(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2016 年 6 月 23 日 (23.06.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/095093 A1

(51) 国际专利分类号:
C12Q 1/68 (2006.01)

(21) 国际申请号: PCT/CN2014/093871

(22) 国际申请日: 2014 年 12 月 15 日 (15.12.2014)

(25) 申请语言: 中文

(26) 公布语言: 中文

(71) 申请人: **天津华大基因科技有限公司 (BGI TIANJIN)** [CN/CN]; 中国天津市滨海新区空港经济区环河北路 80 号空港商务园东区 3 号楼 101、201 室, Tianjin 300308 (CN)。 **深圳华大基因科技有限公司 (BGI SHENZHEN CO., LIMITED)** [CN/CN]; 中国广东省深圳市盐田区北山路 146 号北山工业区综合楼 11F-3, Guangdong 518083 (CN)。

(72) 发明人: **蔡宇航 (CAI, Yuhang)**; 中国天津市滨海新区空港经济区环河北路 80 号空港商务园东区 3 号楼 101、201 室, Tianjin 300308 (CN)。 **陈希 (CHEN, Xi)**; 中国天津市滨海新区空港经济区环河北路 80 号空港商务园东区 3 号楼 101、201 室, Tianjin 300308 (CN)。 **钱朝阳 (QIAN, Zhaoyang)**; 中国天津市滨海新区空港经济区环河北路 80 号空港商务园东区 3 号楼 101、201 室, Tianjin 300308 (CN)。 **管彦芳 (GUAN, Yanfang)**; 中国天津市滨海新区空港经济区环河北路 80 号空港商务园东区 3 号楼 101、201 室, Tianjin 300308 (CN)。 **易鑫 (YI, Xin)**; 中国天津市滨海新区空港经济区环河北路 80 号空港商务园东区 3 号楼 101、201 室, Tianjin 300308 (CN)。 **朱红梅 (ZHU, Hongmei)**; 中国天津市滨海新区空港经济区环河北路 80 号空港商务园东区 3 号楼 101、201 室, Tianjin 300308 (CN)。 **杨玲 (YANG, Ling)**; 中国天津市滨海新区空港经济区环河北路 80 号空港商务园东区 3 号楼 101、201 室, Tianjin 300308 (CN)。 **吴仁花 (WU, Renhua)**; 中国天津市滨海新区空港经济区环河北路 80 号空港商务园东区 3 号楼 101、201 室, Tianjin 300308 (CN)。

(74) 代理人: **北京清亦华知识产权代理事务所（普通合伙） (TSINGYIHUA INTELLECTUAL PROPERTY LLC)**; 中国北京市海淀区清华园清华大学照澜院商业楼 301 室, Beijing 100084 (CN)。

(81) 指定国 (除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国 (除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:
— 包括国际检索报告(条约第 21 条(3))。

(54) **Title**: METHOD FOR SCREENING TUMOR, METHOD AND DEVICE FOR DETECTING VARIATION OF TARGET REGION

(54) **发明名称** : 肿瘤筛查方法、目标区域变异检测方法和装置

(57) **Abstract**: Disclosed is a method for detecting the variation of a target region, comprising: obtaining the nucleic acid of a test sample, wherein the nucleic acid is composed of a plurality of nucleic acid fragments; capturing the nucleic acid fragments using a kit to obtain the target region; sequencing the target region to obtain the sequencing data, wherein the sequencing data is composed of a plurality of reads; and detecting the variation of the target region on the basis of the sequencing data. The kit comprises the probes capable of recognizing the following predetermined region specifically: the gene regions of at least 10 genes of the 547 genes in Table 1. Also disclosed are a device for detecting the variation of the target region and a method for screening tumor.

(57) **摘要**: 公开了一种检测目标区域变异的方法，包括：获取待测样本中的核酸，所述核酸由多个核酸片段组成；利用试剂盒捕获所述核酸片段，获得目标区域；对所述目标区域进行序列测定，获得测序数据，所述测序数据由多个读段组成；基于所述测序数据，检测所述目标区域中的变异；其中，所述试剂盒包含探针，所述探针能够特异性识别以下预定区域：表 1 里的 547 个基因中的至少 10 个基因的基因区域。还公开了检测目标区域变异的装置和对肿瘤进行筛查的方法。

WO 2016/095093 A1

肿瘤筛查方法、目标区域变异检测方法和装置

技术领域

本发明涉及生物医学领域。具体而言，本发明涉及目标区域变异检测方法、目标区域变异检测装置、肿瘤筛查方法和装置。

背景技术

肿瘤分为良性肿瘤和恶性肿瘤，恶性肿瘤亦称为癌症，为由控制细胞生长增殖机制失常而引起的疾病，癌症是威胁人类健康的疾病之一。癌细胞除了生长失控外，还会局部侵入周遭正常组织甚至经由体内循环系统或淋巴系统转移到身体其他部分。

目前针对癌症的筛查诊断、监控等的手段仍有待改进。

发明内容

本发明旨在至少解决现有技术中存在的技术问题之一或者至少提供一种手段。

依据本发明的一方面，本发明提供一种检测目标区域变异的方法，所述方法包括：（1）获取待测样本中的核酸，所述核酸由多个核酸片段组成，所述核酸片段来自断裂的基因组DNA 和/或游离 DNA；（2）利用试剂盒捕获所述核酸片段，获得目标区域；（3）对所述目标区域进行序列测定，获得测序数据，所述测序数据由多个读段组成；（4）基于所述测序数据，检测所述目标区域中的变异；其中，所述试剂盒包含探针，所述探针能够特异性识别以下预定区域：表 1 里的 547 个基因中的至少 10 个基因的基因区域。

表 1

ABL1	C1R	DIS3	FGF19	HSPA4	MIR142	PAX5	RB1	SRSF2
ABL2	C1S	DNMT1	FGF23	IDH1	MITF	PBRM1	REL	SSTR2
ACVR1B	CARD11	DNMT3A	FGF3	IDH2	MLH1	PCBP1	RET	STAG2
ACVR2A	CASP8	DOT1L	FGF4	IFNAR1	MLH3	PCM1	RHEB	STAT4
AJUBA	CBFB	DUSP6	FGF6	IFNAR2	MLL	PDGFRA	RICTOR	STAT5B
AKT1	CBL	EDNRA	FGF7	IGF1	MLL2	PDGFRB	RNASEL	STK11
AKT2	CBLB	EGFR	FGFR1	IGF1R	MLL3	PDK1	RNF43	SUFU
AKT3	CBR1	EGR3	FGFR2	IGF2	MLL4	PHF6	ROBO1	SUZ12
ALK	CCND1	EIF4A2	FGFR3	IKBKB	MPL	PIGF	ROBO2	SYK
ALOX12B	CCND2	ELAC2	FGFR4	IKBKE	MRE11A	PIK3C2A	ROS1	TAF1
ANGPT1	CCND3	ELF3	FH	IKZF1	MS4A1	PIK3C2B	RPA1	TBL1XR1
ANGPT2	CCNE1	EML4	FLCN	IL7R	MSH2	PIK3C2G	RPL22	TBX3
APC	CD79A	EP300	FLT1	INHBA	MSH3	PIK3C3	RPL5	TEK
APCDD1	CD79B	EPCAM	FLT3	IRF4	MSH4	PIK3CA	RPS14	TERT
AR	CDC25C	EPHA2	FLT4	IRS2	MSH5	PIK3CB	RPS6KB1	TET2

ARAF	CDC42	EPHA3	FNTA	ITGB2	MSH6	PIK3CG	RPTOR	TFG
ARFRP1	CDC73	EPHA5	FOXA1	JAK1	MSR1	PIK3R1	RUNX1	TGFBR2
ARHGAP35	CDH1	EPHB1	FOXA2	JAK2	MTOR	PIK3R2	RUNX1T1	TIPARP
ARID1A	CDK12	EPHB2	FOXL2	JAK3	MUC1	PLK1	RXRA	TLR4
ARID1B	CDK2	EPHB6	FPGS	JUN	MUTYH	PML	RXRB	TMEM127
ARID2	CDK4	EPPK1	FUBP1	KAT6A	MYC	PMS1	RXRG	TNFAIP3
ARID5B	CDK6	ERBB2	FYN	KDM5A	MYCL1	PMS2	SDHAF2	TNFRSF14
ASXL1	CDK8	ERBB3	GAB2	KDM5C	MYCN	PNRC1	SDHB	TNFRSF8
ATM	CDKN1A	ERBB4	GATA1	KDM6A	MYD88	POLQ	SDHC	TNFSF11
ATR	CDKN1B	ERCC2	GATA2	KDR	NAV3	PPP2R1A	SDHD	TNFSF13B
ATRX	CDKN2A	ERCC3	GATA3	KEAP1	NBN	PRDM1	SEMA3A	TOP1
AURKA	CDKN2B	ERG	GID4	KIF1B	NCOA1	PRKAA1	SEMA3E	TOP2A
AURKB	CDKN2C	ESR1	GNA11	KIF5B	NCOA2	PRKAR1A	SETBP1	TOP2B
AXIN1	CDX2	ETV1	GNA13	KIT	NCOR1	PRKCA	SETD2	TP53
AXIN2	CEBPA	ETV6	GNAQ	KLF4	NEK11	PRKCB	SF1	TRAF7
AXL	CFLAR	EWSR1	GNAS	KLHL6	NF1	PRKCG	SF3B1	TSC1
B2M	CHD1	EXT1	GNRHR	KRAS	NF2	PRKDC	SH2B3	TSC2
B4GALT3	CHD2	EXT2	GPR124	LCK	NFE2L2	PRSS8	SIN3A	TSHR
BACH1	CHD4	EZH2	GRIN2A	LIMK1	NFE2L3	PSMB1	SLAMF7	TSHZ2
BAK1	CHEK1	FAM123B	GRM3	LRRK2	NFKBIA	PSMB2	SLC4A1	TSHZ3
BAP1	CHEK2	FAM46C	GSK3B	LYN	NKX2-1	PSMB5	SLIT2	TUBA1A
BARD1	CHUK	FANCA	H3F3A	MALAT1	NKX3-1	PTCH1	SMAD2	TUBB
BCL2	CIC	FANCC	H3F3C	MAP2K1	NOTCH1	PTCH2	SMAD3	TUBD1
BCL2A1	CRBN	FANCD2	HCK	MAP2K2	NOTCH2	PTEN	SMAD4	TUBE1
BCL2L1	CREBBP	FANCE	HDAC1	MAP2K4	NOTCH3	PTP4A3	SMARCA1	TUBG1
BCL2L11	CRIPAK	FANCF	HDAC2	MAP3K1	NOTCH4	PTPN11	SMARCA4	TYR
BCL2L2	CRKL	FANCG	HDAC3	MAP3K13	NPM1	PTPRD	SMARCB1	U2AF1
BCL6	CRLF2	FANCI	HDAC4	MAPK1	NR3C1	RAC1	SMARCD1	USP9X
BCOR	CROT	FANCL	HDAC6	MAPK3	NRAS	RAC2	SMC1A	VEGFA
BCORL1	CSF1R	FANCM	HDAC8	MAPK8	NSD1	RAD21	SMC3	VEGFB
BCR	CTCF	FAT3	HGF	MAPK8IP1	NTRK1	RAD50	SMO	VEZF1
BLM	CTLA4	FBXW7	HIF1A	MAX	NTRK2	RAD51	SOCS1	VHL
BMPR1A	CTNNA1	FCGR1A	HIST1H1C	MC1R	NTRK3	RAD51B	SOX10	WHSC1L1

BRAF	CTNNB1	FCGR2A	HIST1H2BD	MCL1	NUP93	RAD51C	SOX17	WISP3
BRCA1	CUL4A	FCGR2B	HIST1H3B	MDM2	PAK3	RAD51D	SOX2	WWP1
BRCA2	CUL4B	FCGR2C	HNF1A	MDM4	PAK7	RAD52	SOX9	XIAP
BRIP1	CYLD	FCGR3A	HRAS	MECOM	PALB2	RAD54L	SPEN	XPA
BTG1	CYP17A1	FCGR3B	HRH2	MED12	PARP1	RAF1	SPOP	XPC
BTK	DAXX	FGF10	HSD17B3	MEF2B	PARP2	RARA	SPRY4	XPO1
C11orf30	DDR1	FGF12	HSD3B2	MEN1	PARP3	RARB	SRC	XRCC3
C1QA	DDR2	FGF14	HSP90AA1	MET	PARP4	RARG	SRD5A2	YES1
ZNF217	ZNF703	ZRSR2	WT1					

在本发明的一个实施例中，所述预定区域为所述 547 个基因中的至少 20 个、30 个、40 个、50 个、60 个、70 个、80 个、90 个、100 个、200 个、300 个、400 个、500 个或者全部 547 个基因的基因区域。本发明的方法中的试剂盒探针能够特异性识别的基因区域组合，是发明人经过多次收集、多次筛选和多次试验组合获得的，这些基因区域组合是常见肿瘤的发生或发育相关区域。所述常见肿瘤包括肺癌、结直肠癌、胃癌、乳腺癌、肾癌、胰腺癌、卵巢癌、子宫内膜癌、甲状腺癌、宫颈癌、食管癌和肝癌。利用本发明一方面的方法能够一次性、简单方便且高特异性的获取多种常见癌症的相关基因序列，并且检测分析这些相关基因序列，检测分析结果可以辅助用于多种常见癌症的早期筛查判断，增加人为的早期干预肿瘤的发生发展的可能性和效果。目前大部分癌症如肺癌、肝癌、胃癌等在医院病理确诊时已是晚期，耽误了较早的治疗时间大大的减少治愈的可能性。

在本发明的一个实施例中，所述预定区域为所述 547 个基因中的表 2 所列的 145 个基因的基因区域。探针能够特异性识别的表 2 的 145 个基因的基因区域，是发明人经过多次收集、多次筛选和多次试验组合获得的，这些基因区域组合与肺癌的发生发展相关。利用本发明方法中的这一试剂盒中的探针，能够一次性的、简单方便且高特异性的获取全部肺癌相关基因序列，而且基于检测这些基因序列获取的信息能够辅助肺癌的早期筛查诊断。

表 2

KRAS	ALK	ROS1	ADAM23	KIAA0907	KRTAP5-5	MAP1B
EGFR	RB1	FGFR3	DNMT3B	GAB1	TSHZ3	ZNF814
TP53	PDGFRA	FGFR4	SDHAP2	OR10Z1	XIRP2	ZFHX4
BRAF	KDR	JAK3	DHX9	CNTNAP3B	NYAP2	ZNF804A
PIK3CA	FBXW7	APC	CSNK2A1	IL32	NUDT11	OR5D18
ERBB2	HRAS	FRG1B	CNTN5	NAV3	SNAPC4	ZNF479
CDKN2A	JAK2	CHEK2	ATXN3	TNRC6A	ZNF598	OR51V1
NRAS	ERBB4	KLK1	CLIP1	FAM135B	KIAA2022	OR4N2
STK11	KIT	NBPF10	OR4M2	VGLL3	DDX11L2	OR4C15

NFE2L2	SMAD4	PARG	OR10G8	KRTAP4-11	MUC6	OR14C36
CTNNB1	FGFR2	FBN2	PAPPA2	ANAPC1	ATXN1	CROCC
MET	DDR2	HSD17B7P2	OR8H2	FAM47C	MUC16	OR2T2
PTEN	ATM	WASH2P	PBX2	AKAP6	BEST3	PCDH11X
AKT1	RET	POTEC	POLDIP2	ZNF804B	DSPP	REG3A
KEAP1	NOTCH1	EEF1B2	SLC6A10P	ZEB1	MB21D2	REG1B
DDX11	EPB41L4A	TBX6	PRB2	OR2T34	NTRK3	LRRIQ3
DNAH8	OR2M2	WDR62	CNTNAP2	LPA	NTRK1	EPHA5
OR2B11	OR4C16	DCAF4L2	CDH10	MMP27	NF1	OR5L2
OR4K2	KCNB2	EPHA3	CDH12	VAV3	INHBA	OR2T33
FAM47A	STAG3L2	PTPRD	RALGAPB	THSD4	FGFR1	GNA15
RYR2	KRTAP4-8	NOTCH2	FOLH1	OR4N4		

在本发明的一个实施例中，所述预定区域为所述 547 个基因中的表 3 所列 60 个基因的基因区域。探针能够特异性识别的表 3 的 60 个基因的基因区域，是发明人经过多次收集、多次筛选和多次试验组合获得的，这些基因区域组合与结直肠癌的发生发展相关。利用本发明方法中的这一试剂盒中的探针，能够一次性的、简单方便且高特异性的获取全部结直肠癌相关基因序列，而且基于检测这些基因序列获取的信息能够辅助结直肠癌的早期筛查诊断。

表 3

KRAS	SRC	TLR3	EP300	TMPRSS13	EPHA5
BRAF	PTEN	MC4R	CYLD	PHF2	EPHA3
APC	AXIN1	MLH1	FBN2	OPRD1	PTPRD
TP53	FLG	AKT1	NF1	LILRB5	NTRK3
PIK3CA	LIG1	CASD1	ASXL1	COL18A1	NTRK1
CTNNB1	MAP2K1	PTCH1	SMAD4	LARP4B	ALK
NRAS	PIK3R1	ADAMTS18	IRF5	DMKN	ROS1
EGFR	ERBB2	MSH2	DOCK3	ROBO2	RET
FBXW7	STK11	BAP1	MYOM1	KCNN3	PDGFRA
ARID1A	IL7R	CTNNA1	NEFH	INHBA	FGFR1

在本发明的一个实施例中，所述预定区域为所述 547 个基因中的表 4 所列 43 个基因的基因区域。探针能够特异性识别的表 4 的 43 个基因的基因区域，是发明人经过多次收集、多次筛选和多次试验组合获得的，这些基因区域组合与妇科生殖道肿瘤的发生发展相关。所说的生殖道肿瘤包括卵巢癌、子宫内膜癌和宫颈癌。利用本发明方法中的这一试剂盒中的探针，能够一次性的、简单方便且高特异性的获取全部生殖道肿瘤相关基因序列，而且基于检测这些基因序列获取的信息能够辅助生殖道肿瘤的早期筛查诊断。

表 4

AFF3	BRCA2	FBXW7	MED12	PDE4DIP	STK11
AKAP9	CDK12	FGFR2	MLL2	PIK3CA	TP53

AKT1	CDKN2A	FGFR3	MLL3	PIK3R1	
APC	CREBBP	FOXL2	MSH6	PPP2R1A	
ARID1A	CSMD3	GNAS	NF1	PTEN	
BCOR	CTNNB1	HRAS	NFE2L2	RB1	
BRAF	EGFR	KIT	NRAS	RNF213	
BRCA1	FAT3	KRAS	NSD1	RNF43	

在本发明的一个实施例中，所述探针的长度为 25-300nt，较佳的，为 50-250nt，更佳的，为 80nt-120nt。

为获得能够在同一反应体系中同时特异性捕获所说的基因区域的探针，在本发明的一个实施例中，探针是通过先获得初始探针集，再筛选所述初始探针集来确定的。获取所述初始探针集包括：确定所述基因区域的参考序列，从所述参考序列的一端开始，在所述参考序列上依次获取 DNA 片段直至所述参考序列的另一端，其中，一条 DNA 片段为一条初始探针，全部所述 DNA 片段构成所述初始探针集，所述 DNA 片段之间完全重叠、部分重叠或完全不重叠，所述初始探针集能够覆盖所述基因区域至少一次。所说的基因区域的参考序列可以从参考基因组上获取，例如从人参考基因组 HG19 上获得对应的基因区域，所有的 HG19 上的对应的基因区域构成所说的基因区域的参考序列，HG19 可以从 NCBI 数据库下载。在本发明的一个实施例中，利用迭代算法设计获取所述初始探针集，包括：确定所述基因区域在参考基因组上的位置，获取所述基因区域的参考序列，从所述参考序列的第一个核苷酸开始拷贝所述参考序列获取第一条 DNA 片段，从所述参考序列的第二个核苷酸开始拷贝所述参考序列获取第二条 DNA 片段，从所述参考序列的第三个核苷酸开始拷贝所述参考序列获取第三条 DNA 片段，这样依次获取后续 DNA 片段直至第 N 条 DNA 片段的一端超出所述参考序列，其中，一条 DNA 片段为一条初始探针，全部所述 DNA 片段构成所述初始探针集，N 为所述初始探针集中包含的初始探针的总数，以获得能够全面覆盖目标基因区域的初始探针集，而且为使最终的探针具高特异性，在本发明的一个实施例中，进一步对所述筛选初始探针集，包括：将所述 DNA 片段（初始探针集）与所述参考序列比对，获得每一条 DNA 片段在参考序列上的比对次数，过滤掉比对次数超过 1 的 DNA 片段。为使最终的探针能在同一反应体系中捕获所说的基因区域，和/或使捕获的基因区域在同一反应条件下被一起洗脱下来，进一步对所述初始探针集进行筛选，包括：去除掉 GC 含量不在 35-70%的 DNA 片段。

在本发明的一个实施例中，步骤（2）和步骤（3），包括：（h）末端修复所述核酸片段，获得末端修复片段；（i）加碱基 A 至所述末端修复片段的两端，获得粘性末端片段；（j）连接接头于所述粘性末端片段的两端，获得接头连接片段；（k）对所述接头连接片段进行第一扩增，获得第一扩增产物；（l）利用所述试剂盒对所述第一扩增产物进行捕获，获得所述目标区域；（m）对所述目标区域进行第二扩增，获得第二扩增产物；以及（n）对所

述第二扩增产物进行序列测定，获得所述测序数据。本发明这一方面方法中的步骤（2）和（3）包括的测序文库构建，特别适用于样本含微量核酸的测序文库的构建，在本发明的一个实施例中，样本为含微量游离 DNA 的血浆样本，包含极其微量的目标游离 DNA，第一扩增使得核酸的量能满足芯片/探针杂交捕获的需求，而因芯片杂交捕获会损耗一定量的核酸，第二扩增能使捕获下的目标片段获得再次扩增以满足上机测序和质控检测的要求。这一文库构建方法特别适用于总游离核酸不低于 10ng 或者常规组织基因组 DNA 不低于 1μg 的样本的测序文库构建。而本发明这一方面方法中的步骤（2）和（3）包括的对所构建的测序文库的测序，测序可以利用已知平台进行，包括但不限于 Illumina 的 Hiseq2000/2500 平台、Life Technologies 的 Ion Torrent 平台和单分子测序平台。测序方式可以选择单端测序，也可以是双末端测序，在本发明的一个实施例中利用双末端测序，所得的测序数据由多对读段对组成。利用本发明的这一方面的方法的步骤（2）和（3）构建的目标区域文库以及对文库测序，测序后的下机数据质量高，基于高质量的下机数据利于后续的准确检测分析。

在本发明的一个实施例中，步骤（4）包括：将所述测序数据与参考序列进行第一比对，获得第一比对结果；将所述第一比对结果与所述参考序列的一部分进行第二比对，获得第二比对结果，所述参考序列的一部分包括目标区域参考序列中的每个已知 InDel 位点，以及所述每个已知 InDel 位点上下游各 1000bp 的参考序列；基于所述第一比对结果和第二比对结果，同时检测所述目标区域中的 SNP、InDel、SV 和 CNV 变异。这里，所说的第二比对为局部比对，第一比对为常规全局比对，可利用但不限于 SOAP 或 BWA 等软件依照其默认设置进行，获得第一比对结果，第一比对结果包括读段在参考序列上的匹配位置及匹配情况信息。而进行第二比对即基于第一比对结果，对与所捕获的基因区域对应的参考序列中的所有已知 INDEL 附近的所有序列信息（reads）进行局部重新比对，能够消除第一比对中的错误，提高后续变异检测的准确性，第二比对可利用 GATK 比对软件（https://www.broadinstitute.org/gatk/ ）进行。在本发明的一个实施例中，通过 GATK UnifiedGenotyper 软件同时检测所说的 SNP 和 INDEL 变异。利用本发明的这一方面的变异检测方法，能够准确检测出突变频率为 1%的低频突变。

为使变异检测结果更准确可信，在本发明的一个实施例中，在所述第一比对之前，对所述测序数据进行过滤，所述过滤包括去除掉不确定碱基比例超过 10%的读段和/或碱基质量值不大于 5 的碱基数的比例不小于 50%的读段。并且任选地，在所述第二比对之前，去除掉第一比对结果中的一个读段对中的两个读段相同的读段对。所说的参考序列的一部分包括目标区域参考序列中的每个已知 InDel 位点，以及所述每个已知 InDel 位点上下游各 1000bp 的参考序列。当待测样本来自于人时，参考序列可选择为人参考基因组，即已知的人基因组序列，例如为 HG19，HG19 可以从 NCBI 下载得来，而所说的目标区域参考序列为与目标区域匹配的那部分参考基因组序列。

在本发明的一个实施例中，步骤（4）还包括，当所检测出的变异中的至少之一满足以下（i）或（ii），则判定所述待测样本为阳性样本：（i）在阴性对照样本中的读段支持数少于 2 条和在阳性对照样本中的突变率大于 1%，（ii）测序深度不小于 10X，至少有 3 个读段的支持，在阴性对照样本中的读段支持数少于 2 条，在阳性对照样本中的突变率大于 1%，以及该变异位点的读段支持量与正常对照样本相同位点的读段支持量具有显著差异。所说的阳性样本指肿瘤样本，两个判定条件是发明人结合目前相关数据库信息和大量文献报道信息、检测统计大量阳性样本和大量阴性样本确定下来的，具有统计意义，后者比前者更为严格，较佳的，这里的阳性或者阴性对照样本超过 30 个，对照样本的数据可以自己对对照样本的核酸进行提取、序列测定来获得，也可以依照他人已公开或公开数据库中的的样本测序数据，多个对照样本数据使统计判定条件/结果具有统计意义，更加可信。依据任一两个判定条件进行判定的结果可以辅助用于临床的肿瘤诊断筛查，可辅助用于了解所测样本个体癌变的种类、可能性及发展情况等。需要说明的是，所说的变异位点在待测样本中的读段支持量与正常对照样本（阴性对照样本）相同位点的读段支持量具有显著差异，其中的读段支持量，可以为支持该变异的读段的数目，也可以是支持该变异的读段在比对上该位点读段中的比例，在本发明的一个实施例中，采用后者来比较，所说的具有显著差异指具有实质差异，例如对于待测样本中的变异位点 A，多个阳性样本（cancer 样本）中的 reads 支持比例都为 5/400(变异 5 条 reads，总 400 条 reads)，即阳性样本中该位点的平均变异频率 1.25%，而在多个阴性对照样本中的 reads 支持比例都为 1/200（变异 1 条 reads，总 200 条 reads），即阴性对照样本中的平均变异频率 0.5%，若待测样本中的该变异位点的变异频率更接近 1.25%，例如达到 0.9%，则达到所说的显著差异或者所说的实质差异。具有显著差异，也可以指统计学上的对数据差异性的评价——显著性差异，例如对待测样本中的变异位点 A 进行多次检测，获得该位点的多组比对结果数据，从每组比对结果数据中都可获得一个读段支持比例，所说的读段支持比例=支持该变异位点的读段数/比对上该位点的总读段数，接着比较待测样本的变异位点 A 的读段支持比例（变异频率）与阴性对照样本中的该位点的突变频率的差异，例如可以利用 z 检验或 t 检验，差异具有显著性（$p \leq 0.05$），即认为达到所说的具有显著差异。

依据本发明的另一方面，本发明还提供一种检测目标区域变异的装置，用以实现或执行上述本发明一方面的或者任一具体实施方式的目标区域变异检测方法，所述装置包括：数据获取单元，用于获取所述目标区域的测序数据，所述测序数据由多个读段和/或多个读段对组成，在所述数据获取单元中进行：获取待测样本中的核酸，所述核酸由多个核酸片段组成，所述核酸片段来自断裂的基因组 DNA 和/或游离 DNA，利用试剂盒捕获所述核酸片段，获得目标区域，对所述目标区域进行序列测定，其中，所述试剂盒包含探针，所述探针能够特异性识别以下预定区域：表 1 里的 547 个基因中的至少 10 个基因的基因区域；

检测单元，用于基于来自数据获取单元的测序数据，检测所述目标区域变异，所述变异包括 SNP、InDel、SV 和 CNV 至少之一。本领域人员可以理解，本发明的这一方面的装置中的全部或部分单元，可选择的、可拆卸的包含一个或多个子单元以执行或实现前述本发明方法的各个具体实施方式。

例如，在本发明的一个实施例中，如图 1 所示，装置 1000 中的检测单元 200 包括第一比对子单元 13、第二比对子单元 15 和变异识别子单元 17，所述第一比对子单元 13 用以将来自数据获取单元 100 的测序数据与参考序列进行第一比对，获得第一比对结果，所述第二比对子单元 15 用以将来自所述第一比对子单元 13 的第一比对结果与所述参考序列的一部分进行第二比对，获得第二比对结果，所述变异识别子单元 17 用以基于来自所述第一比对子单元 13 的第一比对结果和来自所述第二比对子单元 15 的第二比对结果，同时检测所述目标区域中的 SNV、InDel、SV 和 CNV 变异，获得变异位点信息，其中，所述参考序列的一部分包括目标区域参考序列中的每个已知 InDel 位点，以及所述每个已知 InDel 位点上下游各 1000bp 的参考序列。

在本发明的一个实施例中，如图 2 所示，所述装置 1000 的检测单元 200 还包括第一过滤子单元 12，所述第一过滤子单元 12 与所述第一比对子单元 13 连接，用于在所述测序数据进入所述第一比对子单 13 元之前，对所述测序数据进行过滤，所述过滤包括去除掉不确定碱基比例超过 10%的读段和/或碱基质量值不大于 5 的碱基数的比例不小于 50%的读段。任选的，如图 3 所示，所述检测单元 200 还包括第二过滤子单元 14，所述第二过滤子单元分别 14 与所述第一比对子单元 13 和所述第二比对子单元 15 连接，用于在所述第一比对结果进入所述第二比对子单元 15 之前，去除掉来自所述第一比对子单元 13 的第一比对结果中的一个读段对中的两个读段相同的读段对。上述参考序列可以为 HG19，所述第一比对单元中进行的第一比对为全局比对，所述第二比对子单元中进行的第二比对为局部比对。

在本发明的一个实施例中，如图 4 所示，所述装置 1000 中的检测单元 200 还包括判定子单元 19，所述判定子单元 19 用以判定来自所述变异识别子单元 17 中的变异位点是否满足以下，当所述变异位点中的至少一个满足以下则判定所述待测样本为阳性样本：在阴性对照样本中的读段支持数少于 2 条和在阳性对照样本中的突变率大于 1%。在本发明的另一个实施例中，所述检测单元 200 还包括判定子单元 19，所述判定子单元 19 用以判定来自所述变异识别子单元 17 中的变异位点是否满足以下，当所述变异位点中的至少一个满足以下则判定所述待测样本为阳性样本：测序深度不小于 10X，至少有 3 个读段的支持，在阴性对照样本中的读段支持数少于 2 条，在阳性对照样本中的突变率大于 1%，以及变异位点的读段支持量与正常对照样本相同位点的读段支持量具有显著差异。前述对本发明一方面或者任一具体实施方式中的目标区域变异检测方法的技术特征和优点的描述，同样适用于本发明这一方面的检测装置，在此不再赘述。

依据本发明的再一方面，提供一种对肿瘤进行筛查的方法，特别是早期筛查，所述肿瘤包括但不限于肺癌、结直肠癌、胃癌、乳腺癌、肾癌、胰腺癌、卵巢癌、子宫内膜癌、甲状腺癌、宫颈癌、食管癌和肝癌，所述方法包括：获取待测样本中的核酸，所述核酸由多个核酸片段组成，所述核酸片段来自断裂的基因组 DNA 和/或游离 DNA；利用试剂盒捕获所述核酸片段，获得目标区域；对所述目标区域进行序列测定，获得测序数据，所述测序数据由多个读段组成；基于所述测序数据，检测所述目标区域中的变异，基于检测出的变异中的至少之一满足以下（i）或者（ii），判定所述待测样本为阳性样本：（i）在阴性对照样本中的读段支持数少于 2 条和在阳性对照样本中的突变率大于 1%，（ii）测序深度不小于 10X，至少有 3 个读段的支持，在阴性对照样本中的读段支持数少于 2 条，在阳性对照样本中的突变率大于 1%，以及其读段支持量与正常对照样本相同位点的读段支持量具有显著差异；其中，所述试剂盒包含探针，所述探针能够特异性识别以下预定区域：表 1 里的 547 个基因中的至少 10 个基因的基因区域。上述对本发明一方面的目标区域变异检测方法和/装置的相应技术特征和优点的描述，同样适用于本发明这一方面的癌症早期筛查方法，在此不再赘述。本领域技术人员可以理解，本发明这一方面的方法中的全部或部分步骤，可以通过包含对应功能单元的装置来实现。

在本发明的一个实施例中，所述预定区域为所述 547 个基因中的至少 20 个、30 个、40 个、50 个、60 个、70 个、80 个、90 个、100 个、200 个、300 个、400 个、500 个或者全部 547 个基因的基因区域。本发明的方法中的试剂盒探针能够特异性识别的基因区域组合，是发明人经过多次收集、多次筛选和多次试验组合获得的，这些基因区域组合是常见肿瘤的发生或发育相关区域。所述常见肿瘤包括肺癌、结直肠癌、胃癌、乳腺癌、肾癌、胰腺癌、卵巢癌、子宫内膜癌、甲状腺癌、宫颈癌、食管癌和肝癌。利用本发明一方面的方法能够一次性、简单方便且高特异性的获取多种常见癌症的相关基因序列，并且检测分析这些相关基因序列，检测分析结果可以辅助用于多种常见癌症的早期筛查判断，增加人为的早期干预肿瘤的发生发展的可能性和效果。目前大部分癌症如肺癌、肝癌、胃癌等在医院病理确诊时已是晚期，耽误了较早的治疗时间大大的减少治愈的可能性。

在本发明的一个实施例中，所述预定区域为所述 547 个基因中的表 2 所列的 145 个基因的基因区域。探针能够特异性识别的表 2 的 145 个基因的基因区域，是发明人经过多次收集、多次筛选和多次试验组合获得的，这些基因区域组合与肺癌的发生发展相关。利用本发明方法中的这一试剂盒中的探针，能够一次性的、简单方便且高特异性的获取全部肺癌相关基因序列，而且基于检测这些基因序列获取的信息能够辅助肺癌的早期筛查诊断。

在本发明的一个实施例中，所述预定区域为所述 547 个基因中的表 3 所列 60 个基因的基因区域。探针能够特异性识别的表 3 的 60 个基因的基因区域，是发明人经过多次收集、多次筛选和多次试验组合获得的，这些基因区域组合与结直肠癌的发生发展相关。利用本

发明方法中的这一试剂盒中的探针，能够一次性的、简单方便且高特异性的获取全部结直肠癌相关基因序列，而且基于检测这些基因序列获取的信息能够辅助结直肠癌的早期筛查诊断。

在本发明的一个实施例中，所述预定区域为所述 547 个基因中的表 4 所列 43 个基因的基因区域。探针能够特异性识别的表 4 的 43 个基因的基因区域，是发明人经过多次收集、多次筛选和多次试验组合获得的，这些基因区域组合与妇科生殖道肿瘤的发生发展相关。所说的生殖道肿瘤包括卵巢癌、子宫内膜癌和宫颈癌。利用本发明方法中的这一试剂盒中的探针，能够一次性的、简单方便且高特异性的获取全部生殖道肿瘤相关基因序列，而且基于检测这些基因序列获取的信息能够辅助生殖道肿瘤的早期筛查诊断。

在本发明的一个实施例中，所述探针的长度为 25-300nt，较佳的，为 50-250nt，更佳的，为 80nt-120nt。为获得能够在同一反应体系中同时特异性捕获所说的基因区域的探针，在本发明的一个实施例中，探针是通过先获得初始探针集，再筛选所述初始探针集来确定的。获取所述初始探针集包括：确定所述基因区域的参考序列，从所述参考序列的一端开始，在所述参考序列上依次获取 DNA 片段直至所述参考序列的另一端，其中，一条 DNA 片段为一条初始探针，全部所述 DNA 片段构成所述初始探针集，所述 DNA 片段之间完全重叠、部分重叠或完全不重叠，所述初始探针集能够覆盖所述基因区域至少一次。所说的基因区域的参考序列可以从参考基因组上获取，例如从人参考基因组 HG19 上获得对应的基因区域，所有的 HG19 上的对应的基因区域构成所说的基因区域的参考序列，HG19 可以从 NCBI 数据库下载。在本发明的一个实施例中，利用迭代算法设计获取所述初始探针集，包括：确定所述基因区域在参考基因组上的位置，获取所述基因区域的参考序列，从所述参考序列的第一个核苷酸开始拷贝所述参考序列获取第一条 DNA 片段，从所述参考序列的第二个核苷酸开始拷贝所述参考序列获取第二条 DNA 片段，从所述参考序列的第三个核苷酸开始拷贝所述参考序列获取第三条 DNA 片段，这样依次获取后续 DNA 片段直至第 N 条 DNA 片段的一端超出所述参考序列，其中，一条 DNA 片段为一条初始探针，全部所述 DNA 片段构成所述初始探针集，N 为所述初始探针集中包含的初始探针的总数，以获得能够全面覆盖目标基因区域的初始探针集，而且为使最终的探针具高特异性，在本发明的一个实施例中，进一步对所述筛选初始探针集，包括：将所述 DNA 片段（初始探针集）与所述参考序列比对，获得每一条 DNA 片段在参考序列上的比对次数，过滤掉比对次数超过 1 的 DNA 片段。为使最终的探针能在同一反应体系中捕获所说的基因区域，和/或使捕获的基因区域在同一反应条件下被一起洗脱下来，进一步对所述初始探针集进行筛选，包括：去除掉 GC 含量不在 35-70%的 DNA 片段。

在本发明的一个实施例中，所述利用试剂盒捕获所述核酸片段，获得目标区域，以及对所述目标区域进行序列测定，获得测序数据，包括：（h）末端修复所述核酸片段，获得

末端修复片段；（i）加碱基 A 至所述末端修复片段的两端，获得粘性末端片段；（j）连接接头于所述粘性末端片段的两端，获得接头连接片段；（k）对所述接头连接片段进行第一扩增，获得第一扩增产物；（l）利用所述试剂盒对所述第一扩增产物进行捕获，获得所述目标区域；（m）对所述目标区域进行第二扩增，获得第二扩增产物；以及（n）对所述第二扩增产物进行序列测定，获得所述测序数据。本发明这一方面方法中包含的测序文库构建，特别适用于样本含微量核酸的测序文库的构建，在本发明的一个实施例中，样本为含微量游离 DNA 的血浆样本，包含极其微量的目标游离 DNA，第一扩增使得核酸的量能满足芯片/探针杂交捕获的需求，而因芯片杂交捕获会损耗一定量的核酸，第二扩增能使捕获下的目标片段获得再次扩增以满足上机测序和质控检测的要求。这一文库构建方法特别适用于总游离核酸不低于 10ng 或者常规组织基因组 DNA 不低于 1μg 的样本的测序文库构建。而本发明这一方面方法中包含的对所构建的测序文库的测序，测序可以利用已知平台进行，包括但不限于 Illumina 的 Hiseq2000/2500 平台、Life Technologies 的 Ion Torrent 平台和单分子测序平台。测序方式可以选择单端测序，也可以是双末端测序，在本发明的一个实施例中利用双末端测序，所得的测序数据由多对读段对组成。利用本发明的这一方面的方法中的构建目标区域文库以及对文库测序，测序后的下机数据质量高，基于高质量的下机数据利于后续的准确检测分析。

在本发明的一个实施例中，所述基于所述测序数据，检测所述目标区域中的变异，包括：将所述测序数据与参考序列进行第一比对，获得第一比对结果；将所述第一比对结果与所述参考序列的一部分进行第二比对，获得第二比对结果，所述参考序列的一部分包括目标区域参考序列中的每个已知 InDel 位点，以及所述每个已知 InDel 位点上下游各 1000bp 的参考序列；基于所述第一比对结果和第二比对结果，同时检测所述目标区域中的 SNP、InDel、SV 和 CNV 变异。这里，所说的第二比对为局部比对，第一比对为常规全局比对，可利用但不限于 SOAP 或 BWA 等软件依照其默认设置进行，获得第一比对结果，第一比对结果包括读段在参考序列上的匹配位置及匹配情况信息。而进行第二比对即基于第一比对结果，对与所捕获的基因区域对应的参考序列中的所有已知 INDEL 附近的所有序列信息（reads）进行局部重新比对，能够消除第一比对中的错误，提高后续变异检测的准确性，第二比对可利用 GATK 比对软件（https://www.broadinstitute.org/gatk/ ）进行。在本发明的一个实施例中，通过 GATK UnifiedGenotyper 软件同时检测所说的 SNP 和 INDEL 变异。利用本发明的变异检测方法，能够准确检测出突变频率为 1%的低频突变。利用本发明的这一方面的肿瘤筛查方法，能够准确检测出突变频率为 1%的低频突变，利于对肿瘤进行早期筛查，辅助临床诊断，能够对肿瘤进行预防、及早治疗或及时监控跟踪。

为使变异检测结果更准确可信，在本发明的一个实施例中，在所述第一比对之前，对所述测序数据进行过滤，所述过滤包括去除掉不确定碱基比例超过 10%的读段和/或碱基质

量值不大于 5 的碱基数的比例不小于 50%的读段。并且任选地，在所述第二比对之前，去除掉第一比对结果中的一个读段对中的两个读段相同的读段对。所说的参考序列的一部分包括目标区域参考序列中的每个已知 InDel 位点，以及所述每个已知 InDel 位点上下游各 1000bp 的参考序列。当待测样本来自于人时，参考序列可选择为人参考基因组，即已知的人基因组序列，例如为 HG19，HG19 可以从 NCBI 下载得来，而所说的目标区域参考序列为与目标区域匹配的那部分参考基因组序列。

在本发明的一个实施例中，所述基于检测出的变异中的至少之一满足以下（i）或者（ii），判定所述待测样本为阳性样本，其中，所说的阳性样本指肿瘤样本，可以肺癌、结直肠癌、胃癌、乳腺癌、肾癌、胰腺癌、卵巢癌、子宫内膜癌、甲状腺癌、宫颈癌、食管癌或肝癌样本。（i）和（ii）两个判定条件是发明人结合目前相关数据库信息和大量文献报道信息、检测统计大量阳性样本和大量阴性样本确定下来的，具有统计意义，后者比前者更为严格，较佳的，这里的阳性或者阴性对照样本超过 30 个，对照样本的数据可以自己对对照样本的核酸进行提取、序列测定来获得，也可以依照他人已公开或公开数据库中的的样本测序数据，多个对照样本数据使统计判定条件/结果具有统计意义，更加可信。依据任一两个判定条件进行判定的结果可以辅助用于临床的肿瘤诊断筛查，可辅助用于了解所测样本个体癌变的种类、可能性及发展情况等。需要说明的是，所说的变异位点在待测样本中的读段支持量与正常对照样本（阴性对照样本）相同位点的读段支持量具有显著差异，其中的读段支持量，可以为支持该变异的读段的数目，也可以是支持该变异的读段在比对上该位点读段中的比例，在本发明的一个实施例中，采用后者来比较，所说的具有显著差异指具有实质差异，例如对于待测样本中的变异位点 A，多个阳性样本（cancer 样本）中的 reads 支持比例都为 5/400(变异 5 条 reads，总 400 条 reads)，即阳性样本中该位点的平均变异频率 1.25%，而在多个阴性对照样本中的 reads 支持比例都为 1/200（变异 1 条 reads，总 200 条 reads），即阴性对照样本中的平均变异频率 0.5%，若待测样本中的该变异位点的变异频率更接近 1.25%，例如达到 0.9%，则达到所说的显著差异或者所说的实质差异。具有显著差异，也可以指统计学上的对数据差异性的评价——显著性差异，例如对待测样本中的变异位点 A 进行多次检测，获得该位点的多组比对结果数据，从每组比对结果数据中都可获得一个读段支持比例，所说的读段支持比例=支持该变异位点的读段数/比对上该位点的总读段数，接着比较待测样本的变异位点 A 的读段支持比例（变异频率）与阴性对照样本中的该位点的突变频率的差异，例如可以利用 z 检验或 t 检验，差异具有显著性（$p \leq 0.05$），即认为达到所说的具有显著差异。

发明人通过研究结果显示，正常人外周血中的游离血浆 DNA（cfDNA）的浓度为 1-100ng/mL，而肿瘤患者外周血中的循环肿瘤 DNA（ctDNA）含量将明显增加，由于肿瘤细胞分泌、凋亡或坏死所产生基因组片段入血，使肿瘤患者外周血中的 ctDNA 含量平均浓

度可达 180ng/mL，通过对正常人群及肿瘤患者的外周血 ctDNA 的含量变化及突变情况进行定时监控，可应用于下列至少之一：肿瘤的早期诊断检测，遗传性肿瘤预测及状态评估，肿瘤早期发病进展检测，肿瘤术后效果检测评估，肿瘤靶向治疗、化疗治疗基因变异情况分析，肿瘤致病基因微量残留检测，肿瘤耐药性基因变异情况分析。

发明人发现相对与临床传统的肿瘤检测方法，根据本发明或者任一具体实施方式中的方法和/装置，具有下列优势：微创性：受检者只需要提供 5-10mL 外周血样本；实时性：可对受检者进行多次实时采血，早期筛查时可定期检测，监控肿瘤发病风险，肿瘤患者可在手术后、化疗用药/靶向用药后随时检测，以分析手术预后情况及用药的灵敏性、耐药性情况等；高灵敏度：不受限与病灶位置及大小，通过高深度的目标区域捕获测序，可以检测出突变频率为 1%的低频变异，对于肿瘤发病早期以及肿瘤治疗复发后所出现的变异能够及时准确检出；高特异性：在 ctDNA 含量较少的情况下，能够保证较低的假阳性率、假阴性率，确保得到的检测结果能够准确的反应受检者实时外周血状况；高通量：基于新一代测序技术的目标区域捕获测序，能够在很短的时间内同时进行多例样本检测，并且在目标区域捕获芯片的使用下，相同数据量可进行更高深度的数据挖掘。

附图说明

本发明的上述和/或附加的方面和优点从结合下面附图对实施方式的描述中将变得明显和容易理解，其中：

图 1 是本发明的一个实施例中的目标区域变异检测装置的结构示意图；

图 2 是本发明的一个实施例中的目标区域变异检测装置的结构示意图；

图 3 是本发明的一个实施例中的目标区域变异检测装置的结构示意图；

图 4 是本发明的一个实施例中的目标区域变异检测装置的结构示意图。

具体实施方式

本发明中的“变异”、“核酸变异”、“基因变异”可通用，本发明中的“SNP”（SNV）、“CNV”、“插入缺失”（indel）和“结构变异”（SV）同通常定义，但本发明中对各种变异的大小不作特别限定，这样这几种变异之间有的有交叉，比如当插入/缺失的为大片段甚至整条染色体时，也属于发生拷贝数变异（CNV）或是染色体非整倍性，也属于 SV。这些类型变异的大小交叉并不妨碍本领域人员通过上述描述执行实现本发明的方法和/或装置并且达到所描述的结果。

本发明中的“参考序列”为已知基因组序列或者已知基因组序列的至少一部分，本发明中所使用的“第一”、“第二”等仅为方便描述指代，不能理解为指示或暗示相对重要性，也不能理解为有先后顺序关系。本发明的描述中，除非另有说明，“多个”的含义是两个或

两个以上。

下面通过具体的实施例，对本发明进行说明，需要说明的是这些实施例仅仅是为了说明目的，而不能以任何方式解释成对本发明的限制。

获得本发明一方面的试剂盒、实现本发明一方面的方法和/或装置，一般包括目标区域捕获探针/芯片的设计、微量样本建库及杂交上机测序、下机数据的生物信息分析和变异数据解读。

一般方法

除非特别说明，在下面的实施例中，按照下面的一般方法进行：

1．目标区域捕获芯片设计

基于 TCGA、ICGC、COSMIC 等数据库和相关和收集提取大量相关参考文献中的信息，采用迭代算法设计出能够用于或者辅助用于常见癌症或者具体癌症早筛诊断的目标区域捕获探针/芯片。目标区域包括常见癌或者具体癌症的相关的驱动基因（Driver Gene）、高频突变基因、癌症相关 12 条信号通路中重要基因，靶药及化疗药物相关基因等。所设计的探针/芯片可以从复杂的基因组中特异性识别出目标区域，在同一组探针/芯片上以高特异性和高覆盖率捕获目标区域。

2．试验及数据分析

（一）样本制备

抽取受检者外周血 5-10mL，存于 EDTA 抗凝管中，在 4-6 小时内对外周血进行分离，得到血浆游离 DNA（cfDNA），cfDNA 中可能包含来自肿瘤细胞的 DNA（ctDNA）；

cfDNA 定量检测。

（二）文库制备及测序

对 cfDNA 片段进行末端修复；

对 cfDNA 片段末端加 A；

连接 Adapter 文库接头：文库接头（Adapter）是指经过设计的一段碱基序列，作用在于 cfDNA/ctDNA 文库扩增时与引物相结合，使 DNA 扩增进行，并且在上机测序时与测序引物相结合，利于测序引物与待测序位点结合辅助 DNA 测序进行。

文库进行第一轮 PCR 扩增；

扩增后文库质控并与上述探针/芯片杂交；

杂交文库进行第二轮 PCR 扩增；

文库定量及质控；

Illumina HiSeq2500/2000 上机测序，测序深度达 2000X 以上。

（三）目标区域捕获测序下机数据进行生物信息分析

获得下机数据后需进行如下生物信息分析，得到最终的变异结果。

SOAPnuke filter：去除低质量 reads；

与 reference 序列比对，产生 bam 文件；

标记重复序列；

比对结果不好的序列重新比对，并校正质量值；

去除错配序列；

分析下机数据 QC；

寻找变异；

对变异结果进行注释，得到最终数据结果。

（四）变异数据解读

对生物信息分析后的变异数据进行个体化解读，参考构建的肿瘤数据库及相关文献，对受检者检出的变异进行分析，以辅助判断此受检样本的肿瘤相关基因状态，是否有癌症发病风险，以及早期肿瘤组织的良恶性情况等，用以辅助结合临床检验结果给予最适合的预防及治疗方式。

以下结合具体个体样本对利用本发明的方法/装置获得的检测结果进行详细的描述。下面示例，仅用于解释本发明，而不能理解为对本发明的限制。除另有交待，以下实施例中涉及的未特别交待的试剂、序列（接头、标签和引物）、软件及仪器，都是常规市售产品或者开源的，比如购自 Illumina 公司的 hiseq2000 测序平台建库相关试剂盒来进行文库构建等。

实施例 1 常见癌相关基因检测

1. 目标区域捕获芯片设计

基于 TCGA、ICGC、COSMIC 等数据库和收集提取大量相关参考文献中的信息，采用迭代算法设计出能够用于或者辅助用于肿瘤早诊、术后监控以及肿瘤治疗（放化疗、靶向药物治疗等）效果监控的目标区域捕获芯片 CANPer，CANPer 为液相芯片。CANPer 芯片包括了常见高发癌症的相关驱动基因（Driver Gene）、高频突变基因、癌症相关 12 条信号通路中重要基因，共计 547 个基因，300Kb。基因列表详见表 1。

2．以肺癌结节早期患者的外周血血浆作为待检样本，样本来自天津妇幼保健院，内容如下：

（ ）外周血样本分离

1) 采集受检者外周血 1-2 管（5mL/管）于 EDTA 抗凝管中，轻柔上下颠倒（防止细胞破裂）6-8 次充分混匀，在采血当天 4-6 小时内进行以下处理；
2) 在 4℃条件下 1600g 离心 10 分钟，离心后将上清（血浆）分装到多个 1.5mL/2mL 离心管中，在吸取过程中不能吸到中间层白细胞；

3) 在4℃条件下16000g 离心10分钟，去除残余细胞，将上清（血浆）转移到新的1.5mL/2mL离心管中，不能吸到管底白细胞，即得到分离后所需血浆；

4) 血浆样本处理完后，分离得到的血浆及剩余血细胞均保存到-80℃冰箱中，避免反复冻融。

（二） 血浆游离DNA提取（采用QIAamp Circulating Nucleic Acid Kit）

1)加30μ L蛋白酶K至1.5mL离心管中；

2)加入300μ L血浆；

3)加入240μ L Buffer ACL和1.68μ L Carrier RNA（0.2μ g/μ L），涡旋振荡30s，60℃温浴30min，温浴期间适当取出振荡；

4)加入540μ L Buffer ACB，涡旋振荡15-30s，冰上或-20℃冰箱放置5min；

5)取700μ L血浆混合物加入过滤柱中，7500rpm离心30s；

6)过滤柱空甩8000rpm，1min；

7)加入600μ L Buffer ACW1，8000rpm，1min离心洗涤；

8)加入700μ L Buffer ACW2，8000rpm，1min离心洗涤；

9)加入700μ L无水乙醇，8000rpm，1min离心洗涤；

10) 过滤柱空甩14000rpm，3min；

11) 把过滤柱放入新收集管中，打开盖子，56℃金属浴10min；

12) 将柱子放入新离心管汇总，加入60μ L Buffer AVE回溶3min；

13) 14.000rpm离心1min，Qubit（Invitrogen，the Quant-iTTMdsDNA HS Assay Kit）定量质控所提取的cfDNA。

（三） 文库构建（采用KAPA LTP Library Preparation Kit）

1）末端修复

DNA样本	50μ L
Water	8μ L
10x KAPA End Repair Buffer	7μ L
KAPA End Repair Enzyme Mix	5μ L
Total	70μ L
充分混合，20℃孵育30min	

反应后加入AgencourtAMPureXPreagent 120μ L，磁珠纯化后，使用42μ L ddH_2O回溶，带磁珠进行下一步反应；

2）末端加A

End Repaired DNA	42μ L

10x KAPA A-Tailing Buffer	5μ L
KAPA A-Tailing Enzyme	3μ L
Total	50μ L
充分混合，30℃孵育 30min	

反应后加入 PEG/NaCl SPRI Solution 90μ L，充分混合并进行磁珠纯化，下步 Adapter 连接反应体系中 Adapter 和 ddH_2O 加入量按照下面公式进行计算：10 nM*建库起始 DNA 量（ng）*Adaper 用量（μL）=15 μM（Adapter 浓度）*50μL，使用 ddH_2O 用量（μL）=35 μL-Adapter 用量（μL）回溶，进行下一步反应；

3）接头连接

A-Tailed DNA	（35-Adapter 用量）μ L
15μ M Adapter	终浓度 10nM/ng
5x KAPA Ligation Buffer	10μ L
KAPA T4 DNA Ligase	5μ L
Total	50μ L
充分混合，16℃孵育 16h	
加入 3μ L USER 酶，37℃反应 15min	

反应后，加入 PEG/NaCl SPRI Solution 50μ L，进行第一次磁珠纯化，使用 50μ L Tris-HCl（1mM，pH8.0）回溶；

再加入PEG/NaCl SPRI Solution 50μ L，进行第二次磁珠纯化，使用25μ LTris-HCl（1mM，pH8.0）回溶；

4）第一轮 PCR 扩增

Ligased DNA			20μ L
2x KAPA HiFiHotStart Ready Mix			25μ L
Water			3μ L
25μ M Universal Primer			1μ L
25μ M Index Primer			1μ L
Total			50μ L
PCR 反应条件			
Step	Temp	Time	Cycles
Initial denaturation	98℃	45sec	1
Denaturation Annealing Extension	98℃ 65℃ 72℃	15sec 30sec 30sec	8
Final extension Stop reaction	72℃ 4℃	60sec Hold	1 Hold

反应后加入 AgencourtAMPureXPreagent 90μ L，磁珠纯化后，使用 31μ L ddH_2O 回溶，取上清后质控并进行芯片杂交。

5）目标区域捕获芯片杂交

本实施例中采用委托 Roche 合成的上述基因芯片 CANPer-1.75M，参照芯片制造商提供的说明书进行杂交捕获及洗脱。最后使用 21μ L ddH_2O 回溶杂交洗脱磁珠。

6）第二轮 PCR 扩增

Target Captured DNA			20μ L
2x KAPA HiFiHotStart Ready Mix			25μ L
FellowCell Primer 1			2.5μ L
FellowCell Primer 2			2.5μ L
Total			50μ L
PCR 反应条件			
Step	Temp	Time	Cycles
Initial denaturation	98℃	45sec	1
Denaturation Annealing Extension	98℃ 65℃ 72℃	15sec 30sec 30sec	14
Final extension Stop reaction	72℃ 4℃	60sec Hold	1 Hold

反应后加入 AgencourtAMPureXPreagent 108μ L，磁珠纯化后，使用 31μ L EB 回溶，取上清后质控并上机测序。

7）上机测序

本实施例中，采用 Illumina HiSeq2500 PE101+8+101 程序进行上机测序，测序实验操作按照制造商提供的操作说明书（参见 Illumina/Solexa 官方公布 cBot）进行上机测序操作。

（四）下机数据生物信息分析解读

1) SOAPnuke filter：去除 n≥10%和碱基质量值≤5 的碱基数目＞50%的 reads；
2) Bwaaln->sampe | samtools view | samtools sort：与 reference 序列比对，产生 bam 文件；
3) MarkDuplicates.jar：将同一个 pe 的相同的 reads 标记为重复；
4) GenomeAnalysisTK.jar-T RealignerTargetCreator、IndelRealigner：将比对不好的 reads 重新比对；
5) GenomeAnalysisTK.jar-T BaseRecalibrator、PrintReads：校正质量值；
6) Filt_bam：去掉错配（mismatch）碱基≥3 个的 reads；
7) QC：统计芯片的捕获效率、有效 reads 数、平均深度、重复率、覆盖度及未被覆盖的区间等信息；
8) 识别 SNV / InDel / SV / CNV 及筛选其中的高频变异位点：

用 MuTect（http://www.broadinstitute.org/cancer/cga/mutect ）、varScan（http://massgenomics.org/varscan ）流程识别出 SNP 变异；

用 gatk（https://www.broadinstitute.org/gatk/ ）、varScan、ForestSV（http://sebatlab.ucsd.edu/index.php/software-data ）流程识别出 InDel 变异；

用contra.py（http://contra-cnv.sourceforge.net/ ）流程识别出CNV；

用 ForestSV（http://sebatlab.ucsd.edu/index.php/software-data ）流程识别出 SV；

所使用的筛选参数为：测序深度≥10X，在阴性（正常）样本中的变异率≤2%，在阳性样本中的变异率≥1%，在该待测样本数据中支持该变异的 reads 数≥3，与正常对照（体细胞）的读段支持比例具有显著差异 （$p \leq 0.05$）；

9）注释解读

注释变异的功能、reads 支持数、变异频率、氨基酸变异及数据库 Cosmic 中的变异等；根据变异情况辅助判断疾病的可能来源。化疗药物对肿瘤细胞的杀伤效应与特定的一种（一组）基因的表达和/或多态性显著相关，通过相关基因的检测，预测化疗药物的疗效，选择合适的药物进行个体化化疗，已经成为提高疗效、减少无效治疗的合理选择。基于化疗药物以上特点，参考 PharmGKB 数据库，整合目前临床上所有的化疗药物以及与疗效相关的基因及疗效预测评判，形成化疗个体化用药解读数据库。并将化疗数据整合入肿瘤个体化信息流程，完成化疗药物的自动化解读。

靶向药物在肿瘤治疗中具有药效显著、副作用少的特点，但它对靶点（包括蛋白、DNA 等）有依赖性，必须先对患者做靶点分析，才能确定患者能否用药。整合目前 FDA 批准的靶向药物，以及处于临床III、IV的药物。依据 NCCN 临床指南，临床药物基因研究整理药物靶点基因与靶药疗效关系，形成肿瘤个体化靶药解读数据库。

3．结果分析

该样本检测到 EGFR 基因第 451 位氨基酸错义突变，属于 12 号外显子，该变异位点位于蛋白胞外拓扑结构域内，在 COSMIC 数据库中暂无记载，但相同位点所产生的 p.[R451H] 错义突变记载 1 次，报道与肺癌相关（18948947）。功能预测显示该变异为有害性变异，预测可能对基因功能产生影响。

人表皮生长因子受体，原癌基因 c-erbB1 的表达产物，属于受体酪氨酸激酶家族成员。EGFR 主要位于细胞膜表面，通过与配体的结合激活自身酪氨酸磷酸化，自磷酸化促进下游信号传导途径，包括 MPAK，PI3K 和 JNK 通路等，诱导细胞增殖，分化等。在许多实体肿瘤中存在 EGFR 的突变或异常表达。临床研究表明 EGFR 突变阳性（18 号外显子突变、19 号外显子缺失、21 号外显子突变）的患者对 EGFR-TKI 敏感（23344264），而野生型患者基本无效(23883922)；20 号外显子的突变（T790M 为主，插入突变）与 EGFR-TKI 继发性耐药性相关(22263058)。

实施例 2 肺癌相关基因检测

1．目标区域捕获芯片设计

基于 TCGA、ICGC、COSMIC 等数据库和相关和收集提取大量相关参考文献中的信息，采用迭代算法设计出能够用于或者辅助用于肺癌早筛诊断的目标区域捕获芯片 LungPer。LungPer 芯片包括了肺癌相关的驱动基因（Driver Gene）、高频突变基因、癌症相关 12 条信号通路中重要基因，靶药及化疗药物相关基因等，共计 145 个基因，250Kb。基因列表如表 2 所示。

2．以受检者外周血样本检测为例进行操作说明，样本来自天津妇幼保健院。

同实施例 1 中的 2。

3．测序分析

按照一般方法的步骤进行分析，测序结果如表 5 所示，检测结果见表 6。

表 5

样本名称	总有效数据量(Mb)	目标区域有效数据量(Mb)	有效捕获效率	目标区域平均测序深度	覆盖度	重复率
LC_003	3314.35	1119.46	33.78%	4477.84x	98.2%	47.32%

表 6

Gene	cHGVS	pHGVS	Function	阳性样本变异频率（tumor_var_freq，%)
ALK	c.[334G>A]	p.[A112T]	missense（错义）	1.22
ROS1	c.[1478G>A]	p.[R493H]	missense	1.07
EGFR	c.[1351C>T]	p.[R451C]	missense	1.41
TSC1	c.[2066G>A]	p.[R689H]	missense	1.32
ATM	c.[4723C>T]	p.[R1575C]	missense	1.36

该样本检测到 EGFR 基因第 451 位氨基酸错义突变，属于 12 号外显子，该变异位点位于蛋白胞外拓扑结构域内，在 COSMIC 数据库中暂无记载，但相同位点所产生的 p.[R451H] 错义突变记载 1 次，报道与肺癌相关（18948947）。功能预测显示该变异为有害性变异，预测可能对基因功能产生影响。

人表皮生长因子受体，原癌基因 c-erbB1 的表达产物，属于受体酪氨酸激酶家族成员。EGFR 主要位于细胞膜表面，通过与配体的结合激活自身酪氨酸磷酸化，自磷酸化促进下游信号传导途径，包括 MPAK，PI3K 和 JNK 通路等，诱导细胞增殖，分化等。在许多实体肿瘤中存在 EGFR 的突变或异常表达。临床研究表明 EGFR 突变阳性（18 号外显子突变、19 号外显子缺失、21 号外显子突变）的患者对 EGFR-TKI 敏感（23344264），而野生型患者基本无效(23883922)；20 号外显子的突变（T790M 为主，插入突变）与 EGFR-TKI 继发性耐药性相关(22263058)。

从突变分析结果看出该受检者检出与肺癌发生发展相关的重要变异，再结合临床诊断情况，可判断该受检者患结肺癌的风险及良恶性情况。

实施例 3 结直肠癌相关基因检测

一、芯片设计

基于TCGA、ICGC、COSMIC等数据库和相关和收集提取大量相关参考文献中的信息，采用迭代算法设计出能够用于或者辅助结直肠癌早筛诊断的目标区域捕获芯片ColorectalPer。ColorectalPer 芯片包括了结直肠癌相关的驱动基因（Driver Gene）、高频突变基因、癌症相关 12 条信号通路中重要基因，靶药及化疗药物相关基因等，共计 60 个基因，如表 3 所示，共 123Kb。

二、以受检者外周血样本检测为例进行操作说明，样本来自天津妇幼保健院。

同实施例 1 中的 2。

三、测序分析

按照一般方法的步骤进行分析，测序数据统计结果如表 7，变异检测结果如表 8 所示。

表 7

样本名称	总有效数据量(Mb)	目标区域有效数据量(Mb)	有效捕获效率	目标区域平均测序深度	覆盖度	重复率
CRC_002	3424.38	1129.23	32.9%	9180.73x	99.2%	57.87%

表 8

Gene	cHGVS	pHGVS	Function	tumor_var_freq(%)
KRAS	c.[35G>A]	p.[Gly12Asp]	missense	15.07
ARID1A	c.[805C>T]	p.[Gln269*]	nonsense	1.22
ROS1	c.[5557+750T>G]	.	intron	1.01
NRAS	c.[291-59C>A]	.	intron	1.09
MSH2	c.[1663C>T]	p.[Arg555*]	nonsense	9.38

该样本检出 KRAS p.[Gly12Asp]的错义突变，该变异在 COSMIC 数据库中记载 10303 次，约 60%均报道与大肠癌发病相关。KRAS 第 12 号密码子位于 GTP 结构域上，是 KRAS 最常见突变。

KRAS 属于 Ras 基因家族成员，编码 P21 蛋白，在 MAPK 信号通路中起作用，是致癌基因，能够与 GDP/GTP 结合并促进 GTP 酶活性。当 KRAS 发生突变时不能被水解酶水解失活，处于持续激活状态，引起 RAF/MAPK 的上调，传递多种生存通路信号，从而使细胞过度生长、增殖，抵抗 EGFR-TKIs 的作用。其突变可以导致多种恶性肿瘤，包括肺癌，黏蛋白腺瘤，胰腺导管癌和结肠癌等。KRAS 基因被激活最常见的方式是点突变，多发生在 N 端的第 12、13 和 61、146 密码子，其中以第 12 密码子突变最常见。不同突变位点对 P21 蛋白的活化机制不同，第 12 密码子突变可以减弱 P21 内在的 GTP 酶活性，并使细胞凋亡减少，细胞间接触抑制减弱。

该样本检出 MSH2 p.[Arg555*]的无义突变，该变异在 COSMIC 数据库暂无记载。MSH2 第 555 号密码子所在区域功能研究不详，但此无义突变的发生会导致蛋白编码的提前终止，使得基因主要功能区无法表达，可能使基因功能受到损伤或丧失。

所编码的蛋白是 DNA 错配修复系统（MMR）的组分之一，形成 2 种不同的异二聚体：MutSα（MSH2-MSH6 异二聚体）以及 MutSβ（MSH2-MSH3 异二聚体），可与 DNA 错配

部位相结合因此启动 DNA 的修复功能。在错配结合之后，MutSα 或 β 与 MutLα 异二聚体形成一个三元复合物，负责指导下游的 MMR 事件，包括链的识别、切除以及再合成。ATP 的结合及水解在错配修复功能上起重要作用，ATP 酶活性与 MutSα 相关。MutSα 还可以在 DNA 同源重组修复功能上起作用。此基因与遗传性非息肉性结直肠癌类型 I 以及子宫内膜癌相关。

从突变分析结果看出该受检者检出与结直肠癌发生发展相关的重要变异，再结合临床诊断情况，可判断该受检者患结直肠癌的风险及良恶性情况。

实施例 4 妇科生殖道肿瘤相关基因检测

一、目标区域捕获芯片设计

基于TCGA、ICGC、COSMIC等数据库和相关和收集提取大量相关参考文献中的信息，采用迭代算法设计出能够用于或者辅助妇科生殖道肿瘤早筛诊断的目标区域捕获芯片WCNPer。WCNPer 芯片包括了妇科生殖道肿瘤相关的驱动基因（Driver Gene）、高频突变基因、癌症相关 12 条信号通路中重要基因，靶药及化疗药物相关基因等，共计 43 个基因，如表 4 所示，共 300Kb。

二、以受检者的外周血血浆作为研究对象，样本来自天津妇幼保健院，参考实施例 1 进行试验及数据分析。

三、结果分析

测序数据统计结果如表 9，变异检测结果如表 10 所示。

表 9

样本名称	总有效数据量(Mb)	目标区域有效数据量(Mb)	有效捕获效率	目标区域平均测序深度	覆盖度	重复率
FK_001	3294.38	1099.28	33.4%	3753.03x	98.8%	66.3%

表 10

基因（Gene）	cHGVS	pHGVS	功能变化（Function）	变异频率（tumor_var_freq，%)
AFF3	c.[1310_1312delGCA]	.	cds-indel（cds 区中插入缺失）	18.88889
PPP2R1A	c.[1252_1253insGGGGGG]	.	cds-indel	6.25
NF1	c.[168C>T]	.	coding-synon（编码区同义突变）	3.28
BRCA2	c.[1806delA]	p.[G602fs*?]	Frameshift（移码突变）	4.59
BRAF	c.[1406G>T]	p.[G469V]	Missense（错义突变）	3.02
CTNNB1	c.[1804-56G>A]	.	Intron（内含子突变）	2.42
CREBBP	c.[3925G>A]	p.[D1309N]	missense	2.78
TP53	c.[797G>T]	p.[G266V]	missense	3.73
NF1	c.[7292G>A]	p.[R2431H]	missense	2.03
NF1	c.[6670C>T]	p.[Q2224*]	Nonsense（无义突	3.19

			变）	
BCOR	c.[4438C>T]	p.[R1480*]	nonsense	3.15
AFF3	c.[1091+1G>A]	.	splice-5（5’剪接位点突变）	1.99
CSMD3	c.[10381_10382insA]	p.[W3461delins]	stop-gain（无义突变）	2.5
GNAS	c.[*192A>G]	.	utr-3（3’UTR 区突变）	1.94

该样本检出 BRAF p.[G469V]的错义突变，该变异在 COSMIC 数据库中记载 17 次，在肺、大肠、胆道、上呼吸道、食道等肿瘤中检测发现。BRAF 第 469 号密码子位于蛋白激酶结构域的 ATP 结合区，一项黑色素瘤的研究显示该突变为激活突变，可能导致 BRAF 从非活性状态变为活性状态或使得 BRAF 信号通路异常激活，与疾病的发生发展可能相关。

BRAF 基因编码 MAPK 通路中的丝氨酸苏氨酸蛋白激酶，该酶可将信号由 Ras 转导至 MEK1/2，从而参与细胞功能的调控，影响细胞分类、分化和分泌。此基因产生的突变与多种类型的癌症相关，如结直肠癌、肺癌、肝癌、胰腺癌、甲状腺癌、卵巢癌等。在卵巢癌中，BRAF 基因突变频率为 8%，是卵巢癌发生发展过程中的 driver 基因。

该样本检出 TP53 p.[G266V]的错义突变，该变异在 COSMIC 数据库中记载 43 次，在肺部、大肠、胰腺、卵巢等肿瘤中检测发现。TP53 第 266 号密码子位于序列特异的 DNA 结合结构域，是 TP53 发挥功能的重要结构域，该变异可能使 TP53 的完整功能受到影响或丧失，TP53 是肿瘤发生发展中的 Driver 基因，完整功能受到影响或丧失，可能与疾病发生发展相关。

TP53 基因是迄今发现与肿瘤相关性最高的基因之一。作为重要的抑癌基因，在细胞周期调控，DNA 损伤修复、细胞分化、凋亡和衰老等过程中发挥了关键作用。TP53 基因与 50%以上的人类恶性肿瘤有关。临床研究证实肿瘤中 95.1%的 p53 点突变主要发生在高度保守的 175、245、248、249、273 和 282 位点。目前很多肿瘤治疗通过调控 TP53 蛋白实现。TP53 基因在多种癌症中均有临床应用研究。TP53（外显子 5-8）发生突变的乳腺癌患者预后较差，他莫昔芬疗效也明显降低。TP53 的基因突变与功能丧失是卵巢癌中最常见的基因异常之一。

从突变分析结果看出该受检者检出与妇科疾病相关的重要变异，再结合临床诊断情况，可判断该受检者患妇科肿瘤的风险及良恶性情况。

权利要求书

1．一种检测目标区域变异的方法，其特征在于，包括，

（1）获取待测样本中的核酸，所述核酸由多个核酸片段组成，所述核酸片段来自断裂的基因组 DNA 和/或游离 DNA；

（2）利用试剂盒捕获所述核酸片段，获得目标区域；

（3）对所述目标区域进行序列测定，获得测序数据，所述测序数据由多个读段组成；

（4）基于所述测序数据，检测所述目标区域中的变异；其中，

所述试剂盒包含探针，所述探针能够特异性识别以下预定区域：表 1 里的 547 个基因中的至少 10 个基因的基因区域。

2．权利要求 1 的方法，其特征在于，所述预定区域为所述 547 个基因中的至少 20 个基因的基因区域。

3．权利要求 1 的方法，其特征在于，所述预定区域为所述 547 个基因中的至少 30 个基因的基因区域。

4．权利要求 1 的方法，其特征在于，所述预定区域为所述 547 个基因中的至少 40 个基因的基因区域。

5．权利要求 1 的方法，其特征在于，所述预定区域为所述 547 个基因中的表 2 所列 145 个基因的基因区域。

6．权利要求 1 的方法，其特征在于，所述预定区域为所述 547 个基因中的表 3 所列 60 个基因的基因区域。

7．权利要求 1 的方法，其特征在于，所述预定区域为所述 547 个基因中的表 4 所列 43 个基因的基因区域。

8．权利要求 1 的方法，其特征在于，所述预定区域为所述 547 个基因的基因区域。

9．权利要求 1-8 任一方法，其特征在于，所述探针的长度为 25-300nt。

10．权利要求 1-9 任一方法，其特征在于，步骤（2）和步骤（3），包括：

（h）末端修复所述核酸片段，获得末端修复片段；

（i）加碱基 A 至所述末端修复片段的两端，获得粘性末端片段；

（j）连接接头于所述粘性末端片段的两端，获得接头连接片段；

（k）对所述接头连接片段进行第一扩增，获得第一扩增产物；

（l）利用所述试剂盒对所述第一扩增产物进行捕获，获得所述目标区域；

（m）对所述目标区域进行第二扩增，获得第二扩增产物；以及

（n）对所述第二扩增产物进行序列测定，获得所述测序数据。

11．权利要求 1-10 任一方法，其特征在于，步骤（4）包括，

将所述测序数据与参考序列进行第一比对，获得第一比对结果；

将所述第一比对结果与所述参考序列的一部分进行第二比对，获得第二比对结果，所述参考序列的一部分包括目标区域参考序列中的每个已知 InDel 位点，以及所述每个已知 InDel 位点上下游各 1000bp 的参考序列；

基于所述第一比对结果和第二比对结果，同时检测所述目标区域中的 SNP、InDel、SV 和 CNV 变异。

12．权利要求 11 的方法，其特征在于，所述参考序列为 HG19。

13．权利要求 11 的方法，其特征在于，在所述第一比对之前，对所述测序数据进行过滤，所述过滤包括去除掉不确定碱基比例超过 10%的读段和/或碱基质量值不大于 5 的碱基数的比例不小于 50%的读段。

14．权利要求 11-13 任一的方法，其特征在于，所述测序数据包含读段对，在所述第二比对之前，去除掉第一比对结果中的一个读段对中的两个读段相同的读段对。

15．权利要求 11 的方法，其特征在于，所述第一比对为全局比对，所述第二比对为局部比对。

16．权利要求 1-15 任一方法，其特征在于，步骤（4）还包括，当所检测出的变异中的至少之一满足以下（i）或（ii），则判定所述待测样本为阳性样本：

（i)在阴性对照样本中的读段支持数少于 2 条和在阳性对照样本中的突变率大于 1%，

（ii）测序深度不小于 10X，至少有 3 个读段的支持，在阴性对照样本中的读段支持数少于 2 条，在阳性对照样本中的突变率大于 1%，以及变异位点的读段支持量与正常对照样本相同位点的读段支持量具有显著差异。

17．一种检测目标区域变异的装置，其特征在于，包括，

数据获取单元，用于获取所述目标区域的测序数据，所述测序数据由多个读段和/或多个读段对组成，在所述数据获取单元中进行：

获取待测样本中的核酸，所述核酸由多个核酸片段组成，所述核酸片段来自断裂的基因组 DNA 和/或游离 DNA，

利用试剂盒捕获所述核酸片段，获得目标区域，对所述目标区域进行序列测定，

其中，所述试剂盒包含探针，所述探针能够特异性识别以下预定区域：表 1 里的 547 个基因中的至少 10 个基因的基因区域；

检测单元，用于基于来自数据获取单元的测序数据，检测所述目标区域变异。

18．权利要求 17 的装置，其特征在于，所述检测单元包括第一比对子单元、第二比对子单元和变异识别子单元，

所述第一比对子单元用以将来自数据获取单元的测序数据与参考序列进行第一比对，获得第一比对结果，

所述第二比对子单元用以将来自所述第一比对子单元的第一比对结果与所述参考序列的一部分进行第二比对，获得第二比对结果，

所述变异识别子单元用以基于来自所述第一比对子单元的第一比对结果和来自所述第二比对子单元的第二比对结果，同时检测所述目标区域中的 SNV、InDel、SV 和 CNV 变异，其中，

所述参考序列的一部分包括目标区域参考序列中的每个已知 InDel 位点，以及所述每个已知 InDel 位点上下游各 1000bp 的参考序列。

19．权利要求 18 的装置，其特征在于，所述参考序列为 HG19。

20．权利要求 18 的装置，其特征在于，所述检测单元还包括第一过滤子单元，所述第一过滤子单元与所述第一比对子单元连接，用于在所述测序数据进入所述第一比对子单元之前，对所述测序数据进行过滤，所述过滤包括去除掉不确定碱基比例超过 10%的读段和/或碱基质量值不大于 5 的碱基数的比例不小于 50%的读段。

21．权利要求 18-20 任一装置，其特征在于，所述检测单元还包括第二过滤子单元，所述第二过滤子单元分别与所述第一比对子单元和所述第二比对子单元连接，用于在所述第一比对结果进入所述第二比对子单元之前，去除掉来自所述第一比对子单元的第一比对结果中的一个读段对中的两个读段相同的读段对。

22. 权利要求 18 的装置，其特征在于，所述第一比对子单元中的第一比对为全局比对，所述第二比对子单元中的第二比对为局部比对。

23．权利要求 18-22 任一装置，所述检测单元还包括判定子单元，所述判定子单元用以判定来自所述变异识别子单元中的变异位点是否满足以下，当所述变异位点中的至少一个满足以下（i）或（ii）则判定所述待测样本为阳性样本：

（i)在阴性对照样本中的读段支持数少于 2 条和在阳性对照样本中的突变率大于 1%，

（ii）测序深度不小于 10X，至少有 3 个读段的支持，在阴性对照样本中的读段支持数少于 2 条，在阳性对照样本中的突变率大于 1%，以及该变异位点的读段支持量与正常对照样本相同位点的读段支持量具有显著差异。

24．一种对肿瘤进行筛查的方法，其特征在于，包括：

获取待测样本中的核酸，所述核酸由多个核酸片段组成，所述核酸片段来自断裂的基因组 DNA 和/或游离 DNA；

利用试剂盒捕获所述核酸片段，获得目标区域；

对所述目标区域进行序列测定，获得测序数据，所述测序数据由多个读段组成；

基于所述测序数据，检测所述目标区域中的变异，基于检测出的变异中的至少之一满足以下（i）或者（ii），判定所述待测样本为阳性样本：

（i）在阴性对照样本中的读段支持数少于 2 条和在阳性对照样本中的突变率大于

1%，

（ii）测序深度不小于 10X，至少有 3 个读段的支持，在阴性对照样本中的读段支持数少于 2 条，在阳性对照样本中的突变率大于 1%，以及其读段支持量与正常对照样本相同位点的读段支持量具有显著差异；其中，

所述试剂盒包含探针，所述探针能够特异性识别以下预定区域：表 1 里的 547 个基因中的至少 10 个基因的基因区域。

25．权利要求 24 的方法，其特征在于，所述预定区域为所述 547 个基因中的至少 20 个基因的基因区域。

26．权利要求 24 的方法，其特征在于，所述预定区域为所述 547 个基因中的至少 30 个基因的基因区域。

27．权利要求 24 的方法，其特征在于，所述预定区域为所述 547 个基因中的至少 40 个基因的基因区域。

28．权利要求 24 的方法，其特征在于，所述预定区域为所述 547 个基因中的表 2 所列 145 个基因的基因区域。

29．权利要求 24 的方法，其特征在于，所述预定区域为所述 547 个基因中的表 3 所列 60 个基因的基因区域。

30．权利要求 24 的方法，其特征在于，所述预定区域为所述 547 个基因中的表 4 所列 43 个基因的基因区域。

31．权利要求 24 的方法，其特征在于，所述预定区域为所述 547 个基因的基因区域。

32．权利要求 24-31 任一方法，其特征在于，所述探针的长度为 25-300nt。

33．权利要求 32 的方法，其特征在于，所述探针的获得包括，获得初始探针集以及筛选所述初始探针集。

34．权利要求 33 的方法，其特征在于，所述获得初始探针集包括：

确定所述基因区域的参考序列，

从所述参考序列的一端开始，在所述参考序列上依次获取 DNA 片段直至所述参考序列的另一端，其中，

一条 DNA 片段为一条初始探针，全部所述 DNA 片段构成所述初始探针集，所述 DNA 片段之间完全重叠、部分重叠或完全不重叠，所述初始探针集能够覆盖所述基因区域至少一次。

35．权利要求 33 的方法，其特征在于，所述获取初始探针集包括：

确定所述基因区域在参考基因组上的位置，获取所述基因区域的参考序列，

从所述参考序列一端的第一个核苷酸开始拷贝所述参考序列获取第一条 DNA 片段，

从所述参考序列一端的第二个核苷酸开始拷贝所述参考序列获取第二条 DNA 片段，

从所述参考序列一端的第三个核苷酸开始拷贝所述参考序列获取第三条 DNA 片段，

这样依次获取后续 DNA 片段直至第 N 条 DNA 片段的一端超出所述参考序列的另一端，其中，

一条 DNA 片段为一条初始探针，全部所述 DNA 片段构成所述初始探针集，N 为所述初始探针集中包含的初始探针的总数。

36．权利要求 34 或 35 的方法，其特征在于，所述筛选初始探针集包括：

将所述 DNA 片段与所述参考序列比对，获得每一条 DNA 片段在参考序列上的比对次数，过滤掉比对次数超过 1 的 DNA 片段。

37．权利要求 36 的方法，其特征在于，所述筛选初始探针还包括，去除掉 GC 含量不在 35-70%的 DNA 片段。

38．权利要求 24-37 任一方法，其特征在于，所述肿瘤包括肺癌、结直肠癌、胃癌、乳腺癌、肾癌、胰腺癌、卵巢癌、子宫内膜癌、甲状腺癌、宫颈癌、食管癌和肝癌。

39．权利要求 24-38 任一方法，其特征在于，所述利用试剂盒捕获所述核酸片段，获得目标区域，以及对所述目标区域进行序列测定，获得测序数据，包括：

（a）末端修复所述核酸片段，获得末端修复片段；

（b）加碱基 A 至所述末端修复片段的两端，获得粘性末端片段；

（c）连接接头于所述粘性末端片段的两端，获得接头连接片段；

（d）对所述接头连接片段进行第一扩增，获得第一扩增产物；

（e）利用所述试剂盒对所述第一扩增产物进行捕获，获得所述目标区域；以及，

（f）对所述目标区域进行第二扩增，获得第二扩增产物；

（g）对所述第二扩增产物进行序列测定，获得所述测序数据。

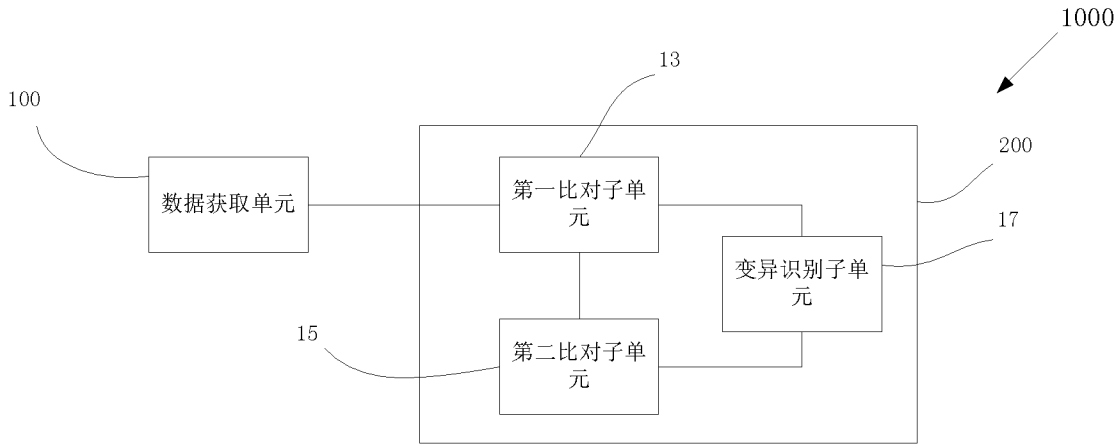

1000
13
100
200
数据获取单元
第一比对子单元
17
变异识别子单元
15
第二比对子单元

图 1

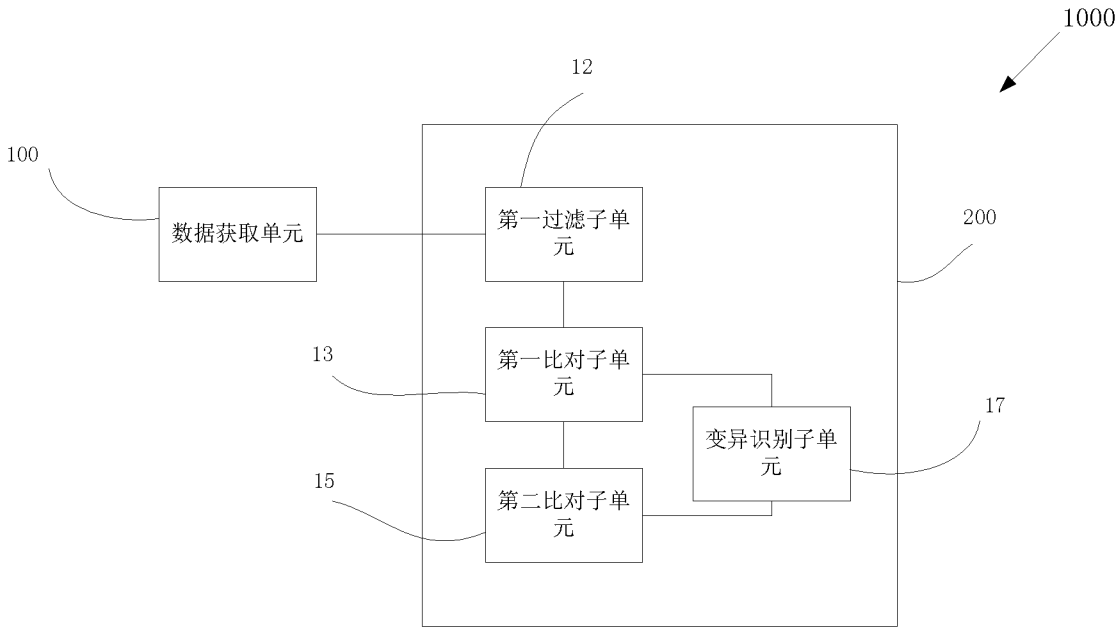

1000
12
100
数据获取单元
第一过滤子单元
200
13
第一比对子单元
17
变异识别子单元
15
第二比对子单元

图 2

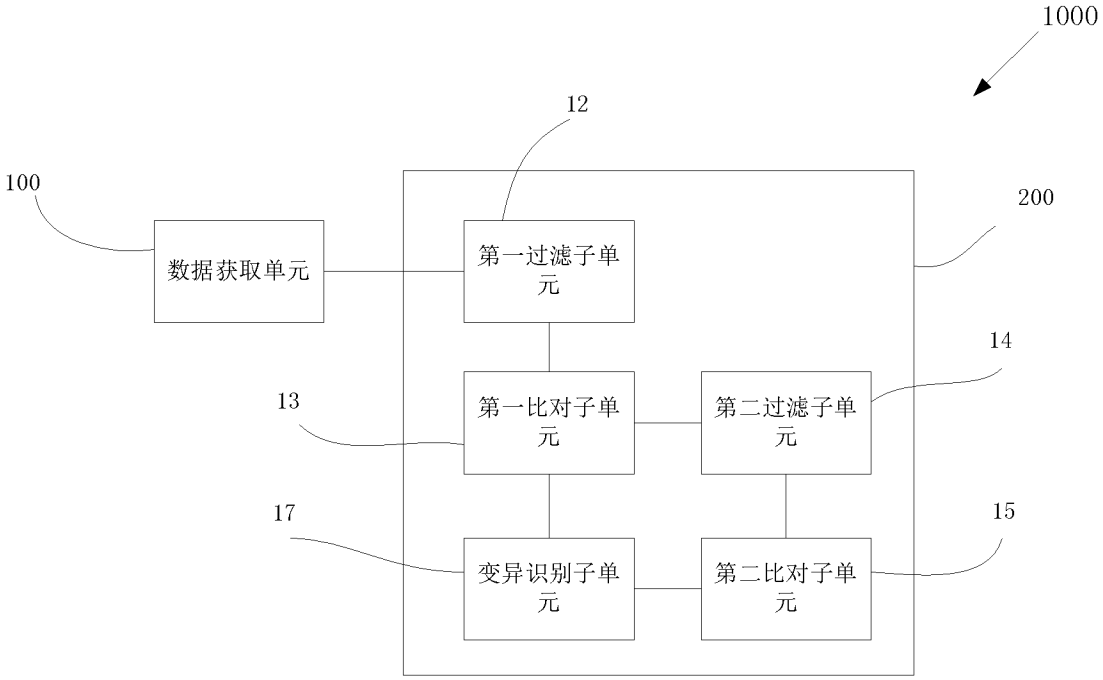

1000
12
100
200
数据获取单元
第一过滤子单元
14
13
第一比对子单元
第二过滤子单元
17
15
变异识别子单元
第二比对子单元

图 3

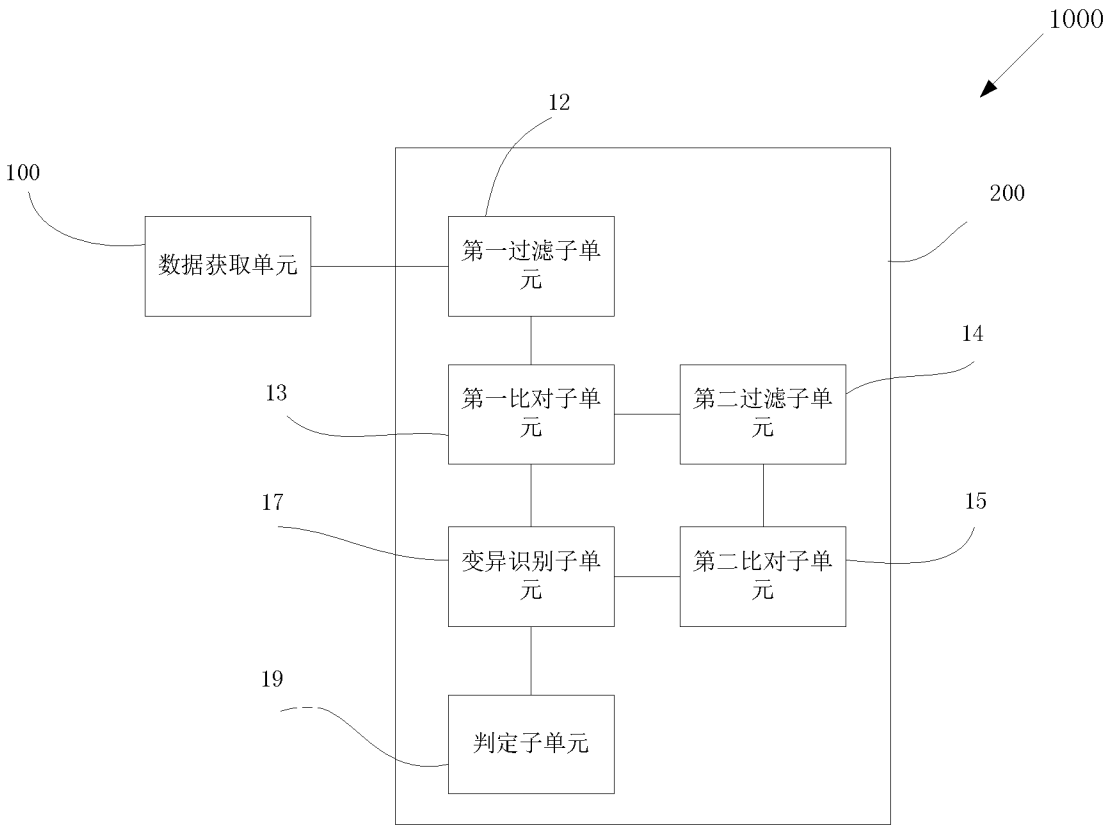

1000
12
100
200
数据获取单元
第一过滤子单元
14
13
第一比对子单元
第二过滤子单元
17
15
变异识别子单元
第二比对子单元
19
判定子单元

图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/093871

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68 (2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DWPI, SIPOABS, CPRSABS, CPEA, TWABS, CNABS, JPABS, MOABS, HKABS, CNMED, TWMED, KRABS, AUABS, DEABS, RUABS, CNKI, ISI web of Knowledge: cfDNA, free, cell free, cell-free, circulating, nucleic acid, tumor, tumour, neoplasm, neoplasma, cancer, carcinoma, sequencing, varia+, mutat+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2014039556 A1 (GUARDANT HEALTH INC. et al.), 13 March 2014 (13.03.2014), see abstract, and claims 1-206	1-39
Y	CN 103119179 A (PRESIDENT AND FELLOWS OF HARVARD COLLEGE), 22 May 2013 (22.05.2013), see abstract, and claims 1-148	1-39

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
11 September 2015 (11.09.2015)	**30 September 2015 (30.09.2015)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **XING, Ying** Telephone No.: (86-10) **62411044**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2014/093871

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
WO 2014039556 A1	13 March 2014	CA 2883901 A1	13 March 2014
		KR 20150067161 A	17 June 2015
		CN 104781421 A	15 July 2015
		GB 201509071 D0	08 July 2015
		EP 2893040 A1	15 July 2015
CN 103119179 A	22 May 2013	EP 2596132 A4	18 December 2013
		WO 2012012693 A3	14 June 2012
		AU 2011280996 A1	28 February 2013
		US 2012053073 A1	01 March 2012
		MX 2013000916 A	05 July 2013
		KR 20130041961 A	25 April 2013
		SG 187159 A1	28 February 2013
		JP 2013538565 A	17 October 2013
		CA 2806291 A1	26 January 2012
		EA 201390150 A1	30 September 2013
		EP 2596132 A2	29 May 2013
		WO 2012012693 A2	26 January 2012

国际检索报告

国际申请号
PCT/CN2014/093871

A. 主题的分类

C12Q 1/68(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

C12Q

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用））

DWPI, SIPOABS, CPRSABS, CPEA, TWABS, CNABS, JPABS, MOABS, HKABS, CNMED, TWMED, KRABS, AUABS, DEABS, RUABS, CNKI, ISI web of Knowledge: 游离, 无细胞, 循环, 核酸, 肿瘤, 癌, 测序, 变异, 突变, cfDNA, free, cell free, cell-free, circulating, nucleic acid, tumor, tumour, neoplasm, neoplasma, cancer, carcinoma, sequencing, varia+, mutat+

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	WO 2014039556 A1 (GUARDANT HEALTH INC等) 2014年 3月 13日 (2014-03-13) 参见摘要，权利要求1-206	1-39
Y	CN 103119179 A (哈佛大学校长及研究员协会) 2013年 5月 22日 (2013-05-22) 参见摘要，权利要求1-148	1-39

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2015年 9月 11日	2015年 9月 30日

ISA/CN的名称和邮寄地址	受权官员
中华人民共和国国家知识产权局(ISA/CN) 北京市海淀区蓟门桥西土城路6号 100088 中国	幸颖
传真号 (86-10)62019451	电话号码 (86-10)62411044

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2014/093871

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
WO	2014039556	A1	2014年 3月 13日	CA	2883901	A1	2014年 3月 13日
				KR	20150067161	A	2015年 6月 17日
				CN	104781421	A	2015年 7月 15日
				GB	201509071	D0	2015年 7月 8日
				EP	2893040	A1	2015年 7月 15日
CN	103119179	A	2013年 5月 22日	EP	2596132	A4	2013年 12月 18日
				WO	2012012693	A3	2012年 6月 14日
				AU	2011280996	A1	2013年 2月 28日
				US	2012053073	A1	2012年 3月 1日
				MX	2013000916	A	2013年 7月 5日
				KR	20130041961	A	2013年 4月 25日
				SG	187159	A1	2013年 2月 28日
				JP	2013538565	A	2013年 10月 17日
				CA	2806291	A1	2012年 1月 26日
				EA	201390150	A1	2013年 9月 30日
				EP	2596132	A2	2013年 5月 29日
				WO	2012012693	A2	2012年 1月 26日