

# Bioinformatics HW5

Ershov Ivan

May 2022

Я выбрал ген PTCH1

## Задание 1.

1.a) Найдите выбранный ген в базе данных IntOGen. В каких типах рака он мутировал?

Согласно сайту [IntOGen](#) выбранный ген мутировал в 6 типах:

1. [Skin basal cell carcinoma](#)
2. [Medulloblastoma](#)
3. [Pilocityc astrocytoma](#)
4. [Esophageal cancer](#)
5. [Osteosarcoma](#)
6. [Mesothelioma](#)

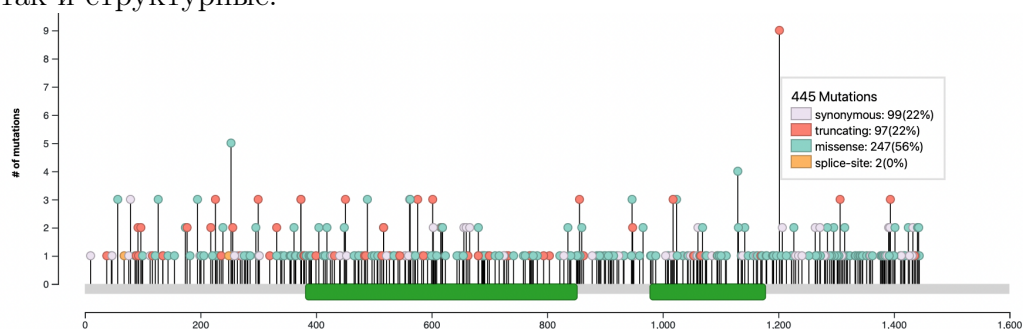
Cancer type	Methods	Samples	Samples (%)
<a href="#">Skin basal cell carcinoma</a>	CBaSE dNdScv MutPanning FML	53	55.21
<a href="#">Medulloblastoma</a>	CBaSE combination dNdScv MutPanning FML	37	5.82
<a href="#">Pilocityc astrocytoma</a>	CBaSE dNdScv MutPanning FML	21	3.38
<a href="#">Esophageal cancer</a>	MutPanning	4	0.42
<a href="#">Osteosarcoma</a>	dNdScv MutPanning	4	1.1
<a href="#">Mesothelioma</a>	MutPanning	2	1.71

Showing 1 to 6 of 6 entries

Previous 1 Next

## 1.6) Какие типы мутаций – точечные или структурные (вставки, делеции, транслокации)?

Всего найдено 445 мутаций, среди которых встречаются как точечные, так и структурные:



У данного гена есть как и вставки:

Mutation (GRCh38)	Protein Position	Samples	Consequence
9:95447312:G>A	1315	2	missense_variant
9:95447370:C>T	1296	2	missense_variant
9:95447401:C>A	1285	2	missense_variant
9:95447440:G>A	1272	2	synonymous_variant
9:95449081:G>A	1264	2	synonymous_variant
9:95449193:G>A	1227	2	missense_variant
9:95449252:G>A	1207	2	synonymous_variant
9:95453538:G>A	1130	2	missense_variant
9:95453539:C>T	1130	2	missense_variant
9:95456377:C>T	1069	2	missense_variant

Showing 21 to 30 of 402 entries

Previous 1 2 3 4 5

так и делеции:

Mutation (GRCh38)	Protein Position	Samples	Consequence
9:95449267:G>-	1202	9	frameshift_variant
9:95447335:G>-	1307	3	frameshift_variant
9:95477699:C>-	451	3	frameshift_variant
9:95480435:G>-	300	3	frameshift_variant
9:95482018:C>-	226	3	frameshift_variant
9:95476812:G>-	517	2	frameshift_variant
9:95506511:T>-	97	2	frameshift_variant
9:95449881:A>-	1170	1	frameshift_variant
9:95458019:C>-	1054	1	frameshift_variant
9:95458123:G>-	1020	1	frameshift_variant

Showing 1 to 10 of 402 entries

Previous 1 2 3 4 5 ... 41 Next

**2) Произведите поиск выбранного гена на портале ICGC: В каких типах рака выбранный ген мутирует?**

Смотрим на сайт [ICGC](#). На странице этого гена, есть две таблицы, в одной (таблица 1) отображены исследования различных типов рака с мутациями в данном гене, в другой (таблица 2) - представлена подробная информация о мутациях. В первой таблице в колонке Tumour Type как раз и написан тип рака:

Project	Site	Tumour Type	Tumour Subtype
<a href="#">LMS-FR</a>	Mesenchymal	Soft Tissue cancer	Leiomyosarcoma
<a href="#">BTCA-SG</a>	Gall Bladder	Biliary Tract cancer	Gall bladder cancer / Cholangiocarcinoma
<a href="#">MELA-AU</a>	Skin	Skin cancer	Melanoma
<a href="#">SKCA-BR</a>	Skin	Melanoma	
<a href="#">ESAD-UK</a>	Esophagus	Esophageal cancer	Esophageal adenocarcinoma
<a href="#">PEME-CA</a>	Brain	Brain cancer	Pediatric Medulloblastoma
<a href="#">LICA-CN</a>	Liver	Liver cancer	Hepatocellular carcinoma HBV-associated
<a href="#">LUSC-KR</a>	Lung	Lung cancer	Adenocarcinoma, Squamous cell carcinoma
<a href="#">BRCA-FR</a>	Breast	Breast cancer	Subtype defined by an amplification of the HER2 gene
<a href="#">OV-AU</a>	Ovary	Ovarian cancer	Serous cystadenocarcinoma
<a href="#">UTCA-FR</a>	Uterus	Uterine cancer	Carcinosarcoma
<a href="#">MALY-DE</a>	Blood	Malignant Lymphoma	Germinal center B-cell derived lymphomas
<a href="#">LIRI-JP</a>	Liver	Liver cancer	Hepatocellular carcinoma (Virus associated)
<a href="#">GACA-CN</a>	Stomach	Gastric cancer	Intestinal- and diffuse-type
<a href="#">THCA-CN</a>	Head and neck	Thyroid cancer	Papillary carcinoma
<a href="#">UCEC-US</a>	Uterus	Endometrial cancer	Uterine corpus endometrial carcinoma
<a href="#">PAEN-IT</a>	Pancreas	Rare Pancreatic Tumors	Enteropancreatic endocrine tumors and rare pancreatic exocrine tumors
<a href="#">BRCA-EU</a>	Breast	Breast cancer	ER+ve, HER2-ve
<a href="#">PAEN-AU</a>	Pancreas	Pancreatic cancer	Endocrine neoplasms
<a href="#">COCA-CN</a>	Colorectal	Colorectal cancer	Adenocarcinoma, non-Western
<a href="#">PACA-CA</a>	Pancreas	Pancreatic cancer	Ductal adenocarcinoma
<a href="#">PBCA-US</a>	Brain	Pediatric Brain Tumor	Multiple subtypes
<a href="#">BRCA-UK</a>	Breast	Breast cancer	Triple Negative/lobular/other
<a href="#">PBCA-DE</a>	Brain	Pediatric Brain Tumors	Medulloblastoma & Pediatric Pilocytic Astrocytoma
<a href="#">GACA-JP</a>	Stomach	Gastric cancer	Multiple histological subtypes (Adenocarcinoma and specific subtypes by WHO classification)

ID	DNA change	Type	Consequences
MU148935	chr9:g.98209594G>A	single base substitution	Missense: <i>PTCH1</i> P1164L, P1249L, P1314L, P1315L 3 UTR: <i>PTCH1</i> Exon: <i>PTCH1</i>
MU65435196	chr9:g.98206399T>A	single base substitution	3 UTR: <i>PTCH1</i> Exon: <i>PTCH1</i> Downstream: <i>PTCH1</i>
MU128949309	chr9:g.98278829T>-	deletion of <=200bp	5 UTR: <i>PTCH1</i> Exon: <i>PTCH1</i> Downstream: <i>RP11-435O5.4</i> Intron: <i>PTCH1</i>
MU113099	chr9:g.98211549G>-	deletion of <=200bp	Frameshift: <i>PTCH1</i> P1201, P1136, P1051, P1202 3 UTR: <i>PTCH1</i> Exon: <i>PTCH1</i>
MU21455623	chr9:g.98221861T>C	single base substitution	Intron: <i>PTCH1</i>
MU1902538	chr9:g.98215774G>A	single base substitution	Upstream: <i>PTCH1</i> Synonymous: <i>PTCH1</i> F1079F, F994F, F1144F, F1145F 3 UTR: <i>PTCH1</i> Downstream: <i>PTCH1</i>
MU81000952	chr9:g.98278975C>T	single base substitution	Missense: <i>PTCH1</i> G43E 5 UTR: <i>PTCH1</i> Upstream: <i>PTCH1</i> Downstream: <i>RP11-435O5.4</i>
MU89924096	chr9:g.98276929->GTGAGA	insertion of <=200bp	Downstream: <i>RP11-435O5.4</i> Intron: <i>PTCH1</i>
MU2186559	chr9:g.98268792->T	insertion of <=200bp	Frameshift: <i>PTCH1</i> N96N?, N31N?, N97N? 5 UTR: <i>PTCH1</i> Upstream: <i>PTCH1</i> Exon: <i>PTCH1</i> - <i>RP11-435O5.5</i> Intron: <i>PTCH1</i>
MU79305842	chr9:g.98228254T>-	deletion of <=200bp	Downstream: <i>PTCH1</i> Intron: <i>PTCH1</i>

## Задание 2.

1) Выберем донора:

1. зайдём на сайт [ICGC](#)
2. выберем любой проект, например [OV-AU](#)
3. перейдем на View in Data Repositories
4. выберем любого донора, например [DO46588](#)

Проверим, действительно ли у этого человека мутация в нужном нам гене:

1. Скачаем с [DO46588](#) "Simple Somatic Mutations" [ICGC](#)
2. поищем в скаченном файле **Ensembl ID**, соответствующий гену (то есть ENSG00000185920, его я нашел [тут](#))

```
Downloads cat simple_somatic_mutation.open.tsv | grep "ENSG00000185920" | head -1
MU6385296 DO46588 OV-AU SP102123 SA505089 SA505820 AOC5-165-1-X AOC5-165-5-4 9 98269274 98269274 n
1 GRCh37 single base substitution G G T 2.544088276857012E-4 29 8 not tested http://www.illumina.com/technol
ot tested intron variant ENSG00000185920 ENST00000437951 75 Illumina HiSeq Paired End http://www.illumina.com/technol
ogy/paired_end_sequencing_assay.ilmn WGS RTA http://support.illumina.com/sequencing/sequencing_software/real-time_analysis_rta.ilmn bwa
http://bio-bwa.sourceforge.net/ qSNP http://qcmg.org/informatics.php GATK http://www.broadinstitute.org/gatk/ EGA EGAS
00001000397
```

(тут очень много разной информации, но главное, что ENSG00000185920 нашелся, а значит, мутация была)

2) Теперь найдем еще 4 гена, с тем же типом рака, что и у нашего донора (то есть с Ovarian cancer):

```
+ Downloads cat NCG_cancerdrivers_annotation_supporting_evidence.tsv | grep "ovarian_cancer" | head -4
672 BRCA1 21720365 WES Gynecologic ovary ovarian_cancer "MutSig, MuSiC" coding
673 BRAF 20668451 "Gene Panel" Gynecologic ovary ovarian_cancer Wood-based
675 BRCA2 21720365 WES Gynecologic ovary ovarian_cancer "MutSig, MuSiC" coding
1499 CTNNB1 20668451 "Gene Panel" Gynecologic ovary ovarian_cancer Wood-based
```

Найдем эти гены на сайте [ICGC](#):

1. [BRCA1](#)
2. [BRAF](#)
3. [BRCA2](#)
4. [CTNNB1](#)

Теперь поищем мутации этих генов в выбранном доноре:

```
+ Downloads cat simple_somatic_mutation.open.tsv | grep "ENSG0000012048"
+ Downloads cat simple_somatic_mutation.open.tsv | grep "ENSG00000139618"
+ Downloads cat simple_somatic_mutation.open.tsv | grep "ENSG00000168036"
+ Downloads cat simple_somatic_mutation.open.tsv | grep "ENSG00000157764"
MU9065383 D046588 OV-AU SP102123 SA505809 SA505820 AOCs-165-1-X AOCs-165-5-4
7 140525067 140525067 1 GRCh37 single base substitution C C T
44 10 not tested intron_variant ENSG00000157764
ENST00000288602 75 Illumina HiSeq WGS PCAWG Consensus SNV-MNV caller
FI25537:FI25536
```

У данного донора рак яичников (Ovarian cancer). Поэтому стало интересно, обязательно ли должны произойти мутации в других генах, связанных с этим типом рака. Однако, как выяснилось, необязательно: среди выбранных 4 генов, ассоциируемых с раком яичников, только один мутировал.