**Резюме проекта НИР (итоговое), выполняемого в рамках ФЦП
«Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007 – 2013 годы»**

Номер контракта: 07.514.11.4007

Тема: «Разработка алгоритмов и программ для сравнительного анализа геномов позволяющие определять функциональные, в частности регуляторные сегменты геномов»

Приоритетное направление: Науки о жизни (НЖ)

Критическая технология: 8 Нано-, био-, информационные, когнитивные технологии

Период выполнения: 13 июля 2011 г. — 30 ноября 2012 г.

Плановое финансирование проекта: _____

Бюджетные средства - 5,7 млн. руб.,

Внебюджетные средства - 0,852 млн. руб.

Исполнитель: Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН), г. Москва

Ключевые слова: алгоритм, предсказание функции, эволюция, геном, транскрипция, регуляция, экспрессия генов, сравнительная геномика

1. Цель исследования, разработки

Целью работы является создание обоснованных методов сравнительно-геномного анализа и разработка алгоритмов и программ для поиска регуляторных модулей в геномах эукариот с использованием сравнительно-геномного анализа.

2. Основные результаты проекта

Выявление и изучение закономерностей взаиморасположения сайтов связывания транскрипционных факторов (ССТФ), сохраняющихся в процессе эволюции или общих для регуляторных участков ко-регулируемых генов, а также полногеномный поиск регуляторных участков, являются актуальными биологическими задачами. Широкие возможности для решения этих задач предоставляют методы биоинформатики, позволяющие использовать алгоритмы машинного обучения для выявления закономерностей системы, а так же осуществлять масштабный поиск регуляторных участков в больших геномах эукариотических организмов.

На первом этапе проекта проведена аналитическая работа — патентного поиск, аналитического обзора информационных источников и выбор направлений исследований.

На втором этапе проекта построены эффективные алгоритмы уточнения предсказаний с учетом дерева, основанные на оригинальной идее скрытых Марковских моделей на деревьях. Разработана методика оценки статистической значимости сравнительно-геномного анализа. Разработана архитектура программного комплекса ЭО ПК CORECLUST для предсказания регуляторных модулей.

На третьем этапе разработан программный комплекс (ПК) для предсказания регуляторных модулей в геномах эукариот. Для реализации ПК CORECLUST разработана программа оптимизации параметров вероятностной модели регуляторных модулей. ПК CORECLUST, который имеет две реализации – локальную и клиент-серверную. Произведено тестирование ПК, разработана программная документация с учетом результатов тестирования. Применение ПК CORECLUST к реальным биологическим данным позволило получить интересные результаты. Проанализирована структура