

Разработка таргетной панели для молекулярно-генетической диагностики рака щитовидной железы

Якушина В.Д.¹, Зайцева М.А.², Павлов А.Е.², Лернер Л.В.³, Лавров А.В.^{1,4}

¹ — Федеральное государственное бюджетное научное учреждение «Медико-генетический научный центр», Москва, Российская Федерация, e-mail: vdyakushina@gmail.ru

² — Компания Парсек Лаб, Санкт-Петербург

³ — ООО «ПреМед» Научно-клинический центр ПреМед — Европейские технологии, Москва, Российская Федерация

⁴ — Государственное бюджетное образовательное учреждение высшего профессионального образования «Российский национальный исследовательский медицинский университет им. Н.И. Пирогова» Министерства здравоохранения Российской Федерации, Москва, Российская Федерация

Рак щитовидной железы (РЩЖ) является самым распространенным злокачественным заболеванием эндокринной системы. Ключевой метод диагностики РЩЖ — цитологическое исследование клеток щитовидной железы, получаемых в результате тонкоигольной аспирационной биопсии (ТАБ). В 20–30% случаев ТАБ выявляет атипию неопределенного значения, что не позволяет клиницисту определить тактику ведения пациента: оперировать или лечить консервативно. В рамках научных исследований в большинстве случаев РЩЖ удается выявить драйверные мутации или иные генетические маркеры. Их определение может повысить точность диагностики РЩЖ. **Цель данной работы** — разработка метода молекулярно-генетической диагностики РЩЖ с применением высокопроизводительного параллельного секвенирования. Был проведен анализ литературы, базы данных (БД) COSMIC, а также результатов исследования проекта The Cancer Genome Atlas, в ходе которого было отобрано 456 точковых соматических мутаций в 25 генах, 23 генные транслокации, а также 3 мутации типа изменения числа копий гена (copy number variations, CNV), характерные для РЩЖ. Для детекции точковых мутаций, коротких инсерций/делеций и CNV в инструменте AmpliSeq Designer был создан дизайн панели, содержащий 221 пару праймеров (в 2 пулах), покрывающий 99,59% выбранных таргетных регионов. В дизайн также были включены регионы гена *RET*, несущие герминативные мутации для детекции наследственного медулярного РЩЖ. С помощью инструмента RNA Gene Fusion designs был разработан дизайн панели для детекции 23 перестроек.

Ключевые слова: рак щитовидной железы, таргетная панель, соматические мутации, гены, геномные перестройки, генетическая диагностика, *KRAS*, *NRAS*, *BRAF*

Введение

Рак щитовидной железы (РЩЖ) лидирует по частоте среди злокачественных новообразований эндокринной системы [1]. Диагностика РЩЖ основана на проведении цитологического анализа биоптатов, полученных методом ТАБ. Существенным недостатком данного диагностического метода является достаточно высокий процент (20–30%) так называемых «неопределенных» диагнозов, соответствующих III, IV и V группе по классификации Bethesda, что ведет к выбору неправильной тактики лечения [2]. Предполагается, что повышение точности диагностики может быть достигнуто за счет использования молекулярно-генетических методов анализа.

В настоящее время доступны тесты, выявляющие мутации генов *BRAF*, *NRAS*, *KRAS*. Однако мутации гена *BRAF* обнаруживают лишь в 40–60% случаев папиллярного РЩЖ, а мутации в генах *RAS* выявляют в 10–20% папиллярного РЩЖ и 40–50% фолликулярного РЩЖ. Таким образом, использование данных тестов позволяет идентифицировать не более 80% папиллярного РЩЖ и не более 50% фолликулярного РЩЖ. При этом в настоящее время известно более 400 мутаций, связанных

с РЩЖ. Учитывая их количество и низкую частоту каждой отдельной мутации целесообразно разработать метод молекулярно-генетической диагностики РЩЖ с применением высокопроизводительного параллельного секвенирования.

Методы

Выбор мутаций для включения в панель

Для формирования списка мутаций была проанализирована англоязычная литература, база данных COSMIC и результаты исследования The Cancer Genome Atlas, в том числе с использованием cBioPortal for Cancer Genomics [3, 4, 5].

Поиск статей был выполнен в базе PubMed по следующим ключевым словам: «genomic thyroid cancer», «thyroid cancer mutation», «thyroid cancer diagnosis».

С портала COSMIC была импортирована база «CosmicMutantExport», 77 v, содержащая все точковые кодирующие мутации, установленные в результате таргетных и полногеномных исследований [3]. В полученных данных были отфильтрованы записи, соответствующие значению «Thyroid» в поле Primary Site.

С сайта cBioPortal for Cancer Genomics были импортированы данные, удовлетворяющие следующим условиям поиска: Cancer Study — Poorly-Differentiated and Anaplastic Thyroid Cancers (MSKCC, JCI 2016), Papillary Thyroid Carcinoma (TCGA, Cell 2014), Thyroid Carcinoma (TCGA, Provisional); Data Type Priority — Mutation and CNA; Gene Set — PPM1D [4, 5].

Дизайн пула праймеров для детекции точковых мутаций и изменений числа копий гена (CNV)

Дизайн пула праймеров осуществляли инструментом версии 5.4.1 по референсному геному GRCh37. Для оценки дизайна были использовали инструменты IGV (Integrative Genomic Viewer) и UCSC Genome Browser.

Дизайн пула праймеров для детекции транслокаций (fusion-мутаций)

Дизайн для детекции транслокаций создавали в специализированном разделе AmpliSeq Designer — RNA Gene Fusion designs. В дизайн также были включены гены для оценки клеточного состава образца: *CALCA*,

KRT7, KRT20, PGK1, PTH, SLC5A5, TG, TTF1, HMBS, TBP, MRPL13, LRP1.

Результаты

Панель для детекции точковых мутаций и CNV

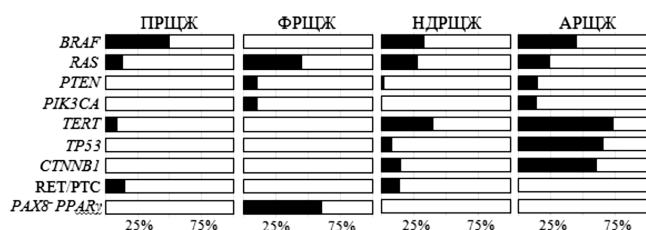
В результате анализа литературы и БД был определен список точковых мутаций и CNV, встречающихся при основных типах РЩЖ (папиллярный, фолликулярный, низкодифференцированный, анапластический и медуллярный). Всего было отобрано 25 генов с точковыми мутациями.

Из БД COSMIC были импортированы мутации, встречающиеся более чем в 1 случае РЩЖ. Всего было импортировано 456 соматических мутаций в 25 генах. Распределение количества мутаций по генам было следующим: *TP53* — 62 варианта, *RET* — 51, *TSHR* — 49, *PIK3CA* — 47, *BRAF* — 31, *CTNNB1* — 30, *PTEN* — 30, *HRAS* — 24, *KRAS* — 21, *APC* — 20, *NRAS* — 14, *SMAD4* — 12, *CDKN2A* — 9, *CDKN2C* — 9, *IDH1* — 7, *EGFR* — 7, *TERT* — 6, *GNAS* — 6, *DICER1* — 5, *CHEK2* — 5, *EIF1AX* — 4, *AKT1* — 3, *MET* — 2, *PPM1D* — 1, *NF2* — 1 [3].

Таблица 1

Гены и экзоны, включенные в таргетную панель

Ген	Экзоны
<i>CDKN2C</i>	1, 2
<i>NRAS</i>	2, 3
<i>IDH1</i>	4, 6
<i>CTNNB1</i>	3, 15
<i>PIK3CA</i>	2, 3, 5, 10, 14, 21
<i>TERT</i>	2, 4
<i>APC</i>	7, 10, 14, 16, 17
<i>EGFR</i>	18, 19, 21
<i>MET</i>	14, 21
<i>BRAF</i>	11, 15
<i>CDKN2A</i>	1, 2
<i>RET</i>	Все
<i>PTEN</i>	1, 5, 6, 7, 8
<i>HRAS</i>	2, 3, 4
<i>KRAS</i>	2, 3, 4
<i>TSHR</i>	9, 10
<i>DICER1</i>	24, 25, 27
<i>AKT1</i>	2, 4
<i>TP53</i>	1, 2, 3, 4, 5, 8
<i>PPM1D</i>	1, 2, 5, 6
<i>SMAD4</i>	5, 6, 7, 8, 9, 10, 12
<i>GNAS</i>	8, 9
<i>CHEK2</i>	3, 4, 12, 14
<i>NF2</i>	4
<i>EIF1AX</i>	2, 6



Распространенность наиболее частых мутаций по основным типам РЩЖ. ПРЩЖ — папиллярный рак щитовидной железы; ФРЩЖ — фолликулярный рак щитовидной железы; НДРЩЖ — низкодифференцированный рак щитовидной железы; АРЩЖ — анапластический рак щитовидной железы.

Экзоны, несущие кластеры мутаций, были включены в дизайн целиком (табл. 1).

Подбор праймеров для детекции отобранных мутаций осуществляли с помощью инструмента AmpliSeq Designer. Был выбран максимально доступный размер ампликона (125–375 п.н.), что обеспечивает лучшее покрытие целевых регионов. Часто расположенные мутации были объединены в кластеры. Для гена *RET* был подобран дизайн, покрывающий экзоны, 3' и 5' UTR области.

Основная доля точковых мутаций при РЩЖ приходится на гены сигнальных путей MAPK и PI3K/Akt/mTOR: *BRAF*, *HRAS*, *NRAS*, *KRAS*, *PIK3CA*, *AKT1* и *PTEN* (рисунок) [6, 7, 8, 9]. Гены, для которых характерны мутации при низкодифференцированном и анапластическом РЩЖ включают *CTNNB1*, *TP53* и *TERT* (рисунок) [1, 8]. Также в литературе описано достаточно большое количество случаев мутаций гена *APC* при РЩЖ. Другая группа генов, которая может быть выделена по результатам анализа литературы включает гены *SMAD4*, *GNAS* и *TSHR*, мутации в которых выявляются как при злокачественных, так и доброкачественных изменениях ЩЖ [1, 6, 10, 11]. К генам, мутации которых встречаются при РЩЖ редко, можно отнести *EGFR*, *CDKN2A*, *CDKN2C*, *DICER1*, *MET*, *NF2*, *IDH1* [1]. Мутации генов *EIF1AX*, *PPM1D* и *CHEK2* были выявлены в рамках проекта The Cancer Genome Atlas (TCGA) в результате полногеномного и полноэкзомного секвенирования клеток папиллярного РЩЖ [3]. Мутации гена *EIF1AX* считаются предположительно драйверными, однако Karunamurthy A. с соавт. (2016) показали, что мутации данного гена могут быть обнаружены и при доброкачественных изменениях ЩЖ.

Для медулярного РЩЖ характерны мутации гена *RET*, как соматические, так и герминативные [12].

Детекция CNV

Помимо точковых мутаций в результате анализа литературы были выбраны 3 CNV, характерные для РЩЖ: *SCNA-22q-del*, *9q21.3-q32 del*, *1q gain* [3].

На сегодняшний день доступны инструменты, позволяющие идентифицировать с относительно высокой чувствительностью только протяженные вариации копийности (более 20 ампликонов) в данных таргетного секвенирования (Ion Reporter, ThermoFisher Scientific). Однако идентификация небольших вариаций копийности, затрагивающих несколько ампликонов, и являющихся соматическими вариантами, является более сложной задачей, поскольку высока вероятность случайного отклонения представленности ампликона, то есть высока вероятность ложной детекции.

Для детекции выбранных CNV на основе оценки глубины покрытия в панель добавлено 20 дополнительных ампликонов для каждого CNV. Для повышения специфичности детекции CNV разработанная панель содержит дополнительные регионы, на границах с целевыми регионами, так называемые «опорные ампликоны». Дополнительные ампликоны позволяют снизить вероятность ложноположительной детекции, так как вероятность случайного достоверного отклонения покрытия нескольких соседних ампликонов невысока. Было оценено оптимальное расположение регионов по хромосомам 22, 9 и 1 с учетом их GC состава.

Детекцию CNV планируется осуществлять с помощью инструмента CONVector разработки ООО «ПАРСЕК ЛАБ» (<https://github.com/parseq/convector>). Инструмент CONVector оценивает покрытие каждого ампликона во всех образцах, анализируемых за один запуск секвенатора. На основании полученных данных CONVector выбирает ампликоны, покрытие которых коррелирует между собой. Далее программа ищет образцы, в которых покрытие по каждому из ампликонов отклоняется от общего покрытия внутри группы коррелирующих ампликонов. CONVector верифицирован на клинических образцах для детекции небольших вариаций копийности с высокой чувствительностью для данных таргетного секвенирования, полученных с помощью технологии AmpliSeq (ThermoFisher Scientific).

Таблица 2

Основные метрики итогового дизайна для детекции точечных мутаций и CNV

Название и ID дизайна	
Версия AmpliSeq Designer	5.4.1
Размер панели	62,49 т.п.н.
Покрытие	99,59%
Размер ампликона	125–375 п.н.
Кол-во пар праймеров	221 пара (в 2 пулах)

С учетом изменений для детекции CNV был получен итоговый дизайн панели для детекции точковых мутаций и CNV, основные метрики которого приведены в табл. 2.

Панель для детекции перестроек

Наиболее распространенными при РЩЖ являются геномные перестройки с участием гена *RET*: *RET-CCDC6* и *RET-NCOA4*. Также при РЩЖ описаны другие перестройки гена *RET*, в том числе *RET-PRKARIA*, *RET-GOLGA5*, *RET-KTN1*, *RET-TRIM24* и *RET-TRIM27*. Перестройки с участием *RET* встречаются в 10–20% случаев папиллярного РЩЖ. Перестройки других генов, характерные для РЩЖ, включили *PAX8-PPAR γ* , *AKAP9-BRAF*, *ETV6-NTRK3*, *STRN-ALK*, *EML4-ALK*, *CRTC1-MAML2*, *MACF1-BRAF*, *TPM3-NTRK1*, *TFG-NTRK1*, *TPR-NTRK1* [13, 14, 15]. Перестройки *PAX8-PPAR γ* наблюдаются в 60% случаев фолликулярного РЩЖ [16].

Из перестроек, описанных в базе COSMIC, в целевую панель было решено включить все, встретившиеся хотя бы в 1 случае РЩЖ (23 перестройки). Среди включенных перестроек наиболее распространенными являются следующие:

- *CCDC6:RET* (*CCDC6* 1 экзон, *RET* 12-20 экзон, COSF1272)
- *NcoA4:RET* (*NcoA4* 1-8 экзон, *RET* 12-20 экзон, COSF1492)
- *NcoA4:RET* (*NcoA4* 1-7 экзон, *RET* 12-20 экзон, COSF1341)
- *PRKARIA:RET* (*PRKARIA* 1-7 экзон, *RET* 12-20 экзон, COSF1512)
- *KTN1:RET* (*KTN1* 1-29 экзон, *RET* 12-20 экзон, COSF1514)
- *ERC1:RET* (*ERC1* 1-11 экзон, *RET* 12-20 экзон, COSF1508)
- *GOLGA5:RET* (*GOLGA5* 1-7 экзон, *RET* 12-20 экзон, COSF1504)
- *PAX8:PPAR γ* (*PAX8* 1-10 экзон, *PPARG* 2-7 экзон, COSF1220)
- *PAX8:PPAR γ* (*PAX8* 1-8 экзон, *PPARG* 2-7 экзон, COSF1224)
- *PAX8:PPAR γ* (*PAX8* 1-9 экзон, *PPARG* 2-7 экзон, COSF1218)
- *AKAP9:BRAF* (*AKAP9* 1-8 экзон, *BRAF* 9-18 экзон, COSF1017)
- *ETV6:NTRK3* (*ETV6* 1-4 экзон, *NTRK3* 14-19 экзон, COSF1535)
- *TPM3:NTRK1* (*TPM3* 1-7 экзон, *NTRK1* 10-17 экзон, COSF1330)
- *STRN:ALK* (*STRN* 1-3 экзон, *ALK* 20-29 экзон, COSF1431)
- *EML4:ALK* (*EML4* 1-13 экзон, *ALK* 20-29 экзон, COSF463)
- *CRTC1:MAML2* (*CRTC1* 1 экзон, *MAML2* 3-6 экзон, COSF1103).

Заключение

Нами была спроектирована диагностическая панель для определения 456 драйверных соматических мутаций в 25 генах и детекции 23 геномных перестроек, ассоциированных с РЩЖ. Панель предназначена для определения мутаций в известном на сегодняшний день драйверных генах папиллярного и фолликулярного РЩЖ, мутаций, характерных для низкодифференцированного, анапластического и медулярного РЩЖ, редких мутаций, описанных в литературе, а также герминативных мутаций, ассоциированных с наследственным медулярным РЩЖ.

Список литературы

1. Hsiao SJ, Nikiforov YE. Molecular approaches to thyroid cancer diagnosis. *Endocr Relat Cancer*. 2014;21(5):T301-T313. doi:10.1530/ERC-14-0166.
2. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda system for reporting thyroid cytopathology: A meta-analysis. *Acta Cytol*. 2012;56(4):333-339. doi:10.1159/000339959.
3. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 2015;43(D1):D805-D811. doi:10.1093/nar/gku1075.
4. Cerami E, Gao J, Dogrusoz U, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov*. 2012;2(5):401-404. doi:10.1158/2159-8290.CD-12-0095.
5. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci Signal*. 2013;6(269):p11-p11. doi:10.1126/scisignal.2004088.
6. Cancer Genome Atlas Research Network N, Akbani R, Aksoy BA, et al. Integrated genomic characterization of papillary thyroid carcinoma. *Cell*. 2014;159(3):676-690. doi:10.1016/j.cell.2014.09.050.
7. Hall, R.D., & Kudchadkar, R.R. (2014). BRAF Mutations: Signaling, Epidemiology, and Clinical Experience in Multiple Malignancies. *Cancer Control*, 21(221). Retrieved from www.henrydomke.com.
8. Landa I, Ibrahimasic T, Boucai L, et al. Genomic and transcriptomic hallmarks of poorly differentiated and anaplastic thyroid cancers. *J Clin Invest*. 2016;126(3):1052-1066. doi:10.1172/JCI85271.
9. Xing M, Pylayeva-Gupta Y, Grabocka E, et al. Clinical utility of RAS mutations in thyroid cancer: a blurred picture now emerging clearer. *BMC Med*. 2016;14(1):12. doi:10.1186/s12916-016-0559-9.
10. Lazzereschi D, Nardi F, Turco A, et al. A complex pattern of mutations and abnormal splicing of Smad4 is present in thyroid tumors. *Oncogene*. 2005;24(34):5344-5354. doi:10.1038/sj.onc.1208603.
11. D'Inzeo S, Nicolussi A, Donini CF, et al. A novel human Smad4 mutation is involved in papillary thyroid carcinoma progression. *Endocr Relat Cancer*. 2012;19(1):39-55. doi:10.1530/ERC-11-0233.
12. Figlioli G, Landi S, Romei C, Elisei R, Gemignani F. Medullary thyroid carcinoma (MTC) and RET proto-oncogene: Mutation spectrum in the familial cases and a meta-analysis of studies on the sporadic form. *Mutat Res — Rev Mutat Res*. 2013;752(1):36-44. doi:10.1016/j.mrrev.2012.09.002.
13. Karunamurthy A, Panebianco F, Hsiao S, et al. Prevalence and phenotypic characteristics of EIF1AX mutations in thyroid nodules. *Endocr Relat Cancer*. 2016;(February):ERC — 16-0043 — . doi:10.1530/ERC-16-0043.
14. Ciampi R, Knauf JA, Kerler R, et al. Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. *J Clin Invest*. 2005;115(1):94-101. doi:10.1172/JCI200523237.

15. Marotta V, Guerra A, Sapio MR, Vitale M. RET/PTC rearrangement in benign and malignant thyroid diseases: a clinical standpoint. *Eur J Endocrinol.* 2011;165(4):499-507. doi:10.1530/eje-11-0499.

16. Eberhardt NL, Grebe SKG, McIver B, Reddi H V. The role of the PAX8/PPARgamma fusion oncogene in the pathogenesis of

follicular thyroid cancer. *Mol Cell Endocrinol.* 2010;321(1):50-56. doi:10.1155/2008/672829.

17. Duan J, Zhang J-G, Deng H-W, Wang Y-P. Comparative Studies of Copy Number Variation Detection Methods for Next-Generation Sequencing Technologies. Salamin N, ed. *PLoS ONE.* 2013;8(3):e59128. doi:10.1371/journal.pone.0059128.

Design of targeted gene panel for molecular diagnostics of thyroid cancer

Yakushina V.D.¹, Zaytseva M.A.², Pavlov A.E.², Lerner L.V.³, Lavrov A.V.^{1,4}

¹ – Federal State Budgetary Institution «Research Centre for Medical Genetics», Moscow, Russian Federation, e-mail: vdyakushina@gmail.ru

² – Parseq Lab, St. Petersburg

³ – PreMed-European Technologies, Ulitsa Petrovka, 17, Building 3, Moscow, 107031, Russian Federation

⁴ – Russian National Research Medical University, Ostrovityanova str., 1, Moscow, 117997, Russian Federation

Thyroid cancer is the most common endocrine malignancy. The key approach for thyroid cancer diagnosis is cytology of fine needle aspiration biopsy (FNA) samples. FNA specimens have indeterminate cytology in 20–30%. This results in wrong clinical diagnosis and inappropriate treatment strategy. Currently known mutations describe vast majority of thyroid cancer cases. Detection of the driver mutations is supposed to improve diagnostic accuracy. The aim of the work is to develop next-generation sequencing based diagnostic panel for thyroid cancer. The analysis of the English-language literature, COSMIC database (DB), as well as results of the research project the Cancer Genome Atlas was performed. In total, 456 point somatic mutations in 25 genes, 23 genetic translocations, and 3 copy number variations (CNV) mutations were identified. Using AmpliSeq Designer, 2 Custom Panels were created – for the detection of point mutations, small indels and CNV (1) and for the detection of translocations (2). The custom Panel for the detection of point mutations, small indels and CNV contains 221 primer pairs in 2 pools, covering 99.59% selected targeted regions. The design also incorporated the regions of the *RET* gene for detection of germline mutations associated with hereditary medullary thyroid cancer. RNA Gene Fusion designs tool in AmpliSeq Designer was used to design the Panel for the detection of 23 translocations.

Key words: targeted panel, somatic mutations, genes, translocations, genetic diagnostics, thyroid cancer, *KRAS*, *NRAS*, *BRAF*, PTC