

Подсистема «Полиморфизм»: Оценка возможного влияния полиморфизмов (SNP) на экспрессию генов и структуру и функцию белков.

Структура документа (оглавление).

1. Цели и задачи подсистемы «Полиморфизм».....	2
2. Использование методов и подходов биоинформатики для оценки влияния полиморфизмов на экспрессию генов, а также структуру и функцию белков: структура подсистемы «Полиморфизм» и детальное руководство по ее применению.....	2
2.1. Информационные компоненты подсистемы «Полиморфизм».....	3
Структурно-функциональная организация транскрипционных регуляторных районов генов.....	3
Описание работы с базой данных TRRD.....	3
База данных ProtAct кинетических параметров ферментативных реакций.	4
Описание работы с ProtAct.....	4
2.2. Программные компоненты подсистемы «Полиморфизм».....	11
2.2.1. Программные компоненты, позволяющие исследовать влияние полиморфизмов на функции транскрипционных регуляторных районов.....	11
Выявление транскрипционных факторов, сайты связывания которых были изменены в результате возникновения полиморфных мутаций (с помощью матриц частот олигонуклеотидов).....	11
Описание работы rSNP-tools.....	13
Оценка нуклеосомного потенциала в полиморфном районе.....	17
Описание работы RECON.....	18
2.2.2. Программные компоненты, позволяющие оценивать возможное влияние полиморфизмов (SNP) на функцию регуляторных районов генов, контролирующих трансляцию.....	18
Оценка изменений трансляционной активности мРНК в SNP-вариантах, связанных с контекстом стартового кодона трансляции и AUG кодонами в составе 5'- НТП мРНК....	18
Описание работы TG-Initiator	20
Оценка стабильности вторичной структуры в 5'-НТП в норме и SNP-вариантах, прогнозирующая изменения трансляционной активности.....	22
Описание работы Leader_hairpin.....	23
2.2.3. Программные компоненты, позволяющие оценивать возможное влияние полиморфизмов (SNP) на первичную и пространственную структуру белков и их функцию.....	27
Программные компоненты ProtStability1D и ProtStability3D: оценка возможного влияния полиморфизмов на термодинамическую стабильность белков по первичной и третичной структуре.....	27
Описание работы ProtStability1D	27
Описание работы ProtStability3D.....	30
Оценка возможного влияния полиморфизмов (SNP) на функциональную активность белков.....	32
Описание работы SNP2WebProAnalyst.....	33

Программные компоненты SNP2PROSITE и SNP2PDBSite: оценка возможного влияния полиморфизмов (SNP) на первичную и третичную структуру функциональных сайтов.....	36
Описание работы SNP2PROSITE	37
Описание работы SNP2PDBSITE	39
3. Полезные ссылки.....	41

1. Цели и задачи подсистемы «Полиморфизм».

Подсистема «Полиморфизм» предназначена для исследования влияния генетического полиморфизма на экспрессию генов и функцию белков.

Подсистема «Полиморфизм» предназначена для анализа функциональной значимости полиморфизмов, выявленных как в кодирующих участках генов (кодирующие экзоны), так и в некодирующих районах (5' и 3' фланкирующих районах, 5' нетранслируемых последовательностях, интронах). Таким образом, в зависимости от района, в котором локализован полиморфизм, оценивается его возможное влияние либо на экспрессию генов (транскрипционный и трансляционный уровень), либо на функцию кодируемого белка. Возможное влияние полиморфизма на экспрессию генов оценивается как на транскрипционном, так и на трансляционном уровне. А влияние на функцию кодируемого белка исследуется на основе оценки функциональной активности полиморфных вариантов белков, влияния мутаций на структуру функциональных сайтов (классификация SNP как мутации, ведущие к повреждениям функциональных сайтов или возникновению новых функциональных сайтов), изменения термодинамической стабильности белков в результате мутаций.

Результаты, полученные с помощью программных компонент подсистемы «Полиморфизм», будут полезны для решения следующих задач:

- Предположить механизм влияния одного из аллелей на развитие болезни, в случае, если полиморфный район выявлен в уже известном гене-кандидате одного из распространенных мультифакториальных заболеваний;
- Сформировать списки потенциальных генов-кандидатов определенных распространенных мультифакториальных заболеваний, которые будут использованы в дальнейших исследованиях по генотипированию. Формирование списков возможно на основе отбора генов, для которых известны полиморфизмы, оцененные с помощью компьютерных методов, как значимые в плане влияния на экспрессию гена либо на функцию белка.

2. Использование методов и подходов биоинформатики для оценки влияния полиморфизмов на экспрессию генов, а также структуру и функцию белков: структура подсистемы «Полиморфизм» и детальное руководство по ее применению

Подсистема «Полиморфизм» представляет собой информационно-программный комплекс, предназначенный для оценки функциональной значимости полиморфизмов за счет решения следующих задач:

- Выявление транскрипционных факторов, потенциальные сайты связывания которых различаются в нормальном и мутантном варианте гена;

- [Оценка нуклеосомного потенциала в полиморфном районе;](#)
- Оценка стабильности вторичной структуры в 5'-НТП в норме и SNP-вариантах, прогнозируя тем самым изменения трансляционной активности;
- Оценка изменения трансляционной активности мРНК в SNP-вариантах, связанные с контекстом стартового кодона трансляции и AUG кодонами в составе 5'-НТП мРНК
- Предсказание уровня экспрессии гена для полиморфного варианта 5'-НТП мРНК;
- Оценка возможного влияния полиморфизмов на термодинамическую стабильность белков;
- Оценка возможного влияния полиморфизмов (SNP) на структуру функциональных сайтов;
- Проекция SNP на эволюционно-консервативные районы генома человека.

Подсистема ПОЛИМОРФИЗМ выполняет следующие функции:

- оценка возможного влияния полиморфизмов (SNP) на экспрессию генов и функцию регуляторных районов генов, контролирующей транскрипцию;
- оценка возможного влияния полиморфизмов (SNP) на первичные и пространственные структуры белков (оценка изменения термодинамической стабильности белков в результате мутаций, оценка изменения функциональной активности мутантных белков, исследование влияния мутаций на структуру функциональных сайтов и выявление мутаций, ведущих к появлению новых функциональных сайтов).

Ниже будут рассмотрены компоненты подсистемы «Полиморфизм» и приведено детальное (пошаговое) описание использования этих компонентов.

2.1. Информационные компоненты подсистемы «Полиморфизм».

Структурно-функциональная организация транскрипционных регуляторных районов генов.

См. раздел «Структурно-функциональная организация транскрипционных регуляторных районов» в обучающем курсе для АСНИ-01, в части, посвященной подсистеме «Регуломика».

Описание работы с базой данных TRRD

См. раздел «Описание работы с TRRD» в обучающем курсе для АСНИ-01, в части, посвященной подсистеме «Регуломика».

База данных ProtAct кинетических параметров ферментативных реакций.

В составе подсистемы “Человек” содержится База Данных ProtAct, содержащая систематизированную информацию о кинетических параметрах ферментативных реакций для ферментов дикого типа и их мутантов. Ферменты - биологические катализаторы белковой природы. В трехмерной структуре молекулы фермента находится активный центр, с которым связывается субстрат (превращаемое в продукт под действием фермента соединение). Роль функциональных групп в активном центре, осуществляющих процесс катализа, выполняют боковые радикалы аминокислот (СООН-группы дикарбоновых аминокислот, NH₂-группы лизина, ОН-группы серина, треонина и другие) и кофакторы, к которым относят ионы металлов и небольшие органические молекулы-коферменты (тиаминовые, флавиновые, пантотеновые, нуклеотидные производные углеводов, спиртов, фосфаты моносахаридов, металлопорфирины и другие). В самом общем виде процесс ферментативного катализа можно условно разделить на стадии а) образования фермент-субстратного комплекса (ES), образующегося в результате диффузии субстрата (S) к ферменту (E) и стерического связывания S с активным центром фермента и б) отделения продукта реакции (P) от активного центра фермента и его диффузию в окружающую среду (схема 1).

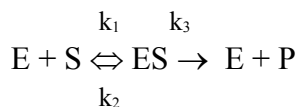


схема 1.

Здесь k_1 , и k_2 , – константы скорости прямой и обратной реакции образования фермент-субстратного комплекса ES, k_3 - константа скорости образования продукта, которая часто обозначается как каталитическая константа k_{cat} (считается, что процесс образования продукта из комплекса ES практически необратим, т.е. скорость образования ES из E и P очень мала).

Зависимость скорости ферментативной реакции (v) от концентрации S определяется уравнением Михаэлиса-Ментен (Ленинджер 1974., Строев 1986., Варфоломеев и Гуревич 1999):

$$v = V_{max}[S]/K_m + [S], \text{ где } V_{max} - \text{максимальная скорость реакции, } K_m - \text{константа Михаэлиса, } [S] - \text{концентрация субстрата.}$$
$$V_{max} = k_3 [E], \text{ где } [E] - \text{концентрация фермента}$$
$$K_m = (k_2 + k_3) / k_1$$

В научной литературе опубликовано огромное количество экспериментальных данных, характеризующих эффективность протекания ферментативных реакций. Для систематизации этой информации и предсказаний, как та или иная предполагаемая мутация повлияет на каталитические свойства определенного фермента, создана БД ProtAct. Кроме значений кинетических параметров, ProtAct содержит сведения о биохимическом или биотехнологическом значении фермента, его функции в биологическом процессе, описание его структуры, описание условий, при которых были определены кинетические параметры (температура, буфер, pH, концентрации субстратов) и библиографическую ссылку.

Описание работы с ProtAct.

Поиск информации о сравнительной активности белка и его мутантов в базе ProtActDB возможен при использовании в качестве входных данных английского названия белка, вида организма, названия процесса или описания реакции. Поиск ведется с использованием

поисковой формы приложения Protein Activity по конкретным полям базы ProtActDB, содержащим информацию данного типа. Запрос по названию белка можно реализовать по полям «Protein full name» и «Protein short name»; по виду – по полю «Organism»; по названию реакции – по полю «Reaction»; по названию процесса – по полю «Process».

Чтобы произвести поиск требуемой информации, необходимо:

Запустить приложение Protein Activity. Установить соединение с сервером базы данных. Для установления соединения с сервером базы данных необходимо сделать следующую последовательность действий:

1.левой кнопкой мыши щёлкните по пункту **File** основного меню программы (расположенного в левом верхнем углу) (Рис. 0).

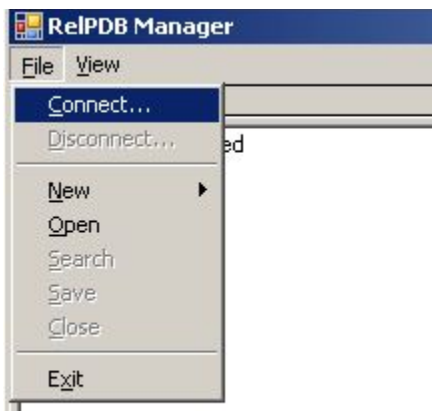


Рисунок 1. Выбор пункта меню Connect...

2. Во всплывшем подменю выберите пункт **Connect...** и щёлкните по нему левой кнопкой мыши. Должно появиться окно параметров соединения (Рис. 2).

3. В окне параметров соединения в поле **SQL Server** укажите имя компьютера, на котором установлен MS SQL сервер, с развернутой базой **ProtAct**. (Если сервер установлен на локальном компьютере, в качестве имени укажите **localhost**).



Рисунок 2. Окно параметров соединения

4. Щелчком левой кнопки мыши выберите тип аутентификации; **Windows authentication** – для аутентификации интегрированной с Windows, **SQL Server authentication** – для внутренней аутентификации SQL сервера. В случае **SQL Server authentication** введите также ваш логин в поле **Login** и пароль в поле **Password**.

Примечание: соответствующий пользователь должен быть зарегистрированным пользователем SQL сервера. Об управлении пользователями в MS SQL Server можно прочитать в соответствующей документации по данной СУБД.

5. Для того, чтобы соединиться с сервером баз данных, нажмите кнопку **Connect**. Если соединение прошло успешно на панели **DB Browser** (расположенной с левой стороны окна приложения), в закладке **PE (Protein Evolution) investigations** надпись **Not connected** должна смениться списком занесённых в базу данных исследований по эволюции белков (Рис. 3).

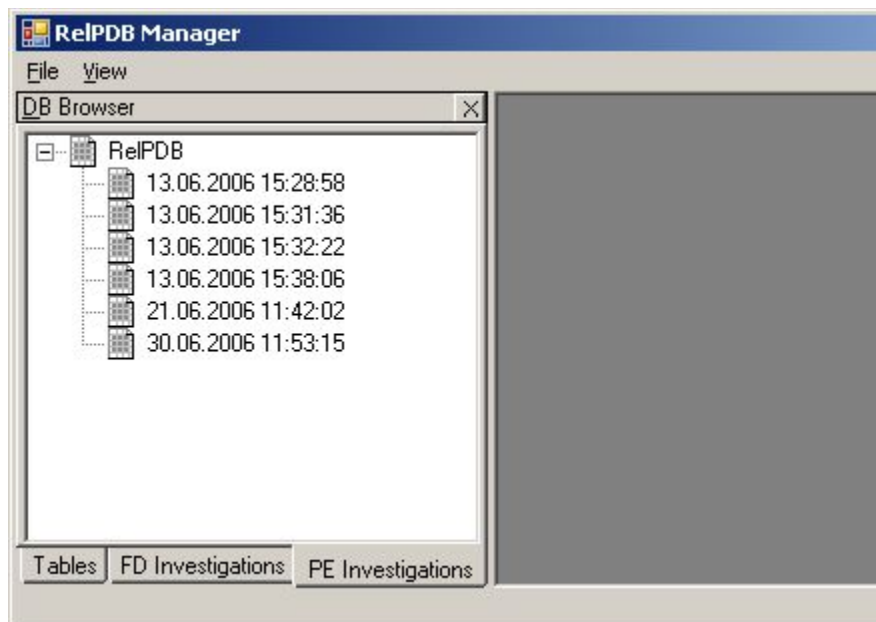


Рисунок 3. Список исследований по эволюции белков (указаны даты занесения исследования в базу).

Открыть форму поиска, для чего необходимо: левой кнопкой мыши щёлкнуть по пункту **File** основного меню программы (расположенного в левом верхнем углу) (Рис. 4);

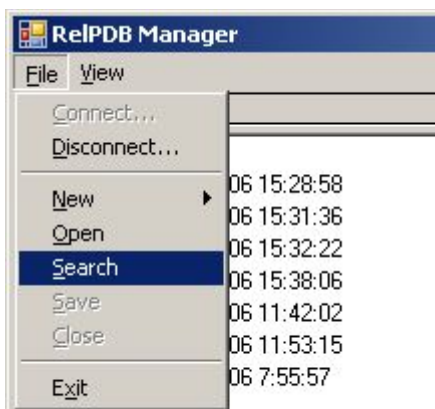


Рисунок 4. Выбор пункта меню Search.

Во всплывшем подменю выбрать пункт **Search** и щёлкнуть по нему левой кнопкой мыши. Должна появиться форма поиска (Рис. 5).

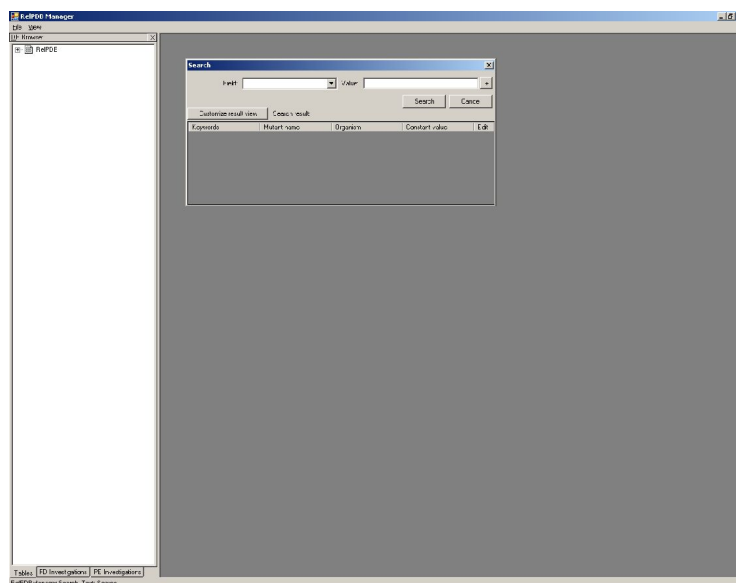


Рисунок 5. Форма поиска.

Выполнение операции начинается с ввода названия белка, вида организма, реакции или процесса в поисковую форму приложения Protein Activity (Рис. 5). В поле **Field** выберите значение «Protein full name» для поиска по полному названию белка или «Protein short name» для поиска по краткому названию белка, «Organism» для поиска по названию вида организма, «Reaction» для поиска по реакции, «Process» для поиска по процессу. В поле **Value** введите название, по которому идет поиск.

Если необходимо осуществить поиск по нескольким полям, нужно нажать кнопку **+**, после чего должна появиться ещё одна строка поиска. Щелчком левой кнопки мыши нужно выбрать комбинационное условие: **And** – и, **or** – или, с которым новая строка включится в запрос (Рис. 6).

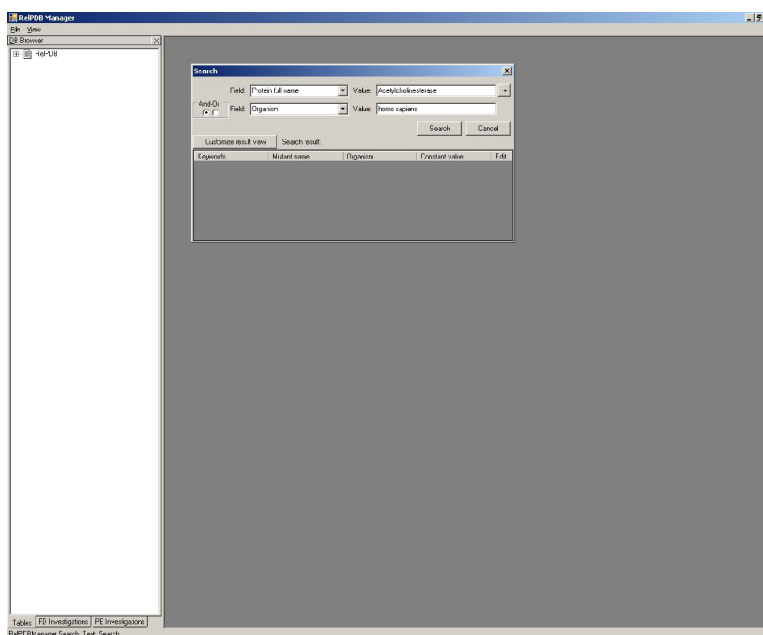


Рисунок 6. Поиск карточки по названию белка и виду организма.

Нажмите кнопку **Search**, чтобы осуществить поиск. В таблице «Search result» появится карточка, удовлетворяющая введенным критериям (Рис. 7).

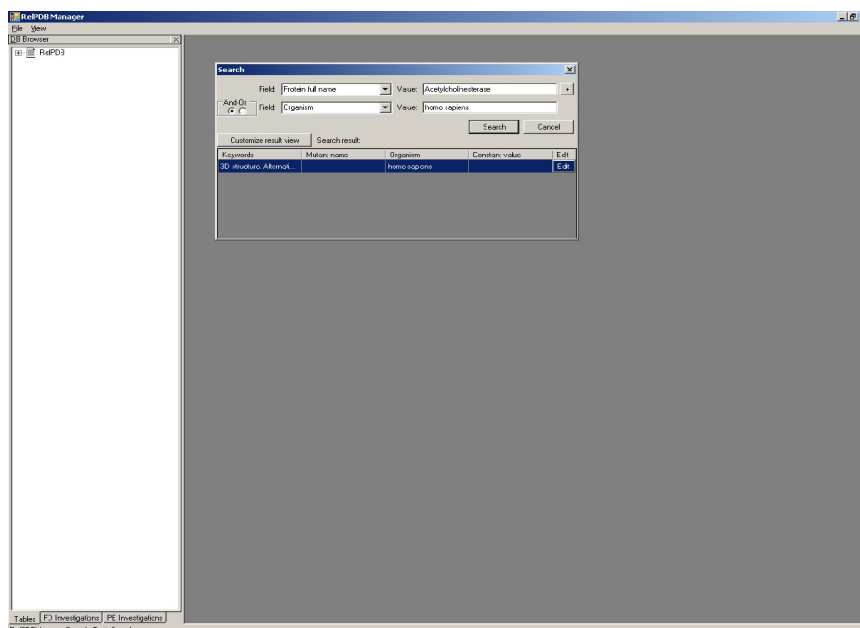


Рисунок 7. Нахождение карточки, удовлетворяющей введенным критериям.

Для того чтобы посмотреть данные активности белка и его мутантов, необходимо нажать на кнопку «Edit» интересующей карточки (Рис. 7), откроется соответствующая карточка, в левом верхнем углу которой имеются закладки Bibliography, Investigation, Data и Constants (Рис. 8).

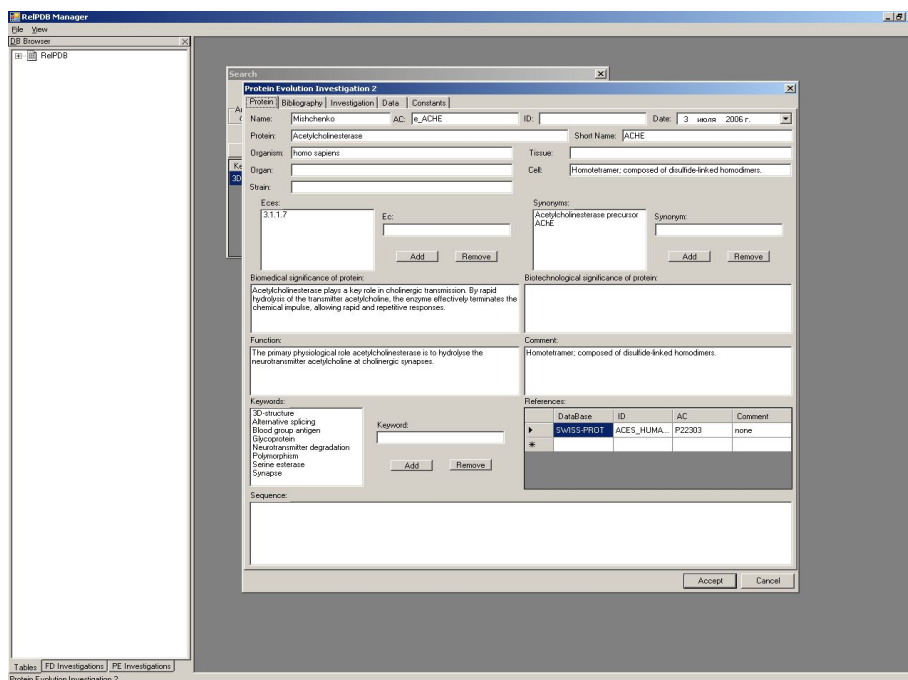


Рисунок 8. Карточка для Acetylcholinesterase, найденная в результате поиска.

Для того чтобы посмотреть данные по активности белка и его мутантов, необходимо в карточке белка нажать на закладку «Constants» левой кнопкой мыши. В данной закладке в колонке **Mutant** указан тип белка (дикий, мутант), в колонке **Substance** – название соединения, для которого определены кинетические параметры, в колонке **Description**, нажав на кнопку **Edit**, можно посмотреть дополнительные условия, в которых проводился эксперимент (информация об основных условиях эксперимента находится в закладке **Investigation**), в колонке **Constant** указаны кинетические параметры, в колонке **Value** – значение параметров, в колонке **Accuracy** – погрешность измерения параметров, в колонке **Measure** – единицы измерения параметров (Рис. 9).

Mutant	Substance	Description	Constant	Value	Accuracy	Measure
WT	acetylcholine	Edit	Km	3.1C	0.00	μM
F29A	acetylcholine	Edit	Km	3.1C	0.00	μM
F33A	acetylcholine	Edit	Km	3.1C	0.00	μM
F33A	acetylcholine	Edit	Km	3.24	0.00	μM
F33A/F33A	acetylcholine	Edit	Km	3.1C	0.00	μM
F33A/F33A	acetylcholine	Edit	Km	3.1E	0.00	μM
F29A/F33A	acetylcholine	Edit	Km	4.9E	0.00	μM
F29A/F33A	acetylcholine	Edit	Km	3.07	0.00	μM
WT		Edit	heat multiplied E.S	3.1C	0.00	1/min
F29A		Edit	heat multiplied E.S	3.3E	0.00	1/min
F33A		Edit	heat multiplied E.S	1.4F	0.00	1/min
F33A		Edit	heat multiplied E.S	1.6E	0.00	1/min
F33A		Edit	heat multiplied E.S	2.4C	0.00	1/min
F33A/F33A		Edit	heat multiplied E.S	7.4F	0.00	1/min
F33A/F33A		Edit	heat multiplied E.S	3.3C	0.00	1/min
F29A/F33A		Edit	heat multiplied E.S	2.24	0.00	1/min
F29A/F33A		Edit	heat multiplied E.S	1.3F	0.00	1/min
WT	acetylcholine	Edit	heat/km multiple	34.00	0.00	1.3E 1/min
F29A	acetylcholine	Edit	heat/km multiple	25.00	0.00	1.3E 1/min
F33A	acetylcholine	Edit	heat/km multiple	16.00	0.00	1.3E 1/min
F33A	acetylcholine	Edit	heat/km multiple	25.00	0.00	1.3E 1/min
F33A	acetylcholine	Edit	heat/km multiple	11.00	0.00	1.3E 1/min

Рисунок 9. Карточка для Acetylcholinesterase. Закладка «Constants», содержащая данные по активности белка дикого типа и мутантов.

В закладке «Bibliography» находятся данные, относящиеся к статье, из которой занесены данные по активности белка. В данной закладке имеются поля для авторов статьи (**Authors**), заголовка статьи (**Title**), журнала (**Journal**), год выхода статьи (**Year**, численное значение), том журнала, в который включена статья (**Volume**, численное значение), выпуск, в который включена статья (**Issue**, численное значение), начальная страница статьи (**Start Page**, численное значение), конечная страница статьи (**End Page**, численное значение), идентификатор PubMed (**PMID**), дополнительная информация, характеризующая статью (**Comment**) (Рис. 10).

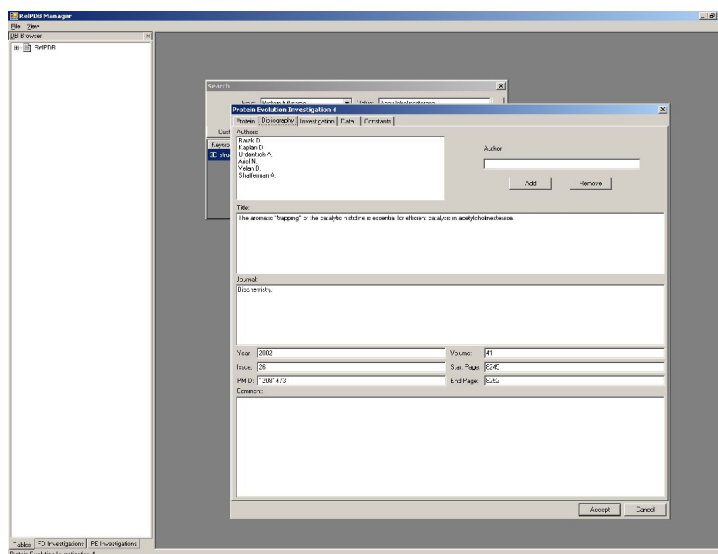


Рисунок 10. Карточка для Acetylcholinesterase. Закладка “Bibliography”.

В закладке «Investigation» собраны данные, характеризующие условия проведения эксперимента. Имеются поля для изучаемого процесса (**Process**), реакции (**Reaction**), используемого метода (**Method**), пункты уточняющие происхождение белка (**Purified, Endogenous, Purified Recombinant, Fusion Protein** - очищенный, эндогенный, очищенный рекомбинантный, химерный белок, соответственно), в поле **Description** приведены данные температуры, состава буфера, значение pH, концентрации соединений, колонки **Mutant** и **Mutations** содержат сведения о мутантах и мутациях (Рис. 11).

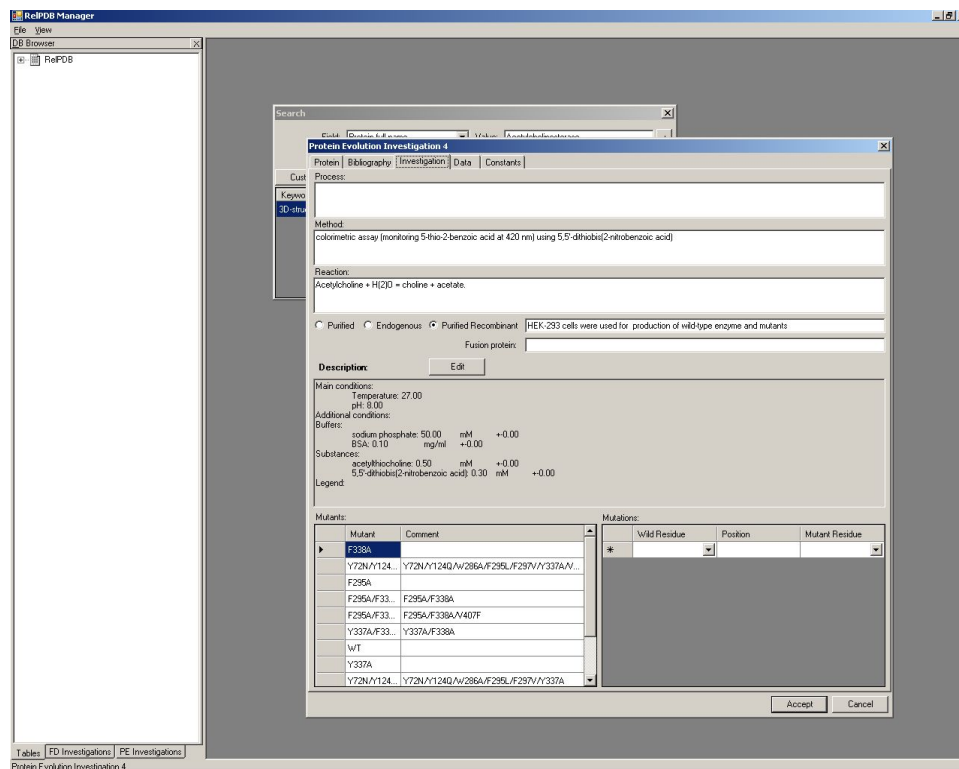


Рисунок 11. Карточка для Acetylcholinesterase. Закладка “Investigation”.

В закладке “Data” содержатся данные экспериментов, представленные в виде графиков (Рис. 12). В колонке **Mutant** указываются мутанты, колонка **Description** содержит их описание, колонка **Comment** предусмотрена для дополнительной информации, поле **Description** указывает название кинетической зависимости, поля **X Values** и **Y Values** содержат значения координат X и Y соответственно для точек графика (численные значения, разделённые пробелом). Для Acetylcholinesterase данная закладка пуста.

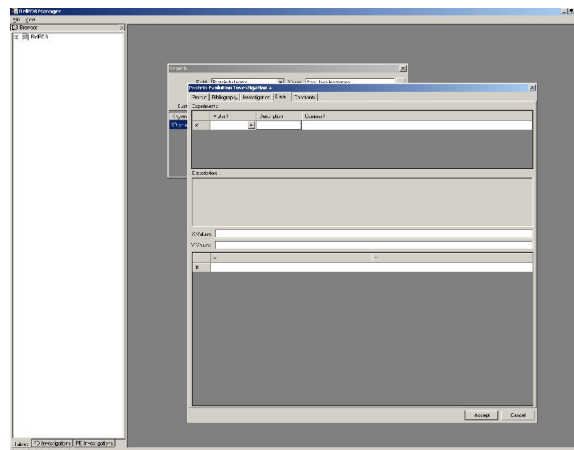


Рисунок 12. Карточка для Acetylcholinesterase. Закладка “Data”.

2.2. Программные компоненты подсистемы «Полиморфизм»

2.2.1. Программные компоненты, позволяющие исследовать влияние полиморфизмов на функции транскрипционных регуляторных районов.

Выявление транскрипционных факторов, сайты связывания которых были изменены в результате возникновения полиморфных мутаций (с помощью матриц частот олигонуклеотидов).

Сегодня, в так называемую Пост-Геномную Эру, исследование изменчивости геномов называемой «полиморфизмами» (Single Nucleotide Polymorphism, SNP; Одиночных Нуклеотидных Полиморфизмов, ОНП) стала новой ключевой задачей генетики: более 2.840.000 ОНП зарегистрировано (Marth et al., 2001) и не менее 1.420.000 картировано (Sachidanandam et al., 2001). В настоящее время результаты экспериментов по ОНП накапливаются в огромном количестве банков данных: общенаучных EMBL Data Library (Cochrane et al., 2006) и Gene Bank (Benson et al., 2006), специализированных HGMD (Krawczak et al., 2000), dbSNP (Sherry et al., 2001), HGBASE (Brookes et al., 2000), ALFRED (Osier et al., 2001), OMIM (Hamosh et al., 2000), а также в так называемых локус-специфических FANCONI (Gillio et al., 1997).

Функциональная роль ОНП в кодирующих районах генов и в сайтах сплайсинга, когда изменяется структура и/или биологическая функция белка, является сейчас более понятной, чем в случае ОНП регуляторных районов, таких как промоторы, энхансеры, интроны, сайленсеры (Ропомаренко et al., 2002). Известно, что регуляция транскрипции генов осуществляется путем взаимодействия белковых факторов транскрипции с сайтами их связывания в этих регуляторных районах ДНК (обозначаемыми ССТФ). В зависимости от локализации в регуляторном районе, ОНП может полностью разрушить ССТФ (Vasiliev et al., 1999; Voccia et al., 1996), создать новый «ложный» ССТФ (Knight et al., 1999; Piedrafita et al.,

1996) или количественно изменить эффективность связывания ДНК-белок (O'Neill et al., 1990; Matsuda et al., 1992; Moi et al., 1992; Tsutsumi-Ishii et al., 1995; Langdon and Kaufman, 1998). Поскольку, транскрипция эукариот регулируется не отдельными факторами транскрипции, а их комплексами, называемых «транскриптосомами» или «транскрипционными машинами» (Gall et al., 1999) и генными сетями (Ananko et al., 2005), лишь для очень немногих ОНП в регуляторных районах генов понятно, как они могут влиять на фенотип. Поэтому распознавание потенциальных ССТФ, мутации в которых ассоциированы с генетической предрасположенностью к заболеваниям и/или к устойчивостью/чувствительностью к лекарствам являются до сих пор важными для понимания молекулярных механизмов фенотипического проявления регуляторных ОНП.

Принцип работы программы rSNP_Guide основан на использовании общеизвестного факта, что все ДНК-связывающие белки имеют два разных уровня эффективности связывания ДНК: (а) так называемое "контекстно-независимое" слабое сродство к любой ДНК, общее для всех таких белков и (б) так называемое "контекстно-специфическое" сильное сродство каждого белка к строго определенным для него участкам ДНК, ССТФ.

Во введенных выше обозначениях и терминах, программа rSNP_Guide прежде всего оценивает теоретически-ожидаемый уровень "контекстно-независимого" ДНК/белок-связывания, который считается общим для всех N анализируемых белков, K мутационных вариантов и каждой из нитей рассматриваемого участка геномной двух-нитевой ДНК:

$$\text{Theor}_0 = M_0(\text{Score}_{dn}(\text{DNA}_k) \mid \text{Score}_{dn}(\text{DNA}_k) < 0; 1 \leq k \leq K, 1 \leq n \leq N, d \in \{0, 1\}); \quad (1)$$

здесь: $M_0(\text{переменная} \mid \text{условие})$ - среднее арифметическое значение переменной при заданном условии); $\text{Score}_{dn}(\text{DNA}_k)$ - для k-ого варианта анализируемого участка прямой (d=0) или комплементарной (d=1) нити ДНК, максимальная количественная оценка его сходства с паттерном известных ССТФ для n-го анализируемого белка.

Как можно видеть, формула (1) находит среднее значение всех отрицательных величин теоретических оценок связывания рассматриваемых ДНК и белков, которые соответствуют отсутствию сильного контекстно-зависимого связывания между ними и, таким образом, наличию только слабого контекстно-независимого такого связывания.

Затем она использует экспериментальные $\{X^\#(\text{DNA}_k)\}$ оценки сродства неизвестного белка к K заданным вариантам мутационных вариантов исследуемого района геномной двух-нитевой ДНК (входные данные) для конструирования всех четырех возможных вариантов наличия/отсутствия сайтов сильного связывания этого белка на каждой из двух нитей рассматриваемого района геномной двух-нитевой ДНК:

$$\{X_{+-d}(\text{DNA}_k)\}_{d=0,1; k=1,K} = \{X_{+-d=0}(\text{DNA}_k) = X^\#(\text{DNA}_k), X_{+-d=1}(\text{DNA}_k) = \text{Theor}_0\}; \quad (2)$$

$$\{X_{-+d}(\text{DNA}_k)\}_{d=0,1; k=1,K} = \{X_{-+d=0}(\text{DNA}_k) = \text{Theor}_0, X_{-+d=1}(\text{DNA}_k) = X^\#(\text{DNA}_k)\}; \quad (3)$$

$$\{X_{++d}(\text{DNA}_k)\}_{d=0,1; k=1,K} = \{X_{++d=0}(\text{DNA}_k) = X^\#(\text{DNA}_k), X_{++d=1}(\text{DNA}_k) = X^\#(\text{DNA}_k)\}; \quad (4)$$

$$\{X_{--d}(\text{DNA}_k)\}_{d=0,1; k=1,K} = \{X_{--d=0}(\text{DNA}_k) = \text{Theor}_0, X_{--d=1}(\text{DNA}_k) = \text{Theor}_0\}. \quad (5)$$

Формулы (2) и (3) соответствуют наличию сайта сильного контекстно-зависимого связывания неизвестного белка только на одной из двух нитей рассматриваемого района геномной двух-нитевой ДНК, формула (4) - на каждой из этих двух нитей, тогда как формула (5) моделирует отсутствие таких сайтов на обеих нитях геномной ДНК.

Далее, с помощью меры сходства Декарта каждого n-ого из N векторов теоретически-ожидаемого сродства белков к ДНК с мутациями $\{\text{Theor}_{dn}(\text{DNA}_k)\}_{d=0,1; k=1,K}$ сравнивается с каждым из четырех векторов (формулы 2-5), сконструированных из данных эксперимента:

$$D_{\xi\psi, n} = \{ \sum_{d=0,1} \sum_{k=1,K} [X_{\xi\psi,d,n}(\text{DNA}_k) - \text{Score}_{dn}(\text{DNA}_k)]^2 \}^{1/2}; \quad (6)$$

$$D_{\xi\psi, --} = \{ \sum_{d=0,1} \sum_{k=1,K} [X_{\xi\psi,d,n}(\text{DNA}_k) - X_{--,d,n}(\text{DNA}_k)]^2 \}^{1/2}; \quad (7)$$

здесь: $\xi, \psi \in \{+, -\}$.

Как можно видеть, формула (6) характеризует соответствие между теоретическими предсказаниями сродства к ДНК для n-ого известного белка и экспериментальными данными о таком сродстве для неизвестного белка. При этом формула (7) устанавливает специфические для каждого рассматриваемого эксперимента пороги наблюдаемого в нем различия между каким-либо из трех вариантов (формулы 2-4) наличия сайтов сильного контекстно-зависимого связывания ДНК с рассматриваемыми регуляторными белками и вариантом отсутствия таких сайтов на обеих нитях ДНК (формула 5).

В заключение программа rSNP_Guide статистически сравнивает полученные оценки сходства между теоретическими предсказаниями и экспериментальными наблюдениями (формула 6) с эксперимент-специфическими порогами различия между наличием и отсутствием сайтов сильного контекстно-зависимого связывания ДНК с неизвестным белком (формула 7) с помощью *t*-критерия Стьюдента:

$$\left\{ \begin{array}{l} d_{\xi\psi, n} = D_{\xi\psi, --} - D_{\xi\psi, n} - t_{\alpha; N-2} \times s.d.(D_{\xi\psi, n}); \quad (8) \\ d_{--, n} = t_{\alpha; N-2} \times s.d.(D_{--, n}) - D_{--, n}; \quad (9) \end{array} \right.$$

ЕСЛИ [$\{d_{\xi\psi, n} > 0\} \& \{d_{--, n} < 0\}$] **ТО** неизвестный белок похож на n-ый известный; (10)

здесь: *s.d.*(переменная) - выборочная оценка стандартного отклонения переменной;

$t_{\alpha; N-2}$ - значение *t*-статистики Стьюдента для уровня статистической значимости α для выборки из *N* независимых величин.

Как можно видеть, правило (10) отбирает из числа *N* рассмотренных известных белков только такие, которые, (а) прежде всего, имеют статистически достоверное сходное (формула 8) с экспериментально наблюдаемым изменением их сродства к мутационным вариантам рассматриваемой ДНК и (б), кроме того, чьи указанными изменения сродства к ДНК при этой статистически достоверно отличаются таковых в случае отсутствия сайтов специфического связывания на обеих нитях исследуемого района ДНК (формула 9).

Описание работы rSNP-tools

Выполнение операции является возможным при использовании в качестве входных данных двух линейных символьных последовательностей ДНК, составленной из букв латинского алфавита { A, T, G, C, a, t, g, c } и имеющих длины не менее 50 и не более 150 символов каждая.

Для выполнения операции надо войти в подсистему «Полиморфизм» и с помощью мыши кликнуть надпись «rSNP_Tools» в первой строке списка операций этой подсистемы. В результате появится HTML-интерфейс операции, который показан на рисунке ниже.

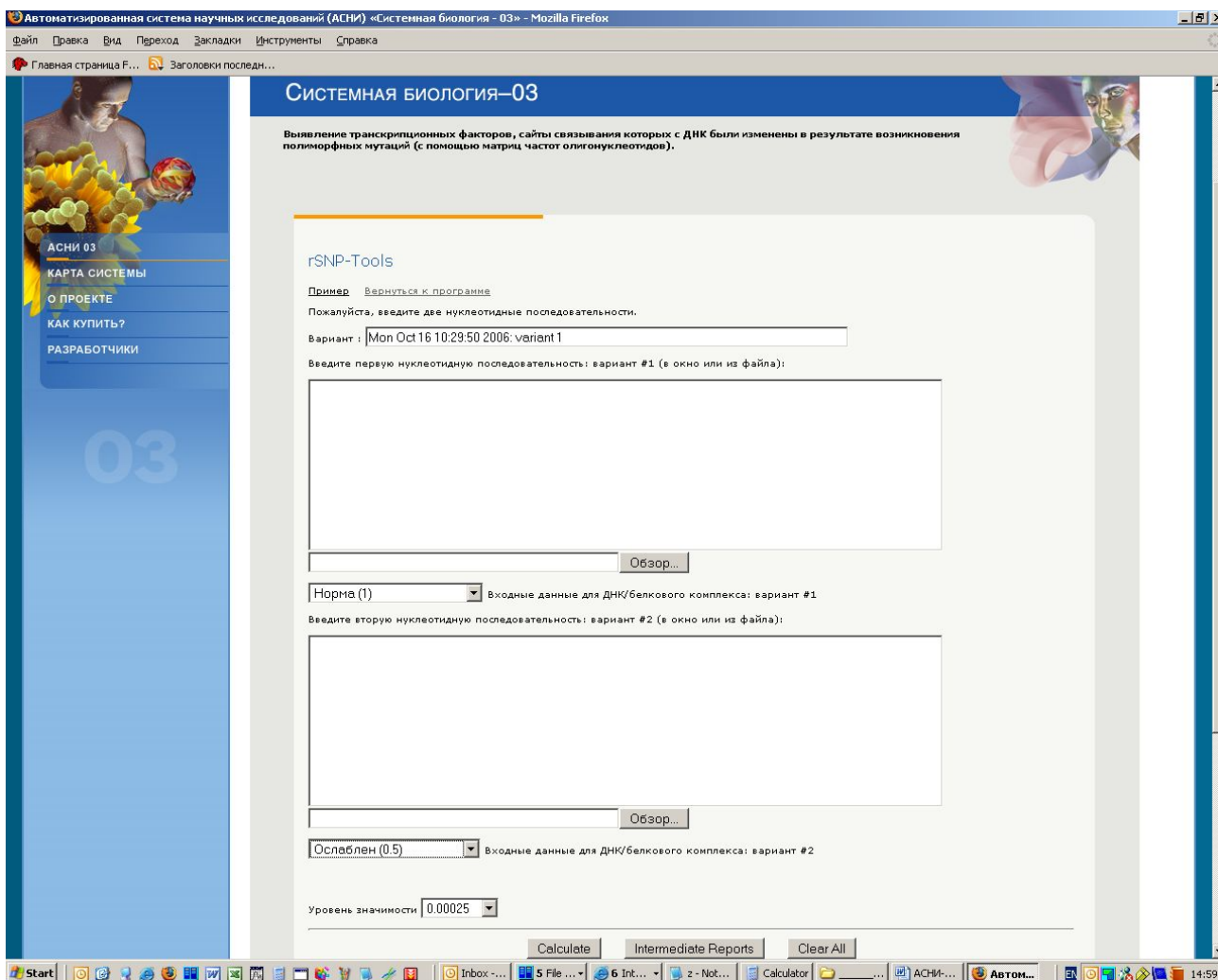


Рисунок 13. HTML-интерфейс операции (начальное состояние).

Выполнение операции начинается с ввода нуклеотидной последовательности первого из двух анализируемых полиморфных вариантов исследуемой ДНК. Для этой цели используется раздел “Введите первую нуклеотидную последовательность: вариант #1 (в окно или из файла):” интерфейса операции, как это показано на рисунке ниже.

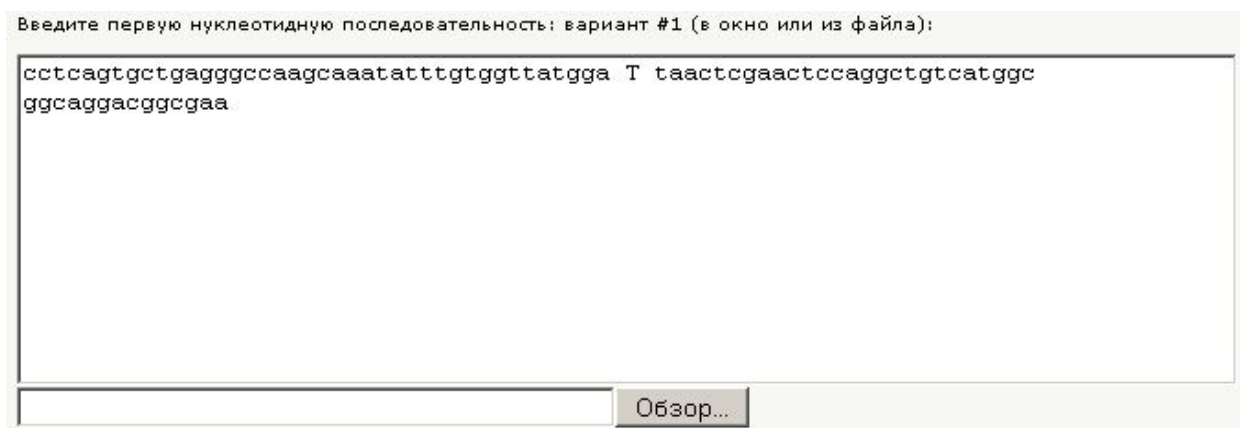


Рисунок 14. Ввод нуклеотидной последовательности первого из двух анализируемых полиморфных вариантов ДНК в режиме ввода данных "Введите первую нуклеотидную последовательность: вариант #1 (в окно или из файла):", который установлен по-умолчанию.

В качестве альтернативного способа ввода нуклеотидной последовательности первого полиморфного варианта анализируемой ДНК из текстового файла используется ИНТЕРНЕТ-браузер путем нажатия клавиши "Обзор".

После ввода нуклеотидной последовательности первого из двух исследуемых полиморфных вариантов анализируемой ДНК с помощью окна «Входные данные для ДНК/белкового комплекса: вариант #1» вводятся входные данные о соответствующем ДНК\белковом комплексе этого варианта ДНК и неизвестного белка, как это показано на рисунке ниже.

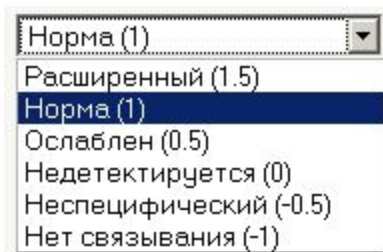


Рисунок 15. Ввод данных о комплексе ДНК-белок с помощью интерфейса.

После завершения ввода данных о первом из двух исследуемых полиморфных вариантах ДНК аналогичным образом осуществляется ввод информации о втором таком полиморфном варианте с помощью второй пары окон интерфейса, маркированных надписью "вариант #2". Пример получаемого в результате этого состояний указанных окон HTML-интерфейса показан на рисунке ниже.

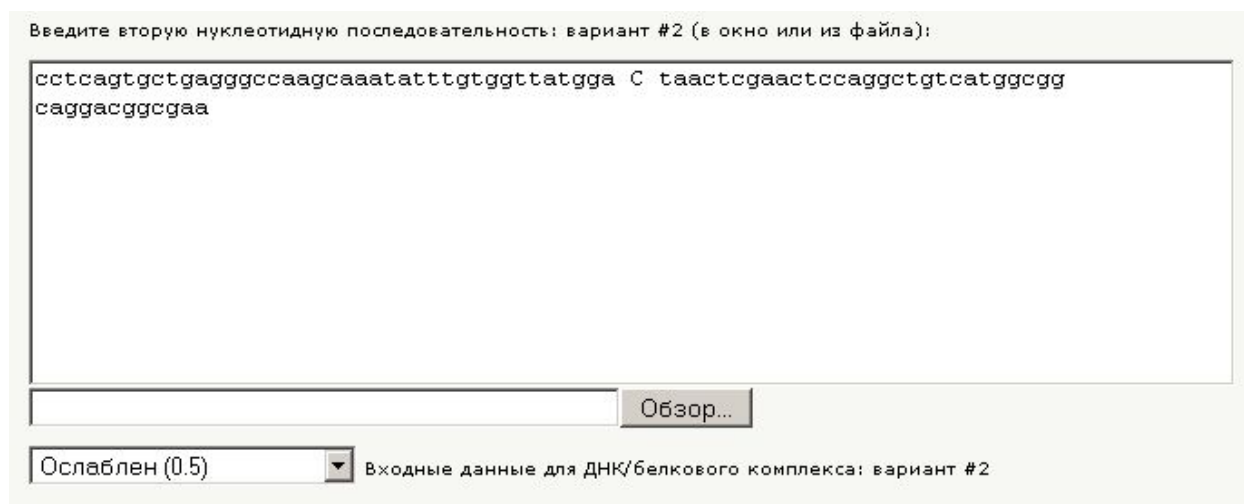


Рисунок 16. Ввод входных данных о нуклеотидной последовательности и ДНК-белковом комплексе для второго из двух анализируемых полиморфных вариантов исследуемой ДНК завершен.

Затем устанавливается уровень статистической значимости выполнения операции с помощью окна “Уровень значимости” интерфейса, как это показано на рисунке ниже.

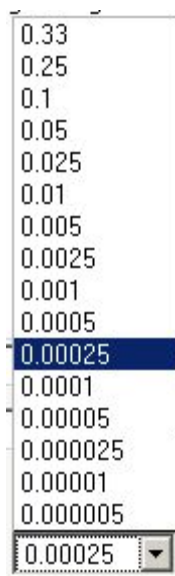


Рисунок 17. Установка уровня значимости в окне "Уровень значимости" интерфейса.

На заключительном шаге необходимо нажать клавишу “Calculate”, в результате чего появляется HTML-страница с предсказанием транскрипционных факторов-кандидатов, сайты связывания которых с исследуемыми полиморфными вариантами ДНК могут изменяться в соответствии с наблюдаемыми изменениями ДНК/белковых комплексов этих двух полиморфных вариантов ДНК. Пример такого предсказания дан ниже на рисунке.

A Simple Example - Mozilla Firefox
 Файл Правка Вид Переход Закладки Инструменты Справка
 Главная страница F... Заголовки последн...

Transcription factors binds to two DNAs: SNPs and/or site-directed mutations
 Time: Mon Oct 16 11:09:24 2006
 Title: Mon Oct 16 10:29:50 2006: variant 1

A nucleotide sequence of the SNP variant #1

```
cctcagtgct gagggccaaag caaatatttg tggttatgga Ttaactcgaa ctcaggctg
tcatggcggc aggacggcga a
```

Select How the unknown binds to this DNA#1: 1

A nucleotide sequence of the SNP variant #2

```
cctcagtgct gagggccaaag caaatatttg tggttatgga Ctaactcgaa ctcaggctg
tcatggcggc aggacggcga a
```

Select How the unknown binds to this DNA#2: 0.5

Significance: 0.00025

RESULTS:

Site-candidatesto be PRESENT: HNF1

Site-candidates to be ABSENT: AP-1; ATF; C/EBP; c-Fos; c-Jun; c-Myb; COUP; CP-1; CRE-BP1; CREB; E2; E2F; EN; ER; Ets; GAGA; GAL4; GATA; GR; HSF; IRF-1; MEF-2
 NF-1; NF-E2; NF-IL6; NF-kB; OCT; PR; RAR; RF-X; RXR; Sp-1; SRF; T3R; TCF-1; TTF-1; USE; YY1;

We have NO IDEA on the Sites: HNF3

Рисунок 18. Пример HTML-страницы предсказания транскрипционных факторов-кандидатов, сайты связывания которых с двумя полиморфными вариантами ДНК могут изменяться в соответствии с наблюдаемыми изменения ДНК/белковых комплексов этих двух вариантов ДНК.

Пользователь может сохранить предсказание стандартными средствами Интернет - браузера, например, с целью накопления таких предсказаний для последующего сравнительного анализа разных полиморфных вариантов исследуемой ДНК.

Для завершения работы с текущей (введенной) парой полиморфных вариантов ДНК и перехода обработке другой пары полиморфных вариантов этой или другой ДНК надо нажать клавишу "Clear all". В результате этого действия осуществляется автоматическая подготовка к повторному выполнению операции.

Требуемый ресурс времени для выполнения операции составляет не более 1 минуты на 1 пару полиморфных вариантов 1 нуклеотидной последовательности ДНК.

Оценка нуклеосомного потенциала в полиморфном районе.

Нуклеосомная укладка нити ДНК является существенным фактором, влияющим на уровень транскрипции. [Agalioti T., et al., 2000] С другой стороны, известны многочисленные случаи полиморфизмов генов типа инсерция\делеция [Mills R.E., et al., 2006]. И инсерционный, и делеционный полиморфизм встречается у человека примерно в равных соотношениях. При этом, размер как инсерций, так и делеций может варьировать от 1 до нескольких тысяч п.о. [Mills R.E., et al., 2006]. В связи с этим крайне важной представляется возможность оценки нуклеосомного потенциала в полиморфном районе. Эту оценку можно осуществить с использованием программной компоненты RECON, основные принципы работы которой представлены в разделе «Построение нуклеосомного потенциала последовательности ДНК» в обучающем курсе для АСНИ-01, в части, посвященной подсистеме «Регуломика».

Описание работы RECON

Для оценки нуклеосомного потенциала в полиморфном районе необходимо произвести запуск программной компоненты RECON, как это описано в разделе «Описание работы Recon» в обучающем курсе для АСНИ-01, в части, посвященной подсистеме «Регуломика».

Исследование необходимо провести дважды – для каждого из двух полиморфных вариантов, а затем сравнить полученные данные

2.2.2. Программные компоненты, позволяющие оценивать возможное влияние полиморфизмов (SNP) на функцию регуляторных районов генов, контролирующих трансляцию.

Оценка изменений трансляционной активности мРНК в SNP-вариантах, связанных с контекстом стартового кодона трансляции и AUG кодонами в составе 5'-НТП мРНК.

Известно, что некоторые параметры мРНК млекопитающих влияют на трансляционную активность. Показано, что 5'НТП регулирует взаимодействие аппарата инициации трансляции и рибосомы (Kozak, 2002). Лидерный район может содержать как негативные (стабильные шпильки, AUG триплеты и рамки считывания), так и позитивные (трансляционные энхансеры) сигналы. К числу известных сигналов общего характера относят контекст стартового кодона трансляции, влияющий на его распознавание рибосомами (Lukaszewicz et al., 2000; Sawant et al., 2001). В ряде случаев 5'НТП содержит сигналы, опосредующие инициацию трансляции по механизмам внутренней инициации трансляции (Pestova et al., 2001). По-видимому, лидерная последовательность может также влиять на цитоплазматическую стабильность мРНК (Gutierrez et al., 1999; Biemelt et al., 2003). В настоящее время критерии, позволяющие определять функциональную активность лидерной последовательности, исследованы недостаточно. Очевидно, что некоторые мутации в 5'-НТП и отдельных позициях белок-кодирующей части могут влиять на паттерн экспрессии гена через изменение трансляционной активности мРНК. Такие случаи известны:

появление в результате мутации нового AUG кодона в составе 5'-НТП. С этого AUG кодона может считываться небольшая рамка считывания или с него может начинаться вариант белка, удлинённый с N-конца. В обоих случаях мутация может существенно изменить либо трансляционную активность матрицы (часть рибосом, поступающих на мРНК, будут транслировать новую рамку считывания), либо свойства кодируемого белка.

Пример:

Matthes et al. Severe hemochromatosis in a Portuguese family associated with a new mutation in the 5'-UTR of the HAMP gene. Blood 2004 104 2181-2183

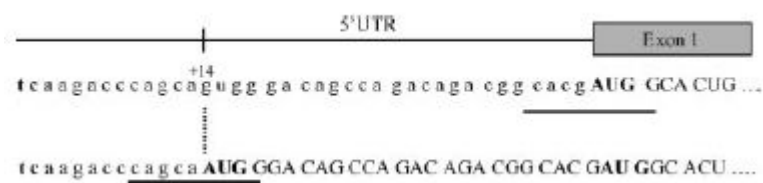


Рисунок 19.

В этом случае произошла мутация G – A, в результате которой появился альтернативный старт трансляции в оптимальном контексте, расположенный в другой рамке считывания. Несмотря на то, что самка ORF гена осталась без изменений, рибосомы не могут

инициировать на нем трансляцию и синтез белка hercidin, контролирующего всасывание железа клетками млекопитающих, не происходит. Что, в свою очередь, приводит к тяжелой форме гемохроматоза.

1. Альтернативный вариант такой мутации – элиминация в результате мутации одного из AUG кодонов в составе 5'-НТП. Эти upstream AUG (uAUG) являются негативными регуляторами трансляции и их элиминация может повысить трансляционную активность мРНК. В случае многих регуляторных белков это может привести к негативным последствиям (например, онкогенезу), что многократно было показано в модельных экспериментах.

Пример 1:

мРНК тромбопоетина в норме характеризуется очень низкой трансляционной активностью, которая поддерживается семью uAUG в составе 5'-НТП. В результате одной из точечных мутаций в минирамке считывания номер 7 (наиболее сильно ингибирующей трансляцию) появился новый стоп-кодон, который уменьшил ее размер и несколько увеличил уровень инициации трансляции на стартовом кодоне тромбопоетина. Что приводит к высокому риску Hereditary thrombocythemia.

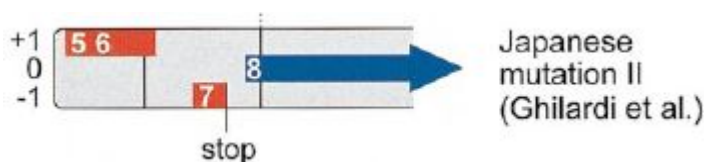


Рисунок 20.

Пример 2.

Другой вариант точечной мутации в этом гене (делеция одного нуклеотида) приводит к сливанию uORF7 и ORF тромбопоетина, что увеличивает белок с N-конца. Получившийся белок функционален, но уровень его синтеза также значительно превышает нормальные физиологические концентрации.



Рисунок 21.

2. В результате мутации может измениться контекст стартового кодона трансляции или одного из AUG кодонов в составе лидерного района. Контекст определяет эффективность распознавания стартового AUG кодона рибосомами, что непосредственно определяет уровень трансляции рамки считывания. Приближение к консенсусу увеличивает эффективность распознавания, и наоборот. Критические позиции контекста -3 выше AUG и +4 (сразу следующая за AUG). Остальные позиции консенсуса (GCC(A/G)CCAUGG) оказывают менее выраженный эффект.

Пример 1.

Jacobson et al. A Graves' disease-associated Kozak sequence single-nucleotide polymorphism enhances the efficiency of CD40 gene translation: a case for translational pathophysiology. *Endocrinology*. 2005 Jun;146(6):2684-91

В этом случае показано, что замена в минорной позиции -1 оптимального С (СТCGCCATG) на неоптимальный Т (СТCGCTATGG) привела к 15% снижению трансляционной активности мРНК и высокому риску Graves' disease.

Пример 2.

Аналогичным образом замена в критической позиции -3 гена BRCA1 оптимального G (AAAGAAATGG) на неоптимальный Т (AAATAAATGG) привела к 70% снижению трансляции и частому появлению рака молочной железы.

3) Во многих случаях 5'- и 3'-НТП содержат сигналы, связывающие регуляторные белки и влияющие на специфичность трансляции мРНК. SNP могут повреждать эти сигналы, что – в свою очередь – может влиять на различные процессы жизнедеятельности клетки и организма в целом. Следует отметить, что этот раздел молекулярной генетики исследован очень плохо и структура большинства сигналов практически не изучена. Причиной этого является тот факт, что в структуре сигнала участвуют как элементы первичной структуры, так и элементы вторичной структуры и анализ таких сигналов методически очень сложен. В настоящее время известно много случаев, когда наследственные заболевания связаны с точечными мутациями в 5'-НТП, которые не имеют отношения к известным трансляционно-активным характеристикам (контекст стартового кодона трансляции, uAUG, uORF, существенные изменения стабильности вторичной структуры). По-видимому, такие негативные мутации связаны со специфическими сигналами посттранскрипционного контроля.

Программа TG_initiator предназначена для сравнения трансляционной структуры 5'-НТП SNP-вариантов мРНК млекопитающих. В программу заложена информация о трансляционно-значимых параметрах лидерных районов (контекстная компонента – триплеты AUG, лидерные рамки считывания (размер, положение), контекст стартового кодона). Если в результате мутации происходит изменение одного или нескольких трансляционно-значимых элементов, программа сообщит об этом, а также предскажет вероятное направление изменения трансляционной активности (увеличение или уменьшение). Следует учесть, что оба варианта отклонения могут быть опасны для организма: как снижение, так и повышение уровня экспрессии гена может привести к дисфункции и негативным последствиям.

Описание работы TG-Initiator

Оценка изменений трансляционной активности мРНК в полиморфных вариантах, связанных с контекстом стартового кодона трансляции и AUG кодонами в составе 5'-НТП с использованием программы TG-Initiator выполняется при использовании в качестве входных данных символьных последовательностей 5'-нетранслируемого района и кодирующей последовательности мРНК, составленных из букв латинского алфавита { A, T, U, G, C, a, t, u, g, c }. Для проведения расчетов необходимо ввести нуклеотидные последовательности 5'-НТП мРНК размерами не менее 10 нуклеотидов и не более 1000 нуклеотидов, а последовательности CDS размерами не менее 300 нуклеотидов и не более 9000.

Необходимо войти в подсистему «Полиморфизм» и с помощью мыши кликнуть надпись «Оценка изменений трансляционной активности мРНК в SNP-вариантах, связанных с контекстом стартового кодона трансляции и AUG кодонами в составе 5'- НТП мРНК» в списке операций этой подсистемы. В результате появится HTML-интерфейс этой операции, который показан на рисунке ниже.

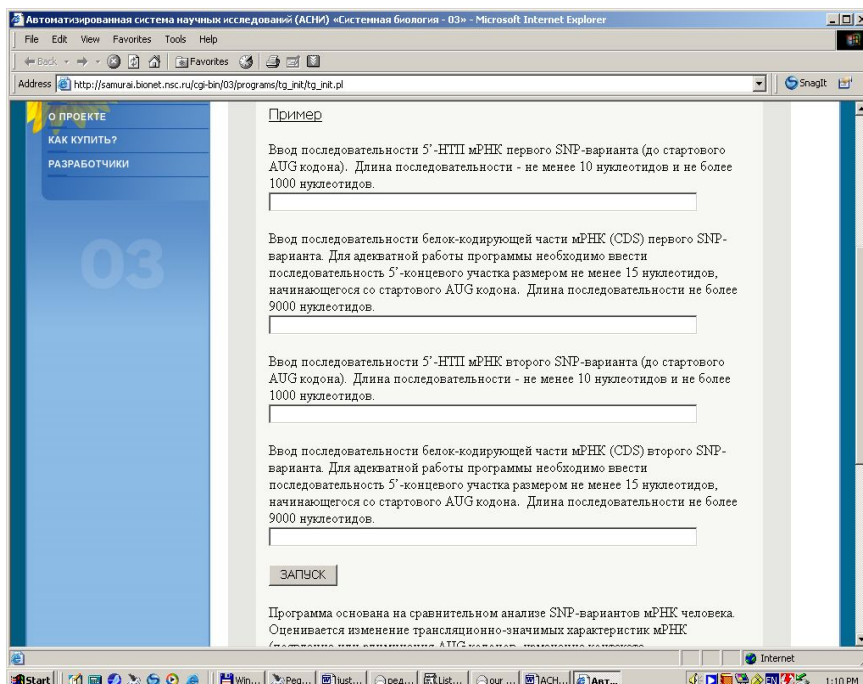


Рисунок 22. Интерфейс операции: начальное состояние.

Как можно видеть, этот интерфейс имеет 4 поля для ввода нуклеотидных последовательностей соответственно 5'-НТП и CDS каждого из двух анализируемых полиморфных вариантов мРНК.

Выполнение операции начинается с ввода нуклеотидных последовательностей 5'-НТП и CDS исследуемых полиморфных вариантов мРНК. Для этой цели используется HTML-интерфейс операции, как это показано на рисунке выше для случая режима ввода данных в соответствующие окна на экране, который установлен по-умолчанию.

При этом последовательности 5'-НТП и CDS полиморфного варианта мРНК необходимо ввести в разделы для первого и второго SNP-вариантов нуклеотидные последовательности 5'-НТП и белок-кодирующих частей.



Рисунок 23. Ввод нуклеотидных последовательностей 5'-НТП и CDS двух исследуемых полиморфных вариантов мРНК в режиме ввода данных "from Screen", который установлен по-умолчанию.

После ввода последовательностей, необходимо нажать кнопку **Запуск**, в результате чего появляется HTML-страница с результатами расчетов, как это показано на рисунке ниже.

```
Codon position 78
codon with context (-6, +4):
before- GGAGCCAAGC
after - GGAGCCATGG
message: SNP may change mRNA translational properties: mutation resulted in an appearance of a new AUG codon within a 5'-UTR. Ribosome could recognize this uAUG and initiate translation from the uORF it starts. Since this uAUG is located in the optimal context, initiation of translation at AUGs located within this uORF (if any) is likely to be suppressed. Initiation of translation at downstream AUGs located beyond the stop codon of the uORF could occur by reinitiation mechanism. Note that a reinitiation efficiency increases with a decrease of the uORF's size and with an increase of a distance to the next start codon. Generally, an appearance of a AUG triplet within 5'-UTR is a negative feature potentially reducing eukaryotic mRNA translation rate.
```

```
Codon position 87
codon with context (-6, +4):
before- CGGACGTGGG
after - GGGACGTGAG
message: SNP may change mRNA translational properties: mutation resulted in an appearance of a new nonsense codon which changes the size of a leader-located ORF (uORF). In general, the influence of this mutation on mRNA translatability depends on the combination of various factors including the efficiencies of uORF translation and the reinitiation of translation beyond uORF.
```

```
Codon position 105
codon with context (-6, +4):
before- ATTCGGATGG
after - ATTCGGATGT
message: SNP may change mRNA translational properties: mutation resulted in a change of start codon context. In this case the optimal guanine (pos. +4) is mutated to a suboptimal nucleotide. Since the second important position (-3) is occupied by a suboptimal pyrimidine nucleotide the start codon recognition efficiency may decrease
```

Рисунок 24. Пример результата выполнения операции.

Как можно видеть, в качестве выходных данных генерируется на так называемом «ограниченном естественном языке» текстовое описание и экспертная оценка изменений трансляционной активности мРНК в полиморфных вариантах в результате мутаций. Пользователь имеет возможность сохранить результаты работы программы с помощью стандартных средств интернет - браузера.

Оценка стабильности вторичной структуры в 5'-НТП в норме и SNP-вариантах, прогнозирующая изменения трансляционной активности.

Хорошо известно, что мРНК в цитоплазме способна формировать различные вторичные структуры за счет комплементарных взаимодействий. Предсказание этих структур (в особенности с учетом РНК-белковых взаимодействия) – это сложная задача, не решенная до конца в настоящее время. Тем не менее, особенности вторичной структуры в лидерном районе мРНК, влияющие на эффективность инициации трансляции, могут быть оценены с помощью методов компьютерного анализа. Экспериментально показано, что вторичная структура в 5'-НТП негативно сказывается на трансляционную активность мРНК (Kozak, 2002). При этом особенно негативное влияние (в аспекте взаимосвязи ингибирующего эффекта и стабильности вторичной структуры) проявляли шпильки, расположенные на небольшом расстоянии от 5'-конца мРНК, что, по-видимому, может быть связано с ингибированием взаимодействия кеп- и кеп-связывающего комплекса трансляционных факторов. Следует отметить, что в клетках разных организмов этот эффект может варьировать, что связано с особенностью организации аппарата инициации трансляции

(например, активность геликаз). Так, мРНК млекопитающих и однодольных растений характеризуются значительно более высоким содержанием G+C и, соответственно, более высоким потенциалом формирования вторичной структуры, чем мРНК двудольных растений.

Соответственно, изменение стабильности вторичной структуры 5'-НТП в результате SNP может изменить уровень трансляционной активности.

Программа `Leader_hairpin` предназначена для оценки изменений структурных параметров 5'-НТП мРНК генов человека в результате мутации. Программа оценивает как полную вторичную структуру, так и ее элементы, расположенные в 5'-концевой части лидерной последовательности. Если 5'-концевая часть неструктурирована, это может означать высокую эффективность взаимодействия мРНК и кеп-связывающего комплекса и, как следствие, высокую эффективность этой стадии процесса инициации трансляции.

Если программа предсказывает существенное изменение стабильности вторичной структуры в 5'-концевом участке мРНК, это может означать значительные изменения в параметрах экспрессии гена за счет отклонений в уровне трансляционной активности матрицы.

Описание работы `Leader_hairpin`

Провести оценку стабильности вторичной структуры в 5'-НТП в полиморфных вариантах, прогнозирующая изменения трансляционной активности с помощью программы `Leader_hairpin` можно, введя нуклеотидную последовательность так называемого «5'-концевого» фрагмента (другие, синонимичные обозначения этого фрагмента: «5'UTR», «5'НТП») мРНК, локализованного непосредственно перед стартовым AUG-кодоном. Минимальная анализируемая длина (предпочтительно не менее 15 нуклеотидов) 5'UTR с необходимостью устанавливается пользователем с помощью интерфейса операции в качестве оптимизируемого параметра ее выполнения (как это описано ниже). При этом нуклеотидная последовательность должна быть составлена из символов латинского алфавита «A, T, U, G, C, a, t, u, g, c».

Для выполнения операции надо войти в подсистему «Полиморфизм» и выбрать опцию «Оценка стабильности вторичной структуры в 5'-НТП в норме и SNP-вариантах, прогнозирующая изменения трансляционной активности» в списке операций этой подсистемы. В результате появится HTML-интерфейс операции для ввода нуклеотидных последовательностей, который показан на рисунке ниже.

Leader_hairpin

Пример

Заголовок :

Введите нуклеотидную последовательность 5' -нетранслируемого района до стартового AUG кодона (только полноразмерные экспериментально-верифицированные лидерные последовательности; не вводите стартовый AUG кодон). Нуклеотидная последовательность должна быть в чистом виде или в фаста-формате; пробелы или неканонические символы использовать нельзя (только a,t,g,c в любом регистре).

Введите 5'-концевой участок нуклеотидной последовательности белок-кодирующей части (около 100 нуклеотидов, начиная со стартового AUG кодона).

Введите нуклеотидную последовательность 5'-нетранслируемого района (SNP-вариант; только полноразмерные экспериментально-верифицированные лидерные последовательности; не вводите стартовый AUG кодон). Нуклеотидная последовательность должна быть в чистом виде или в фаста-формате; пробелы или неканонические символы использовать нельзя (только a,t,g,c в любом регистре).

Введите 5'-концевой участок нуклеотидной последовательности белок-кодирующей части (SNP-вариант; около 100 нуклеотидов, начиная со стартового AUG кодона).

Размер 5'-концевого участка

Рисунок 25. Интерфейс операции: ввод нуклеотидных последовательностей мРНК (начальное состояние).

Сначала надо с помощью верхнего окна ввода нуклеотидных последовательностей мРНК ввести анализируемый полиморфный вариант 5'-нетранслируемый район мРНК (5'-НТП) мРНК, который локализован непосредственно перед стартовым AUG-кодоном, как это показано на рисунке ниже.

Введите нуклеотидную последовательность 5'-нетранслируемого района до стартового AUG кодона (только полноразмерные экспериментально-верифицированные лидерные последовательности; не вводите стартовый AUG кодон). Нуклеотидная последовательность должна быть в чистом виде или в фаста-формате; пробелы или неканонические символы использовать нельзя (только a,t,g,c в любом регистре).

gсggсaaааааАgссggagссаagсggacgtgggagссggtgattcgg

Browse...

Введите 5'-концевой участок нуклеотидной последовательности белок-кодирующей части (около 100 нуклеотидов, начиная со стартового AUG кодона).

atgggtggacacggaaagcccactctgccccctctccccactcgaggссggcgatctagagagссcgttatctgaagagtt
cctgссааgaaatgggaaacatссааgagatttcgcaatссатсggсgaggatagttctggaaгсtttgгсtttacggaat
accagtattaggaagctgtcctggctcagatggctcggtcatcacggacacgcttcaccagcttcgagссcctcctcg
gtgacttatcctgtggtccccggcagcgtggacgagctctccagtgagcattgaacatсgaatgtagaatctgсgggga
caaggсctcaggctatcattacggagctccacgcgtgtgaaggctgсaaaggгсttcttcggсgaacgatctgactcaagc

Browse...

Введите нуклеотидную последовательность 5'-нетранслируемого района (SNP-вариант; только полноразмерные экспериментально-верифицированные лидерные последовательности; не вводите стартовый AUG кодон). Нуклеотидная последовательность должна быть в чистом виде или в фаста-формате; пробелы или неканонические символы использовать нельзя (только a,t,g,c в любом регистре).

gсggсaaааааАgссggagссаagсggacgtgggagссggtgattcgg

Рисунок 26. Ввод нуклеотидной последовательности полиморфного варианта 5'UTR мРНК непосредственно перед стартовым AUG-кодоном.

В качестве альтернативного варианта выполнения этого действия, нуклеотидные последовательности мРНК могут быть введены из текстовых файлов в стандартном формате Fasta с помощью кнопки “Browse”, расположенных непосредственно ниже окна текстового ввода 5'UTR. Необходимо установить минимальную анализируемую длину 5'UTR непосредственно перед стартовым AUG-кодоном мРНК (предпочтительно не менее 15 нуклеотидов), как показано на рисунке ниже.

Размер 5'-концевого участка

Calculate

Ранее было экспериментально показано, что стабильные шпильки в 5'-концевой участке мРНК (размером около 15 нуклеотидов) значительно снижают уровень трансляционной активности (Kozak, 1989). Основываясь на этих данных, программа Leader_hairpin предсказывает структурированные участки, 5'-границы которых локализованы в границах 5'-концевого участка (с 1 по 15 нуклеотид; правая граница может меняться пользователем). Программа также может предсказывать изменения в эффективности трансляции эукариотических мРНК, произошедшие в результате точечных мутаций (SNP): если SNP изменяет стабильность вторичной структуры (в особенности в 5'-концевой части матрицы), это может сопровождаться заметными изменениями в трансляционной активности.

Рисунок 27. Интерфейс управления выполнением операции: установка минимальной анализируемой длины 5'UTR непосредственно перед стартовым AUG-кодоном мРНК (по-умолчанию - предпочтительно не менее 15 нуклеотидов).

После завершения ввода «входных данных» надо нажать кнопку “Calculate”.

В результате этого действия получаются выходные данные, как показано на рисунке ниже.

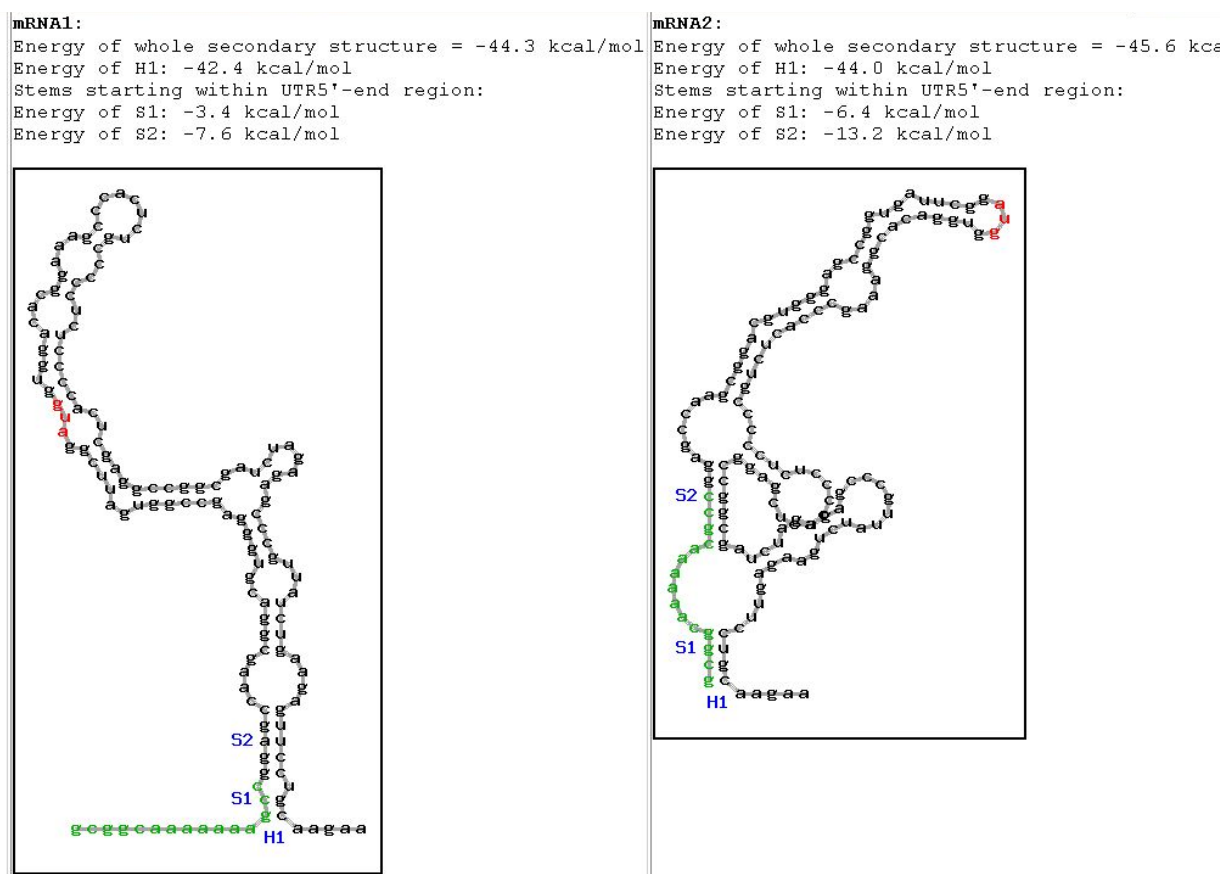


Рисунок 28. Выходные данные – пример результата выполнения операции.

В верхней части этого рисунка последовательно выведены количественные характеристики предсказанной вторичной структуры всего района мРНК - свободная энергия Гиббса всей вторичной структуры, свободная энергия Гиббса вторичной структуры только 5'НТП, свободная энергия Гиббса всех стеблей в 5'НТП. Ниже строк характеристик располагается графическое представление всей вторичной структуры 5'НТП мРНК. Зеленым цветом обозначена вторичная структура, которая потенциально может влиять на скорость инициации трансляции.

На рисунке с выдачей программы можно видеть, что стабильность вторичной структуры в 5'-концевом участке мРНК (энергия шпилек H1 и H2) у SNP-вариантов, различающихся по одной нуклеотидной позиции, сильно отличаются. Это может означать существенные различия в трансляционной активности мРНК.

С помощью стандартных средств Интернет - браузера пользователь может сохранить это предсказание, например, с целью накопления таких предсказаний для разных полиморфных вариантов исследуемой мРНК для последующего сравнительного анализа.

Для завершения работы с текущим (введенным) полиморфным вариантом 5'-района мРНК и перехода обработке другого такого варианта этой или другой мРНК надо удалить нуклеотидные последовательности из соответствующих окон. В результате этого действия осуществляется автоматическая подготовка к повторному выполнению операции.

2.2.3. Программные компоненты, позволяющие оценивать возможное влияние полиморфизмов (SNP) на первичную и пространственную структуру белков и их функцию

Программные компоненты ProtStability1D и ProtStability3D: оценка возможного влияния полиморфизмов на термодинамическую стабильность белков по первичной и третичной структуре

Аминокислотные замены белка, соответствующие некоторому полиморфизму, влияют на его термодинамическую стабильность, то есть способность сохранять биологически активную конформацию под действием температуры окружающей среды. Изменение термодинамической стабильности при конкретной аминокислотной замене зависит от различий между физико-химическими свойствами нормальной и мутантной аминокислот, а также от физико-химических свойств близлежащих аминокислот, а также от третичной структуры того участка, в котором происходит рассматриваемая замена. Наличие или отсутствие изменений в биологической активности белка при заданных аминокислотных заменах, могут служить характеристикой фенотипического проявления конкретного полиморфизма.

Многие мутации могут приводить к вариациям функциональной активности либо полной потере функции белком, либо приобретению новой функции, что часто связано с развитием заболеваний человека. Нарушения функции белков могут быть связаны с изменением термодинамической стабильности мутантных форм.

Среди существующих подходов можно особо выделить работу (Capriotti et al., 2004), авторы которой проводили предсказание изменения свободной энергии мутантной формы белка с помощью нейронной сети с обратным распространением ошибок.

Программные компоненты ProtStability1D и ProtStability3D предназначены для теоретической оценки влияния полиморфизмов на термодинамическую стабильность белков по первичной и по третичной структуре (Ivanisenko et al, 2006). Работа обеих программ основана на общем алгоритме, который состоит из двух этапов. На первом около заданной позиции строится окно определенной длины, для первичной структуры или сфера определенного радиуса для третичной структуры. Запоминаются аминокислоты, попавшие в пределы сферы или окна, определяются их физико-химические свойства. На втором этапе по данным, полученным на первом этапе, производится предсказание термодинамической стабильности. Предсказание основывается на методе кластеризации данных КРАБ и методе нейронных сетей.

Программы позволяют производить оценки влияния полиморфизмов на термодинамическую стабильность белков, как по третичной структуре нормального белка, так и опираясь лишь на аминокислотную последовательность.

Описание работы ProtStability1D

Оценка возможного влияния полиморфизмов на термодинамическую стабильность белков по первичной структуре помощью программы ProtStability1D является возможным при использовании в качестве входных данных линейной символьной последовательности белка в однобуквенном коде, составленной из букв латинского алфавита {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y} и имеющей длину не менее 35 и не более 30000 символов.

Для подготовки к выполнению операции следует войти в подсистему «Полиморфизм» и выбрать функцию «Оценка возможного влияния полиморфизмов на термодинамическую стабильность белков по первичной структуре» в списке операций этой подсистемы. В результате появится HTML-интерфейс этой операции, который показан на рисунке ниже.

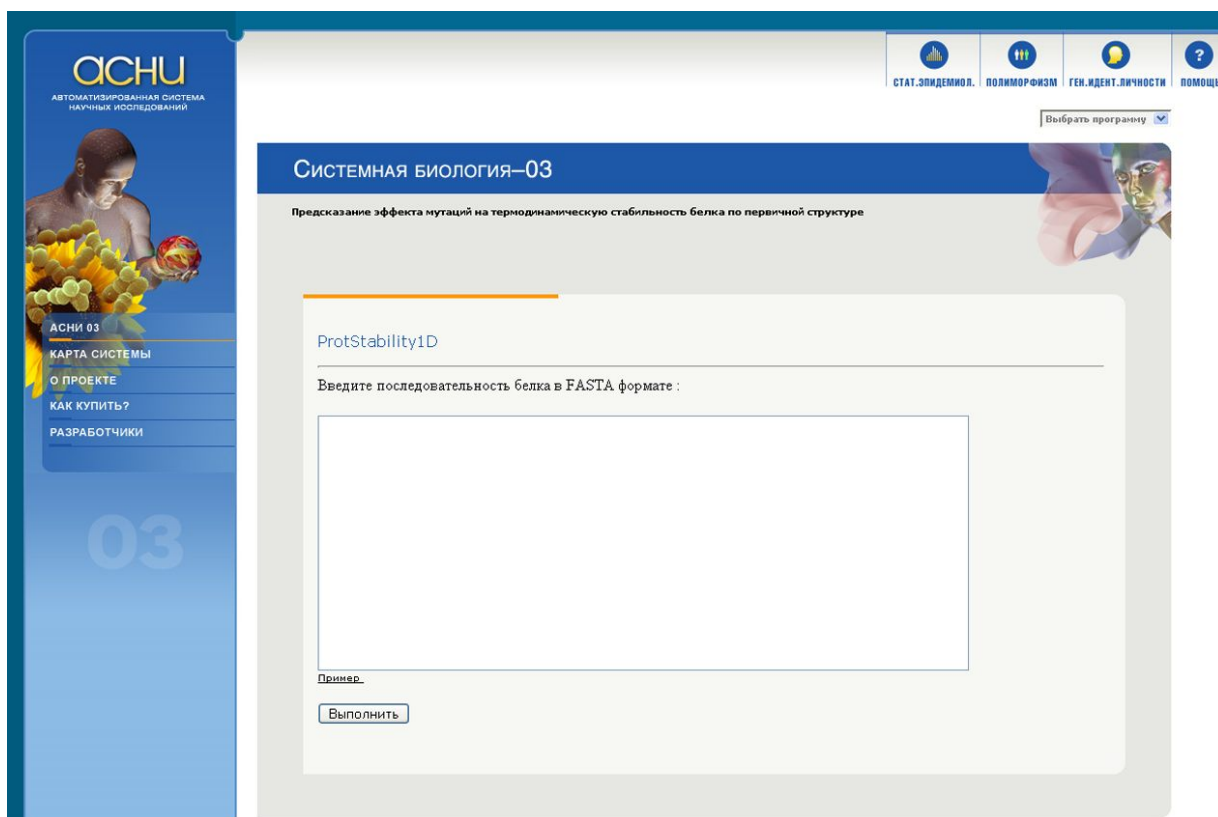


Рисунок 29. HTML-интерфейс операции «оценка возможного влияния полиморфизмов на термодинамическую стабильность белков», подготовленной к началу ее выполнения.

Выполнение операции начинается с ввода аминокислотной последовательности исследуемого белка. Для этой цели используется раздел “ Введите последовательность белка в FASTA формате” HTML-интерфейса операции, как это показано на рисунке ниже:

ASHU
АВТОМАТИЗИРОВАННАЯ СИСТЕМА
НАУЧНЫХ ИССЛЕДОВАНИЙ

СТАТ. ЭПИДЕМИОЛ. ПОЛИМОРФИЗМ ГЕН. ИДЕНТ. ЛИЧНОСТИ ПОМОЩЬ

Выбрать программу

СИСТЕМНАЯ БИОЛОГИЯ—03

Предсказание эффекта мутаций на термодинамическую стабильность белка по первичной структуре

ProtStability1D

Введите последовательность белка в FASTA формате :

```
>HSP53
HEEPQSDPSVEPPLSQETFSDLWKLLENVNLSPPLFSQAMDDLMLSPDDIEQWFTEDPGPDEAPRMEAAAP
PVAFAPAARTRAAPAPAPSWPLSSVPSQTYGSGYGFRLGFLHSGTAKSVTCTTSPALNKHFCQLAKTCP
VQLVLDSTRPFGTRVAHAITYQSQHTEVVRKCFRNERCSDSDGLAPPOHLIRVEGNLRVEYLDNRNFR
HSVVVYEPFPEVGSDCITLHNVYKNSSCGGRNRPRLTITILESSGMLLGRNSFEVRVACAPGRDART
EEMLRKKGERNHLELPQSTKRALPMNTSSSPQKPKPLDGEYFTLQIRGRERFEMFRELMEALELKDQAQ
GKEPQGSRAHSSHLKSKKQGSTSRHKKLMFKTEGPDSD
```

Пример

Выполнить

Рисунок 30. Ввод белковой последовательности.

После ввода данных, как было описано выше, необходимо нажать клавишу “Выполнить”, в результате чего появляется HTML-страница с оценкой возможного влияния полиморфизмов на термодинамическую стабильность исследуемого белка, как это показано ниже на рисунке.

244	-	+	+	+	+	N	+	+	+	+	-	+	-	+	+	+	+	+	+
245	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-
246	-	-	-	-	-	-	-	-	-	0	N	-	-	-	-	-	-	-	-
247	-	+	+	+	+	+	+	+	+	+	N	+	-	+	+	+	+	+	+
248	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-
249	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-
250	+	-	-	-	-	-	-	+	+	-	-	N	+	-	-	-	-	-	-
251	-	-	-	-	-	-	N	-	-	-	-	+	-	-	-	-	-	-	-
252	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-
253	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-
254	+	+	+	-	+	+	N	+	+	+	+	+	+	+	+	+	+	-	+
255	-	-	+	+	-	+	+	N	+	-	-	+	+	+	+	+	+	+	+
256	-	0	0	-	0	-	-	0	0	0	0	-	-	-	0	-	N	0	0
257	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-
258	-	-	+	N	+	-	+	+	+	+	+	+	+	+	+	-	-	+	+
259	-	-	N	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-
260	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	+	-	-
261	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	N	-	-	-
262	-	+	-	-	-	N	+	+	+	-	-	+	-	+	+	-	-	+	-
263	-	-	-	-	-	-	-	-	-	-	+	-	N	-	-	-	-	+	-
264	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	+	-	-	+
265	-	-	-	-	-	-	-	-	-	N	-	-	-	-	+	-	-	-	-
266	-	-	-	-	-	N	-	-	-	-	+	-	-	-	-	+	+	-	-
267	-	-	-	-	-	-	-	-	-	-	+	-	-	-	N	-	-	-	-
268	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-
269	-	-	-	-	-	-	-	-	-	0	-	-	-	-	-	N	-	-	-
270	-	-	-	-	N	-	-	-	-	+	-	-	-	-	-	-	-	+	-
271	-	-	-	N	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
272	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	N	-	-
273	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-
274	-	-	-	-	+	-	-	-	-	0	-	-	-	-	-	-	N	-	-
275	-	N	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-
276	N	+	+	+	+	+	+	+	+	+	-	+	+	+	-	+	+	+	+
277	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-
278	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-
279	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	+	-	+	-
280	+	+	-	+	+	-	+	+	0	+	-	+	+	N	-	+	+	+	+
281	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
282	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-
283	-	-	-	-	-	-	-	-	0	-	-	0	-	N	-	-	-	-	-
284	-	+	-	-	+	-	-	+	-	0	-	-	-	-	+	-	N	+	+

Рисунок 31. Пример HTML-страницы, содержащей оценку возможного влияния полиморфизмов на термодинамическую стабильность исследуемого белка. Стрелками указаны предсказания для замен аминокислот в позициях 248, 273 и 281.

В данном случае анализу подвергалась аминокислотная последовательность противоопухолевого белка p53 человека, EMBL ID=HSP53. Известно, что наличие мутаций в кодонах 248, 273 и 281 белка p53 приводит к нарушению его способности связываться с ДНК, активировать транскрипцию (Zhang et al. 1993; Kaku et al. 2001), приводит к нарушению веретена деления (Gualberto et al. 1998). Как видно из рисунка с выдачей программы, предсказано нарушение термодинамической стабильности белка p53 для большинства замен остатков в позициях 248, 273 и 281.

С помощью стандартных средств интернет-браузера пользователь может сохранить это предсказание, например, с целью накопления таких предсказаний для разных последовательностей белка и их последующего сравнительного анализа. Для возврата на главную страницу необходимо воспользоваться опцией “Back”, а затем «Refresh» в интернет-браузере.

Описание работы ProtStability3D

Оценка возможного влияния полиморфизмов на термодинамическую стабильность белков по третичной структуре с помощью программы ProtStability3D является возможным при использовании в качестве входных данных файла третичной структуры белка в общепринятом в качестве стандарта формате базы данных Protein Data Bank (www.pdb.org) или идентификатора третичной структуры белка и указания идентификатора его исследуемой полипептидной цепи в базе данных Protein Data Bank.

Для подготовки к выполнению операции следует войти в подсистему «Полиморфизм» и выбрать функцию «Оценка возможного влияния полиморфизмов на термодинамическую

стабильность белков по третичной структуре» в списке операций этой подсистемы. В результате появится HTML-интерфейс этой операции, который показан на рисунке ниже.

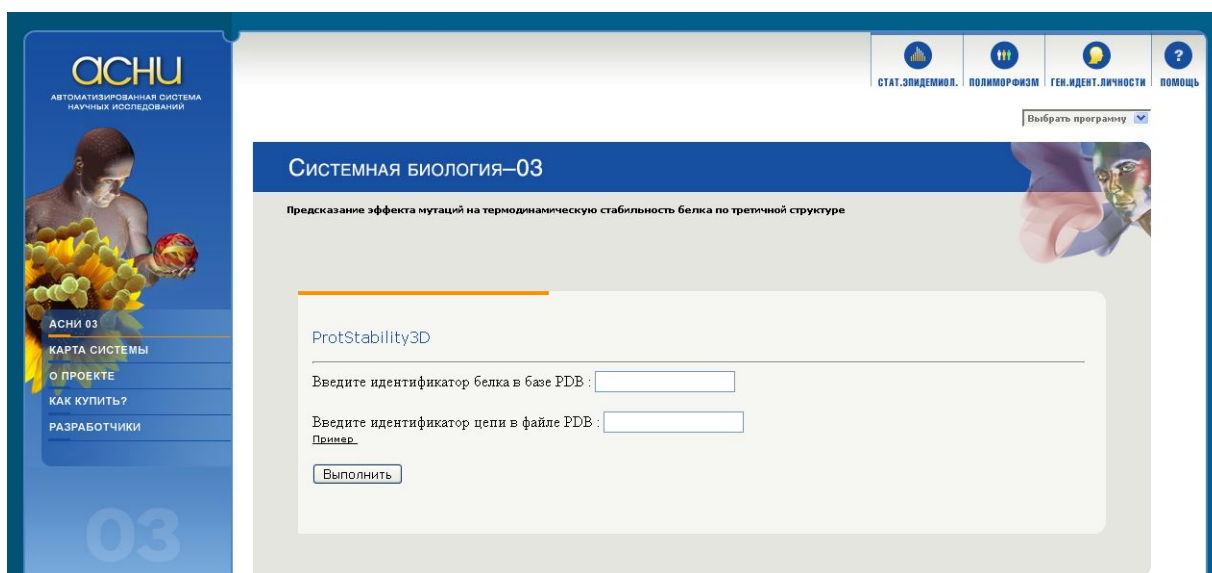


Рисунок 32. HTML-интерфейс операции «оценка возможного влияния полиморфизмов на термодинамическую стабильность белков», подготовленной к началу ее выполнения.

Выполнение операции начинается с ввода идентификатора структуры белка в базе PDB и идентификатора цепи белка. Для этой цели используются разделы “ Введите идентификатор белка в базе PDB” и “ Введите идентификатор цепи в файле PDB” HTML-интерфейса операции, как это показано на рисунке ниже.

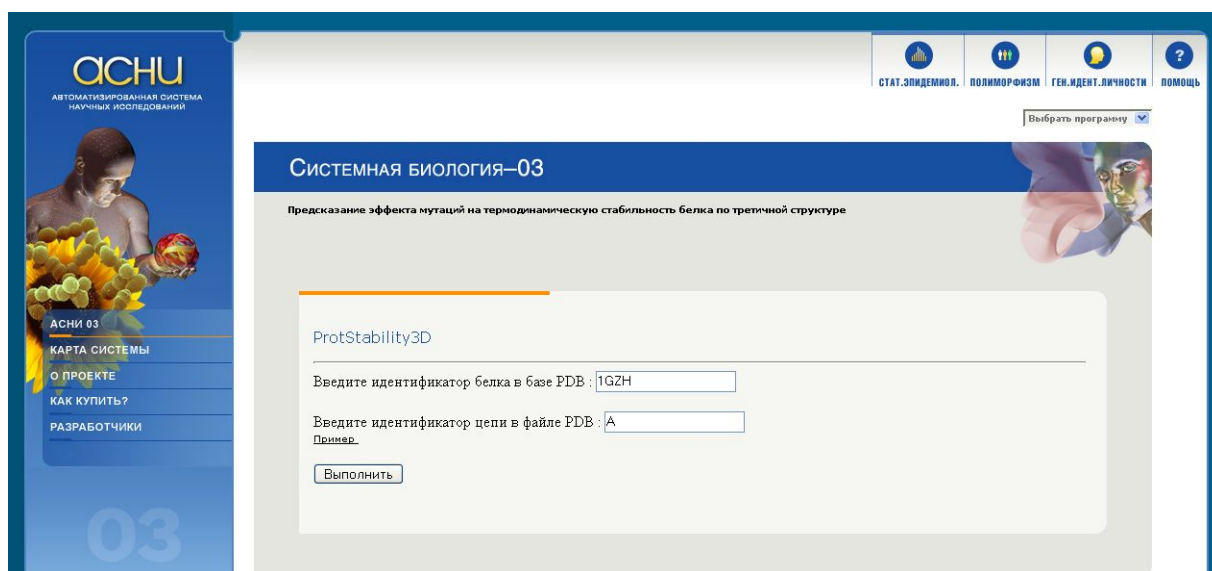


Рисунок 33. Ввод идентификаторов белка и исследуемой цепи.

После ввода данных, как было описано выше, необходимо нажать клавишу “Выполнить”, в результате чего появляется HTML-страница с оценкой возможного влияния полиморфизмов на термодинамическую стабильность исследуемого белка, как это показано ниже на рисунке.

246	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-
247	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-
248	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-
249	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-
250	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-
251	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
252	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
253	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-
254	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-
255	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-
256	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-
257	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-
258	+	+	+	N	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
259	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
260	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-
261	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-
262	-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-
263	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-
264	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-	-
265	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	-	-
266	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
267	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	+	-
268	+	+	+	+	+	+	+	+	+	+	+	N	+	+	+	+	+	+	+
269	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-
270	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
271	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
272	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-
273	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-
274	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-
275	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
276	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
277	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
278	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	-	+	-
279	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
280	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	N	+	+	+
281	-	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
282	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-
283	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-
...																			

Рисунок 34. HTML-страница с фрагментом выходных данных. Стрелками указаны предсказания замен остатков в позициях 248, 273 и 281.

В данном случае анализу подвергалась третичная структура противоопухолевого белка p53 человека (PDB ID=1GZH, цепь «A»). Известно, что замены в аминокислотных позициях 248, 273, 281 белка p53 приводят к нарушению его способности связываться с ДНК и активировать транскрипцию (Zhang et al. 1993; Kaku et al. 2001), приводит к нарушению веретена деления (Gualberto et al. 1998).

Как видно из рисунка с выдачей программы, предсказано нарушение термодинамической стабильности белка p53 для большинства замен остатков в позициях 248, 273 и 281.

С помощью стандартных средств интернет-браузера пользователь может сохранить это предсказание, например, с целью накопления таких предсказаний для разных последовательностей белка и их последующего сравнительного анализа.

Оценка возможного влияния полиморфизмов (SNP) на функциональную активность белков.

Мутации в кодирующей последовательности гена могут изменять аминокислотную последовательность белковой молекулы, что нарушает ее структуру и физико-химические свойства. В итоге изменяется функциональная активность белка. На основе статистического анализа можно предсказать количественно изменение функциональной активности заданного белка при той или иной мутации. Некоторые мутантные аллели, соответствующие единичным нуклеотидным заменам, присутствуют в популяции постоянно, с относительно высокой частотой, то есть являются единичными нуклеотидными полиморфизмами. Некоторая часть этих полиморфизмов (3-5% у человека) приходится на кодирующие последовательности, что приводит полиморфизмам по функциональной активности кодируемых белков.

Проблема взаимосвязи структуры и функции биологических молекул всегда была центральной в молекулярной биологии. Для белков эта взаимосвязь наиболее интересна и,

по-видимому, наиболее сложна. Компьютерные методы анализа взаимосвязи структура-активность белков дают возможности для решения проблемы поиска сайтов в белковых последовательностях, аминокислоты которых определяют величину активности белков и позволяют строить количественные модели, связывающие значение активности с физико-химическими свойствами данных аминокислот. Такие модели служат необходимым компьютерным инструментом в руках исследователей для рационального планирования белково-инженерных экспериментов.

Из анализа научной литературы можно выделить два основных подхода, используемых для решения данной проблемы, каждый из которых имеет свои положительные стороны и недостатки. Изменения активностей белков в результате мутаций, как правило, сопряжено с конформационными изменениями белка, которые могут быть выявлены методами молекулярной динамики (Prevost M. et al., 1991). Следующий подход к анализу взаимосвязи структура-активность базируется на установлении статистических зависимостей между активностью белков и характеристиками сайтов, определяемых через физико-химические, структурные или эволюционные свойства аминокислот.

В ИЦИГ СОРАН активно развиваются подходы для этой актуальной области молекулярной биологии. Разработана программа SNP2WebProAnalyst, которая автоматически генерирует и проверяет гипотезы по связи между физико-химическими свойствами аминокислот участков в выровненных последовательностях белков и их активностями (Ivanisenko V.A. et al., 2005a, 2006).

Алгоритм компьютерного анализа количественной взаимосвязи структура-активность в семействах гомологичных и/или мутантных белков базируется на автоматической генерации и проверки гипотез по связи между физико-химическими свойствами аминокислотных участков в выровненных последовательностях белков и белковых активностей. Программа сканирует множественное выравнивание белков подвижной рамкой. Рамка задает сайты последовательностей белков для анализа. Для каждого сайта рассчитываются физико-химические свойства и строится регрессионная модель (множественный регрессионный анализ или нейронная сеть), связывающая активности белков с физико-химическими свойствами. Далее оценивается статистическая значимость модели. Построенная модель используется для предсказания активности мутантных или вновь секвенированных белков.

Данный метод позволяет планировать получение искусственных белков с заранее заданной функцией путем направленного мутагенеза и может применяться в молекулярной биологии. Также метод может быть применен в эволюционной биологии для исследования влияния предполагаемых мутаций на эволюцию белков.

Таким образом, данный метод позволяет установить взаимосвязь между вариабельностью генетического материала и фенотипическим разнообразием. Метод может применяться в популяционной биологии и биоинформатике для исследования полиморфизмов на уровне функциональной активности отдельных белков.

Описание работы SNP2WebProAnalyst

Оценка изменения функциональной активности мутантных белков с помощью программы SNP2WebProAnalyst возможно при использовании в качестве входных данных линейных символьных последовательностей белка в однобуквенном коде в Fasta – формате (www.ebi.ac.uk/fasta/), составленных из букв латинского алфавита {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y} и имеющих длину не менее 35 и не более 30000 символов. Так же в качестве входных данных необходимо указать количественные значения активностей белков.

Для подготовки к выполнению операции следует войти в подсистему «Полиморфизм» и выбрать функцию «Оценка возможного влияния полиморфизмов (SNP) на функциональную активность белков» в списке операций этой подсистемы. В результате появится HTML-интерфейс этой операции, который показан на рисунке ниже.

Системная биология-03

Предсказание активности мутантных белков

SNP2WebProAnalyst

Введите белковые последовательности в FASTA формате :

Введите активность :

Пример

Выполнить

Рисунок 35. HTML-интерфейс операции, подготовленной к началу ее выполнения.

Выполнение операции начинается с ввода символьных последовательностей белков, включая один анализируемый белок, в fasta-формате. Для этой цели используется раздел “Введите белковые последовательности в FASTA формате:” HTML-интерфейса операции.

Затем, используется окно «Введите активность:» HTML-интерфейс ввода количественных значений активностей белков в том же порядке, в котором были введены последовательности. При этом в качестве неизвестного количественного значения активности анализируемого белка, которое необходимо предсказать в результате выполнения настоящей операции, надо ввести символ «?» (вопросительный знака) как это показано на рисунке ниже.

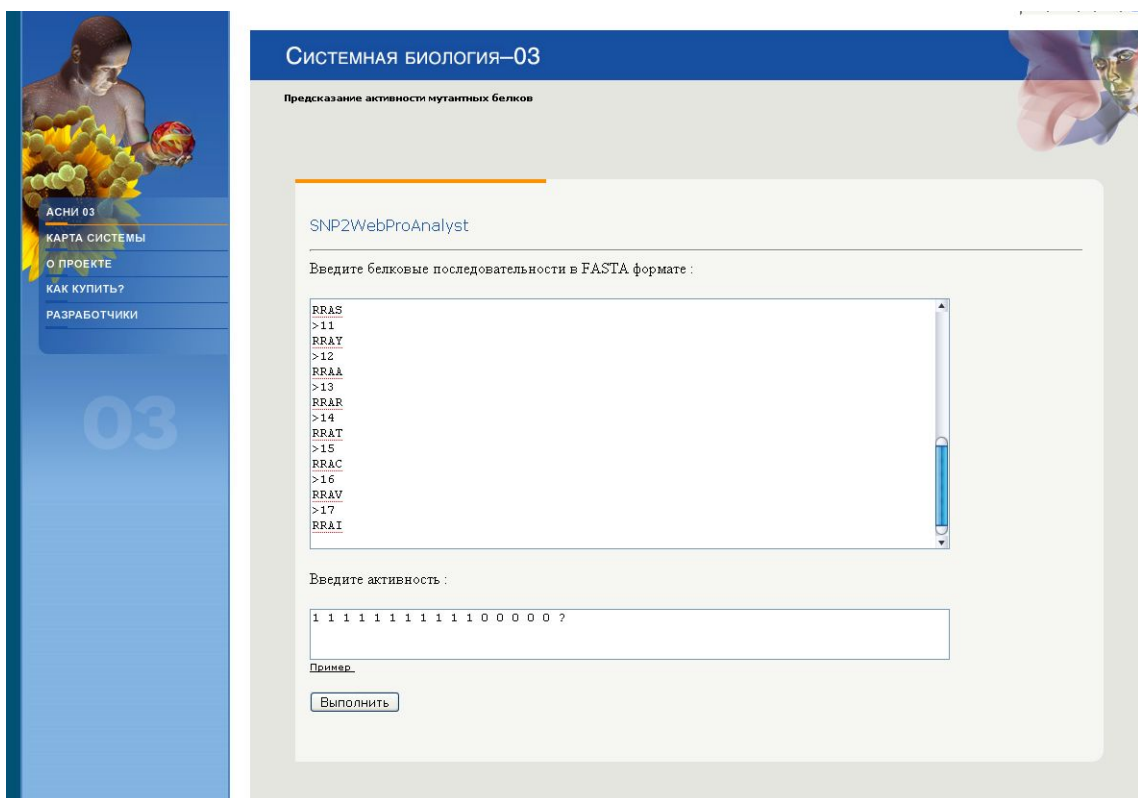


Рисунок 30. Ввод последовательностей белков и значений их активностей

После ввода данных как было описано на рисунке, выше необходимо нажать кнопку HTML-интерфейса “Выполнить”, в результате чего появляется HTML-страница с предсказанием значений активностей белков, как это показано ниже на рисунке.

```
Multiple linear regression analysis.
>1) Site 1-4 Multiple coefficient correlation R=0.148 F=0.145 P=0.866 SACC=1.000
Properties:
1) Hydrophobicity (Bogardt)
2) Volume (Bogardt)
[Seq name ] [Sequence] [Properties] [Activity measured Activity predicted]
[1 ] [NRAN] [ 12.63 39.38 ] [1.0000 0.7741]
[2 ] [REAN] [ 16.48 42.33 ] [1.0000 0.7395]
[3 ] [RTAN] [ 12.50 39.02 ] [1.0000 0.7752]
[4 ] [RWAN] [ 37.02 55.53 ] [1.0000 0.5542]
[5 ] [RRTN] [ 12.38 52.75 ] [1.0000 0.7789]
[6 ] [RRAD] [ 21.45 47.20 ] [1.0000 0.6950]
[7 ] [RRAG] [ 17.75 39.38 ] [1.0000 0.7273]
[8 ] [RRAH] [ 22.85 51.67 ] [1.0000 0.6830]
[9 ] [RRAL] [ 31.48 55.27 ] [1.0000 0.6049]
[10 ] [RRAS] [ 17.55 43.90 ] [1.0000 0.7300]
[11 ] [RRAY] [ 34.78 59.00 ] [1.0000 0.5754]
[12 ] [RRAA] [ 22.85 43.35 ] [0.0000 0.6814]
[13 ] [RRAE] [ 22.73 57.08 ] [0.0000 0.6852]
[14 ] [RRAT] [ 17.55 47.88 ] [0.0000 0.7307]
[15 ] [RRAC] [ 27.15 46.38 ] [0.0000 0.6427]
[16 ] [RRAV] [ 29.48 51.30 ] [0.0000 0.6224]
[17 ] [RRAI] [ 37.98 55.27 ] [ ? 0.5455]

Regression equation:
Y=-0.009*X1+ 0.000*X2+ 0.882
95% confidence intervals for coefficients of regression
K1 K2 K0
Lower -0.060 -0.060 -1.385
High 0.042 0.060 3.149
t-test statistics for H0
0.386 0.007 0.840
P-values for H0
0.545 0.936 0.376
F-test statistic for H0 = 0.145045
F-value for H0 = 0.866364
```

Рисунок 37. Пример HTML-страницы с предсказанием активностей белков. Вариант RRAI указан стрелкой.

В данном случае анализу подвергалась выборка из 17 фрагментов длиной 4 остатка от позиции 31 до позиции 34 аминокислотных последовательностей мутантных вариантов белка TNF человека, из которых 11 растворимы в клетках E.coli (активность установлена равной «1»), 5 нерастворимы (активность установлена равной «0»), и 1 вариант (RRAI) с неизвестной (в рамках этого контрольного примера) активностью, которую надо оценить с помощью настоящей операции для контроля правильности ее выполнения. На рисунке можно видеть, что для варианта RRAI (указан стрелкой) получена пониженная оценка растворимости (0.5455).

С помощью стандартных средств интернет-браузера пользователь может сохранить это предсказание, например, с целью накопления таких предсказаний для разных последовательностей белка и их последующего сравнительного анализа

Программные компоненты SNP2PROSITE и SNP2PDBSite: оценка возможного влияния полиморфизмов (SNP) на первичную и третичную структуру функциональных сайтов

Базы данных ProSite и PDBSite содержат информацию о первичной и третичной структуре функциональных сайтов белков. Аминокислотные замены, возникающие в белке, кодируемом мутантным аллелем, который соответствует единичному нуклеотидному полиморфизму (SNP), могут приходиться на остатки, составляющие функциональный сайт белка или его пространственное окружение. Из информации, содержащейся в базах ProSite и PDBSite можно определить, какие именно аминокислотные замены несовместимы с функционированием данного сайта. Таким образом, данный метод позволяет качественно оценивать влияние мутаций на функцию белка (Pintus et al, 2006). Метод применим в популяционной биологии и биоинформатике для исследования полиморфизмов, связанных с нарушением или существенным изменением функции белков. Метод реализован в виде двух программных компонент, работающих с первичной и третичной структурой функциональных сайтов, соответственно. Первая компонента использует базу данных PROSITE, другая – базу PDBSite (Ivanisenko et al, 2004, 2005b).

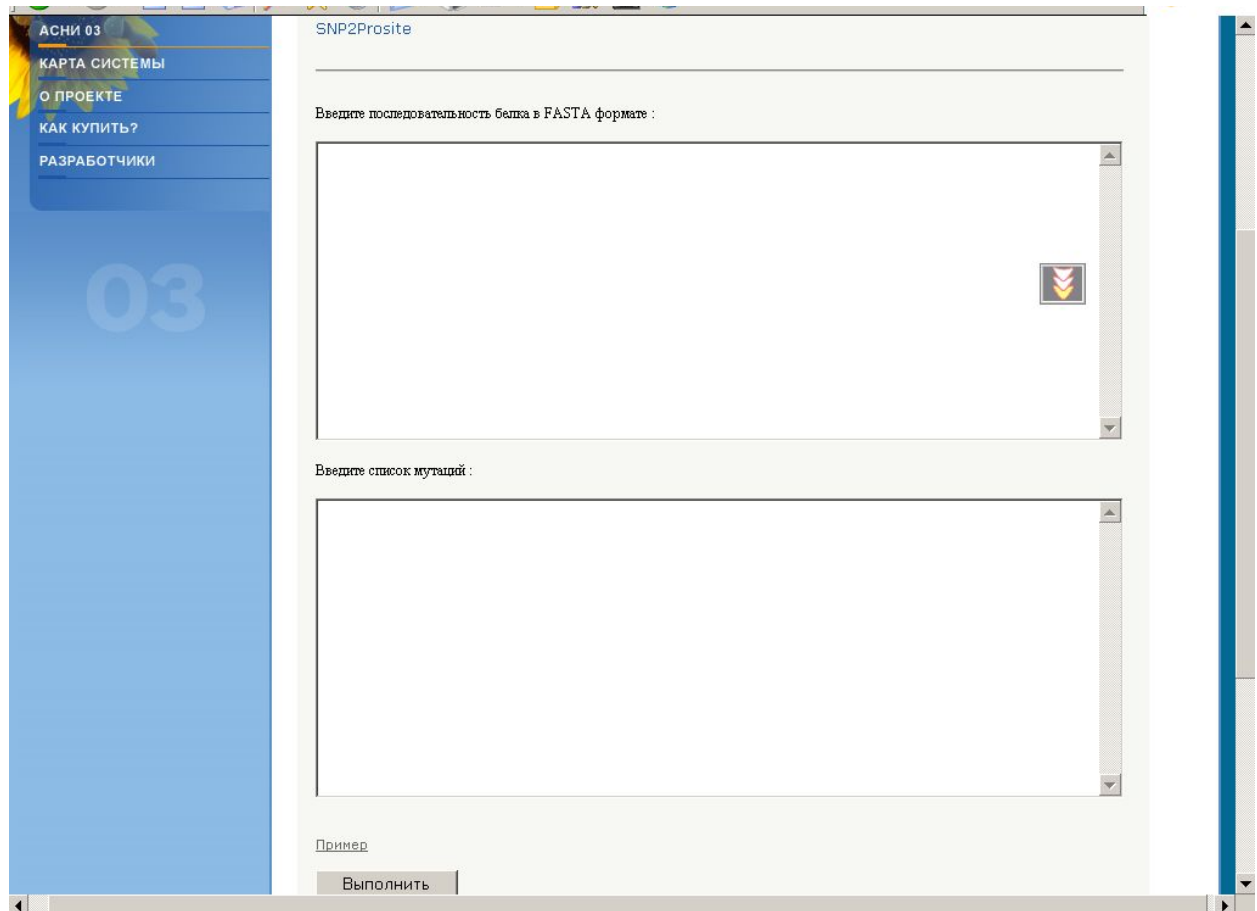
Программа, работающая с первичной структурой сайтов и использующая базу PROSITE, выбирает из списка аминокислотных замен, предоставленных исследователем, те, которые приводят к изменениям трех видов:

1. Изменение аминокислотной последовательности функционального сайта белка, несовместимое с его активностью.
2. Изменение аминокислотной последовательности функционального сайта белка, сохраняющее его активность.
3. Изменение аминокислотной последовательности района, не относящегося к области функционального сайта, приводящее к появлению последовательности, характерной для нового функционального сайта.
4. Программа, работающая с третичной структурой сайтов и использующая базу PDBSite, выбирает из списка аминокислотных замен, предоставленных исследователем, те, которые приводят к изменениям трех видов:
5. Изменение третичной структуры функционального сайта белка, несовместимое с его активностью.
6. Изменение третичной структуры функционального сайта белка, сохраняющее его активность.
7. Изменение третичной структуры района, не относящегося к области функционального сайта, приводящее к появлению структуры, характерной для нового функционального сайта.

Описание работы SNP2PROSITE

Исследование влияния мутаций на функциональные сайты белков посредством анализа их первичной структуры с помощью программы *SNP2PROSITE* возможно при использовании в качестве входных данных линейной символьной последовательности белка в однобуквенном коде, в Fasta – формате (www.ebi.ac.uk/fasta/), составленной из букв латинского алфавита {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y} и имеющей длину не менее 35 и не более 30000 символов. Кроме того, в качестве входных данных необходимо привести список анализируемых аминокислотных замен.

Для подготовки к выполнению операции надо войти в подсистему «Полиморфизм» и выбрать функцию «Оценка возможного влияния полиморфизмов (SNP) на структуру функциональных сайтов» в списке операций этой подсистемы. В результате появится HTML-интерфейс этой операции, как показано на рисунке ниже.



The screenshot displays the SNP2Prosite web interface. On the left is a blue navigation menu with the following items: 'АСНИ 03', 'КАРТА СИСТЕМЫ', 'О ПРОЕКТЕ', 'КАК КУПИТЬ?', and 'РАЗРАБОТЧИКИ'. The main content area is titled 'SNP2Prosite' and contains two large text input fields. The first field is labeled 'Введите последовательность белка в FASTA формате :'. The second field is labeled 'Введите список мутаций :'. Below the second field, there is a small 'Пример' (Example) link and a 'Выполнить' (Execute) button. The interface is clean and functional, designed for data entry.

Рисунок 38. HTML-интерфейс операции (начальное состояние).

Выполнение операции начинается с ввода символьной последовательности белка в fasta-формате. Для этой цели используется раздел “Введите последовательность белка в FASTA формате:” HTML-интерфейса операции. Также производится отдельный ввод анализируемых аминокислотных замен, разделенных символом пробела. Для этой цели используется раздел «Введите список мутаций:» HTML-интерфейса операции (смотри рисунок ниже).

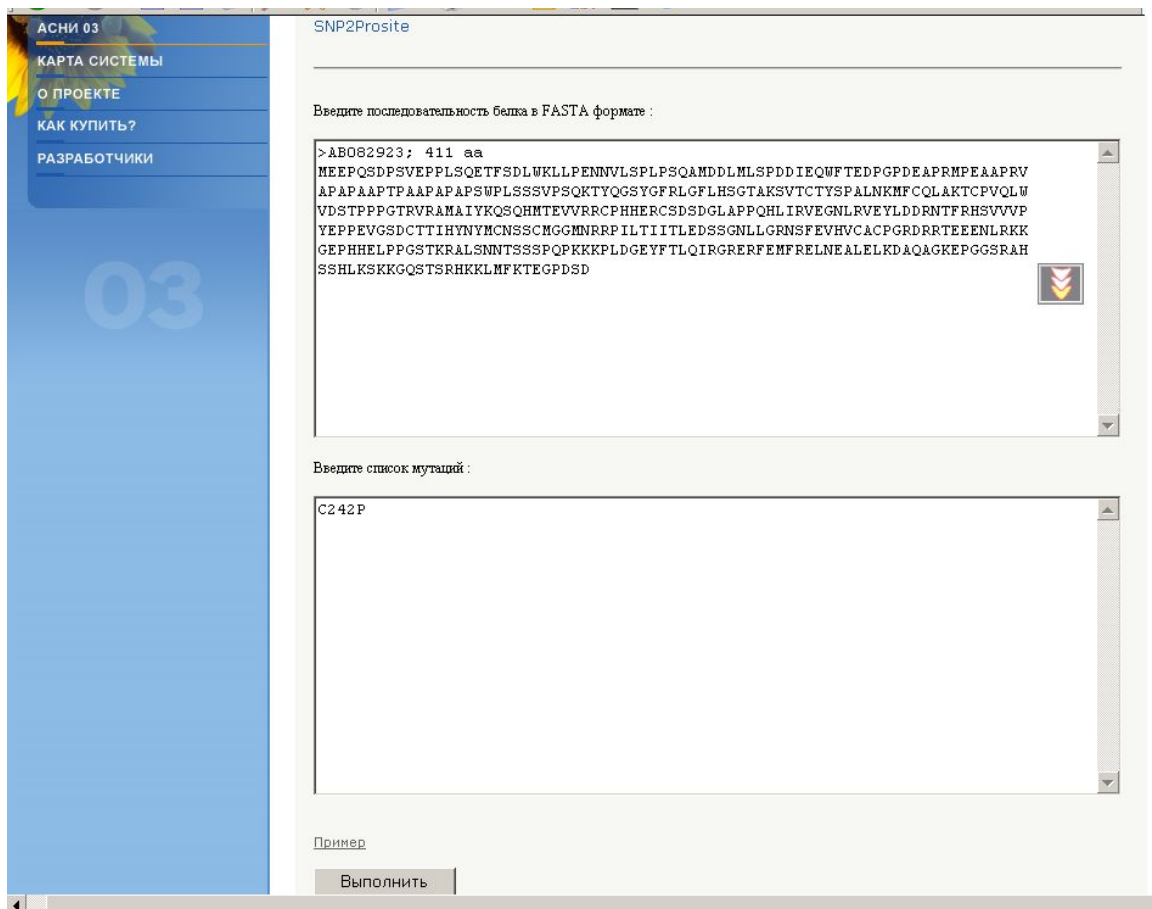


Рисунок 39. Ввод аминокислотной последовательности белка и списка анализируемых аминокислотных замен.

После ввода данных, необходимо нажать кнопку HTML-интерфейса “Выполнить” в результате чего появляется HTML-страница с предсказанием возникших, разрушенных или модифицированных функциональных сайтов в первичной структуре белка, как это показано ниже на рисунке.

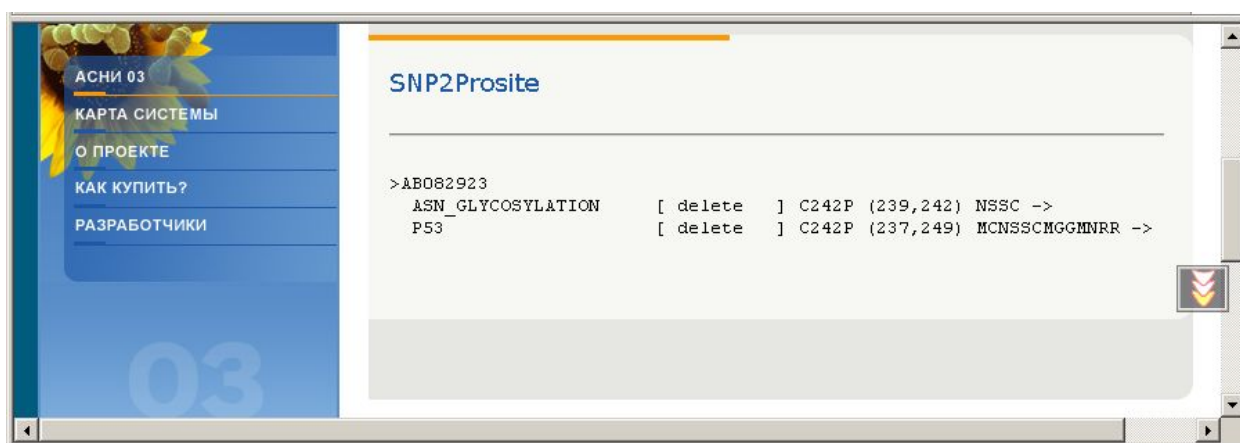


Рисунок 40. Пример HTML-страницы, содержащей возникшие (new), разрушенные (delete) или модифицированные (modified) функциональные сайты в первичной структуре белка.

В данном случае анализу подвергалась мутация С242Р в аминокислотной последовательности белка р53 человека. При этом критерием правильности выполнения операции является оценка указанной мутации в качестве «повреждающей функциональные сайты» этого белка, поскольку это известно экспериментально как для сайта гликозилирования 239-nssC-242, так и для сайта связывания иона цинка 237-mcnssCmggmrrg-249. На рисунке, содержащем выдачу программы, можно видеть, что влияние контрольной замены С242Р на белок р53 оценено как исчезновение (delete) и сайта гликозилирования (239-nssC-242), и сайта связывания иона цинка – 237-mcnssCmggmrrg-249.

С помощью стандартных средств интернет - браузера пользователь может сохранить это предсказание, например, для накопления таких предсказаний для разных мутаций белка для их последующего сравнительного анализа.

Описание работы [SNP2PDBSITE](#)

Исследование влияния мутаций на функциональные сайты белков посредством анализа их третичной структуры с помощью программы [SNP2PDBSITE](#) является возможным при использовании в качестве входных данных идентификатора третичной структуры исследуемого белка в базе данных Protein Data Bank (PDB), идентификатора полипептидной цепи, а также списка мутаций в этом белке, которые надо проанализировать с помощью настоящей операции.

Для подготовки к выполнению операции надо войти в подсистему «Полиморфизм» и выбрать функцию «Оценка возможного влияния полиморфизмов (SNP) на третичную структуру функциональных сайтов белков» в списке операций этой подсистемы. В результате появится HTML- интерфейс этой операции, как показано на рисунке ниже.

СТАТ.ЭПИДЕМИОЛ. ПОЛИМОРФИЗМ ГЕН.ИДЕНТ.ЛИЧНОСТИ ПОМОЩЬ

Выбрать программу

СИСТЕМНАЯ БИОЛОГИЯ-03

Предсказание эффекта мутации на функциональные сайты белка по пространственной структуре

SNP2PDBSite

Введите идентификатор белка в базе PDB :

Введите идентификатор цепи в файле PDB :

Например :

A 110	R -> C
A 110	R -> L
A 113	F -> C

Введите мутации :

[Пример](#)

Рисунок 41. HTML-интерфейс операции, подготовленной к началу ее выполнения.

Выполнение операции начинается с ввода идентификатора структуры белка и идентификатора полипептидной цепи. Для этой цели, соответственно, используются разделы «Введите идентификатор белка в базе PDB:» и «Введите идентификатор цепи в файле PDB:» HTML-интерфейса операции, как это показано на рисунке ниже.

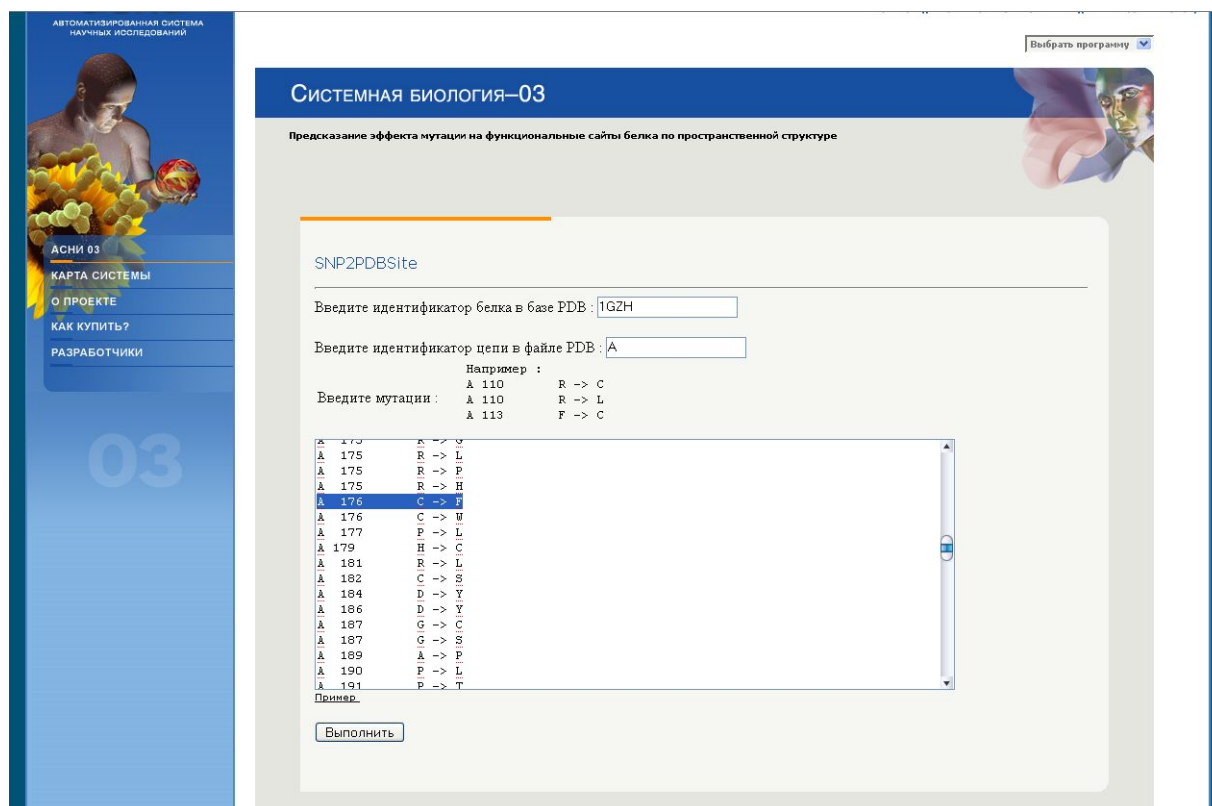


Рисунок 42. Ввод идентификаторов структуры белка и полипептидной цепи, а также списка анализируемых аминокислотных замен.

Затем с помощью окна «Введите мутации:» вводится список анализируемых аминокислотных замен, образец формата которого показан на интерфейсе непосредственно над окном ввода этого списка (смотри рисунок выше).

После ввода данных, необходимо нажать кнопку HTML-интерфейса “Выполнить” в результате чего появляется HTML-страница с предсказанием возникших, разрушенных или модифицированных функциональных сайтов в третичной структуре белка, как это показано ниже на рисунке.

```

SNP A 176 C->F
Altered site:
PDBSite ID 2BIQAC1
  Site type: SITE_TYPE METAL_BINDING
  Site description: SITE_DESCR ZN BINDING SITE FOR CHAIN A
  Site of the query protein: 176C 179H 238C 242C
  
```

Рисунок 43. Пример HTML-страницы , содержащей возникших (arisen), разрушенных (deleted) или модифицированных (altered) функциональных сайтов в третичной структуре белка.

В данном случае анализу подвергалась третичная структура белка p53 человека, замена C176F в котором встречается во многих типах опухолевых тканей. Для этой мутации экспериментально известно, что она искажает нормальный сайт связывания иона цинка.

Нормальная работа этого сайта необходима для своевременной задержки деления клетки и проведения репарации в случае повреждений ДНК. Мутантная форма С176F белка р53 неспособна противодействовать делению клетки с повреждениями ДНК, что может приводить к возникновению опухолей.

Рисунок, на котором представлен фрагмент выдачи программы, показывает, что замена С176F приводит к искажению сайта связывания иона цинка.

С помощью стандартных средств интернет - браузера пользователь может сохранить это предсказание, например, для накопления таких предсказаний для разных мутаций белка для их последующего сравнительного анализа.

3. Полезные ссылки.

- 1) Варфоломеев С.Д., Гуревич К.Г. Биокинетика. 1999. P. 85-100.
- 2) Кочетов А. В., Шумный В. К. Влияние структуры мРНК на процесс инициации трансляции в клетках растений // Успехи совр. биол. 1998. V. 118. P. 754-770.
- 3) Ленинджер А. Биохимия. 1974. P. 174-197.
- 4) Строев Е.А. Биологическая химия. 1986. P. 122-163.
- 5) Agalioti,T., Lomvardas,S., Parekh,B., Yie,J., Maniatis,T., Thanos,D. Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter.//Cell. 2000. V 103. P. 667-678.
- 6) Benson D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J. and Wheeler D.L. GenBank. // Nucl. Acids Res. 2006. V. 34. Database issue. P. D16-D20.
- 7) Biemelt S., Sonnewald U., Galmbacher P., Willmitzer L. et al. Production of human papillomavirus type 16 virus-like particles in transgenic plants //J. Virol. 2003 V. 77. P. 9211-9220.
- 8) Brendel V., Kleffe J. Prediction of locally optimal splice sites in plant pre-mRNA with applications to gene identification in Arabidopsis thaliana genomic DNA// Nucleic Acids Res. 1998. V. 26. P. 4748-4757.
- 9) Boccia LM, Lillicrap D, Newcombe K, and Mueller CR. Binding of the Ets factor GA-binding protein to an upstream site in the factor 9 promoter is a critical event in transactivation. // Mol. Cell. Biol. 1996. V. 16. P. 1929-1935.
- 10) Brookes A, Lehvaeslaiho H, Siegfried M, Boem J, Yuan Y, Sarkar C, Bork P, and Ortigao F. HGBASE: a database of SNPs and other variations in and around human genes. // Nucleic Acids Res. 2000. V. 28. P. 356-360.
- 11) Capriotti E, Fariselli P, Casadio R. A neural-network-based method for predicting protein stability changes upon single point mutations.// Bioinformatics. 2004. V. 20. SI63-I68.
- 12) Chiapello H., Lisacek F., Caboche M., Henaut A. Codon usage and gene function are related in sequences of Arabidopsis thaliana // Gene. 1998. V. 209. GC1-GC38.
- 13) Cochrane G. et al. EMBL Nucleotide Sequence Database: developments in 2005. // Nucl. Acids Res. 2006. V. 34. Database issue. P. D10-D15.
- 14) DellaPenna D. Plant metabolic engineering // Plant Physiol. 2001. V. 125. P. 160-163.
- 15) Diehn S. H., Chiu W-L., De Rocher E. J., Green P. J. Premature polyadenylation at multiple sites within a Bacillus thuringiensis toxin gene-coding region // Plant Physiol. 1998. V. 117. P. 1433-1443.
- 16) Duret L., Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis // Proc. Natl. Acad. Sci. U S A. 1999. V. 96. P. 4482-4487.

- 17) Gall J., Bellini M., Wu Z. and Murphy C. Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. // *Mol. Biol. Cell.* 1999. V. 10. P. 4385-4402.
- 18) Gallie D. R. Controlling gene expression in transgenics // *Curr. Opin. Plant Biol.* 1998. V. 1. P. 166-172.
- 19) Gillio A.P., Verlander P.C., Batish S.D., Giampietro P.F. and Auerbach A.D. Phenotypic consequences of mutations in the Fanconi anemia FAC gene: an International Fanconi Anemia Registry study.// *Blood.* 1997. V. 90. P. 105-110.
- 20) Grec S., Wang Y., Le Guen L. et al. Cryptic polyadenylation sites within the coding sequence of three yeast genes expressed in tobacco // *Gene.* 2000. V. 242. P. 87-95.
- 21) Gualberto et al. // *Proc. Natl. Acad. Sci. USA.* 1998. V. 95. N.9. P. 5166-5171.
- 22) Gutierrez R. A., Macintosh G. C., Green P. J. Current perspectives on mRNA stability in plants: multiple levels and mechanisms of control // *Trends Plant Sci.* 1999. V. 4. P. 429-438.
- 23) Ivanisenko VA, Pintus SS, Grigorovich DA, Kolchanov NA. PDBSiteScan: a program for searching for active, binding and posttranslational modification sites in the 3D structures of proteins.//*Nucleic Acids Res.* 2004. V. 32. P. W549-54.
- 24) Ivanisenko VA, Eroshkin AM, Kolchanov NA. WebProAnalyst: an interactive tool for analysis of quantitative structure-activity relationships in protein families.// *Nucleic Acids Res.* 2005a. V. 33. P. W99-104.
- 25) Ivanisenko VA, Pintus SS, Grigorovich DA, Kolchanov NA. PDBSite: a database of the 3D structure of protein functional sites. // *Nucleic Acids Res.* 2005b. V. 33. P. D183-7
- 26) Ivanisenko V.A., S.S. Pintus, P.S. Demenkov, M.A. Krestyanova, E.K. Litvenko, D.A. Grigorovich, V.A. Debelov Fastprot: a Computational Workbench for Recognition of the Structural and Functional Determinants in Protein Tertiary Structures.// *Bioinformatics of Genome Regulation and Structure II.* (Eds. N.Kolchanov, R. Hofstaedt, L.Milanesi) Springer Science+Business Media, Inc. 2006, P. 305-316.
- 27) Jacobs G. H., Rackham O., Stockwell P. A., et al. Transterm: a database of mRNAs and translational control elements // *Nucleic Acids Res.* 2002. V. 30. P. 310-311.
- 28) Haseloff J., Siemering K. S., Prasher D. C., Hodge S. Removal of cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic Arabidopsis plants brightly // *Proc. Natl. Acad. Sci. USA.* 1997. V. 94. P. 2122-2127.
- 29) Higo K., Ugawa Y., Iwamoto M., Korenaga T. Plant cis-acting regulatory DNA elements (PLACE) database:1999 // *Nucleic Acids Res.* 1999. V. 27. P. 297-300.
- 30) Hamosh A, Scott A, Amberger J, Valle D, and McKusick V. Online Mendelian Inheritance in Man (OMIM). // *Hum. Mutat.* 2000. V. 15. P. 57-61.
- 31) Kaku et al. // *Biochem Biophys Res Commun.* 2001. V.280. N.1. P. 204-211.
- 32) Kanehisa M., Goto S., Kawashima S., Nakaya A. The KEGG databases at GenomeNet // *Nucleic Acids Res.* 2002. V. 30. P. 42-46.
- 33) Khanna H. K., Raina S. K. Elite Indica transgenic rice plants expressing modified Cry1Ac endotoxin of *Bacillus thuringiensis* show enhanced resistance to yellow stem borer (*Scirpophaga incertulas*) // *Transgenic Res.* 2002. V. 11. P. 411-423.
- 34) Knight J.C., Udalova I., Hill A.V., Greenwood B.M., Peshu N., Marsh K. and Kwiatkowski D. A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria. // *Nat. Genetics.* 1999. V. 22. P. 145-150.
- 35) Kochetov A. V., Grigorovich D., Kolchanov N. A., Sarai A. Database on mRNA located eukaryotic expression signals influencing translation efficiency and

- specificity // *Genome Informatics*. Ser. 12 / Edt: Matsuda et al. Universal Academy Press: Tokyo, 2001. P. 492-493.
- 36) Kochetov A. V., Ischenko I. V., Vorobiev D. G., Kel A. E. et al. Eukaryotic mRNAs encoding abundant and scarce proteins are statistically dissimilar in many structural features // *FEBS Lett.* 1998. V. 440. P. 351-355.
 - 37) Kochetov A. V., Ponomarenko M. P., Kisselev L. L., Kolchanov N. A. Prediction of eukaryotic mRNA translational properties // *Bioinformatics*. 1999. V. 15. P. 704-712.
 - 38) Kolchanov N. A., Ignatieva E. V., Ananko E. A. et al. Transcription Regulatory Regions Database (TRRD): its status in 2002. // *Nucleic Acids Res.* 2002a. V. 30. P. 312-317.
 - 39) Komar A. A, Lesnik T., Reiss C. Synonymous codon substitutions affect ribosome traffic and protein folding during in vitro translation // *FEBS Lett.* 1999. V. 462. P. 387-391.
 - 40) Kozak M. Pushing the limits of the scanning mechanism for initiation of translation // *Gene*. 2002. V. 299. P. 1-34.
 - 41) Krawczak M, Ball E, Fenton I, Stenson P, Abeyasinghe S, Thomas N, and Cooper D. Human gene mutation database-a biomedical information and research resource. // *Hum. Mutat.* 2000. V. 15. P. 45-51.
 - 42) Langdon S. and Kaufman R. Gamma-globin gene promoter elements required for interaction with globin enhancers. // *Blood*. 1998. V. 91. P. 309-318.
 - 43) Larrick J. W., Thomas D. W. Producing proteins in transgenic plants and animals // *Curr. Opin. Biotechnol.* 2001. V. 12. P. 411-418.
 - 44) Le Hir H., Nott A., Moore M. J. How introns influence and enhance eukaryotic gene expression // *Trends Biochem. Sci.* 2003. V. 28 P. 215-220.
 - 45) Lescot M., Dehais P., Thijs G. et al.. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences // *Nucleic Acids Res.* 2002. V. 30. P. 325-327.
 - 46) Lukaszewicz M., Feuermann M., Jerouville B., Stas A., Boutry M. *In vivo* evaluation of the context sequence of the translation initiation codon in plants // *Plant Sci.* 2000. V. 154. P. 89-98.
 - 47) Marth G., Yeh R., Minton M., Donaldson R., Li Q., Duan S., Davenport R., Miller R. and Kwok P.-Y. Single-nucleotide polymorphisms in the public domain: how useful are they? // *Nat. Genetics*. 2001. V. 27. P. 371-372.
 - 48) Matsuda M., Sakamoto N. and Fukumaki Y. Delta-thalassemia caused by disruption of the site for an erythroid-specific transcription factor, GATA-1, in the delta-globin gene promoter. // *Blood*. 1992. V. 80. P. 1347-1351.
 - 49) McCarthy J. E. G. Posttranscriptional control of gene expression in yeast // *Microbiol Mol. Biol. Rev.* 1998. V. 62. P. 1492-1553.
 - 50) Mills R.E., Luttig C.T., Larkins C.E., Beauchamp A., Tsui C., Pittard W.S., Devine S.E. An initial map of insertion and deletion (INDEL) variation in the human genome. // *Genome Res.* 2006, 16(9), 1182-90.
 - 51) Moi P., Loudianos G., Lavinha J., Murru S., Cossu P., Casu R., Oggiano L., Longinotti M., Cao A. and Pirastu M. Delta-thalassemia due to a mutation in an erythroid-specific binding protein sequence 3' to the delta-globin gene. // *Blood*. 1992. V. 79. P. 512-516.
 - 52) Mueller L.A., Zhang P., Rhee S.Y. AraCyc. A Biochemical Pathway Database for Arabidopsis // *Plant Physiol.* 2003. V. 132. P. 453-460.
 - 53) O'Neill D., Kaysen J., Donovan-Peluso M., Castle M. and Bank A. Protein-DNA interactions upstream from the Human A gamma globin gene. // *Nucleic Acids Res.* 1990. V. 18. P. 1977-1982.

- 54) Osier M, Cheung K-H, Kidd J, Pakstis A, Miller P, and Kidd K. ALFRED: an allele frequency database for diverse populations and DNA polymorphisms -- an update. // *Nucleic Acids Res.* 2001. V. 29. P. 317-319.
- 55) Pesole G., Liuni S., Grillo G. et al. UTRdb and UTRsite: specialized databases of sequences and functional elements of 5' and 3' untranslated regions of eukaryotic mRNAs. Update 2002 // *Nucleic Acids Res.* 2002. V. 30. P. 335-340.
- 56) Pestova T. V., Kolupaeva V. G., Lomakin I. B. et al. Molecular mechanisms of translation initiation in eukaryotes // *Proc. Natl. Acad. Sci. U S A.* 2001. V. 98. P. 7029-7036.
- 57) Piedrafita F.J., Molander R.B., Vansant G., Orlova E.A., Pfahl M. and Reynolds W.F. An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element. // *J. Biol. Chem.* 1996. V. 271. P. 14412-14420.
- 58) Pintus SS, Fomin ES, Ivanisenko VA, Kolchanov NA. Phylogenetic analysis of the family of p53. // *Biofizika.* 2006 V. 51. N. 4. P.:640-9.
- 59) Ponomarenko J.V., Orlova G.V., Merkulova T.I., Gorshkova E.V., Fokin O.N., Vasiliev G.V., Frolov A.S. and Ponomarenko M.P. rSNP_Guide: an integrated database-tools system for studying SNPs and site-directed mutations in transcription factor binding sites. // *Hum Mutat.* 2002. V. 20. N. 4. P. 239-248.
- 60) Prevost M, Wodak SJ, Tidor B, Karplus M. Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile-96-Ala mutation in barnase. // *Proc Natl Acad Sci U S A.* 1991. V. 88. N. 23. P. 10880-4
- 61) Rouwendal G. J. A., Mendes O., Wolbert E. J. H., de Boer A. D. Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage // *Plant Mol. Biol.* 1997. V. 33. P. 989-999.
- 62) Sachidanandam R. et al. The International SNP Map Working Group. A map of human genome sequence variation containing 1,42 million single nucleotide polymorphisms. // *Nature.* 2001. V. 409. P. 928-933.
- 63) Sawant S. V., Kiran K., Singh P. K., Tuli R. Sequence architecture downstream of the initiator codon enhances gene expression and protein stability in plants // *Plant Physiol.* 2001. V. 126. P. 1630-1636.
- 64) Schacherer F., Choi C., Gotze U. et al. The TRANSPATH signal transduction database: a knowledge base on signal transduction networks // *Bioinformatics.* 2001. V. 17. P. 1053-1057.
- 65) Sherry S, Ward M, Kholodov M, Baker J, Phan L, Smigielski EM, and Sirotkin K. dbSNP: the NCBI database of genetic variation. // *Nucleic Acids Res.* 2001. V. 29. P. 308-311.
- 66) Tolstrup N., Rouze P., Brunak S. A branch point consensus from Arabidopsis found by non-circular analysis allows for better prediction of acceptor sites // *Nucleic Acids Res.* 1997. V. 25. P. 3159-3163.
- 67) Tsutsumi-Ishii Y., Tadokoro K., Hanaoka F. and Tsuchida N. Response of heat shock element within of human HSP70 promoter of mutated p53 genes. // *Cell Growth and Differentiation.* 1995. V. 6. P. 1-8.
- 68) Wingender E., Chen X., Fricke E. et al. The TRANSFAC system on gene expression regulation // *Nucleic Acids Res.* 2001. V. 29. P. 281-283.
- 69) van Hoof A., Green P. J. Rare codons are not sufficient to destabilize a reporter gene transcript in tobacco // *Plant Mol. Biol.* 1997. V. 35. P. 383-387.
- 70) Vasil'ev G, Merkulov V, Kobzev V, Merkulova T, Ponomarenko M, and Kolchanov N. Point mutations within 663-666 bp of intron 6 of the human TDO2 gene, associated

with a number of psychiatric disorders, damage the YY-1 transcription factor binding signal. // FEBS Lett. 1999. V. 462. P. 85-88.

71) Zhang et al. // Oncogene. 1993. V.8. N.9. P. 2555-9255