3.28%
COSF1242	KIF5B	24	3219+372	-	RET	8	1713-416	-	1	1.64%
COSF1256	KIF5B	15	2183	-	RET	11	2177	-	1	1.64%
COSF1263	KIF5B	24	3219	-	RET	11	2070	-	1	1.64%
COSF1242	KIF5B	24	3219+372	-	RET	7	1713-416	-	1	1.64%
COSF1241	KIF5B	23	3002+152	y	RET	12	2327-12	-	1	1.64%
COSF1239	KIF5B	15	2183+2335	-	RET	12	2327-1785	-	1	1.64%
COSF1238	KIF5B	15	2183+550	y	RET	12	2327-741	-	1	1.64%
COSF1237	KIF5B	15	2183+2204	-	RET	12	2327-318	-	1	1.64%
COSF1240	KIF5B	16	2372+334	-	RET	12	2327-450	-	1	1.64%
COSF1252	KIF5B	?	?	-	RET	?	?	-	2	3%
总的突变									61	100%

图4E

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF1059	KIF5B	24	3219	-	ALK	20	4080	-	4	40%
COSF1382	KIF5B	15	2183	-	ALK	20	4080	-	2	20%
COSF1258	KIF5B	17	2490	-	ALK	20	4080	-	2	20%
COSF1061	KIF5B	15	2183+2477	-	ALK	19	4005	-	1	10%
COSF1485	KIF5B	?	?	-	ALK	?	?	-	1	10%
总的突变									10	100%

图4F

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF166	EWSRI	7	1112	-	FLII	6	920	-	736	54.04%
COSF168	EWSRI	7	1112	-	FLII	5	854	-	345	25.33%
COSF170	EWSRI	10	1364	-	FLII	6	920	-	41	3.01%
COSF172	EWSRI	10	1364	-	FLII	5	854	-	33	2.42%
COSF177	EWSRI	7	1112	-	FLII	8	1046	-	15	1.1%
COSF181	EWSRI	7	1112	-	FLII	7	986	-	9	.66%
COSF185	EWSRI	10	1364	-	FLII	8	1046	-	4	.29%
COSF178	EWSRI	9	1331	-	FLII	4	650	-	4	.29%
COSF184	EWSRI	9	1331	-	FLII	7	986	-	3	.22%
COSF228	EWSRI	7	1112	-	FLII	9	1094	-	3	.22%
COSF1303	EWSRI	10	1364	-	FLII	7	986	-	2	.15%
COSF183	EWSRI	8	1293	-	FLII	7	986	-	1	.07%
COSF205	EWSRI	8	1293	-	FLII	6	920	-	1	.07%
COSF179	EWSRI	?	?	-	FLII	?	?	-	165	12%
总的突变									1362	100%

图4G

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF155	EWSR1	7	1112	-	ERG	8	967	-	27	22.13%
COSF154	EWSR1	7	1112	-	ERG	11	1141	-	15	12.3%
COSF161	EWSR1	7	1112	-	ERG	9	1036	-	5	4.1%
COSF162	EWSR1	10	1364	-	ERG	8	967	-	4	3.28%
COSF156	EWSR1	7	1112	-	ERG	10	1093	-	3	2.46%
COSF159	EWSR1	?	?	-	ERG	?	?	-	68	56%
总的突变									122	100%

图4H

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF1201	CD74_ENST000000009530	6	627	-	ROS1	34	5757	-	32	84.21%
COSF1203	CD74_ENST000000009530	6	627	-	ROS1	32	5448	-	3	7.89%
COSF1619	CD74_ENST000000009530	?	?	-	ROS1	?	?	-	3	8%
总的突变									38	100%

图4I

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF1220	PAX8	10	1355	-	PPARG	2	216	-	22	25.58%
COSF1224	PAX8	8	1064	-	PPARG	2	216	-	7	8.14%
COSF1218	PAX8	9	1253	-	PPARG	2	216	-	4	4.65%
COSF1225	PAX8	7	943	-	PPARG	2	216	-	1	1.16%
COSF1216	PAX8	?	?	-	PPARG	?	?	-	52	60%
总的突变									86	100%

图4J

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF199	NPM1	5	487	-	ALK	20	4080	-	301	94.36%
COSF448	NPM1	5	487+192	-	ALK	20	4080-616	-	1	.31%
COSF449	NPM1	5	487+201	-	ALK	20	4080-24	-	1	.31%
COSF450	NPM1	5	487+230	-	ALK	20	4080-1134	-	1	.31%
COSF451	NPM1	5	487+295	-	ALK	20	4080-954	-	1	.31%
COSF458	NPM1	5	487+897	-	ALK	20	4080-1511	-	1	.31%
COSF453	NPM1	5	487+443	-	ALK	20	4080-1040	-	1	.31%
COSF454	NPM1	5	487+546	-	ALK	20	4080-583	-	1	.31%
COSF455	NPM1	5	487+726	-	ALK	20	4080-660	-	1	.31%
COSF456	NPM1	5	487+778	-	ALK	20	4080-208	-	1	.31%
COSF457	NPM1	5	487+865	-	ALK	20	4080-629	-	1	.31%
COSF201	NPM1	5	487	-	ALK	20	4080	-	1	.31%
COSF452	NPM1	5	487+358	-	ALK	20	4080-1128	-	1	.31%
COSF420	NPM1	?	?	-	ALK	?	?	-	6	2%
总的突变									319	100%

图4K



突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF482	KIAA1549	15	5097	-	BRAF	9	1202	-	318	53.45%
COSF484	KIAA1549	14	4779	-	BRAF	9	1202	-	105	17.65%
COSF486	KIAA1549	15	5097	-	BRAF	11	1376	-	39	6.55%
COSF1229	KIAA1549	17	5302	-	BRAF	10	1239	-	4	.67%
COSF1227	KIAA1549	14	4779	-	BRAF	11	1376	-	3	.5%
COSF1477	KIAA1549	14	4779+34	-	BRAF	10	1239	-	1	.17%
COSF510	KIAA1549	17	5302	-	BRAF	10	1239	-	1	.17%
COSF1284	KIAA1549	15	5097	-	BRAF	10	1239	-	1	.17%
COSF1472	KIAA1549	18	5448	-	BRAF	9	1202	-	1	.17%
COSF1475	KIAA1549	12	4401	-	BRAF	9	1202	-	1	.17%
COSF512	KIAA1549	18	5448	-	BRAF	9	1202	-	1	.17%
COSF588	KIAA1549	?	?	-	BRAF	?	?	-	120	20%
总的突变									595	100%

图4L

突变ID	5'配偶体基因				3'配偶体基因				突变数	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF501	SS18	10	1286	-	SSX2	6	439	-	334	93.3%
COSF587	SS18	10	1286+?	-	SSX2	6	439-?	-	5	1.4%
COSF530	SS18	10	1286+9829	-	SSX2	6	439	-	2	.56%
COSF567	SS18	10	1286	y	SSX2	6	439-1995	-	1	.28%
COSF579	SS18	10	1286+1984	-	SSX2	6	439-684	-	1	.28%
COSF580	SSX2	5	438+280	-	SS18	11	1287-12012	-	1	.28%
COSF583	SS18	10	1286+1455	-	SSX2	6	439-1412	-	1	.28%
COSF584	SS18	10	1286+4538	-	SSX2	6	439-606	-	1	.28%
COSF585	SSX2	5	438+420	-	SS18	11	1287-12550	-	1	.28%
COSF586	SSX2	5	438+1318	-	SS18	11	1287-9312	-	1	.28%
COSF524	SSX2	4	388+1711	-	SS18	11	1287-12732	-	1	.28%
COSF523	SS18	10	1286+701	-	SSX2	5	389-6	-	1	.28%
COSF517	SS18	10	1286	-	SSX2	4	313	-	1	.28%
COSF574	SS18	10	1286+10808	-	SSX2	4	293	-	1	.28%
COSF576	SSX2	?	?	-	SS18	?	?	-	6	2%
总的突变									358	100%

图4M

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF504	SS18	10	1286	-	SSX1	6	467	-	569	95.95%
COSF514	SS18	10	1286	-	SSX1	5	417	-	1	.17%
COSF519	SS18	10	1286	-	SSX1	6	467-103	-	1	.17%
COSF581	SS18	10	1286+8793	-	SSX1	6	467-?	-	1	.17%
COSF528	SS18	9	1152	-	SSX1	4	321	-	1	.17%
COSF577	SS18	10	1286+1941	-	SSX1	6	467-1313	-	1	.17%
COSF578	SSX1	5	466+575	-	SS18	11	1287-12019	-	1	.17%
COSF506	SS18	9	1152	-	SSX1	5	417	-	1	.17%
COSF521	SS18	10	1286	y	SSX1	4	321	-	1	.17%
COSF575	SSX1	?	?	-	SS18	?	?	-	15	3%
COSF515	SS18	?	?	-	SSX1	?	?	-	1	0%
总的突变									593	100%

图4N

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF1197	SLC34A2	4	429	-	ROSI	32	5448	-	8	66.67%
COSF1260	SLC34A2	13	2076	-	ROSI	32	5448	-	1	8.33%
COSF1618	SLC34A2	?	?	-	ROSI	?	?	-	3	25%
总的突变									12	100%

图40

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF248	PAX3	7	1539	-	FOXO1	2	1016	-	357	92.49%
COSF906	FOXO1	1	1015	-	PAX3	8	1540	-	6	1.55%
COSF343	PAX3	?	?	-	FOXO1	?	?	-	23	6%
总的突变									386	100%

图4P

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSFI512	PRKARIA	7	825	-	RET	12	2369	-	30	100%
总的突变										100%
									30	100%

图4Q

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF1268	EZR	10	1259	-	ROS1	34	5757	-	6	85.71%
COSF1396	EZR	10	1259+207	-	ROS1	34	1259+210	-	1	14.29%
总的突变									7	100%

图4R

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF115	TMPRSS2	2	150	-	ETVI	6	850	-	2	20%
COSF134	TMPRSS2	2	150	-	ETVI	7	904	-	1	10%
COSF122	TMPRSS2	?	?	-	ETVI	?	?	-	7	70%
总的突变									10	100%

图4S



突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变频率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF572	ETV6	5	1283	-	NTRK3	15	1908	-	118	81.38%
COSF1535	ETV6	4	737	-	NTRK3	14	1719	-	15	10.34%
COSF1537	ETV6	5	1283	-	NTRK3	14	1719	-	1	.69%
COSF824	ETV6	4	737	-	NTRK3	15	1908	-	1	.69%
COSF889	NTRK3	14	1907	-	ETV6	6	1284	-	1	.69%
COSF495	ETV6	?	?	-	NTRK3	?	?	-	9	6%
总的突变									145	100%

图4T

突变ID	5'配偶体基因				3'配偶体基因				突变的数目	突变率
	基因名称	最后观察到的外显子	推论的断点	插入的序列	基因名称	第一个观察到的外显子	推论的断点	插入的序列		
COSF288	PAX7	7	1753	-	FOXO1	2	1016	-	98	98.99%
COSF344	PAX7	?	?	-	FOXO1	?	?	-	1	1%
总的突变									99	100%

图4U

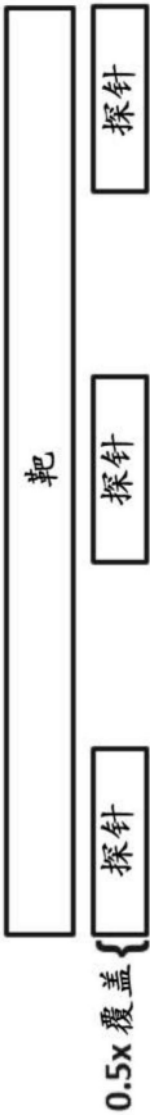
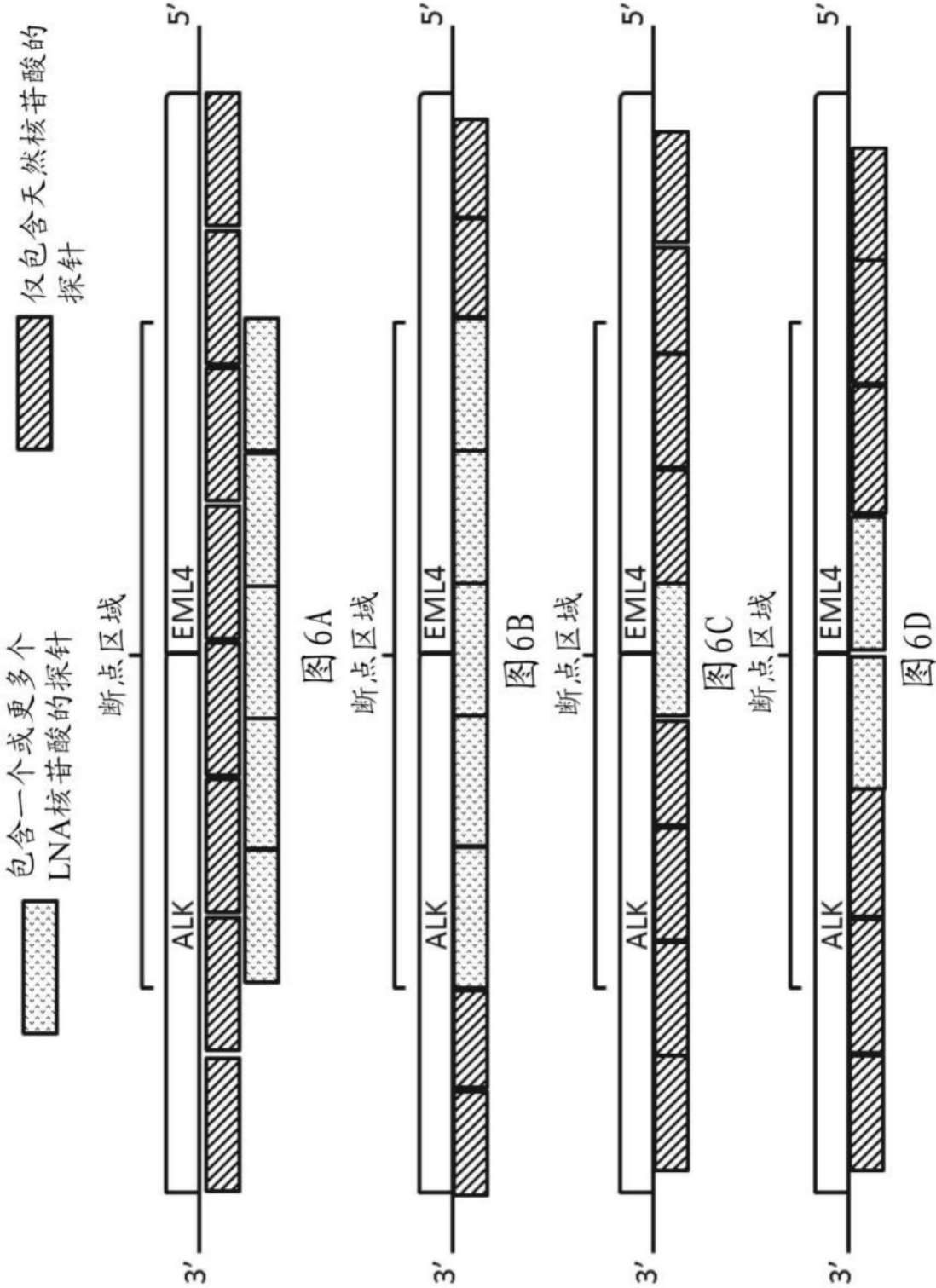


图5A



图5B



68基因族群

具有完整的外显子和部分的内含子覆盖的基因

APC	AR	ARID1A	BRAF	BRCA1	BRCA2	CCND1	CCND2	CCNE1
CDK4	CDK6	CDKN2A	CDKN2B	EGFR*	ERBB2	FGFR1	FGFR2	HRAS
KIT	KRAS	MET	MYC	NF1	NRAS	PDGFRA	PIK3CA	PTEN
RAF1	TP53							

粗体的基因是还被分析了拷贝数变异(CMV)的那些。

\*包括外显子19和20中的得失位

具有决定性的外显子覆盖的基因

AKT1	ALK	ARAF	ATM	CDH1	CTNNB1	ESR1	EZH2	FBXW7
FGFR3	GATA3	GNA11	GNAQ	GNAS	HNF1A	IDH1	IHD2	JAK2
JAK3	MAP2K1	MAP2K2	MLH1	MPL	NFE2L2	NOTCH1	NPM1	NTRK1
PTPN11	RET	RHEB	RHOA	RIT1	ROS1	SMAD4	SMO	SRC
STK11	TERT*	VHL						

\*包括启动子区域

重排

ALK	NTRK1	RET	ROS1
-----	-------	-----	------

图7

基因	染色体	起始	终止	长度(bps)	覆盖的外显子
ALK	chr2	29446405	29446655	250	内含子 19
ALK	chr2	29446062	29446197	135	内含子 20
ALK	chr2	29446198	29446404	206	20
ALK	chr2	29447353	29447473	120	内含子 19
ALK	chr2	29447614	29448316	702	内含子 19
ALK	chr2	29448317	29448441	124	19
ALK	chr2	29449366	29449777	411	内含子 18
ALK	chr2	29449778	29449950	172	18

图8

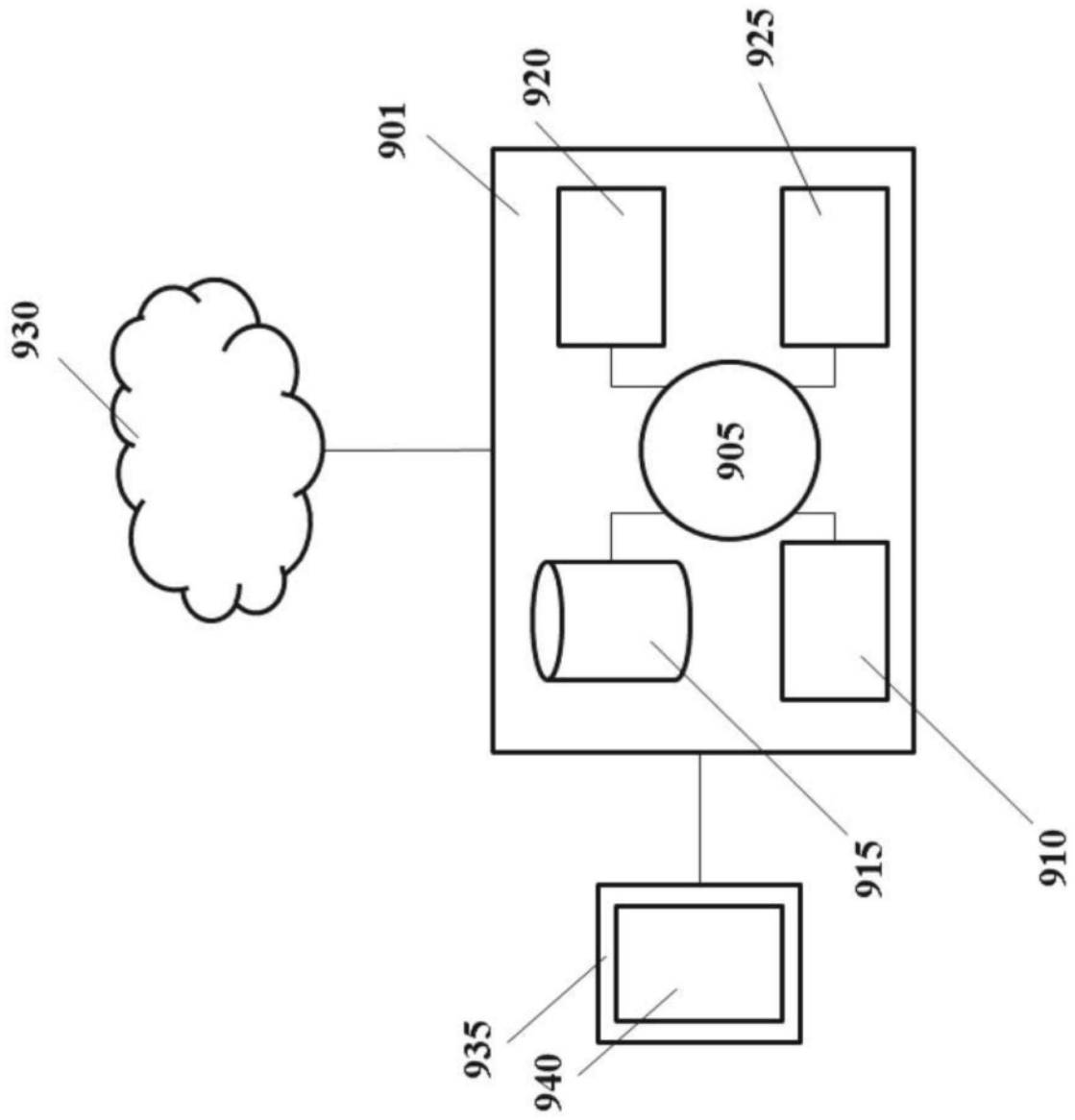


图9