

附件 2.

学校编码: 10384

密级

学号: 24520121153102

厦 门 大 学

硕 士 学 位 论 文

食管鳞癌的体细胞水平基因组变异

Exomic somatic variation of esophageal squamous cell carcinoma

李慧莉

指导教师姓名: 李奇渊 副教授

专 业 名 称: 病理学与病理生理学

论文提交日期: 2015 年 4 月

论文答辩日期: 2015 年 5 月

2015 年 4 月

食管鳞癌的体细胞水平基因组变异

李慧莉

指导教师

李奇渊

副教授

厦门大学

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下, 独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果, 均在文中以适当方式明确标明, 并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外, 该学位论文为()课题(组)的研究成果, 获得()课题(组)经费或实验室的资助, 在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称, 未有此项声明内容的, 可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月

摘要

目的: 食管鳞状细胞癌 (esophageal squamous cell carcinoma, ESCC) 是全球常见的致死性恶性肿瘤。为进一步探讨其分子机制, 我们利用全基因组外显子测序技术寻找食管鳞癌中多发的突变, 包括点突变、结构变异等, 并且进一步鉴定食管鳞癌中多发的单核苷酸变异 (Single nucleotide variation, SNV)。在体细胞水平探讨食管鳞癌发生发展的分子机制。

方法: 1. 通过外显子测序生成并收集数据库 ESCC 测序数据。2. 128 对食管鳞癌外显子测序数据采用新一代基因序列比对方法 (Burrows Wheeler Alignment, BWA) 进行比对进而运用突变计算软件 (Genome Analysis Toolkit, GATK)、(Mutation Detect, MuTect) 发现 SNV。3. 用数学分析方法来识别那些比预计随机发生更加频繁发生突变的基因 (Mutation Significance covariates, MutSigCV) 进行突变筛选。4. 运用 R 语言统计软件比较 SNV 计算方法和不同研究发现的 SNV。5. 用剪接体揭示结构变异算法 (Clipping Reveals Structure, CREST) 分析 ESCC 外显子序列结构变异 (Structural Variation, SV)。6. 对 ESCC 突变基因进行功能分析 (Gene Ontology, GO)。

结果: 1. 外显子测序生成的 10 对 ESCC 测序数据, 编号为“ZHSHAN10”合并收集的 118 对 ESCC 测序数据, 编号为“SRP30009”共 128 对标本的信息。2. 128 对 ESCC 测序数据比对和突变计算后发现 18863 个 SNV。3. MutSigCV 计算后筛选得到 175 个突变基因。4. *PLEC*、*EPPK1*、*N4BP3*、*KRT35* 等突变频率较高的多发性点突变基因尚未报道。5. ESCC 的 *PLEC* 基因突变主要发生在 (8q24.3) 位

置。6. CREST 算出 1601 个结构变异，其中 14 个结构变异发生频率较高包括：*TITIN*、*MUC16*、*BRCA1*、*FAT1*、*TESK1* (>10%)。7. GO 分析得出 ESCC 的突变基因与 RAS 家族、细胞周期、钙粘蛋白、网蛋白、蛋白磷酸化、甲基化、锌指结构、G 蛋白信号调节、角蛋白相关蛋白、核转录因子、肿瘤坏死因子、凋亡、超激活、染色体开放阅读框等相关。

结论：我们通过分析食管鳞癌的外显子测序数据，在体细胞水平对中国人食管鳞癌的基因组变异进行探讨，并比较不同研究食管鳞癌基因组研究的方法、发现突变的结果等，进而对食管鳞癌多发性点突变、结构变异等信息及其相关的功能通路情况进行分析。为进一步阐明 ESCC 的分子机制、寻找疾病的治疗靶点奠定了理论基础。

关键词：体细胞突变；结构变异；通路功能分析

Abstract

Objective: Esophageal squamous cell carcinoma is a global common lethal cancer with poor survival rates, The aim of this study was to further investigate the molecular mechanisms about the somatic variation of esophageal squamous cell carcinoma through identifying a significant related SNV (multiple single nucleotide mutation in esophageal squamous carcinoma) including: point mutations, insertions deletions. Until we identify and establish the SNV with respect to the development of esophageal squamous cell carcinoma, we would be able to explore the molecular mechanisms of esophageal squamous cell carcinoma at the cellular level and ultimately to provide the scientific basis for molecular targetings therapy of esophageal squamous cell carcinoma in clinic.

Methods: We have collected 128 paired exome sequencing data of esophageal squamous cell carcinoma. Then the related SNV were screened and annotated. Using bioinformatics data analysis method we also analysis the SNV, SV and do the algorithm comparison of 118 cases of ESCC sequencing data which were previously downloaded from the research data of Lin et al. Aiming to find the significant correlated SNV and to compare, as well as to analyses the functional pathway related to ESCC.

Results: In the study about SNV of ESCC, 128 cases with clinical information of ESCC patients and all the SNV were in our statistics. We identified 18863 genes that

are recurrently mutated in esophageal squamous cell carcinoma, of which 175 SNV (0.93%) were screened by using algorithm of MutSigCV. No study of SNV about gene PLEC, EPPK1, N4BP3, KRT35 so far have found in the esophageal squamous cell carcinomas. We also found that the PLEC mutations occur in the position (8q24.3). In the study of structural variations, we analysed 1601 structural variants by using CREST (Clipping REveals STructure) algorithm. Then we also measured 14 fusion gene with the occurrence frequency more than 10%: *TITIN*, *MUC16*, *BRCA1*, *FAT1*, and *TESK1* et al. Finally, the functional pathway about the significant SNV and SV were analysed to examine their molecular mechanism in the esophageal squamous cell carcinoma development.

Conclusion: We analysed exon sequencing data about esophageal squamous cell carcinoma and then discussed the somatic variation of the esophageal squamous cell carcinoma of Chinese at the cellular level. We also gathered and compared the different algorithm, SNV from different research data of ESCC. The information of SNV, SV and the functional pathway related to ESCC have been achieved by this study. These findings were of great importance for future studies in finding and identifying SNV, as well as the functional verification of esophageal squamous cell carcinoma genomic.

Keywords: somatic mutations; structural variation; functional pathway analysis

目 录

中文摘要.....	I
英文摘要.....	III
第 1 章 前言	1
第 2 章 实验材料、方法和数据分析	6
2.1 实验材料	6
2.1.1 组织标本.....	6
2.1.2 PCR 引物.....	6
2.1.3 主要试剂.....	6
2.1.4 主要仪器.....	6
2.1.5 数据分析软件.....	6
2.2 实验方法	7
2.2.1 组织标本 DNA 提取.....	7
2.2.2 基因组 DNA 的 PCR 反应.....	8
2.2.3 全外显子建库测序.....	8
2.2.4 数据质控流程.....	9
2.2.4.1 数据说明	10
2.2.4.2 数据过滤方法.....	11
2.3 数据分析	11
2.3.1 测序数据比对与突变计算.....	11
2.3.2 突变基因功能注释.....	12
2.3.3 突变筛选.....	12

2.3.4 突变 Sanger 法测序验证.....	12
2.3.5 结构变异分析.....	13
第 3 章 结果	14
3.1 收集 ESCC 标本信息	16
3.2 外显子测序数据质量控制	16
3.2.1 测序数据统计.....	16
3.2.2 数据覆盖度统计.....	19
3.3 突变计算参数质量控制	21
3.4 SNV 汇总	23
3.5 发现的 SNV 比较	24
3.5.1 SNV 计算方法比较.....	24
3.5.2 SNV 碱基突变分布频率比较.....	25
3.5.3 不同研究发现的 SNV 比较.....	27
3.5.4 发现的 SNV 类型.....	29
3.6 SNV Sanger 法测序验证.....	31
3.7 128 对 ESCC 多发点突变基因功能.....	33
3.8 结构变异算法参数质量控制	34
3.9 ESCC 结构变异.....	35
3.9.1 ESCC 结构变异汇总.....	35
3.9.2 复发频率高于 10%的结构变异.....	36
第 4 章 讨论	38
4.1 ESCC 的体细胞水平突变研究情况.....	38
4.2 突变研究比较	39
4.2.1 突变计算方法、发现的 SNV、碱基突变分布比较.....	39
4.3 突变影响的基因	40
4.3.1 <i>PLEC</i> 基因.....	40

4.3.2	<i>RAB</i> 家族	41
4.3.3	<i>TPSD1</i> 基因	42
4.3.4	<i>ABCB4</i> 基因	43
4.3.5	<i>GAL3ST3</i> 基因	44
4.3.6	<i>MUC4</i> 基因	44
4.3.7	<i>CTTN</i> 基因	45
4.3.8	<i>PTEN</i> 基因	46
4.4	ESCC 多发点突变基因参与的功能通路	47
4.4.1	<i>PI3K/AKT/mTOR</i> 信号转导通路 与食管鳞癌	48
4.4.2	<i>EGFR</i> 在 ESCC 中的基础研究	49
4.4.3	ESCC 结构变异	51
4.5	ESCC 体细胞水平基因组变异探讨的临床意义	51
第 5 章	结论	55
	参考文献	56
	致 谢	65
	硕士期间发表文章	66
	附表	67

Table of Contents

Abstract in Chinese	I
Abstract in English	III
Chapter 1 Introduction	1
Chapter 2 Materials, Methods, Steps and data analysis	6
2.1 Materials	6
2.1.1 Tissue samples.....	6
2.1.2 PCR primer.....	6
2.1.3 Reagents.....	6
2.1.4 Instruments.....	6
2.1.5 Data analysis software.....	6
2.2 Methods	7
2.2.1 DNA extraction from tissue samples.....	7
2.2.2 PCR for genomic DNA.....	8
2.2.3 All Exon sequencing library construction.....	8
2.2.4 Processes of data quality control.....	9
2.2.4.1 Data description.....	10
2.2.4.2 Data filtering methods.....	11
2.3 Data analysis	11
2.3.1 sequence alignment and mutation calculation.....	11
2.3.2 annotation for gene function.....	12
2.3.3 SNV screening.....	12
2.3.4 Sanger sequence for mutation validation.....	12
2.3.5 Data analysis for structural variation.....	13

Chapter 3 Results	14
3.1 Specimens information collected for ESCC samples	16
3.2 The results of exon sequencing data quality control	16
3.2.1 statistical analysis of sequencing data.....	16
3.2.2 statistics of data coverage.....	19
3.3 Data quality control about mutation calculated parameters ...	21
3.4 SNV summary	23
3.5 SNV comparison	24
3.5.1 SNV detected methods comparison.....	24
3.5.2 SNVmutation frequency distribution comparison.....	25
3.5.3 SNVdetected comparison.....	27
3.5.4 SNV type comparison.....	29
3.6 Sanger sequence for mutation validation	31
3.7 The gene function pathway about SNV for128 pairs of ESCC samples	33
3.8 Data quality control about structural variation calculated parameters	34
3.9 Structural variation	35
3.9.1 Fusion gene summary.....	35
3.9.2 The fusion gene of recurrence frequency more than 10%....	36
Chapter 4 Discusses	38
4.1 The study of ESCC genomica variation at the somatic cell levels	38
4.2 The comparison about studies	39
4.2.1 The comparison about the SNV,SV algorithm and the distribution frequency of them.....	39

4.3 The related gene of SNV	40
4.3.1 <i>PLEC</i> gene.....	40
4.3.2 <i>PTEN</i> gene.....	41
4.3.3 <i>RAB</i> family.....	42
4.3.4 <i>TPSD1</i> gene.....	43
4.3.5 <i>ABCB4</i> gene.....	44
4.3.6 <i>GAL3ST3</i> gene.....	44
4.3.7 <i>MUC4</i> gene.....	45
4.3.8 <i>CTTN</i> gene.....	46
4.4 The functional pathway about the SNV	47
4.4.1 <i>PI3K/AKT/mTOR</i> signal pathway related to ESCC.....	48
4.4.2 <i>EGFR</i> basis research in ESCC.....	49
4.4.3 The Structural Variation about ESCC.....	51
4.5 The clinical significancy in the study about the genomicvariation of ESCC at the somatic cell levels	51
Chapter 5 Conclusion	55
References	56
Acknowledgements	65
Published articles	66
Appendices	67

第 1 章 前言

食管癌是全球第 6 大致死性肿瘤，其发病在不同地理位置、不同种族、不同社会文化、不同环境间存在巨大差异。其病理类型主要是食管鳞状细胞癌（esophageal squamous cell carcinoma, ESCC）^[1, 2]。包括中国在内的东亚地区，食管鳞癌在全部食管恶性肿瘤中占有 90% 以上^[3, 4]。每年新增患者人数超过 48 万例，患者的 5 年生存率差（15%~25%）^[5, 6]，且多数食管癌患者目前不能手术切除或伴有转移性疾病。据国际癌症研究机构（International Agency for Research on Cancer, IARC）的估计，我国食管癌的发病率和死亡率均排在我国全部恶性肿瘤的第 4 位，年死亡人数超过 21 万人^[7]。此外，环境因素对食管鳞癌的发病具有极其重要的意义^[8]。无论是在西方国家抑或是包括中国在内的亚洲国家，吸烟与饮酒都被认为是食管鳞癌发病最主要的环境因素^[2]。其他环境因素包括不均衡的饮食、营养缺乏、环境暴露以及感染性病原体（如乳头瘤病毒和真菌）等因素都被认为可能在食管鳞癌的发病过程中发挥重要作用^[9]。

肿瘤在基因水平发生的改变包括扩增，缺失，易位等。肿瘤的基因变异可能涉及不同癌基因的激活^[10]。癌基因变异被认为可由生长因子、生长因子受体、信号转导因子、转录因子及程序性细胞死亡条件因子等的调节异常而产生^[11]。体细胞中积累的各种基因变异可能与肿瘤的发生、发展相关^[10]。体细胞水平基因突变、结构变异等可导致蛋白结构的改变，出现蛋白活性的提高和调节功能的丧失；或者通过错误的调控使蛋白表达、蛋白丰度显著提高，这些都可能激活原癌基因^[12]。因此在体细胞水平进行基因组变异的探讨有助于鉴别更多癌基

因激活导致的变异。有研究表明基因结构变异可导致不同的分子事件发生，包括基因的激活、失活、转录调节异常、扩增、缺失等^[13]。大片段基因的丢失或失活对肿瘤的形成具有重要作用。基因序列的变化可能涉及癌基因或抑癌基因的激活进而调控细胞的生长与分化，使受影响的细胞产生克隆瘤性增殖而导致肿瘤的发生^[14]。在对大片段碱基序列改变的研究中发现染色体异位可导致在错误的时间或错误的细胞中基因出现高表达。继而产生有活性的结构变异蛋白^[15, 16]。

Xiong 等人的研究表明 PTKs 由原癌基因或癌基因编码，由 PTKs（蛋白质酪氨酸激酶）和 PTPs（蛋白质酪氨酸磷酸酶）精确调控的蛋白质酪氨酸磷酸化水平在非小细胞肺癌（non-small cell of lung cancer, NSCLC）发生过程中起重要作用^[17]。研究还发现第一个编码 PTPs 的抑癌基因磷酸酶-张力蛋白基因（PTEN）的第 10 号染色体同源丢失。PTEN 基因的变异导致 PTEN-PI3K/Akt 信号通路的失衡可导致 NSCLC 的恶性转化和较差的预后。Li 和 Steck 阐明脑肿瘤、前列腺癌及乳腺癌肿瘤细胞株和一些原发肿瘤中有不同形式的体细胞突变基因的发生。^[18-20]。这些研究说明体细胞中的基因突变以及结构变异对于肿瘤的发生发展至关重要。

因为不同研究在体细胞水平对多种肿瘤基因突变研究为肿瘤分子机制的探讨提供了很多信息，所以我们有必要在体细胞水平对食管鳞癌基因组变异进行探讨以了解食管鳞癌的分子机制，指导食管鳞癌的早期诊断，判断预后以及研究药物靶点治疗等。

目前，国内外的肿瘤研究领域遍及基因芯片、基因测序技术。由于基因组外显子区域仅占整个基因组的 1%左右，外显子直接编码蛋白质合成可以揭示个

Degree papers are in the “[Xiamen University Electronic Theses and Dissertations Database](#)”.

Fulltexts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.