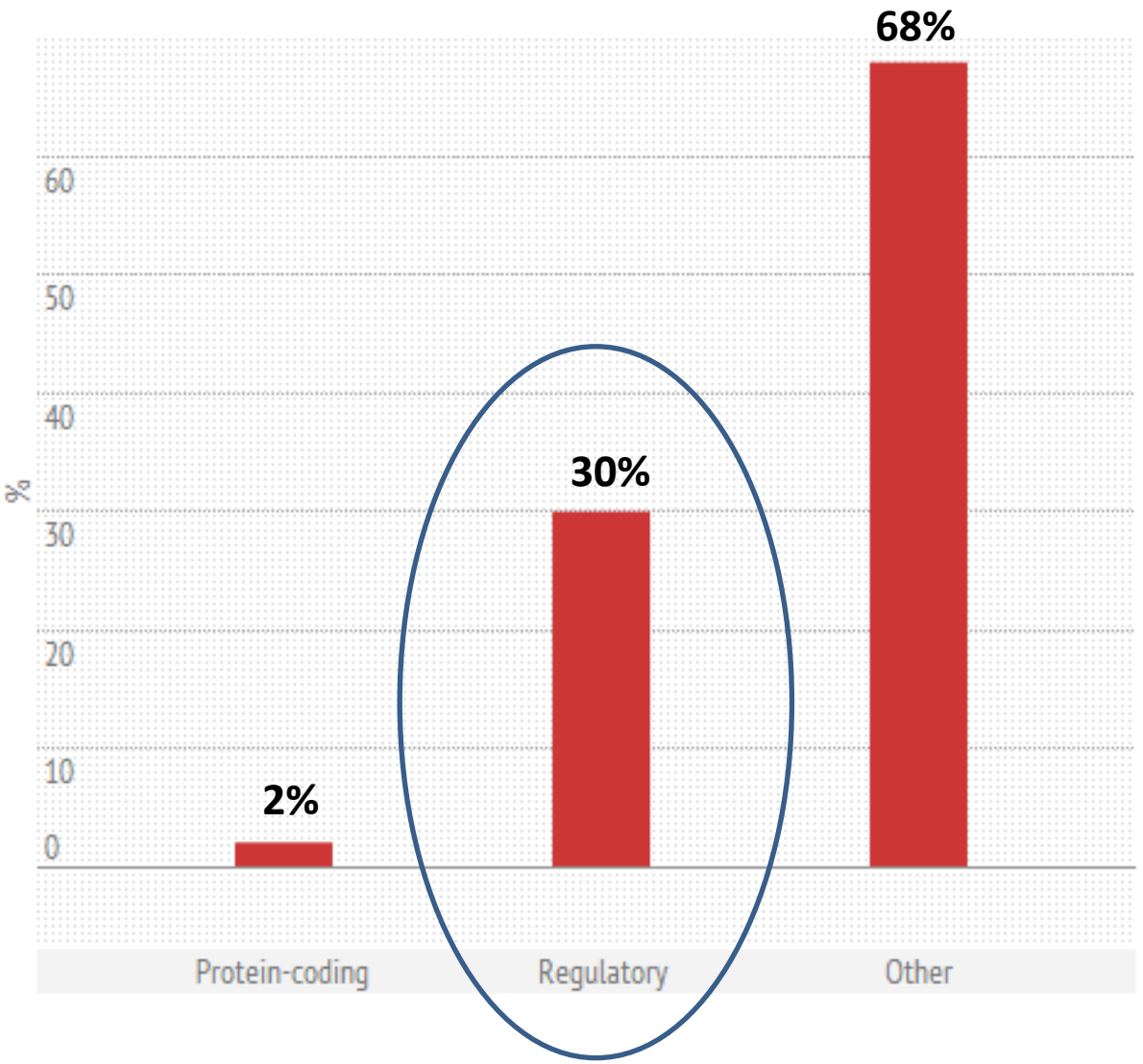
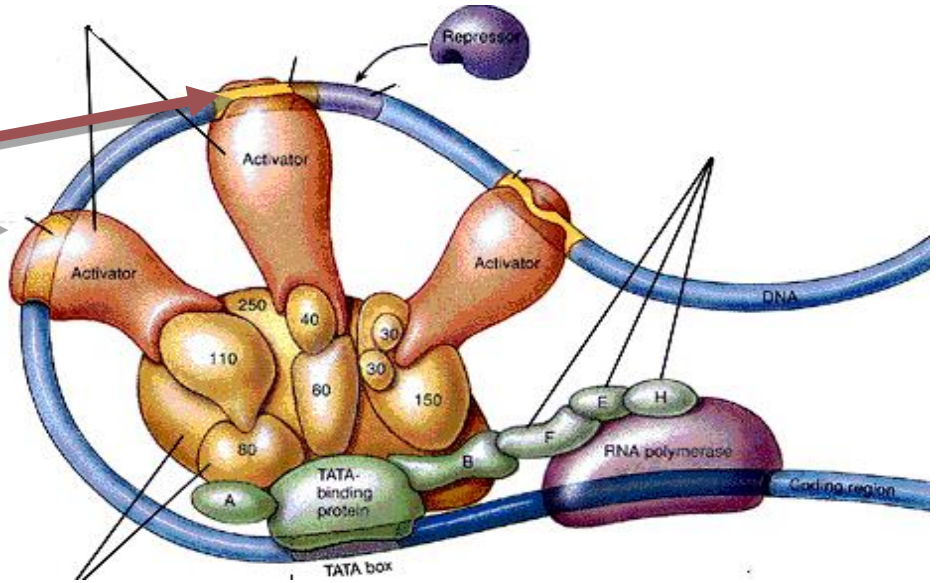
РЕАЛИЗАЦИЯ СРЕДСТВ ИЕРАРХИЧЕСКОГО АНАЛИЗА РЕГУЛЯТОРНЫХ РАЙОНОВ ГЕНОВ ДЛЯ ИНТЕГРИРОВАННОЙ СИСТЕМЫ EXPERT DISCOVERY И UGENE

Юрий Васькин

Геном человека



Сайты связывания

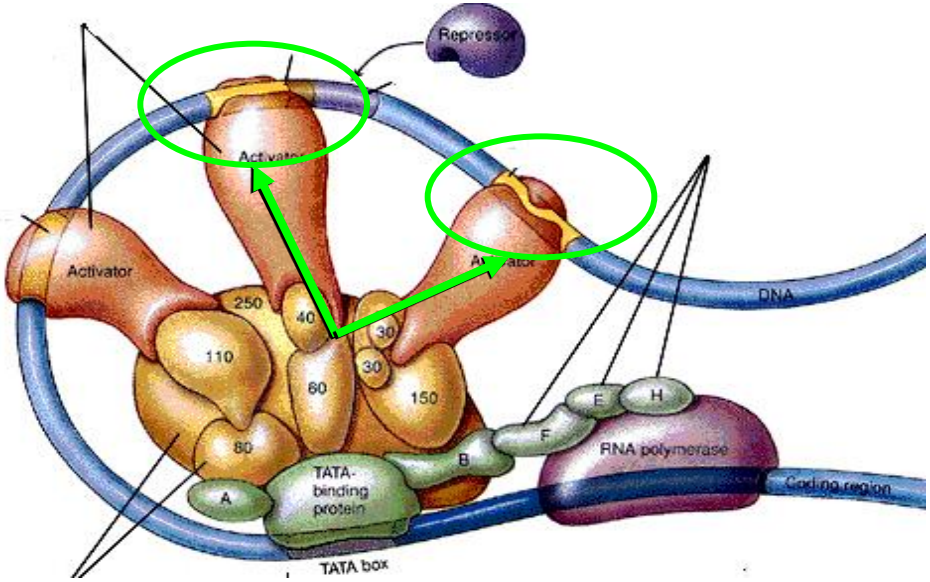


Ген

Проблема: автоматическое распознавание регуляторных областей

Актуальность: знание структуры = контроль работы гена

Существующие методы распознавания сайтов: Weight Matrix, SITECON, Motif Finders, etc...



Недостаток: не учитывают взаиморасположение

Цель: разработка системы для иерархического анализа регуляторных областей

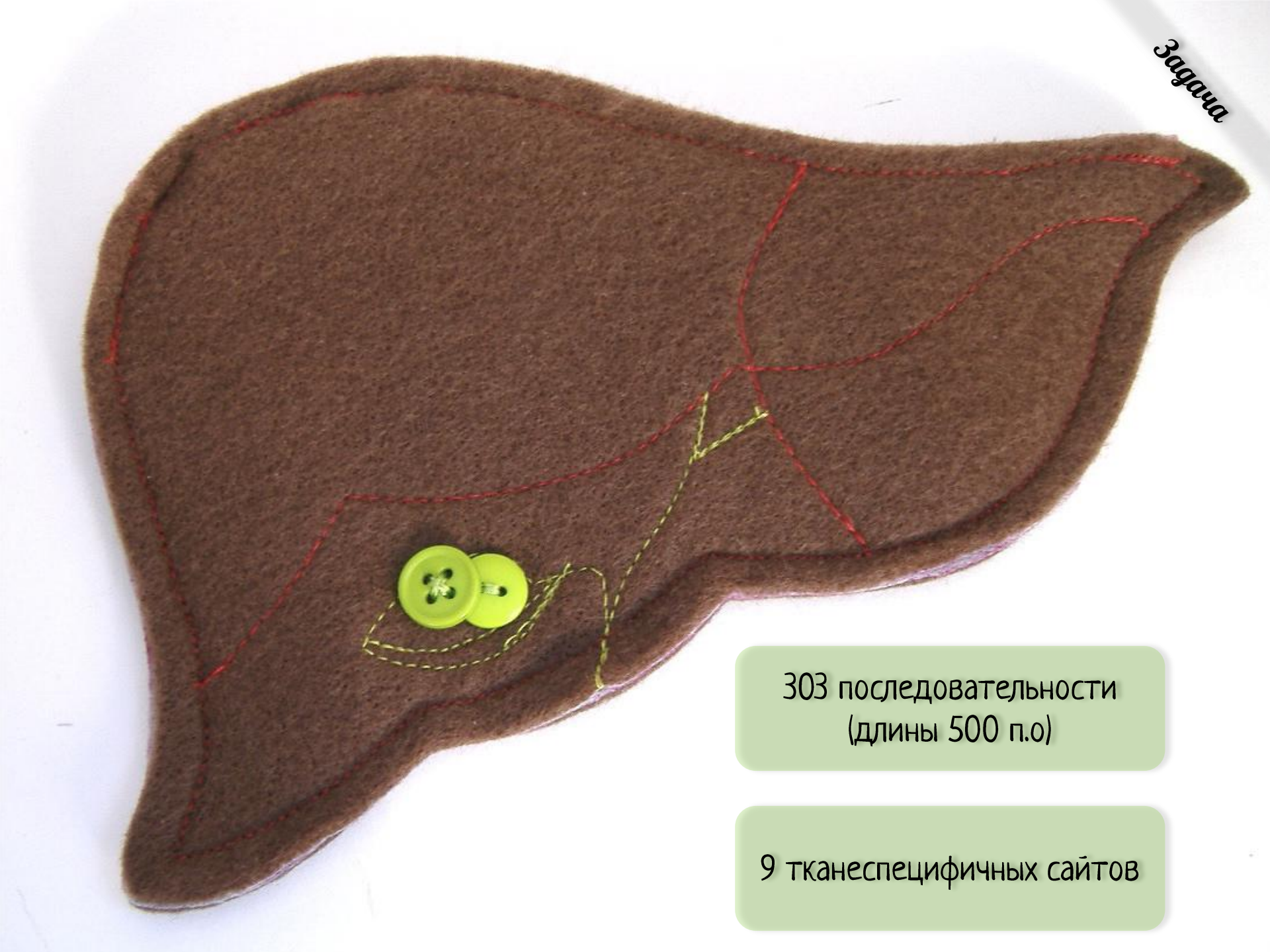
Структура и требования

Автоматическое построение модели

Распознавание

Визуализация и интерфейс

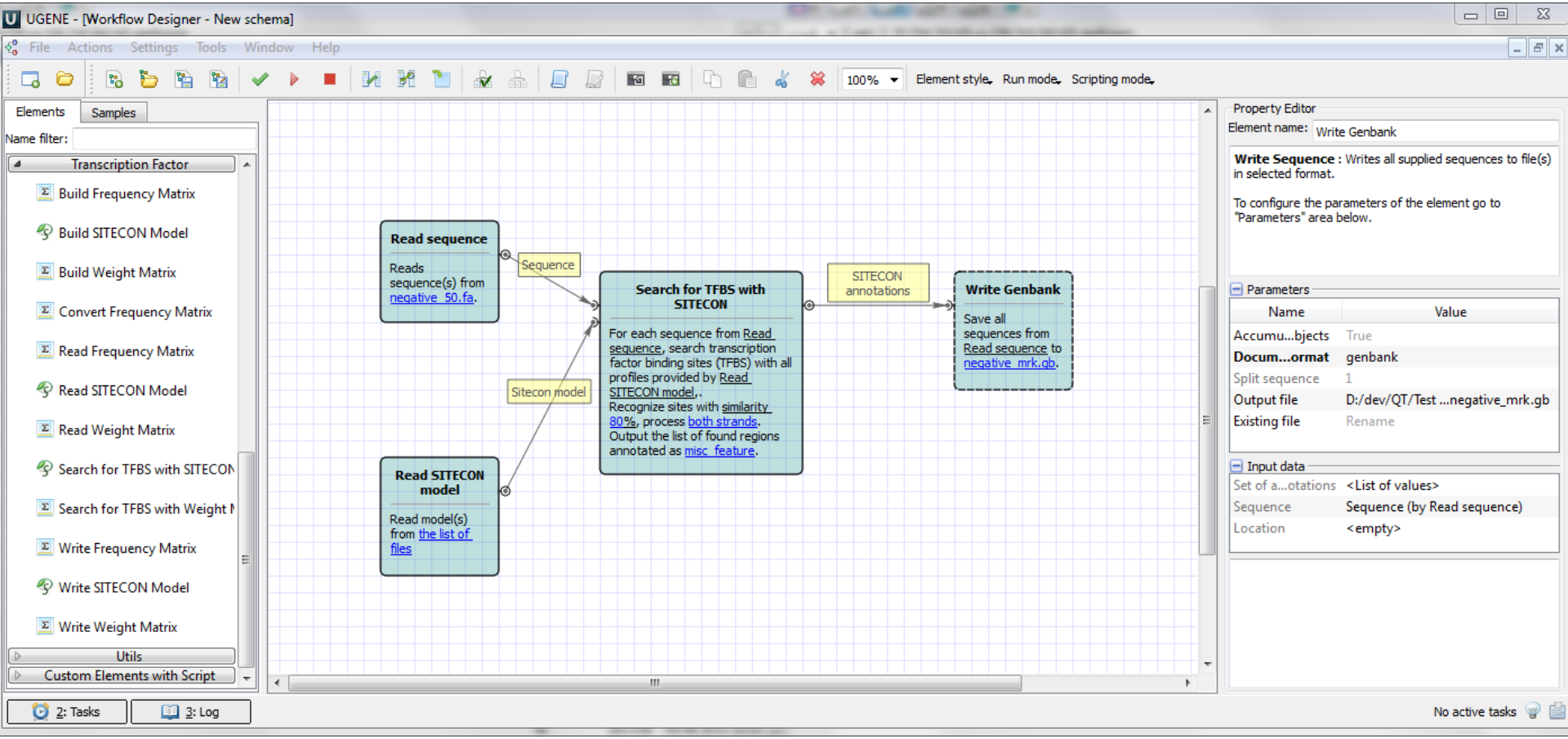
Биологические результаты



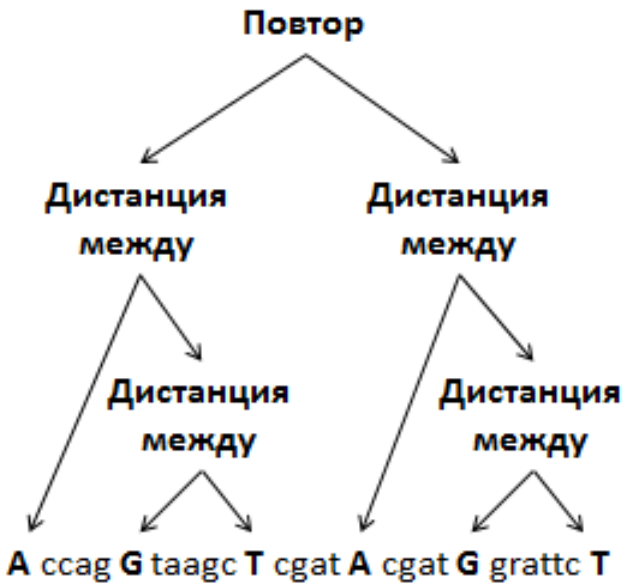
303 последовательности
(длины 500 п.о)

9 тканеспецифичных сайтов

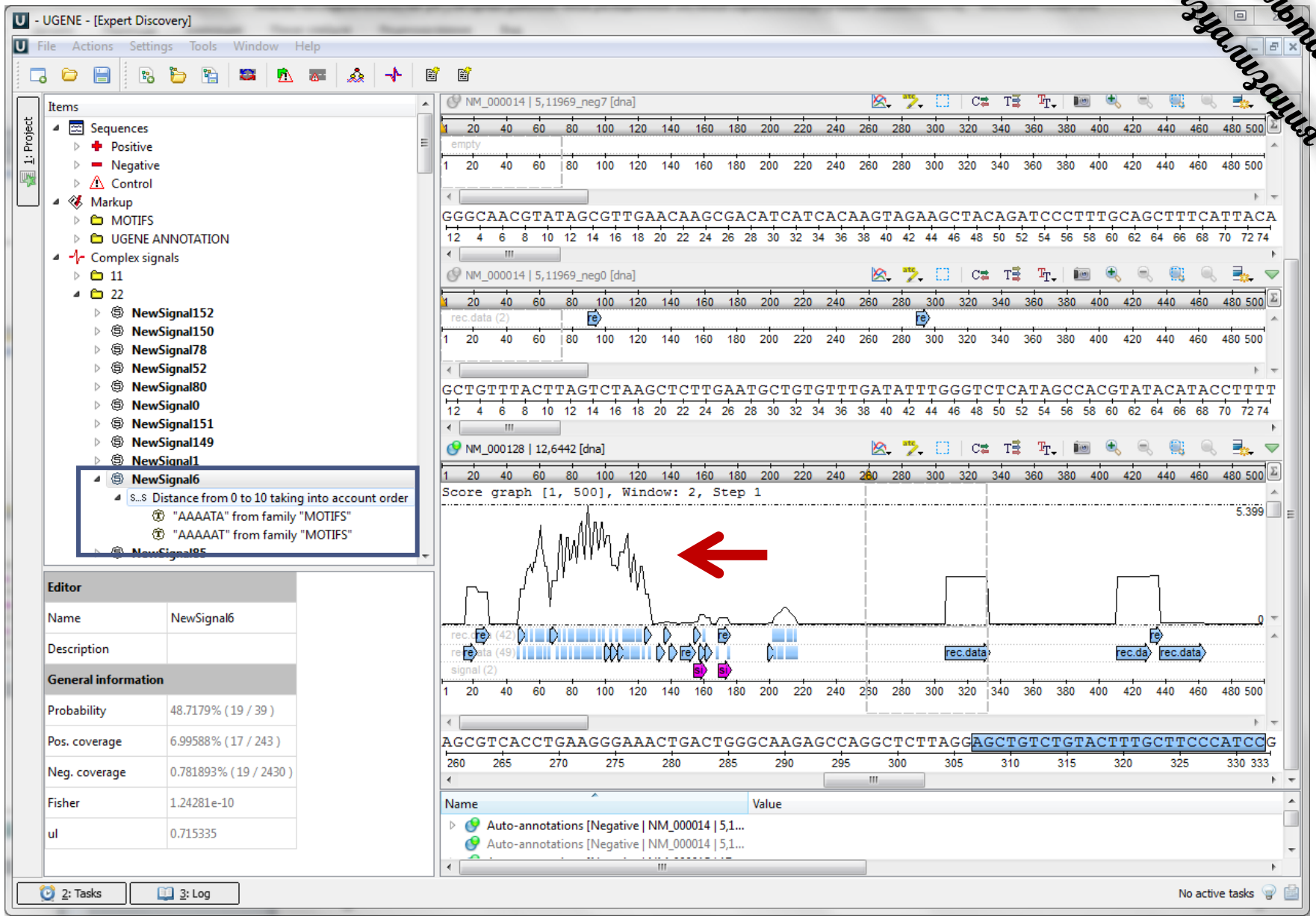
UGENE
Интегрированные инструменты молекулярного биолога



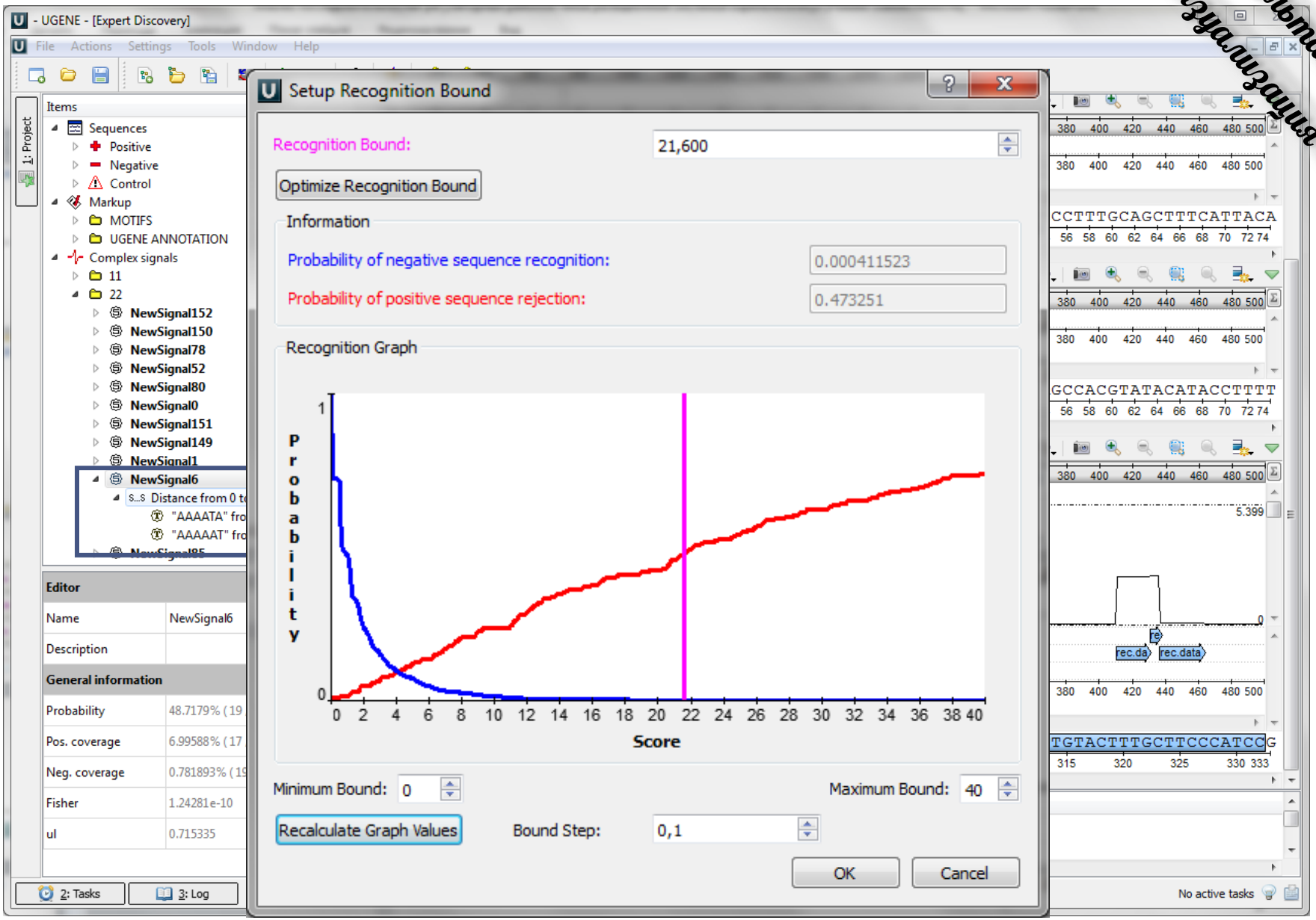
Комплексный сигнал



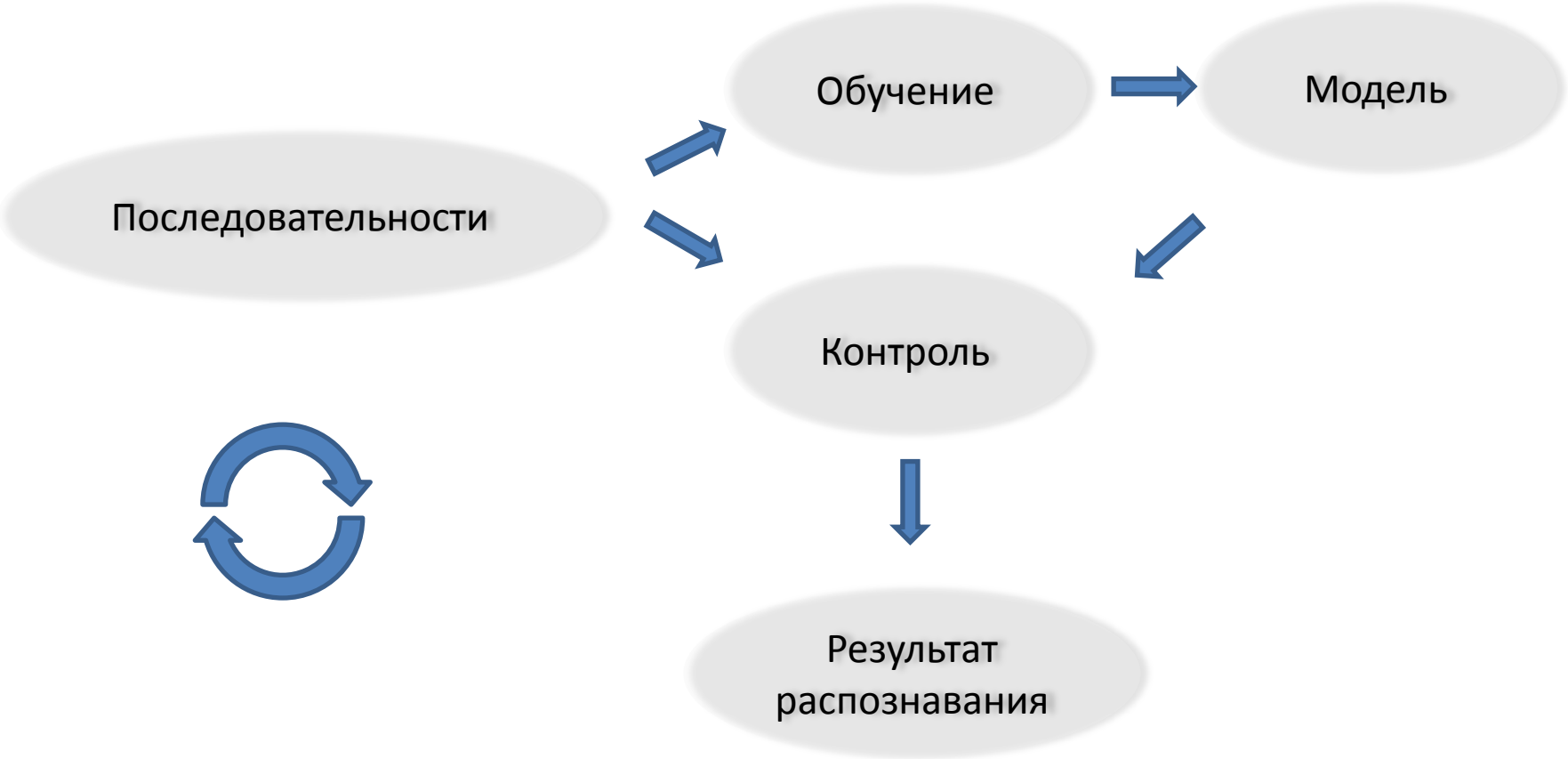
Алгоритм: Семантический вероятностный вывод


Результат
Визуализация
- UGENE - [Expert Discovery]
File Actions Settings Tools Window Help
1: Project
Items
Sequences
Positive
Negative
Control
Markup
MOTIFS
UGENE ANNOTATION
Complex signals
11
22
NewSignal152
NewSignal150
NewSignal78
NewSignal52
NewSignal80
NewSignal0
NewSignal151
NewSignal149
NewSignal1
NewSignal6
Distance from 0 to 10 taking into account order
"AAAATA" from family "MOTIFS"
"AAAAAT" from family "MOTIFS"
Editor
Name NewSignal6
Description
General information
Probability 48.7179% (19 / 39)
Pos. coverage 6.99588% (17 / 243)
Neg. coverage 0.781893% (19 / 2430)
Fisher 1.24281e-10
ul 0.715335
NM_000014 | 5,11969_neg7 [dna]
NM_000014 | 5,11969_neg0 [dna]
NM_000128 | 12,6442 [dna]
Score graph [1, 500], Window: 2, Step 1
5.399
rec.data
Name
Value
Auto-annotations [Negative | NM_000014 | 5,1...
2: Tasks
3: Log
No active tasks

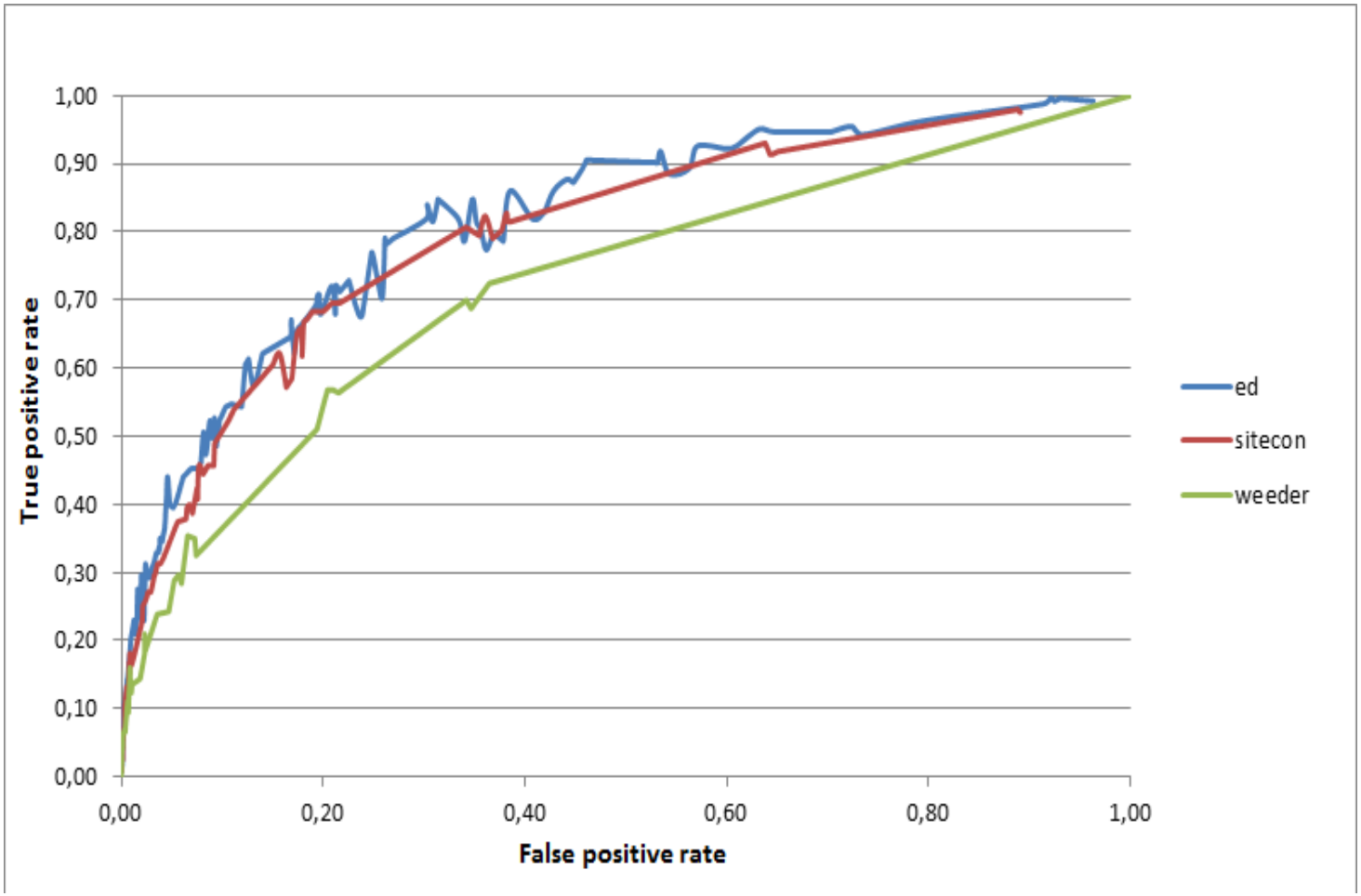
Результат
Визуализация



Проверка модели: Cross-validation



Receiver operating characteristic



Результаты

Система, способная:

- Интегрировать сигналы разных типов
- Качественно распознавать
- Генерировать значимые результаты
- Отображать данные различных типов

Публикации

1. “Анализ последовательностей регуляторных районов генов реляционной системой ExpertDiscovery, встроенной в пакет UGENE”. “Вестник НГУ”, Том 10, Выпуск 1, Новосибирск, 2011

2. “ExpertDiscovery and UGENE integrated system for intelligente analysis of regulatory regions of genes”. In Silico Biol. 2011-2012;11(3-4):97-108. doi: 10.3233/ISB-2012-0448.

3. “Analysis and Prediction of Regulatory Regions of Eukaryotic Genes by integrated UGENE and ExpertDiscovery Systems”. Special issue: ECML/PKDD 2011 (5th Workshop on Data Mining in Functional Genomics and Proteomics: Current Trends and Future Directions), Athens, 2011

Конференции

1. L Международная научная студенческая конференция «Студент и научно-технический прогресс» (Диплом III степени);

2. Международная конференция «Современные проблемы математики, информатики, биоинформатики»;

3. Международная конференция «Постгеномные методы в биологии, лабораторной и клинической медицине: геномика, протеомика, биоинформатика».

4. VIII Международная конференция по биоинформатике, системной биологии, регуляции и структуре геномов.