

采用高通量测序技术检测肺癌致病突变及靶向药物突变位点

胡才智^{1*} 胡春玲²

¹重庆市大足区疾病预防控制中心 重庆 402360

²重庆市大足区第二人民医院 重庆 402360

摘要 Ion Torrent高通量测序平台可通过测序癌症病人的基因组,找出致病突变位点进而选择对应的靶向药物以更加有效地治疗并减少毒副作用。利用Ion Torrent平台共检测了55例非小细胞肺癌组织的甲醛固定石蜡包埋样本(FFPE样本),对测序结果进行分析并找出致病突变位点。其中,30(54.5%)例样本有抑癌基因*TP53*的突变,突变类型一共包括36种,是突变频率最高的一个基因。突变频率次高的基因是*EGFR*,检测到17(30.9%)例病人共8种突变类型。突变频率第三高的是*ERBB4*(*HER4*,与*EGFR*同属*HER*家族),检测到14(25.5%)例病人共两类突变。此外,还有其他突变频率较高的基因,如*MLH1*(8例,14.5%)、*KRAS*(6例,10.9%)、*PIK3CA*(5例,9.1%)等。本研究表明通过提取肺癌病人FFPE样本中的DNA并使用Ion Torrent平台测序可以找到致病突变,可用于指导医生进行更好的靶向药物治疗。(图1表2参29)

关键词 二代测序; Ion Torrent测序仪; 肺癌; 突变; 靶向药物

CLC R730.43

Using high throughput sequencing technology to detect lung cancer pathogenic and drug target mutations

HU Caizhi^{1*} & HU Chunling²

¹Center for Disease Control and Prevention of Dazu District of Chongqing, Chongqing 402360, China

²Second People's Hospital of Dazu District of Chongqing, Chongqing 402360, China

Abstract Cancer is one of the three major killers of human health in the Twenty-first Century, and lung cancer has the highest incidence and mortality. Along with the development of personalized medicine, the treatment of lung cancer is no longer confined to the traditional radiotherapy and chemotherapy. According to the patient's individual differences, targeted drugs can effectively treat cancer and reduce side effects. Ion torrent high-throughput sequencing platform can sequence the genome of cancer patients, and then find the causative mutation to select the corresponding targeted therapeutic drugs. We examined the formalin-fixed and paraffin-embedded tissues of 55 cases of non-small cell lung cancer using ion torrent platform, and analyzed the sequencing results to find the pathogenic mutations. Among them, 30 (54.5%) samples showed a total of 36 types of mutations in *TP53* gene, which had the highest mutation frequency. *TP53* is an important tumor suppressor gene that is found to mutate and lose function in a variety of tumors. The second high mutation frequency gene was *EGFR*, which was detected in 17 (30.9%) samples with a total of 8 mutation types. *EGFR* is a common mutation gene in non-small cell lung cancer, and Gefitinib is an effective targeted therapeutic drug. The third high mutation frequency gene was *ERBB4* detected in 14 (25.5%) samples with a total of 2 kinds of mutations. *ERBB4*, also known as *HER4*, belongs to the same family *HER* with *EGFR*. It is also closely related with the occurrence of cancer. Other genes also showed a certain frequency of mutation, including *MLH1* (8 cases, 14.5%), *KRAS* (6 cases, 10.9%), *PIK3CA* (5 cases, 9.1%), in which *KRAS* mutation has targeted therapeutic drugs, Selumetinib. The results showed that DNA extraction from lung cancer patients FFPE samples and sequencing by ion torrent platform can sort out the mutations to guide doctors in accurate use of targeted therapeutic drugs.

Keywords next generation sequencing; ion torrent PGM; lung cancer; mutation; targeted therapeutic drugs

近年来,癌症的发病率和死亡率居高不下,主要原因是其难以预防。癌症早期多无明显全身症状,发现时大多已经

到了中晚期,治愈的可能性很小。在所有的癌症种类中,由于空气污染和吸烟等,肺癌发病率持续增长,是目前对人类危害最大的一类。据统计,对全球范围内的男性而言,肺癌是癌症死亡的主要原因;对女性而言,在发展中国家,因肺癌致

收稿日期 Received: 2016-03-09 接受日期 Accepted: 2016-05-13

*通讯作者 Corresponding author (E-mail: polymerase2012@163.com)

死的女性仅次于乳腺癌,而在发达国家,肺癌已经超越乳腺癌成为女性的癌症死亡的主要原因^[1-3]。一般来讲肺癌可以分为非小细胞肺癌和小细胞肺癌两大类,非小细胞肺癌又包括大细胞癌、腺癌、鳞状细胞癌等类型,与小细胞癌相比其癌细胞生长分裂较慢,扩散转移相对较晚。非小细胞肺癌约占有肺癌的80%,大多数患者在发现时已处于中晚期,5年生存率很低^[4-5]。在人类研究癌症的过程中发现,原癌基因和抑癌基因与癌症的发生发展密切相关,这些基因的突变很可能引起细胞癌变。其实癌症在形成实体肿块以前,与癌症相关的基因突变的过程已经发生,而且这一过程可能已长达十余年甚至更长时间。传统的治疗方法,包括放疗化疗等,疗效差异很大且对病人的身心都会有伤害。近年来,靶向药物的出现为肺癌的治疗开辟了一条新的道路^[6-8]。靶向治疗疗效确切、针对性强,并且副作用小,是相关临床和基础的研究热点。目前对肺癌靶向治疗的药物主要是酪氨酸激酶的小分子抑制剂,包括EGFR突变的三大药物,易瑞沙(吉非替尼)、特罗凯(厄洛替尼)、凯美纳(盐酸埃克替尼),还有针对*EML4-ALK*融合基因表达阳性的克唑替尼^[9-12]。

在分析不同肺癌病人的发病原因时,DNA测序技术能够从基因组层面上寻找致病突变。自从1970年代双脱氧链终止测序法发明后,DNA测序技术发展迅速,但是由于测序通量不够大,测序费用也较昂贵等原因限制了其应用范围^[13-16]。2005年,高通量测序技术出现以来,逐步解决了测序通量不够大等一系列问题,并且测序成本不断下降,使得其应用越来越广泛^[17]。如今,以数据为导向的工业化高通量测序模式,极大地加速了人们对于疾病的认识,为疾病的预防、诊断及治疗提供了新的思路。赛默飞公司(ThermoFisher)的Ion torrent测序平台,推出了一种肺癌测序检测试剂盒,一次可以检测50个基因,包括致病基因和靶向用药的基因^[3, 18-20]。利用这一高通量测序平台,我们检测了55例肺癌的FFPE样本,通过对这些测序结果进行分析后找到不同的突变位点,结合肺癌靶向药物的相关信息,可以指导医生对肺癌病人的治疗,从而提高药物的疗效并减轻病人的痛苦。

1 材料与方法

1.1 实验材料

肺癌病人甲醛固定石蜡包埋(FFPE)组织,共55例样本(表1),都是非小细胞肺癌(NSCLC),主要为腺癌和鳞癌。其中腺癌有35例,占比63.6%,鳞癌有20例,占比36.4%。女性14人(25.5%),男性41人(74.5%)。来自城市的52例(94.5%),来自农村的3例(5.5%)。有吸烟历史的33例(60.0%),无吸烟历史的22例(40.0%)。年龄分段为:50岁(含)以下为4人(7.3%),51-60岁(含)为16人(29.1%),61-70岁(含)为24人(43.6%),71-80岁(含)为10人(18.2%),80岁以上为1人(1.8%)。

1.2 实验流程

样本DNA的提取:样本总DNA的提取采用AxyPrep™ Multisource Genomic DNA Miniprep Kit试剂盒(AXYGEN公司,美国)。

扩增文库的制备:扩增文库目标区域扩增子是采用 Ion

Ampliseq™ Cancer Hotspot Panel V2 (Thermo Fisher公司,美国)准备,通过Ion Ampliseq™ Library Kit 2.0 (Thermo Fisher公司,美国)和Ion Xpress™ Barcode Adapter 1-16 Kit (Thermo Fisher公司,美国)制备。

乳液PCR和测序:将制备好的扩增文库通过Ion One-Touch™ 2 System (Thermo Fisher公司,美国)和Ion PGM™ Hi-Q™ OT2 200 Template Kit (Thermo Fisher公司,美国)进行乳液PCR模板的制备,并将乳液PCR制备的ISP模板通过Dynabeads MyOne™ Streptavidin C1 Beads (Thermo Fisher公司,美国)和Ion OneTouch ES system (Thermo Fisher公司,美国)进行富集。

ISP模板制备完成后,采用Ion PGM™ Hi-Q™ Sequencing Kit (Thermo Fisher公司,美国)和Ion 318™ Chip (Thermo Fisher公司,美国)通过Ion Torrent Personal Genome Machine (PGM) (Thermo Fisher公司,美国)进行测序。

1.3 数据处理流程

数据处理流程如图1所示,测序仪得到的原始读长(Reads)经过第一步质控,如去掉低质量的,去掉长度较短的,得到高质量的读长;第二步是将得到的这些读长与人类基因组按照某种算法比对,去掉比对质量不高的读长;第三步则是用变异位点识别软件或者脚本找出单核苷酸变异(SNV)、多核苷酸变异(MNV)、插入缺失(INDEL)等变异位点,再根据覆盖度、等位基因频率、质量值等方面进行过滤,最终得到可能致病的这些突变位点。

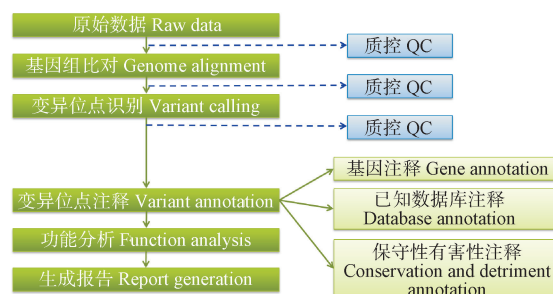


图1 测序数据处理流程。QC: 质量控制。

Fig. 1 Workflow of data processing. QC: quality control.

在第三步质控过程中,主要使用了Ion torrent平台的Torrent Variant Caller插件,其参数设置如下,每个样本的最小等位基因频率控制在0.02 (SNP)、0.1 (INDEL)、0.01 (Hotspot);最小质量值为6 (SNP)、6 (INDEL)、6 (Hotspot);最小覆盖度为6 (SNP)、15 (INDEL)、6 (Hotspot);链偏好性的最大值为0.95 (SNP)、0.9 (INDEL)、0.95 (Hotspot)。其他的参数都是插件推荐的默认值。

变异位点注释阶段是将Torrent Variant Caller插件得到的包含所有突变位点的VCF文件上传到Ion Reporter服务器进行注释,得到突变位点的详细注释信息。再将千人基因组中中国汉人的SNP位点的频率高于0.05的位点(被认为是非致病位点)去掉,得到每个样本的致病突变位点,最后统计55例样本的突变位点及突变基因频率。

表1 病理样本信息

Table 1 Information of pathological samples

序号 No.	性别 Sex	地区 District	吸烟史 Smoking	诊断年龄 Age	肿瘤部位 Tumor location
001	女 Female	城市 City	无 No	57	右肺中叶腺癌, 淋巴结颈根部异物巨细胞肉芽肿 Lung adenocarcinoma of right middle lobe and giant cell granuloma at neck lymph node
002	女 Female	城市 City	无 No	62	肺浸润性腺癌伴纵膈淋巴结转移 Invasive lung adenocarcinoma with mediastinal lymph node metastasis
003	男 Male	城市 City	有 Yes	64	非角化型中分化鳞状细胞癌 Non-keratinized intermediately differentiated lung squamous cell carcinoma
004	女 Female	城市 City	无 No	35	浸润性肺腺癌 Invasive lung adenocarcinoma
005	女 Female	城市 City	无 No	58	左肺上叶肺腺癌 Lung adenocarcinoma of left upper lobe
006	男 Male	城市 City	有 Yes	56	右上肺鳞癌 Lung squamous cell carcinoma of right upper lobe
007	男 Male	城市 City	有 Yes	68	右下肺中分化鳞癌 Intermediately differentiated lung squamous cell carcinoma of right lower lobe
008	男 Male	城市 City	有 Yes	76	左下肺鳞癌 Lung squamous cell carcinoma of left lower lobe
009	女 Female	城市 City	有 Yes	73	左肺上叶腺癌 Lung adenocarcinoma of left upper lobe
010	男 Male	城市 City	有 Yes	70	左下肺鳞癌 Lung squamous cell carcinoma of left lower lobe
011	男 Male	城市 City	有 Yes	66	右肺上叶肺腺癌, 纵膈淋巴结转移 Lung adenocarcinoma of right upper lobe with mediastinal lymph node metastasis
012	男 Male	城市 City	有 Yes	61	右肺上叶鳞癌 Lung squamous cell carcinoma of right upper lobe
013	男 Male	城市 City	有 Yes	57	左肺上叶肺中分化腺癌, 4、10号淋巴结转移 Intermediately differentiated lung adenocarcinoma of left upper lobe with 4# and 10# lymph nodes metastasis
014	男 Male	城市 City	有 Yes	38	右肺上叶低分化腺癌, 10号淋巴结转移 Poorly differentiated lung adenocarcinoma of right upper lobe with 10# lymph node metastasis
015	男 Male	城市 City	有 Yes	68	右肺腺癌, 淋巴结转移癌, 乳头为主 Lung adenocarcinoma of right lobe with lymph node(nipple mainly) metastasis
016	女 Female	城市 City	无 No	46	右肺上叶腺癌 Lung adenocarcinoma of right upper lobe
017	男 Male	城市 City	有 Yes	64	右肺下叶腺癌, 7、10、11号淋巴结转移 Lung adenocarcinoma of right lower lobe with 7#, 10# and 11# lymph nodes metastasis
018	男 Male	城市 City	有 Yes	68	右肺中叶腺癌, 纵膈淋巴结转移 Lung adenocarcinoma of right middle lobe with mediastinal lymph node metastasis
019	男 Male	城市 City	无 No	71	左肺上叶腺癌 Lung adenocarcinoma of left upper lobe
020	女 Female	城市 City	无 No	68	右肺中叶腺癌 Lung adenocarcinoma of right middle lobe
021	男 Male	城市 City	有 Yes	60	右肺上叶中分化腺癌 Intermediately differentiated lung adenocarcinoma of right upper lobe
022	女 Female	城市 City	无 No	67	左肺下叶腺癌, 4号淋巴结转移 Lung adenocarcinoma of left lower lobe with 4# lymph nodes metastasis
023	男 Male	城市 City	有 Yes	56	左下肺中分化腺癌 Intermediately differentiated lung adenocarcinoma of left lower lobe
024	男 Male	城市 City	无 No	53	右肺下叶肺腺癌, 肺门淋巴结转移 Lung adenocarcinoma of right lower lobe with hilus pulmonis lymph nodes metastasis
025	女 Female	城市 City	无 No	72	右肺中叶非角化中-低分化鳞癌, 9号淋巴结转移 Intermediately-poorly differentiated non-keratinized lung squamous cell carcinoma of right middle lobe with 9# lymph node metastasis
026	男 Male	城市 City	有 Yes	66	左肺上叶腺癌, 4、5、6、10号淋巴结转移 Lung adenocarcinoma of left upper lobe with 4#, 5#, 6# and 10# lymph nodes metastasis
027	男 Male	城市 City	有 Yes	64	右肺下叶中分化鳞癌, 4、7、10号淋巴结转移 Intermediately differentiated lung squamous cell carcinoma of right lower lobe with 4#, 7# and 10# lymph nodes metastasis
028	男 Male	城市 City	有 Yes	75	左肺上叶中分化鳞癌 Intermediately differentiated lung squamous cell carcinoma of left upper lobe
029	女 Female	城市 City	无 No	61	右肺上叶中分化腺癌 Intermediately differentiated lung adenocarcinoma of right upper lobe
030	男 Male	城市 City	有 Yes	54	左肺上叶中分化鳞癌, 左肺下叶浸润性腺癌 Intermediately differentiated lung squamous cell carcinoma of left upper lobe and invasive lung adenocarcinoma of left lower lobe

续表1 Continued

序号 No.	性别 Sex	地区 District	吸烟史 Smoking	诊断年龄 Age	肿瘤部位 Tumor location
031	男 Male	城市 City	有 Yes	63	左肺上叶腺癌 Lung adenocarcinoma of left upper lobe
032	男 Male	城市 City	有 Yes	71	右肺中叶中分化鳞癌 Intermediately differentiated lung squamous cell carcinoma of right middle lobe
033	男 Male	城市 City	有 Yes	69	右肺上叶中分化浸润性腺癌 Intermediately differentiated invasive lung adenocarcinoma of right upper lobe
034	男 Male	城市 City	有 Yes	51	左上肺腺癌, 5#淋巴结转移 Lung adenocarcinoma of left upper lobe with 5# lymph node metastasis
035	男 Male	城市 City	有 Yes	61	右下肺中-低分化鳞癌, 7、10号淋巴结转移 Intermediately-poorly differentiated lung squamous cell carcinoma of right lower lobe with 7# and 10# lymph nodes metastasis
036	女 Female	城市 City	无 No	73	右肺低分化鳞癌 Poorly differentiated lung squamous cell carcinoma of right lobe
037	男 Male	城市 City	无 No	82	左上肺腺癌 Lung adenocarcinoma of left upper lobe
038	男 Male	城市 City	无 No	65	右肺上叶腺癌 Lung adenocarcinoma of right upper lobe
039	男 Male	城市 City	无 No	52	左肺上叶腺癌 Lung adenocarcinoma of left upper lobe
040	男 Male	城市 City	无 No	55	右肺下叶腺癌 Lung adenocarcinoma of right lower lobe
041	男 Male	城市 City	有 Yes	71	右肺下叶腺癌 Lung adenocarcinoma of right lower lobe
042	男 Male	城市 City	无 No	56	左上肺中-低分化腺癌 Intermediately-poorly differentiated lung adenocarcinoma of left upper lobe
043	男 Male	城市 City	无 No	63	左肺上叶腺癌 Lung adenocarcinoma of left upper lobe
044	女 Female	城市 City	无 No	66	右肺下叶腺癌 Lung adenocarcinoma of right lower lobe
045	女 Female	农村 Country	无 No	51	右上肺高分化鳞癌 Well differentiated lung squamous cell carcinoma of right upper lobe
046	男 Male	城市 City	有 Yes	63	右肺下叶中分化鳞癌 Intermediately differentiated lung squamous cell carcinoma of right lower lobe
047	男 Male	城市 City	无 No	63	左上肺腺癌 Lung adenocarcinoma of left upper lobe
048	男 Male	城市 City	有 Yes	53	右下肺腺癌 Lung adenocarcinoma of right lower lobe
049	男 Male	农村 Country	有 Yes	59	右肺中叶腺癌 Lung adenocarcinoma of right middle lobe
050	女 Female	农村 Country	有 Yes	66	左上肺鳞癌 Lung squamous cell carcinoma of left upper lobe
051	男 Male	城市 City	有 Yes	73	左下肺腺癌 Lung adenocarcinoma of left lower lobe
052	男 Male	城市 City	无 No	72	左肺下叶鳞癌 Lung squamous cell carcinoma of left lower lobe
053	男 Male	城市 City	有 Yes	60	右肺下叶鳞癌 Lung squamous cell carcinoma of right lower lobe
054	男 Male	城市 City	有 Yes	70	左肺上叶鳞癌 Lung squamous cell carcinoma of left upper lobe
055	男 Male	城市 City	有 Yes	49	左肺上叶鳞癌 Lung squamous cell carcinoma of left upper lobe

2 结果

55例样本共检测到93个突变位点, 包括外显子和内含子区域. 其中80 (86%) 个突变位点位于外显子区域, 剩余13 (14%) 个突变位点位于内含子区域, 可能会参与到mRNA的剪切过程, 进而影响基因的功能. 为保证突变位点是可靠的, 我们检查了这些突变位点覆盖度, 发现这93个突变位点的平均覆盖度是1 984, 覆盖度最小的一个位点是38号样本

的CTNNB1 c.101G > A位点, 为398, 覆盖度最大的一个位点是128号样本的MLH1 c.1151T > A位点, 为3 994.

这93个突变位点属于30个基因, 其中, 突变频率最高的基因是TP53, 一共有30 (54.5%) 例样本有TP53基因的突变, 包括36种突变类型. TP53基因编码的蛋白为肿瘤抑制蛋白, 也称为p53蛋白或p53肿瘤蛋白, 属于最早发现的抑癌基因之一^[21-22]. 任何导致TP53失活的突变都会引起其抑癌功能的丧失而导致癌症发生, 这也是其突变类型很多的原因.

突变频率第二高的基因是*EGFR* (Epidermal growth factor receptor), 一共检测到17 (30.9%) 例样本发生突变, 包括8种突变类型. *EGFR*是表皮生长因子受体 (*HER*) 家族成员之一, 在细胞生理过程中发挥重要的调节作用. *EGFR*的突变有几十种亚型, 但最主要是两种, 第一种是L858R, 也就是*EGFR*蛋白的第858个氨基酸从L突变成了R, 我们在55例样本中检测到了4例 (7.3%), 第二种是“19号外显子缺失”, 也就是*EGFR*蛋白中负责抑制它活性一部分被切掉了, 这种突变在55例样本中检测到了11例 (20%). 一般来说, 这两种突变占到所有肺癌*EGFR*突变的90%, 因此如果病人被诊断为*EGFR*突变肺癌, 那多半就是这两种突变之一. 此外, 还有一种突变是*EGFR*蛋白的第719位的甘氨酸突变为丝氨酸 (G719S), 在我们的55例样本中检测到了一例. *EGFR*突变是非小细胞肺癌常见的一类突变, 吉非替尼和厄罗替尼是其有效的靶向药物, 已被FDA批准用于治疗晚期非小细胞肺癌^[11-12].

突变频率第三高的基因是*ERBB4*, 有14 (25.5%) 例样本发生了突变, 包括两类突变分别是*ERBB4* c.718G > A和*ERBB4* c.884-22C > T. *ERBB4*又叫*HER4*, 是一种受体酪氨酸激酶, 与*EGFR*同属*HER*家族. *ERBB4*通过释放可溶性胞内结构域来调节目的基因表达, 能诱导多种细胞反应, 包括有丝

分裂和分化.

此外, 其他突变频率较高的基因还包括*MLH1* (8例, 14.5%)、*KRAS* (6例, 10.9%)、*PIK3CA* (5例, 9.1%)、*PTEN* (4例, 7.3%) 等. 其中, *MLH1*基因编码的蛋白是由7个蛋白组成的DNA错配修复系统的一个组成部分蛋白, 在人类启动DNA错配修复过程中起作用. *KRAS*基因突变与人类癌症发生直接相关^[23]. *PIK3CA*基因编码I类磷脂酰肌醇-3-激酶 (Phosphatidylinositol 3-kinases, PI3Ks) 的p110催化亚单位, 即PI3Kp110a. PI3K是*EGFR*下游信号分子, 可被生长因子受体酪氨酸激酶 (如*EGFR*) 激活, 使丝/苏氨酸激酶(AKT)磷酸化而上调该通路的活性并产生多种生物学效应, 包括调节细胞增殖、存活和细胞周期调控等^[24]. *PTEN*基因是继*TP53*基因后另一个较为广泛地与肿瘤发生关系密切的抑癌基因, 它在细胞周期调控及细胞快速生长抑制中起着重要的作用^[25].

国际癌症基因组联盟 (International Cancer Genome Consortium, ICGC) 目前已公布了超过一万个人的癌症基因组数据, 是癌症研究的一个重要的数据库, 其中包括了中国 (10例)、韩国人 (36例) 和美国人 (178例) 的肺癌的数据. 将从55例样本中找到的突变位点与ICGC肺癌数据库进行比较发现, 我们找到的大部分都是新位点. 考虑人种因素分析, 首先与中国人的ICGC数据相比, 只有*EGFR* c.2573T >

表2 55例肺癌样本突变位点及突变基因

Table 2 Variant sites and variant genes of 55 lung cancer samples

突变位点 Mutation	外显子/内含子 Location	样本数 Sample No.	位点频率 VSF	基因 Gene	样本数 Sample No.	基因频率 VGF	SFCLC	SFKLC	SFALC
<i>AKT1</i> c.122G > T	EX3	1	1.8%	<i>AKT1</i>	2	3.6%	-	-	-
<i>AKT1</i> c.47-12G > A	IN2	1	1.8%				-	-	-
<i>ALK</i> c.3600G > C	EX23	3	5.5%	<i>ALK</i>	3	5.5%	-	-	-
<i>APC</i> c.4393A > T	EX16	1	1.8%	<i>APC</i>	1	1.8%	-	-	-
<i>ATM</i> c.*29C > G	EX63	1	1.8%	<i>ATM</i>	1	1.8%	-	-	-
<i>ATM</i> c.5793T > C	EX39	1	1.8%				-	-	-
<i>BRAF</i> c.1398A > G	EX11	1	1.8%	<i>BRAF</i>	1	1.8%	-	-	-
<i>CDKN2A</i> c.182_200delAGCTGCTGCTGCTCCACGG	EX2	1	1.8%				-	-	-
<i>CDKN2A</i> c.197A > G	EX2	1	1.8%	<i>CDKN2A</i>	4	7.3%	-	-	-
<i>CDKN2A</i> c.343G > T	EX2	1	1.8%				-	-	-
<i>CDKN2A</i> c.374_374delA	EX2	1	1.8%				-	-	-
<i>CDKN2A</i> c.405G > A	EX2	1	1.8%				-	-	-
<i>CTNNB1</i> c.101G > A	EX3	1	1.8%	<i>CTNNB1</i>	2	3.6%	-	-	-
<i>CTNNB1</i> c.98C > A	EX3	1	1.8%				-	-	-
<i>EGFR</i> c.2155G > T	EX18	1	1.8%	<i>EGFR</i>	17	30.9%	-	-	-
<i>EGFR</i> c.2235_2249delGGAATTAAGAGAAGC	EX19	8	14.5%				-	-	-
<i>EGFR</i> c.2236_2250delGAATTAAGAGAAGCA	EX19	1	1.8%				-	-	-
<i>EGFR</i> c.2238_2252delATTAAGAGAAGCAAC	EX19	1	1.8%				-	-	-
<i>EGFR</i> c.2240_2257delTAAGAGAAGCAACATCTC	EX19	1	1.8%				-	-	-
<i>EGFR</i> c.2248G > C	EX19	1	1.8%				-	-	-
<i>EGFR</i> c.2260A > G	EX19	1	1.8%				-	-	-
<i>EGFR</i> c.2573T > G	EX21	4	7.3%				0.4	0.0278	-
<i>ERBB2</i> c.2263_2264delTTinsCC	EX19	1	1.8%	<i>ERBB2</i>	1	1.8%	-	-	-
<i>ERBB2</i> c.2264T > C	EX19	1	1.8%				-	-	-
<i>ERBB4</i> c.718G > A	EX6	1	1.8%	<i>ERBB4</i>	14	25.5%	-	-	-
<i>ERBB4</i> c.884-22C > T	IN7	13	23.6%				-	-	-
<i>FGFR1</i> c.474G > A	EX5	1	1.8%	<i>FGFR1</i>	1	1.8%	-	-	-
<i>FGFR3</i> c.1965+22G > A	IN14	1	1.8%	<i>FGFR3</i>	1	1.8%	-	-	-
<i>IDH1</i> c.315C > T	EX4	1	1.8%	<i>IDH1</i>	1	1.8%	-	-	-
<i>JAK2</i> c.1864+14C > T	IN14	1	1.8%	<i>JAK2</i>	2	3.6%	-	-	-
<i>JAK2</i> c.1864+57A > G	IN14	1	1.8%				-	-	-
<i>JAK3</i> c.1702-24T > C	IN12	1	1.8%	<i>JAK3</i>	1	1.8%	-	-	-
<i>KDR</i> c.1413-8T > C	IN10	1	1.8%	<i>KDR</i>	1	1.8%	-	-	-

续表2 Continued

突变位点 Mutation	外显子/内含子 Location	样本数 Sample No.	位点频率 VSF	基因 Gene	样本数 Sample No.	基因频率 VGF	SFCLC	SFKLC	SFALC
<i>KIT</i> c.1638A > G	EX10	3	5.5%	<i>KIT</i>	4	7.3%	-	-	-
<i>KIT</i> c.1674G > A	EX11	1	1.8%				-	-	-
<i>KRAS</i> c.34G > T	EX2	1	1.8%				-	0.0278	-
<i>KRAS</i> c.35G > A	EX2	1	1.8%	<i>KRAS</i>	6	10.9%	-	-	-
<i>KRAS</i> c.35G > C	EX2	1	1.8%				-	-	-
<i>KRAS</i> c.35G > T	EX2	3	5.5%				0.1	-	-
<i>MLH1</i> c.1151T > A	EX12	8	14.5%	<i>MLH1</i>	8	14.5%	-	-	-
<i>NRAS</i> c.159G > A	EX3	1	1.8%	<i>NRAS</i>	1	1.8%	-	-	-
<i>PDGFRA</i> c.2123T > C	EX15	1	1.8%	<i>PDGFRA</i>	1	1.8%	-	-	-
<i>PIK3CA</i> c.*37C > G	EX21	1	1.8%	<i>PIK3CA</i>	5	9.1%	-	-	-
<i>PIK3CA</i> c.1618C > G	EX10	1	1.8%				-	-	-
<i>PIK3CA</i> c.1633G > A	EX10	3	5.5%				-	-	0.0562
<i>PIK3CA</i> c.3075C > T	EX21	1	1.8%				-	-	-
<i>PTEN</i> c.-9C > G	EX1	1	1.8%				-	-	-
<i>PTEN</i> c.-10delC	EX1	2	3.6%	<i>PTEN</i>	4	7.3%	-	-	-
<i>PTEN</i> c.493-34TC > CA	IN5	1	1.8%				-	-	-
<i>RBI</i> c.473T > A	EX4	1	1.8%	<i>RBI</i>	1	1.8%	-	-	-
<i>RET</i> c.1879+27A > G	IN10	1	1.8%	<i>RET</i>	1	1.8%	-	-	-
<i>SMAD4</i> c.955+58C > T	IN8	3	5.5%	<i>SMAD4</i>	4	7.3%	-	-	-
<i>SMAD4</i> c.955+64A > C	IN8	1	1.8%				-	-	-
<i>SMARCB1</i> c.215C > A	EX2	1	1.8%				<i>SMARCB1</i>	2	3.6%
<i>SMARCB1</i> c.500+41G > A	IN4	1	1.8%	-	-	-			
<i>STK11</i> c.593_593delC	EX4	1	1.8%	<i>STK11</i>	1	1.8%			
<i>TP53</i> c.1027G > T	EX10	1	1.8%				-	-	-
<i>TP53</i> c.214C > G	EX4	2	3.6%	<i>TP53</i>	30	54.5%	-	-	-
<i>TP53</i> c.215C > G c.216_216delC	EX4	1	1.8%				-	-	-
<i>TP53</i> c.219G > T	EX4	1	1.8%				-	-	-
<i>TP53</i> c.298C > T	EX4	1	1.8%				-	-	-
<i>TP53</i> c.313G > T	EX4	1	1.8%				-	-	-
<i>TP53</i> c.329G > C	EX4	1	1.8%				-	-	-
<i>TP53</i> c.394A > G	EX5	1	1.8%				-	-	-
<i>TP53</i> c.408A > G	EX5	2	3.6%				-	-	-
<i>TP53</i> c.455C > T	EX5	1	1.8%				-	-	-
<i>TP53</i> c.469G > T	EX5	1	1.8%				-	-	0.0169
<i>TP53</i> c.488A > G	EX5	1	1.8%				-	-	0.0169
<i>TP53</i> c.523C > T	EX5	1	1.8%				-	-	-
<i>TP53</i> c.524G > A	EX5	1	1.8%				-	-	-
<i>TP53</i> c.524G > T	EX5	1	1.8%				-	-	-
<i>TP53</i> c.527G > T	EX5	1	1.8%				-	-	0.0056
<i>TP53</i> c.535C > T	EX5	1	1.8%				-	-	0.0056
<i>TP53</i> c.536A > T	EX5	1	1.8%				-	-	0.0112
<i>TP53</i> c.560-1G > T	IN5	1	1.8%				-	-	-
<i>TP53</i> c.578A > G	EX6	1	1.8%				-	-	0.0056
<i>TP53</i> c.592G > T	EX6	1	1.8%				-	-	-
<i>TP53</i> c.614A > G	EX6	1	1.8%	-	-	0.0056			
<i>TP53</i> c.673-1G > T	IN6	1	1.8%	-	-	0.0056			
<i>TP53</i> c.729_730delGGinsTT	EX7	1	1.8%	-	-	-			
<i>TP53</i> c.742C > T	EX7	1	1.8%	-	-	0.0056			
<i>TP53</i> c.742_744delCGGinsTGT	EX7	1	1.8%	-	-	-			
<i>TP53</i> c.743G > A	EX7	1	1.8%	-	-	0.0056			
<i>TP53</i> c.746G > T	EX7	1	1.8%	-	-	-			
<i>TP53</i> c.747G > T	EX7	1	1.8%	-	-	0.0056			
<i>TP53</i> c.757_758delAC	EX7	1	1.8%	-	-	-			
<i>TP53</i> c.832C > T	EX8	1	1.8%	-	-	0.0056			
<i>TP53</i> c.833C > T	EX8	1	1.8%	-	-	0.0056			
<i>TP53</i> c.845G > C	EX8	1	1.8%	-	-	-			
<i>TP53</i> c.856G > A	EX8	1	1.8%	-	-	-			
<i>TP53</i> c.914_915insA	EX8	1	1.8%	<i>VHL</i>	1	1.8%	-	-	-
<i>VHL</i> c.409G > C	EX2	1	1.8%				-	-	-

ICGC: 国际癌症基因组联盟; VSF: 突变位点频率; VGF: 突变基因频率; SFCLC: ICGC数据库中国人肺癌频率; SFKLC: ICGC数据库韩国人肺癌频率; SFALC: ICGC数据库美国人肺癌频率。

ICGC: International Cancer Genome Consortium; VSF: Variant sites frequency; VGF: Variant gene frequency; SFCLC: Site frequency of Chinese lung cancer in ICGC; SFKLC: Site frequency of Korean lung cancer in ICGC; SFALC: Site frequency of American lung cancer in ICGC.

G和KRAS c.35G > T这两个位点在ICGC中有对应的频率,分别为0.4和0.1,其余位点均未在ICGC中出现,这很可能是因为ICGC中国人的肺癌样本数太少(10人),得到的突变位点数据不全面造成的。再看ICGC韩国人的数据,共有36人,我们找到的位点有EGFR c.2573T > G和KRAS c.34G > T这两个位点在ICGC中有对应的频率,和中国人的情况类似。最后看美国人的数据发现,TP53基因的多个突变位点在美国人的ICGC肺癌数据库中出现,这可能是由于在肺癌中,作为重要的肿瘤抑制基因TP53基因突变发生频率较高,并且突变类型多样,而其他突变位点则几乎没有在ICGC美国人的数据库中出现,这很可能是人种差异造成的。

3 讨论

目前,从肺癌病人的FFPE样本提取DNA并使用Ion Torrent平台测序已经是一项比较成熟的技术。相较于经典的Sanger测序技术检测基因突变的方法,二代测序技术只需要更少的DNA量,一般在10 ng左右,并且在灵敏性和特异性上面都有极大的提升。通过二代测序找到病人的致病突变位点,主要的难点还在于对测序数据的分析处理,特别是对测序数据多次进行的质量控制,包括读长的长度,单个碱基的质量值,比对的质量等等,这些都是通过多次试验并不断优化得到的。由于Ion Torrent测序仪测序原理的局限性,在对基因组的较长的同聚物区域(如AAAAAAA)测序时,其准确度会下降,可能会有假阳性的情况出现。一般这种情况需要通过其他方式如Sanger测序法来进行验证,从而排除假阳性位点。另外,为了避免测序过程中的其他错误产生假阳性,一般需要控制覆盖度在100x以上。

在找到致病突变位点之后,如何根据突变位点综合选择靶向药物是另外一个难点。目前针对EGFR开发的分子靶向药物主要分为两类,一是小分子酪氨酸激酶抑制剂(TKIs),如吉非替尼和厄洛替尼,它们可以抑制EGFR胞内区酪氨酸激酶活性;另一类是单克隆抗体,如西妥昔单抗和帕尼单抗,可以与EGFR胞外区结合,阻断依赖于配体的EGFR活化。这两类药物通过不同途径阻断EGFR介导的细胞内信号转导通路,从而抑制肿瘤生长、转移和血管生成,并促进肿瘤细胞凋亡,提高放疗化疗敏感性。然而,大部分对EGFR-TKIs治疗有效的患者最终都会对EGFR-TKIs产生耐药性。进一步临床研究还表明,EGFR基因外显子20的体细胞突变是EGFR-TKIs继发耐药的主要机制之一,该突变主要是第790位密码子出现C > T的转换,引起EGFR蛋白编码的相应氨基酸由苏氨酸转变为甲硫氨酸(T790M)。这一突变仅见于药物治疗后复发者,突变使得非小细胞肺癌患者对吉非替尼和厄洛替尼产生耐药性。当然,除了这些,还有一些新的靶向药物开发出来,例如阿法替尼和盐酸埃克替尼等。KRAS突变是在使用EGFR-TKIs前需要考虑的因素,它主要引起原发耐药^[26]。肺腺癌中大部分KRAS突变与吸烟有关,携带KRAS突变的患者不能采用EGFR-TKIs治疗。除了EGFR突变以外,ALK重排也常出现在非小细胞肺癌中。ALK是位于2号染色体上胰岛素受体家族的一员,编码跨膜受体酪氨酸激酶。ALK的激活首先通过融合基因的形成。ALK胞内激酶区与EML4的

N末端融合,然后编码具有激酶活性的胞质嵌合蛋白,随后驱动肿瘤生长^[27-29]。目前,克唑替尼是针对ALK重排的非小细胞肺癌患者的十分有效的靶向药物。美国国立综合癌症网络(National Comprehensive Cancer Network)每年都会发布各种恶性肿瘤临床实践指南,其中包括非小细胞肺癌的治疗。除了靶向药物,该指南对其它致病位点如何用药也给予了详细的指导。将突变位点检测结果与该指南结合,可以优化治疗方案,减轻病人的痛苦。

我们此次收集并测序的样本中男性占了大约3/4,包括了腺癌和鳞癌,其中腺癌的样本数稍多,由于样本量比较小,所以并未分开统计。在各年龄段中,50-70岁的肺癌病人最多,占72.7%,这与中国肺癌人群的年龄段分布相当。当然,目前中国年轻人肺癌的发病率也有所上升,其原因还需要进一步的研究。此前的研究已经证实,吸烟与肺癌发生存在正相关性,在这批样本中,吸烟的共33人,无吸烟史的共22人,吸烟的病人占比较高。另外,这批样本主要来自城市居民,农村居民的情况可能会有不同。总共55例样本找到的这些位点,可以部分地展示中国人非小细胞肺癌发病原因,但是还不能完全代表整体情况。随着更多的中国人的肺癌测序数据公布,建立中国人自己的肺癌数据库,将更有利于我们国家对肺癌的防治。

此外,在癌症发生早期,癌细胞还没有形成明显的肿块,难以从形态学上直接观察到,但已有微量肿瘤DNA释放到循环血液中。目前已经有研究机构通过高通量基因测序技术检测循环血液中这些基因的突变,可以筛查早期或超早期癌症,评估放疗化疗的疗效,预测靶向药物的疗效,监测术后早期复发等。这项无创肿瘤基因检测技术将极大地提升癌症的治疗效果和减少病人的痛苦,也是癌症检测未来的发展方向。

参考文献 [References]

- 1 Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global Cancer Statistics, 2012 [J]. *CA Cancer J Clin*, 2015, **65** (2): 87-108
- 2 Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics [J]. *CA Cancer J Clin*, 2011, **61** (2): 69-90
- 3 Qiu T, Guo H, Zhao H, Wang L, Zhang Z. Next-generation sequencing for molecular diagnosis of lung adenocarcinoma specimens obtained by fine needle aspiration cytology [J]. *Sci Rep*, 2015, **5**: 11317
- 4 Leighl NB, Rehkman N, Biermann WA, Huang J, Mino-Kenudson M, Ramalingam SS, West H, Whitlock S, Somerfield MR. Molecular testing for selection of patients with lung cancer for epidermal growth factor receptor and anaplastic lymphoma kinase tyrosine kinase inhibitors: American Society of Clinical Oncology endorsement of the College of American Pathologists/International Association for the study of lung cancer/association for molecular pathology guideline [J]. *J Clin Oncol*, 2014, **32** (32): 3673-3679
- 5 Bunnell CA, Shulman LN. Will we be able to care for cancer patients in the future? [J]. *Oncology*, 2010, **14**: 1343-1348
- 6 吴一龙. 肺癌分子靶向治疗的个体化[J]. 循证医学, 2004, **4** (2): 67-68 [Wu YL. Individualized molecular targeted therapy for lung cancer [J]. *J Evid-Based Med*, 2004, **4** (2): 67-68]
- 7 Wu K, House L, Liu W, Cho WC. Personalized Targeted Therapy for

- Lung Cancer [J]. *Chin J Lung Cancer*, 2013, **16** (8): 11471-11496
- 8 Sharma SV, Settleman J. Oncogenic shock: Turning an activated kinase against the tumor cell [J]. *Cell Cycle*, 2006, **5** (24): 2878-2880
 - 9 Belani, CP, Goss G, Blumenschein G Jr. Recent clinical developments and rationale for combining targeted agents in non-small cell lung cancer (NSCLC) [J]. *Cancer Treat Rev*, 2012, **38** (3): 173-184
 - 10 Kobayashi S, Boggon TJ, Dayaram T, Jänne PA, Kocher O, Meyerson M, Johnson BE, Eck MJ, Tenen DG, Halmos B. EGFR mutation and resistance of non-small-cell lung cancer to Gefitinib [J]. *New Engl J Med*, 2005, **352** (8): 786-792
 - 11 Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, Herman P, Kaye FJ, Lindeman N, Boggon TJ, Naoki K, Sasaki H, Fujii Y, Eck MJ, Sellers WR, Johnson BE, Meyerson M. EGFR mutations in lung cancer: Correlation with clinical response to gefitinib therapy [J]. *Science*, 2004, **304** (5676): 1497-1500
 - 12 Stella GM, Luisetti M, Inghilleri S, Cemmi F, Scabini R, Zorzetto M, Pozzi E. Targeting EGFR in non-small-cell lung cancer: Lessons, experiences, strategies [J]. *Respir Med*, 2012, **106** (2): 173-183
 - 13 Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LDW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF, Wilson RK. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome [J]. *Nature*, 2008, **456** (7218): 66-72
 - 14 Mardis ER. Next-generation DNA sequencing methods [J]. *Annu Rev Genomics Hum Genet*, 2008, **9**: 387-402
 - 15 Rusk N. Torrents of sequence [J]. *Nat Methods*, 2011, **8** (1): 44
 - 16 Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes [J]. *Nature*, 2009, **461** (7261): 272-276
 - 17 Metzker ML. Sequencing technologies - the next generation [J]. *Nat Rev Genet*, 2010, **11** (1): 31-46
 - 18 Beadling C, Neff TL, Heinrich MC, Rhodes K, Thornton M, Leamon J, Andersen M, Corless CL. Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping [J]. *J Mol Diagn*, 2013, **15** (2): 171-176
 - 19 Xu Z, Huo X, Tang C, Ye H, Nandakumar V, Lou F, Zhang D, Jiang S, Sun H, Dong H, Zhang G, Liu Z, Dong Z, Guo B, Yan H, Yan C, Wang L, Su Z, Li Y, Gu D, Zhang X, Wu X, Wei X, Hong L, Zhang Y, Yang J, Gong Y, Tang C, Jones L, Huang XF, Chen SY, Chen J. Frequent KIT mutations in human gastrointestinal stromal tumors [J]. *Sci Rep*, 2014, **4**: 5907
 - 20 Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers [J]. *BMC Genomics*, 2012, **13**: 341
 - 21 Guimaraes DP, Hainaut P. TP53: a key gene in human cancer [J]. *Biochimie*, 2002, **84** (1): 83-93
 - 22 Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use [J]. *Cold Spring Harb Perspect Biol*, 2010, **2** (1): a001008
 - 23 Westcott PM, Halliwill KD, To MD, Rashid M, Rust AG, Keane TM, Delrosario R, Jen KY, Gurley KE, Kemp CJ, Fredlund E, Quigley DA, Adams DJ, Balmain A. The mutational landscapes of genetic and chemical models of Kras-driven lung cancer [J]. *Nature*, 2015, **517** (7535): 489-492
 - 24 Koren S, Reavie L, Couto JP, De Silva D, Stadler MB, Roloff T, Britschgi A, Eichlisberger T, Kohler H, Aina O, Cardiff RD, Bentiress-Alj M. PIK3CA (H1047R) induces multipotency and multi-lineage mammary tumours [J]. *Nature*, 2015, **525** (7567): 114-118
 - 25 Wrighton KH. Tumour suppressors: role of nuclear PTEN revealed [J]. *Nat Rev Cancer*, 2011, **11** (3): 154
 - 26 Linardou H, Dahabreh IJ, Kanaklopiti D, Siannis F, Bafaloukos D, Kosmidis P, Papadimitriou CA, Murray S. Assessment of somatic K-RAS mutations as a mechanism associated with resistance to EGFR-targeted agents: a systematic review and meta-analysis of studies in advanced nonsmall-cell lung cancer and metastatic colorectal cancer [J]. *Lancet Oncol*, 2008, **9** (10): 962-972
 - 27 Lee CC, Jia Y, Li N, Sun X, Ng K, Ambing E, Gao MY, Hua S, Chen, C Kim, S, Michellys PY, Lesley SA, Harris JL, Spraggon G. Crystal structure of the ALK (anaplastic lymphoma kinase) catalytic domain [J]. *Biochem J*, 2010, **430** (3): 425-437
 - 28 Morris SW, Naeve C, Mathew P, James PL, Kirstein MN, Cui X, Witte DP. ALK, the chromosome 2 gene locus altered by the t(2;5) in nonHodgkin's lymphoma, encodes a novel neural receptor tyrosine kinase that is highly related to leukocyte tyrosine kinase (LTK) [J]. *Oncogene*, 1997, **14** (18): 2175-2188
 - 29 Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashin K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer [J]. *Nature*, 2007, **448** (7153): 561-566