**Задание по практике по биоинформатике**  
студента 4 курса специальности «Биоинженерия и биоинформатика»

«Исследование транскрипционного фактора, связанного с заболеванием семейная гиперхолестеринемия»

Гориченко Ивана Вадимовича

07.12.2021г.

Факторы транскрипции — белки, контролирующие процесс синтеза мРНК, а также других видов РНК на матрице ДНК путём связывания со специфичными участками ДНК. Транскрипционные факторы выполняют свою функцию либо самостоятельно, либо в комплексе с другими белками.

**Структура гена.**

Поиск в GenBank (https://www.ncbi.nlm.nih.gov/genbank/) дал 19 результатов, из них 9 для человека по болезни Gaucher’s disease.

Рассмотрим структуру белков, кодируемых генами из списка для семейной гиперхолестеринемии (СГХС), в таблице показаны гены, чей счет не менее 37.

Таблица. Гены СХГС из GeneCards

|  |  |  |  |
| --- | --- | --- | --- |
| Ген | Краткое описание | Категория | Счет |
| LDLR | Low Density Lipoprotein Receptor | Protein Coding | 235.00 |
| APOB | Apolipoprotein B | Protein Coding | 216.31 |
| PCSK9 | Proprotein Convertase Subtilisin/Kexin Type 9 | Protein Coding | 149.20 |
| BRCA2 | BRCA2 DNA Repair Associated | Protein Coding | 109.99 |
| LDLRAP1 | Low Density Lipoprotein Receptor Adaptor Protein 1 | Protein Coding | 94.77 |
| LDLR-AS1 | LDLR Antisense RNA 1 | RNA Gene | 70.18 |
| LOC106560211 | APOB 5' Regulatory Region | Biological Region | 56.96 |
| FLNC | Filamin C | Protein Coding | 53.78 |
| APOE | Apolipoprotein E | Protein Coding | 50.09 |
| MIR6886 | MicroRNA 6886 | RNA Gene | 50.03 |
| APOA2 | Apolipoprotein A2 | Protein Coding | 47.78 |
| SMARCA4 | SWI/SNF Related, Matrix Associated, Actin Dependent Regulator Of Chromatin, Subfamily A, Member 4 | Protein Coding | 41.43 |
| ABCA1 | ATP Binding Cassette Subfamily A Member 1 | Protein Coding | 40.82 |
| MYH7 | Myosin Heavy Chain 7 | Protein Coding | 39.13 |
| GHR | Growth Hormone Receptor | Protein Coding | 39.09 |
| LMNA | Lamin A/C | Protein Coding | 39.03 |
| LPL | Lipoprotein Lipase | Protein Coding | 38.92 |
| RELN | Reelin | Protein Coding | 38.19 |
| TNNT2 | Troponin T2, Cardiac Type | Protein Coding | 37.77 |
| MYLK | Myosin Light Chain Kinase | Protein Coding | 37.45 |

Для дальнейшего анализа был взят ген SREBF1 (Sterol Regulatory Element Binding Transcription Factor 1), у которого счет составляет 16,54.

## Из интернет-ресурса GenBank было взято графическое представление гена SREBF1 (https://www.ncbi.nlm.nih.gov/gene/6720).

## Структура белка.

С базы данных UniProt была получена информация о кодируемом геном SREBF1 белке.

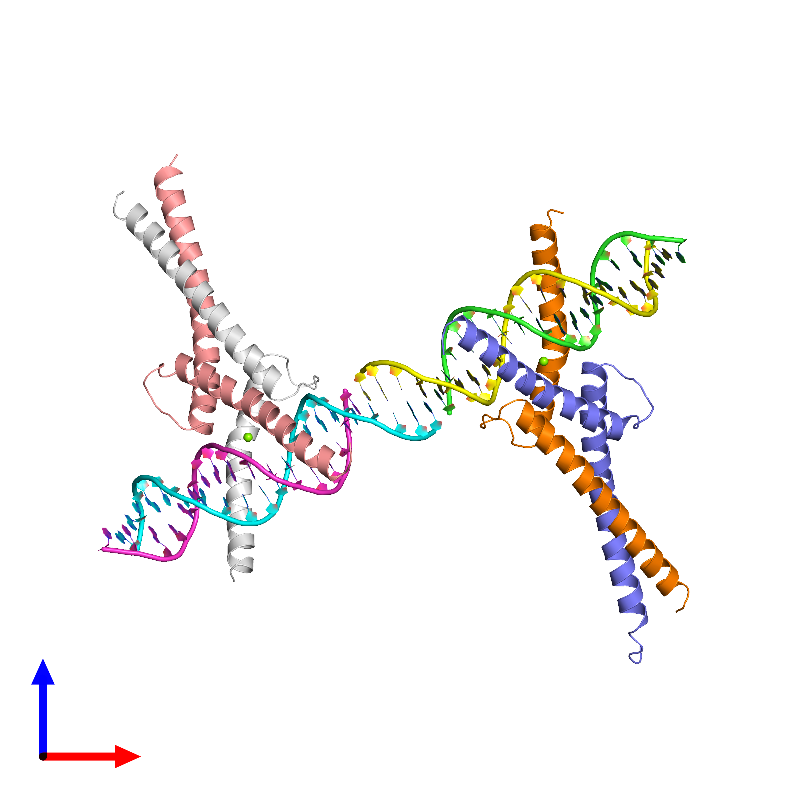
Название белка: Sterol regulatory element-binding protein 1.

Идентификационный номер белка: P36956

Функция белка: Предшественник формы фактора транскрипции, который встроен в мембрану эндоплазматического ретикулума. Низкие концентрации стерола способствуют переработке этой формы, высвобождая форму фактора транскрипции, которая транслоцируется в ядро и активирует транскрипцию генов, участвующих в биосинтезе холестерина и гомеостазе липидов.

С сайта базы данных «PDB» была взято изображение пространственной структуры белка (https://www.ebi.ac.uk/pdbe/entry/pdb/1am9).

Структура белка SREBF1



Структура белка представлена альфа-спиралями. Видна двойная спираль ДНК в центре рисунка.

**Тип ДНК-связывающего домена ТФ.**

Классификация белка SREBF1 по данным TRANSFAC (<http://factor.genexplain.com/cgi-bin/transfac_factor/getTF.cgi?AC=T08912>)

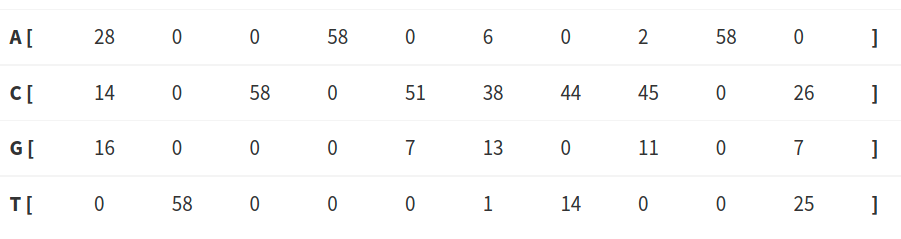
Белок SREBF1 относится к классу - Class D basic helix-loop-helix protein 1 (bHLHd1).

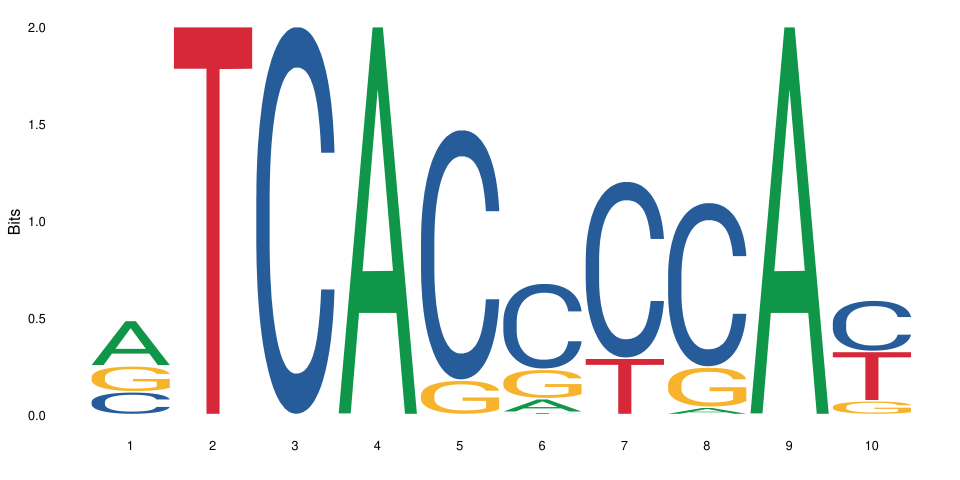
**Мотив связывания.**

Для анализа мотива связывания белка SREBF1 был использован интернет-ресурс JASPAR (<https://jaspar.genereg.net/matrix/MA0595.1/?revcomp=0>).

Для нашего белка матрица мотива представлена на рисунке ниже.

Рисунок. Матрица мотива связывания SREBF1.

Рисунок. Логотип мотива связывания ТФ.

**Последовательности ДНК с координатами на хромосомах, их положение в геноме.**

Для визуализации положения гена SREBF1 была использована база данных «Gene» (<https://www.ncbi.nlm.nih.gov/gene/6720>).

Локализация гена: 17p11.2.

Число экзонов: 23

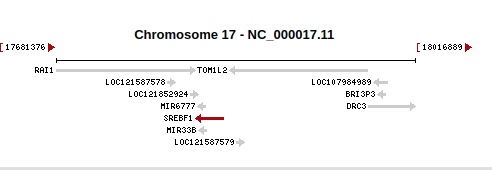
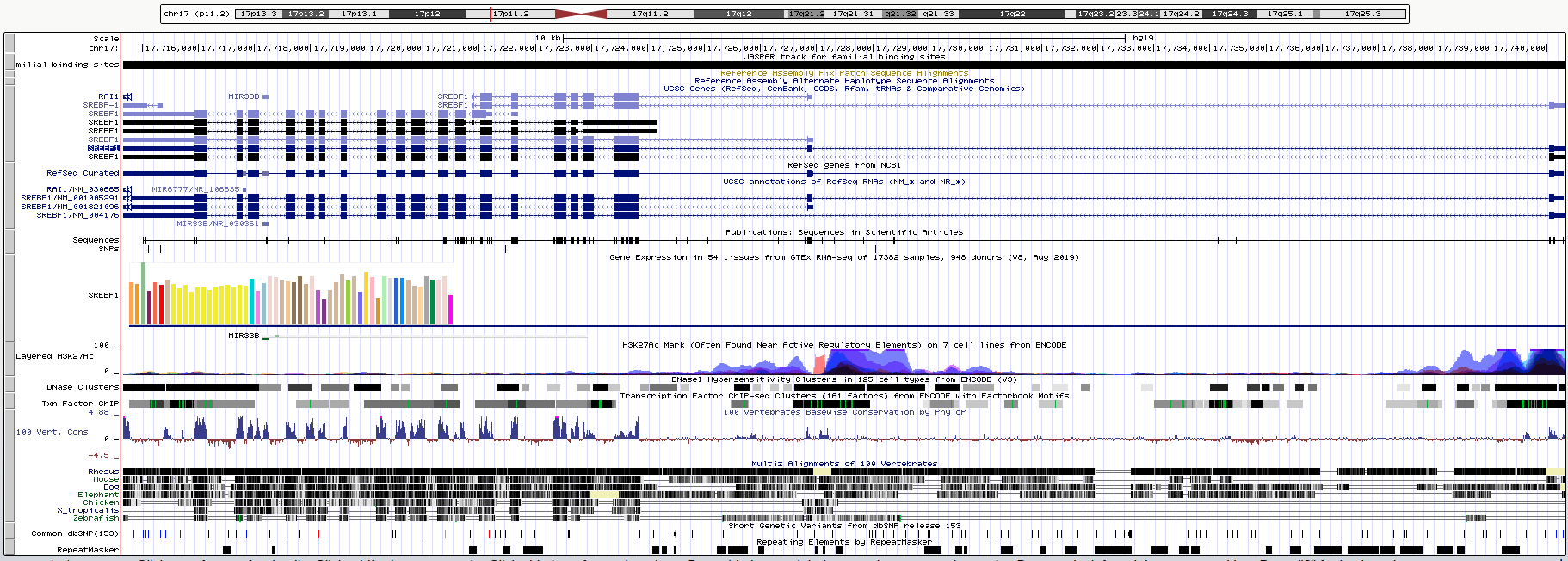
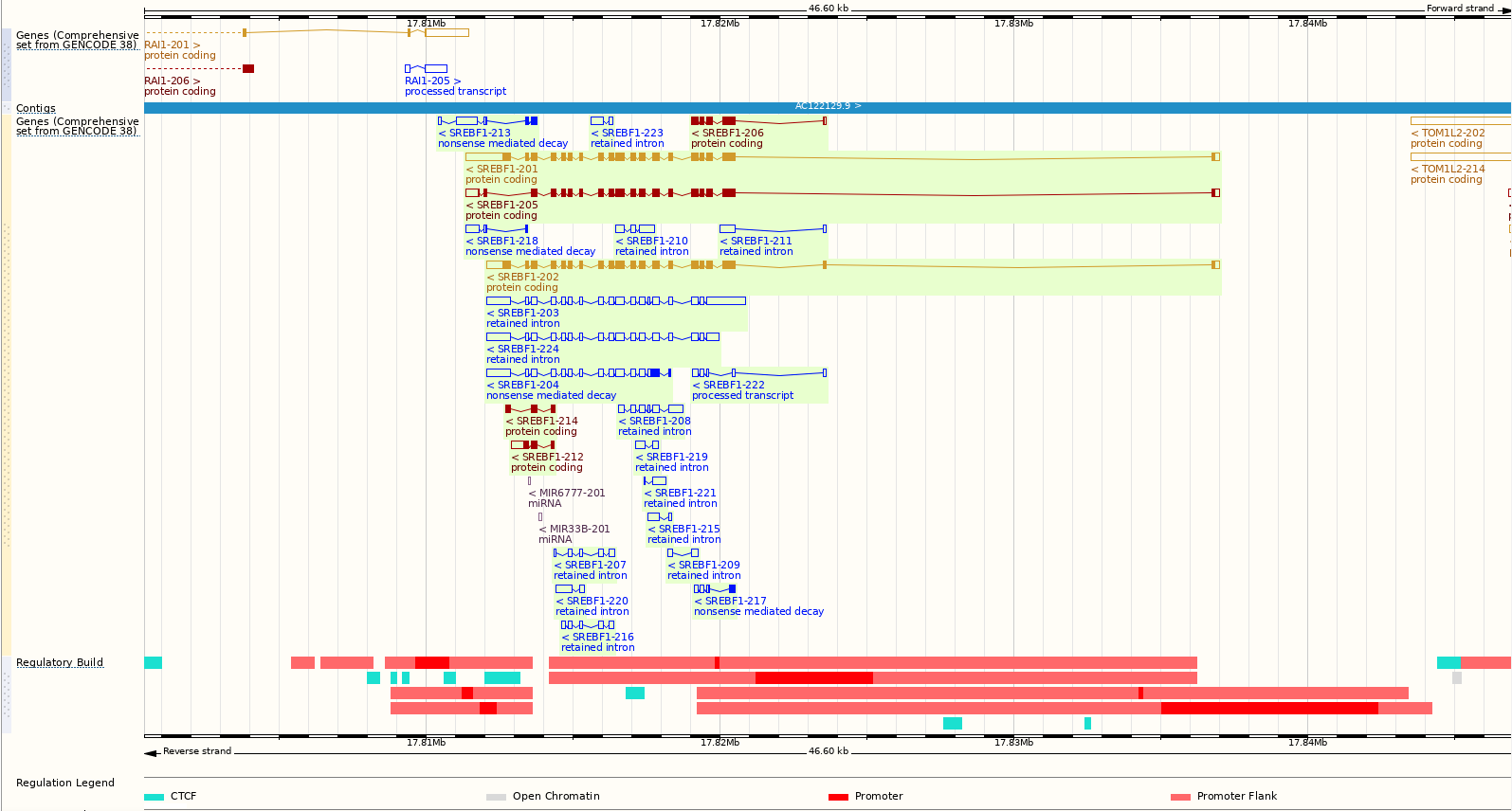
А также был использован геномный браузер **«UCSC»** (https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=chr17%3A17714663%2D17740325&hgsid=1231810057\_FmrcOuT3E7kDfm3De3Ke4fjvdAUN) и **«Еnsembl»** (http://www.ensembl.org/Homo\_sapiens/Gene/Summary?db=core;g=ENSG00000072310;r=17:17810399-17837002).

Рисунок. Расположение гена SREBF1 («UCSC»).

Рисунок. Расположение гена SREBF1 («Еnsembl»).

**Филогения гена SREBF1.**

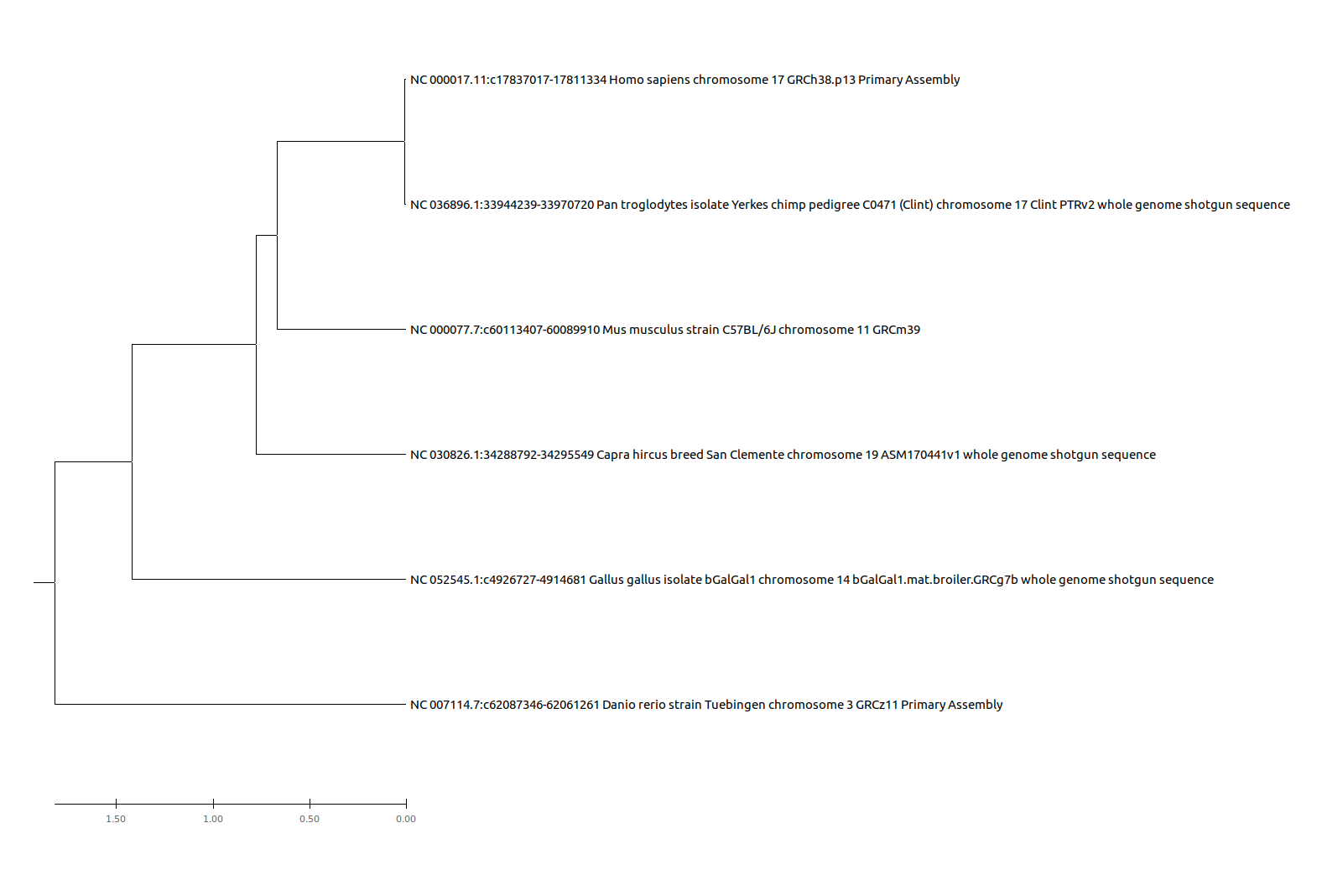
Для построения филогенетического дерева были взяты варианты гена SREBF1 от разных организмов: человека (Homo sapiens), мыши (Mus musculus), овцы (Capra hircus), шимпанзе (Pan troglodytes), Данио-рерио (Danio rerio), курицы (Gallus gallus).

Последовательности различных вариантов гена были взяты из источников:

* <https://www.ncbi.nlm.nih.gov/nuccore> — человек.
* <https://www.ncbi.nlm.nih.gov/nuccore/NC_000077.7?report=fasta&from=60089910&to=60113407&strand=true> — мышь.
* <https://www.ncbi.nlm.nih.gov/nuccore/NC_052545.1?report=fasta&from=4914681&to=4926727&strand=true> — курица.
* <https://www.ncbi.nlm.nih.gov/nuccore/NC_030826.1?report=fasta&from=34288792&to=34295549> — овца.
* <https://www.ncbi.nlm.nih.gov/nuccore/NC_036896.1?report=fasta&from=33944239&to=33970720> - шимпанзе.
* <https://www.ncbi.nlm.nih.gov/nuccore/NC_007114.7?report=fasta&from=62061261&to=62087346&strand=true> - данио-рерио.

Все последовательности были собраны в один fasta файл и выравнены. Для выравнивания последовательностей и построения филогенетического дерева использовалась программа «MEGA» (<https://www.megasoftware.net/>). Ниже представлено получившееся дерево.

Рисунок. Филогенетическое дерево гена SREBF1.

Исходя из филогении можно сделать вывод, что родство генов имеет прямую зависимость от классификации организмов, так ген шимпанзе наиболее схож с человеческим (человек и шимпанзе относятся к одной трибе), ген мыши уже расположен дальше от гена человека, затем ген овцы (мышь, овца, шимпанзе и человек относятся к одному классу), ген курицы расположен еще дальше (курица и человек относятся к одному надклассу Четвероногих), и наиболее отдаленный ген — ген рыбы данио-рерио, классификация человека и данио-рерио сходяится только на инфратипе Челюстноротые.