

BREAST CANCER: ANALYSIS OF DRIVER SOMATIC MUTATIONS DETECTED BY NEXT-GENERATION SEQUENCING

Tsukanov KYu¹, Krasnenko AYu¹, Korostin DO¹, Churov AV², Stetsenko IF¹, Plotnikov NA¹, Zarubina SA³, Belova VA³, Kovyrshina AV³, Vorotnikov IK⁴, Mescheryakov AA⁴, Ilyinsky VV^{5,3,1} ✉

¹ Genotek Ltd., Moscow, Russia

² Karelian Research Centre of the Russian Academy of Sciences, Petrozavodsk, Russia

³ Vavilov Institute of General Genetics of the Russian Academy of Sciences, Moscow, Russia

⁴ N. N. Blokhin Russian Cancer Research Center, Moscow, Russia

⁵ Institute of Biomedical Chemistry, Moscow, Russia

Breast cancer (BC) is one of the most common malignancies. There is a need for novel approaches to screening for genetic mutations in patients with BC that will help to reduce high mortality rates caused by this disease and improve treatment outcomes. In this study we employed next generation sequencing to screen a few key genes associated with the risk of breast cancer for mutations. We also evaluated their pathogenicity using the previously proposed bioinformatics-based algorithm and analyzed the associations between some of the detected mutations and the clinical manifestations of the disease. Our study recruited 16 female patients with BC (mean age was 50.7 ± 11.3 years). A total of 58 mutations were detected in the oncogenes *BRCA1*, *BRCA2*, *ATM*, *CDH1*, *CHEK2* and *TP53*. Bioinformatic analysis of the sequencing data revealed 14 mutations that affect the sequence of the encoded proteins. Most deleterious mutations were harbored by the genes *BRCA1/2*, *ATM* and *TP53*.

Keywords: breast cancer, next-generation sequencing, somatic mutation, oncogenes, *BRCA1*, *BRCA2*, *TP53*

Funding: this work was supported by the Ministry of Education and Science of the Russian Federation (Project ID RFMEFI60716X0152).

✉ **Correspondence should be addressed:** Valery Ilyinsky
Nastavnichesky per., d.17, str. 1, pod. 14, 15, Moscow, 105120; info@genotek.ru

Received: 14.12.2017 **Accepted:** 22.12.2017

РАК МОЛОЧНОЙ ЖЕЛЕЗЫ: АНАЛИЗ СПЕКТРА СОМАТИЧЕСКИХ ДРАЙВЕРНЫХ МУТАЦИЙ С ПРИМЕНЕНИЕМ ВЫСОКОПРОИЗВОДИТЕЛЬНОГО СЕКВЕНИРОВАНИЯ

К. Ю. Цуканов¹, А. Ю. Красненко¹, Д. О. Коростин¹, А. В. Чуров², И. Ф. Стеценко¹, Н. А. Плотников¹, С. А. Зарубина³, В. А. Белова³, А. В. Ковыршина³, И. К. Воротников⁴, А. А. Мещеряков⁴, В. В. Ильинский^{5,3,1} ✉

¹ ООО «Генотек», Москва

² Карельский научный центр Российской академии наук, Петрозаводск

³ Институт общей генетики имени Н. И. Вавилова РАН, Москва

⁴ Национальный медицинский исследовательский центр онкологии имени Н. Н. Блохина, Москва

⁵ Научно-исследовательский институт биомедицинской химии имени В. Н. Ореховича, Москва

Рак молочной железы (РМЖ) представляет собой одну из наиболее распространенных форм злокачественных опухолей. Развитие новых подходов к скринингу генетических изменений у больных с опухолями молочной железы поможет значительно снизить общую высокую смертность от рака этого типа и повысить эффективность противоопухолевой терапии. Целью настоящей работы являлось выявление методом высокопроизводительного секвенирования спектра мутаций в составе ключевых онкогенов при РМЖ оценка их патогенности с применением ранее разработанного биоинформатического алгоритма, а также оценка взаимосвязи некоторых мутаций с особенностями клинического проявления заболевания. В исследовании приняло участие 16 пациенток с РМЖ (средний возраст — 50,7 ± 11,3 года). Было обнаружено 58 мутаций в онкогенах *BRCA1*, *BRCA2*, *ATM*, *CDH1*, *CHEK2* и *TP53*. Среди выявленных генетических вариантов с применением биоинформатических подходов найдено 14 мутаций, оказывающих влияние на последовательность кодируемого белка. Большая часть патогенных мутаций идентифицирована в генах *BRCA1/2*, *ATM* и *TP53*.

Ключевые слова: рак молочной железы, высокопроизводительное секвенирование, соматическая мутация, онкогены, *BRCA1*, *BRCA2*, *TP53*

Финансирование: работа выполнена при поддержке Министерства образования и науки РФ (идентификатор соглашения RFMEFI60716X0152).

✉ **Для корреспонденции:** Ильинский Валерий Владимирович
Наставнический пер., д. 17, стр. 1, под. 14, 15, г. Москва, 105120; info@genotek.ru

Статья получена: 14.12.2017 **Статья принята к печати:** 22.12.2017

Breast cancer (BC) is the second most common type of cancer and the second leading cause of death in women; it is also the most incident cancer worldwide [1]. The risk of BC increases with age: the majority of new cases are reported in women who are 60 to 65 years old. High BC mortality is explained by late diagnosis established when the disease has already progressed to the advanced stage. Metastatic BC is particularly dangerous, since it is resistant even to combination treatments based on chemotherapy, hormones and targeted drugs. The 5-year survival rate in patients with BC is 55 %. This brings the need for novel approaches towards more effective screening as well as targeted therapy of BC based on the molecular genetic profiling of tumors.

The rapid development of next generation sequencing (NGS) has yielded a bulk of information about genetic variants [2]. A lot of mutations are associated with BC, including somatic and germinal mutations in the genes *PIK3CA*, *STK11/LKB1*, *CDH1*, *ATM*, *CHEK2*, *BRIP1*, and *PALB2* and mutant variants of the highly penetrant genes associated with hereditary BC, such as *TP53*, *PTEN*, *MLH1*, *BRCA1*, and *BRCA2* [3].

The majority of tumor mutations are somatic; they have an important role in the pathogenesis of cancer and confer *de novo* resistance to treatment. Thus, a lot of ongoing studies utilize NGS in an attempt to profile mutant variants in tumors. As a result, it has been identified a significant amount of new mutations with unknown function. To describe these polymorphisms, mathematical algorithms are necessary that can automatically process huge data arrays, predict potentially pathogenic mutations and distinguish them from harmless variants. The resulting data can be used when developing screening or diagnostic tools (including liquid biopsy) and selecting adequate targeted therapies.

In this work we analyze a range of mutations identified in key BC oncogenes by NGS, using a previously developed bioinformatic pipeline for the functional annotation of mutations and assessment of their pathogenicity.

METHODS

We obtained tumor samples from 16 patients of Blokhin Russian Cancer Research Center, Moscow. The participants' age range was 27 to 76 years, with a mean of 50.7 ± 11.3 years. All patients had breast malignancies and received combination therapy. The inclusion criteria were as follows: age of 18 to 70 years, sex (all patients were females), histologically and cytologically confirmed breast cancer. The exclusion criteria were a medical history of other tumor types and pregnancy.

Disease stages were determined according to the TNM classification [4]. The study was carried out in the patients with stages T1–3N0–3M0–1.

All patients gave voluntary informed consent. The study complied with the principles of confidentiality. Patients' clinicopathologic features are summarized in Table 1.

DNA isolation and quality control

DNA was isolated from the samples of tumor tissue using DNeasy Blood and Tissue Kit (Qiagen, USA). Tumor tissue was cut into small pieces, and buffer ATL was added to the samples. The samples were then treated with proteinase K, incubated at 56 °C until fully lysed, and treated with RNase A. Next, we added 200 µl buffer AL and 96 % ethanol. The resulting mixture was transferred to spin columns and centrifuged at 8,000 g for 1 min. The samples were washed with AW1 and AW2 buffers

to remove salts (guanidine and SDS). The columns were eluted twice with 30 µl Low-TE buffer; the samples were incubated and centrifuged according to the manufacturer's protocol. Quality control of the obtained DNA was performed on Qubit 3.0 (Thermo Fisher Scientific, USA). The samples were also run on 1 % agarose gel electrophoresis with ethidium bromide.

Sequencing of targeted oncogenes

DNA libraries were prepared using NEBNext Ultra DNA Library Kit for Illumina (New England Biolabs, USA). The libraries were dual-indexed by PCR using NEBNext Ultra DNA Library Prep Kit for Illumina and NEBNext Multiplex Oligos for Illumina (Dual Index Primers Set 1, New England Biolabs). Quality control of the obtained DNA libraries was performed on Agilent Bioanalyzer 2100 (Agilent Technologies, USA) using High Sensitivity Kit by the same manufacturer according to the official protocol.

For targeted enrichment of the coding regions of tumor genomes we used MYbaits Onconome KL v1.5 Panel (MYcroarray, USA). The enriched fragments were sequenced with 100 b. p. paired-end reads on HiSeq 2500 (Illumina, USA). Sample preparation and sequencing were done according to Illumina's protocols.

Bioinformatic analysis

Sequencing data were analyzed using an original algorithm developed previously [5]. First, the quality of reads was checked: sequences with read quality below 10 were removed from NGS data using Cutadapt software [6]. Then the reads were mapped to the reference genome hg19 (GRCh37.p13) using the Burrows–Wheeler Aligner algorithm [7]. PCR-duplicates were removed by running the `rmdup` command in SAMtools [8].

Mutations were called with MuTect [9]. DNA sequences covered by at least 12 reads were considered the most significant.

To assess the functional effect of the discovered mutations, they were annotated in SnpEff and their effect on the encoded protein was predicted based on the analysis of genomic coordinates [10].

RESULTS

Using Illumina-based NGS, we have screened 16 breast tumors for mutations harbored by cancer-associated genes *BRCA1*, *BRCA2*, *ATM*, *CDH1*, *CHEK2*, *MRE11A*, *NBN*, *PALB2*, *PTEN*, *RAD50*, *RAD51C*, *TP53*, and *SEC23B*. Our original bioinformatic algorithm has detected 58 point mutations in the genes *BRCA1*, *BRCA2*, *ATM*, *CDH1*, *CHEK2* and *TP53*, including 19 homozygous and 39 heterozygous variants. The list of unique mutations is provided in Table 2.

The figure below shows the frequency of mutations in the genes with the highest abundance of mutations, namely *ATM*, *TP53* and *BRCA1*. The most frequent mutations were c.376-283T>C (*TP53*), c.3994-193T>C, c.8010+186C>T (*ATM*), and c.5215+66G>A (*BRCA1*).

Based on the bioinformatic analysis and annotation of the identified polymorphisms, we selected those mutations that could significantly affect the regulatory or protein sequences. To assess pathogenicity and conservation of the mutations, we used data from COSMIC (Catalogue of Somatic Mutations in Cancer) [11] and dbNSFP [12]. Additionally, SIFT (Sorting Intolerant From Tolerant) and PolyPhen2 tools were used to

Table 1. Clinicopathologic features of patients with breast cancer (n = 16)

Parameter	Value. abs. (%)
Age, years	50.7 ± 11.3
Surgical intervention:	
yes	16 (100)
no	0 (0)
Stage according to TNM	
T1	10 (62.5)
T2	5 (31.3)
T3	1 (6.2)
Metastases in lymph nodes	
no. M0	10 (62.5)
yes. M1	6 (37.5)
Expression of estrogen receptors (ER):	
ER+	11 (68.8)
ER-	5 (31.2)
Expression of progesterone receptors (PR):	
PR+	10 (62.5)
PR-	6 (37.5)
Expression of HER2/neu:	
Her2+	9 (56.3)
Her2-	7 (43.7)
Expression of Ki-67:	
low (< 14 %)	13 (81.3)
high (≥ 14 %)	3 (18.7)
Adjuvant chemotherapy:	
yes	10 (62.5)
no	6 (37.5)
Adjuvant hormonal therapy:	
yes	9 (56.3)
no	7 (43.7)
Radiation therapy	
yes	0 (0)
no	16 (100)

predict pathogenicity of the mutations and assess their effect on the function of the encoded protein [13, 14]. Information about mutation frequencies was obtained from the 100 Genomes project and the Exome Aggregation Consortium [15, 16].

Altogether, we singled out 14 mutations affecting the protein sequence: *BRCA2* — c.4828G>A (p.Val1610Met), c.5070A>C (p.Lys1690Asn); *TP53* — c.524G>A (p.Arg175His), c.469G>T (p.Val157Phe); *CHEK2* — c.1289C>T (p.Thr430Ile); *ATM* — c.146C>G (p.Ser49Cys), c.4258C>T (p.Leu1420Phe), c.1192G>C (p.Asp398His); *CDH1* — c.790C>T (p.Gln264), c.1342C>T (p.Gln448); *BRCA1* — c.1865C>T (p.Ala622Val), c.384G>A (p.Met128Ile), and c.54G>T (p.Met18Ile).

DISCUSSION

In Russia, the PCR-based methods for the detection of known mutations in BC-associated genes have become most widespread. However, today there are more advanced methods of genetic screening, the most promising being next generation sequencing that can be used for identifying genetic variants in malignant tumors and is especially suitable in exploring the

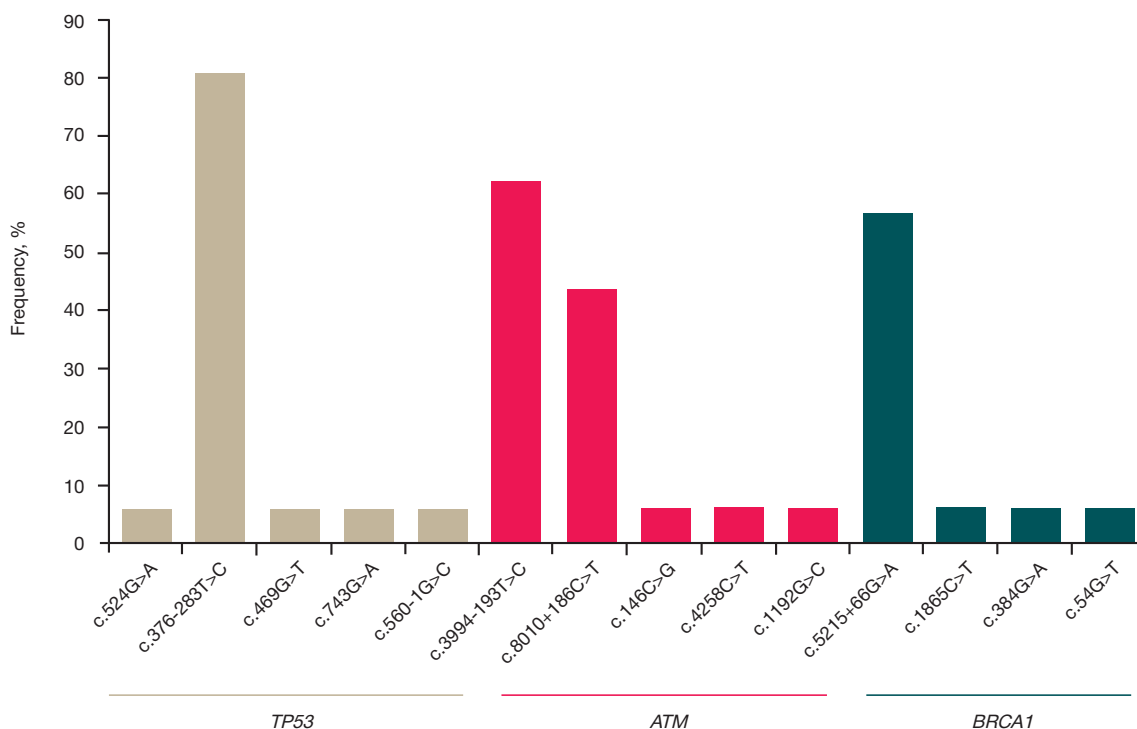
variability of highly heterogeneous regions of tumor genomes. In this work we applied NGS to study a number of mutations of key oncogenes associated with BC and tested a previously developed algorithm for bioinformatic analysis of sequencing data.

One of the most well-studied genes playing a significant role in BC pathogenesis is *TP53*. It is involved in the regulation of the cell cycle, apoptotic activity and DNA repair. Mutations in *TP53* lead to the disruption of these regulatory mechanisms and may trigger formation of cancer. *TP53* is a tumor suppressor; mutant variants of this gene are detected in half of all cancers and in more than 30 % of BC cases. In turn, sporadic breast cancer is characterized by a varying frequency of *TP53* mutations between 25 % and 86 %, depending on the disease stage and the screening technique applied. The prognostic value of *TP53* mutations in BC has been sufficiently studied [17]. Among the mutations identified in our study the most frequent was c.376-283T>C discovered in 13 of 16 patients (81 %).

Patients with BC and with some of its types in particular have relatively high frequency of *BRCA1* and *BRCA2* mutations. *BRCA1* and *BRCA2* are involved in the regulation of many cell processes maintaining genomic stability and homologous

Table 2. Single nucleotide variants (mutations) identified in patients with breast cancer (n = 16)

Sample	Abundance of mutations in the sample, %	Mutations	Gene	Coordinates
1	18.7	c.4828G>A	<i>BRCA2</i>	Chr13:32913320
		c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.-73G>A	<i>CHEK2</i>	Chr22:29137870
2	25.5	c.5215+66G>A	<i>BRCA1</i>	Chr17:41215825
		c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.8010+186C>T	<i>ATM</i>	Chr11:108204881
		c.524G>A	<i>TP53</i>	Chr17:7578406
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
3	29.4	c.8755-272A>G	<i>BRCA2</i>	Chr13:32953182
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
4	26.5	c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.8010+186C>T	<i>ATM</i>	Chr11:108204881
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
5	26.5	c.5215+66G>A	<i>BRCA1</i>	Chr17:41215825
		c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.8010+186C>T	<i>ATM</i>	Chr11:108204881
		c.1289C>T	<i>CHEK2</i>	Chr22:29091797
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
6	20.6	c.5215+66G>A	<i>BRCA1</i>	Chr17:41215825
		c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.8010+186C>T	<i>ATM</i>	Chr11:108204881
7	25.5	c.146C>G	<i>ATM</i>	Chr11:108098576
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
8	25.5	c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.790C>T	<i>CDH1</i>	Chr16:68844202
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
9	28.5	c.5070A>C	<i>BRCA2</i>	Chr13:32913562
		c.-73G>A	<i>CHEK2</i>	Chr22:29137870
		c.469G>T	<i>TP53</i>	Chr17:7578461
10	37.3	c.5215+66G>A	<i>BRCA1</i>	Chr17:41215825
		c.4258C>T	<i>ATM</i>	Chr11:108160350
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
11	26.5	c.5215+66G>A	<i>BRCA1</i>	Chr17:41215825
		c.1865C>T	<i>BRCA1</i>	Chr17:41245683
		c.384G>A	<i>BRCA1</i>	Chr17:41256196
		c.54G>T	<i>BRCA1</i>	Chr17:41276060
		c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.-73G>A	<i>CHEK2</i>	Chr22:29137870
		c.743G>A	<i>TP53</i>	Chr17:7577538
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
12	28.5	c.5215+66G>A	<i>BRCA1</i>	Chr17:41215825
		c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.8010+186C>T	<i>ATM</i>	Chr11:108204881
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
13	22.6	c.5215+66G>A	<i>BRCA1</i>	Chr17:41215825
		c.1192G>C	<i>ATM</i>	Chr11:108119786
		c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.8010+186C>T	<i>ATM</i>	Chr11:108204881
		c.1342C>T	<i>CDH1</i>	Chr16:68849439
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
14	22.6	c.5215+66G>A	<i>BRCA1</i>	Chr17:41215825
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
15	26.5	c.5215+66G>A	<i>BRCA1</i>	Chr17:41215825
		c.560-1G>C	<i>TP53</i>	Chr17:7578290
		c.376-283T>C	<i>TP53</i>	Chr17:7578837
16	28.5	c.3994-193T>C	<i>ATM</i>	Chr11:108158134
		c.8010+186C>T	<i>ATM</i>	Chr11:108204881
		c.376-283T>C	<i>TP53</i>	Chr17:7578837



Distribution of mutation frequencies in the genes *ATM*, *TP53* and *BRCA1* in patients with breast cancer

recombination during repair of double-strand DNA breaks. Mutations occurring in these genes often disrupt their normal function and are a major causative factor of hereditary BC, increasing the risk of cancer in an individual. About a quarter of all hereditary BC cases are associated with mutations in *BRCA1/2* [17].

Mutations in *BRCA1* account for 80 % of all *BRCA1* and *BRCA2* mutations in Russians with BC. One of the most common mutant variants identified in Russian patients is 5382insC (*rs80357906*) that causes a reading frame shift and the loss of function of the encoded protein. The majority of the polymorphisms identified in our study were mutations in *BRCA1* and *BRCA2*, the most common being c.5215+66G>A (*rs3092994*) in *BRCA1*, detected in 9 of 16 patients (52.9 %).

Our findings on *ATM*, *TP53* and *BRCA1* mutations are on the whole consistent with the literature, which reports *TP53* variants to be the most common mutations in BC [17]. Our results of the diversity of *BRCA1/2* variants are also comparable with the literature data. Importantly, mutations in these genes are associated with poor prognosis and development of invasive ductal breast cancer. The existences of these mutations are considered at assessment of volume of surgical intervention [17]. In our study, of 12 patients with BC who had mutations in *BRCA1* and *BRCA2*, 8 were diagnosed with invasive ductal carcinoma. Of those 8, six had the mutation c.5215+66G>A in *BRCA1*.

We have analyzed next generation sequencing data using the original bioinformatic approach and discovered many

driver mutations in the samples of malignant breast tumors. Using different databases, we have selected and annotated functionally significant mutations. Altogether, we have discovered 14 mutations affecting the amino acid sequence of the encoded proteins. Each of the studied samples had at least one such mutation. The original bioinformatic protocol allowed us to automatically process DNA sequencing data obtained with NGS.

CONCLUSIONS

A combination of next generation sequencing and modern algorithms for bioinformatic analysis is a good and clinically attractive method of screening for genetic polymorphisms and assessing the functional effect of mutations detected in the tumor. To date, NGS enables molecular classification of breast tumors and can be used to determine their subtypes depending on the spectrum of the identified mutations and the expression profiles of the studied genes. NGS data can facilitate the choice of adequate targeted therapies. One of the major tasks of cancer genetics is development of convenient tools for the detection of breast cancer biomarkers that can be used by clinicians for more accurate diagnosis and effective treatment. We believe that advances in the field should include improvement of bioinformatic approaches, adoption of the systems for automatic analysis of tumor genetic profiles and introduction of NGS into clinical routine.

References

1. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast Cancer Statistics, 2017, Racial Disparity in Mortality by State. *CA Cancer J Clin.* 2017 Nov; 67 (6): 439–48. DOI: 10.3322/caac.21412.
2. Yohe S, Thyagarajan B. Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med.* 2017 Nov; 141 (11): 1544–57. DOI: 10.5858/arpa.2016-0501-RA.
3. Melchor L, Benítez J. The complex genetic landscape of familial

- breast cancer. *Hum Genet.* 2013 Aug; 132 (8): 845–63. DOI: 10.1007/s00439-013-1299-y.
4. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ et al. Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin.* 2017 Jul 8; 67 (4): 290–303. DOI: 10.3322/caac.21393.
 5. Tsukanov KYu, Krasnenko AYu, Plakhina DA, Korostin DO, Churov AV, Druzhilovskaya OS et al. A bioinformatic pipeline for NGS data analysis and mutation calling in human solid tumors. *Biomed Khim.* 2017; 63 (5): 413–7.
 6. Martin M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.journal.* 2011; 17: 10–2.
 7. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009 Jul 15; 25 (14): 1754–60. DOI: 10.1093/bioinformatics/btp324.
 8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009 Aug 15; 25 (16): 2078–9. DOI: 10.1093/bioinformatics/btp352.
 9. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013 Mar; 31 (3) :213–9. DOI: 10.1038/nbt.2514.
 10. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012 Apr-Jun; 6 (2): 80–92. DOI: 10.4161/fly.19695.
 11. Forbes SA, Beare D, Bindal N, Bamford S, Ward S, Cole CG et al. COSMIC: High-Resolution Cancer Genetics Using the Catalogue of Somatic Mutations in Cancer. *Curr Protoc Hum Genet.* 2016 Oct 11; 91: 10.11.1-10.11.37. DOI: 10.1002/cphg.21.
 12. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat.* 2016 Mar; 37 (3): 235–41. DOI: 10.1002/humu.22932.
 13. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003 Jul 1; 31 (13): 3812–4.
 14. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013 Jan; Chapter 7: Unit7.20. DOI: 10.1002/0471142905.hg0720s76.
 15. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO et al. A global reference for human genetic variation. *Nature.* 2015 Oct 1; 526 (7571): 68–74. DOI: 10.1038/nature15393.
 16. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016 Aug 18; 536 (7616): 285–91. DOI: 10.1038/nature19057.
 17. Grishina KA, Muzaffarova TA, Khaylenko VA, Karpukhin AV. [Molecular genetic markers of breast cancer]. Tumors of female reproductive system. 2016; 12 (3): 36–42. DOI: 10.17650/1994-4098-2016-12-3-26-42. Russian.

Литература

1. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast Cancer Statistics, 2017, Racial Disparity in Mortality by State. *CA Cancer J Clin.* 2017 Nov; 67 (6): 439–48. DOI: 10.3322/caac.21412.
2. Yohe S, Thyagarajan B. Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med.* 2017 Nov; 141 (11): 1544–57. DOI: 10.5858/arpa.2016-0501-RA.
3. Melchor L, Benítez J. The complex genetic landscape of familial breast cancer. *Hum Genet.* 2013 Aug; 132 (8): 845–63. DOI: 10.1007/s00439-013-1299-y.
4. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ et al. Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin.* 2017 Jul 8; 67 (4): 290–303. DOI: 10.3322/caac.21393.
5. Цуканов К. Ю., Красненко А. Ю., Плахина Д. А., Коростин Д. О., Чуров А. В., Дружилловская О. С. и др. Биоинформатический протокол для обработки NGS-данных и идентификации мутаций в солидных опухолях человека. *Биомедицинская химия.* 2017; 63 (5): 413–7. DOI: 10.18097/PBMC20176305413.
6. Martin M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet.journal.* 2011; 17: 10–2.
7. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009 Jul 15; 25 (14): 1754–60. DOI: 10.1093/bioinformatics/btp324.
8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009 Aug 15; 25 (16): 2078–9. DOI: 10.1093/bioinformatics/btp352.
9. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol.* 2013 Mar; 31 (3) :213–9. DOI: 10.1038/nbt.2514.
10. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012 Apr-Jun; 6 (2): 80–92. DOI: 10.4161/fly.19695.
11. Forbes SA, Beare D, Bindal N, Bamford S, Ward S, Cole CG et al. COSMIC: High-Resolution Cancer Genetics Using the Catalogue of Somatic Mutations in Cancer. *Curr Protoc Hum Genet.* 2016 Oct 11; 91: 10.11.1-10.11.37. DOI: 10.1002/cphg.21.
12. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat.* 2016 Mar; 37 (3): 235–41. DOI: 10.1002/humu.22932.
13. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003 Jul 1; 31 (13): 3812–4.
14. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013 Jan; Chapter 7: Unit7.20. DOI: 10.1002/0471142905.hg0720s76.
15. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO et al. A global reference for human genetic variation. *Nature.* 2015 Oct 1; 526 (7571): 68–74. DOI: 10.1038/nature15393.
16. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016 Aug 18; 536 (7616): 285–91. DOI: 10.1038/nature19057.
17. Гришина К. А., Музаффарова Т. А., Хайленко В. А., Карпукhin А. В. Молекулярно-генетические маркеры молочной железы. Опухоли женской репродуктивной системы. 2016; 12 (3): 36–42. DOI: 10.17650/1994-4098-2016-12-3-26-42.