Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики

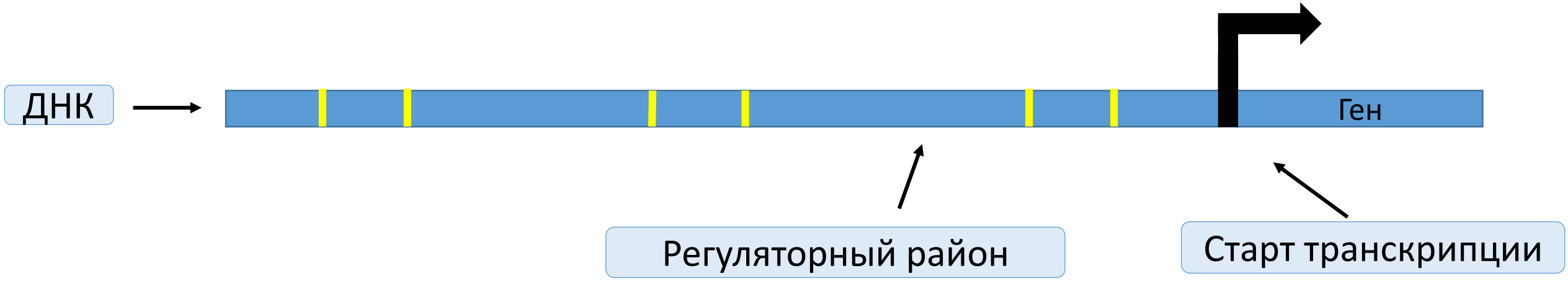
Математическая модель динамики экспрессии генов dar в эмбрионе плодовой мушки, учитывающая сайты связывания транскрипционных факторов

Выполнила: Дымова А.В., гр. 63601/4
Научный руководитель: Козлов К.Н.

Регуляция экспрессии гена

Экспрессия генов - это процесс преобразования наследственной информации гена в функциональный продукт (РНК или белок)

Транскрипция



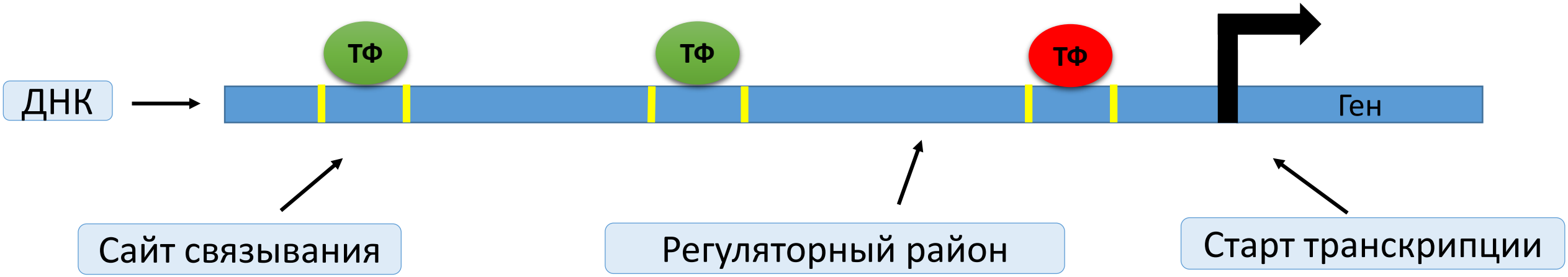
Регуляция экспрессии гена

Экспрессия генов - это процесс преобразования наследственной информации гена в функциональный продукт (РНК или белок)

Районы регуляции: регуляторные модули, районы открытого хроматина

Транскрипция

Транскрипционные факторы (ТФ): **активаторы** и **репрессоры**

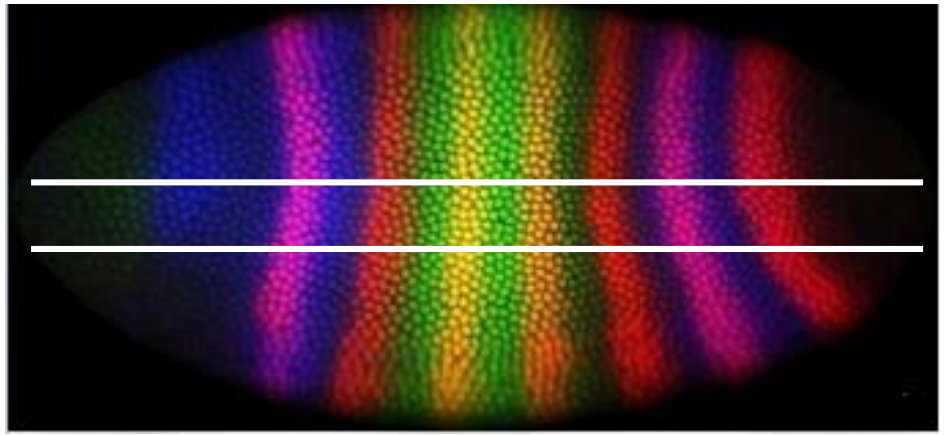


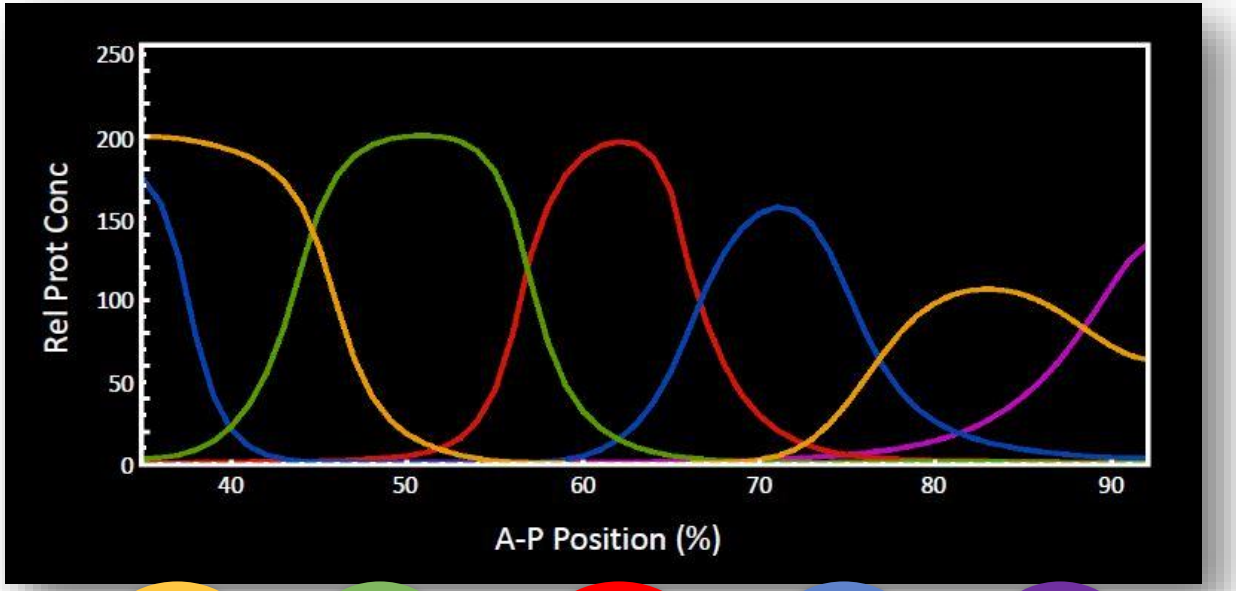
Механизмы регуляции требуют дальнейшего изучения

Модельный организм – плодовая мушка

Стадия синцитиальной бластодермы: циклы деления ядер - 13, 14А
гены *gap – Kruppel (Kr), giant (gt), hunchback (hb), knirps (kni)*
контролируют развитие сегментов в эмбрионе

Центральная 10% полоса





hb ***Kr*** ***kni*** ***gt*** ***tll***

Цель и задачи

Разработка модифицированной математической модели сети генов *gap*, учитывающая сайты связывания транскрипционных факторов

- Рассмотреть сайты связывания, найденные в регуляторных модулях, но не входящие в открытые участки хроматина
- Получить параметры и проверить предсказательную силу модели, которая использует вычисленные концентрации ТФ, учитывает данные о мРНК и двоякое действие ТФ Hb на ген *Kr*
- Применить модель для исследования значимости сайтов связывания

Расширенный список сайтов

8 ТФ – Hb, Gt, Kr, Kni, Bcd, Cad, Tll, Hkb

Сайты связывания, попадающие в район открытого хроматина [1]

+

Сайты связывания из регуляторных модулей

=

1395 сайтов связывания (было 889)

<u>Добавлены</u> <u>сайты</u> из важных регуляторных модулей согласно базе данных RedFly

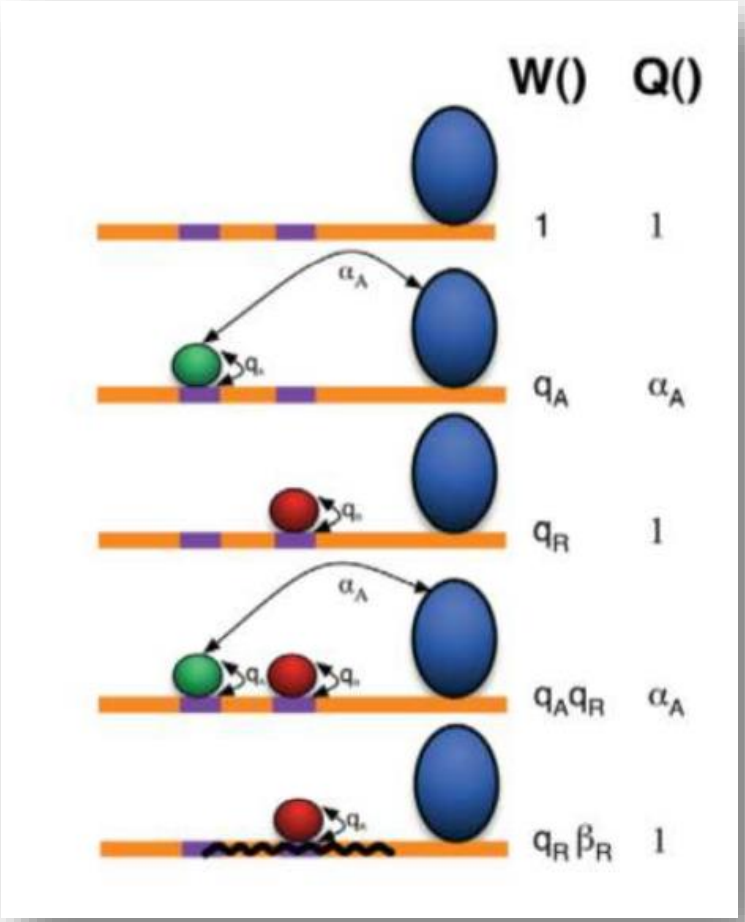
Уровень активации гена

$$E = \frac{Z_{ON}}{Z_{ON} + Z_{OFF}} = \frac{\sum_\sigma W(\sigma)Q(\sigma)}{\sum_\sigma W(\sigma)Q(\sigma) + \sum_\sigma W(\sigma)},$$

σ – молекулярная конфигурация регуляторного района гена

W(σ) – статистический вес конфигурации σ

Q(σ) – статистический вес взаимодействия между ТФ и связанным базальным транскрипционным механизмом [2]



α – сила активации

β – сила репрессии

Динамика концентраций мРНК и белка

$$du_i^a / dt = R_u^a E_i^a(t) + D_u^a(n)[(\mathrm{u}_{i-1}^a - \mathrm{u}_i^a) + (\mathrm{u}_{i+1}^a - \mathrm{u}_i^a)] - \lambda_u^a u_i^a$$

$$dv_i^a / dt = R_v^a u_i^a(t - \tau_v^a) + D_v^a(n)[(\mathrm{v}_{i-1}^a - v_i^a) + (\mathrm{v}_{i+1}^a - v_i^a)] - \lambda_v^a v_i^a$$

n - число циклов деления

R_v^a, R_u^a - максимальный уровень синтеза

$D_v^a(n), D_u^a(n)$- коэффициенты диффузии

λ_v^a, λ_u^a - коэффициент линейной деградации

τ_v^a - время задержки

T^{ab} матрицу межгенных взаимодействий, элементы которой характеризуют действие ТФ *b* на экспрессию гена *a*

Время интегрирования
$0 \le t \le 71.1$ (мин)
Нулевые начальные условия, кроме Hb

Граничные условия типа Дирихле

i – номер ядра
a – ген

Определение параметров

$$Error = RSS + 5 * 10^4 * wPGP + 0.001 * Penalty$$ → Функционал качества

$$RSS = \sum_{\forall g,n,t:\exists data_n^g(t)} (v_n^g(t) - data_n^g(t))^2$$ → Сумма квадратов разностей между решением модели и экспериментальными данными

$$wPGP = 0.5 + 0.5 * (penalty - reward)$$ → Weighted Pattern Generation Potential – больше реагирует на сдвиг чем на амплитуду [3]

$$Penalty = \sum_{\forall a,b} (T^{ab})^2$$ → Сумма квадратов регуляторных параметров

Метод глобальной стохастической минимизации полностью параллельной разностной эволюции

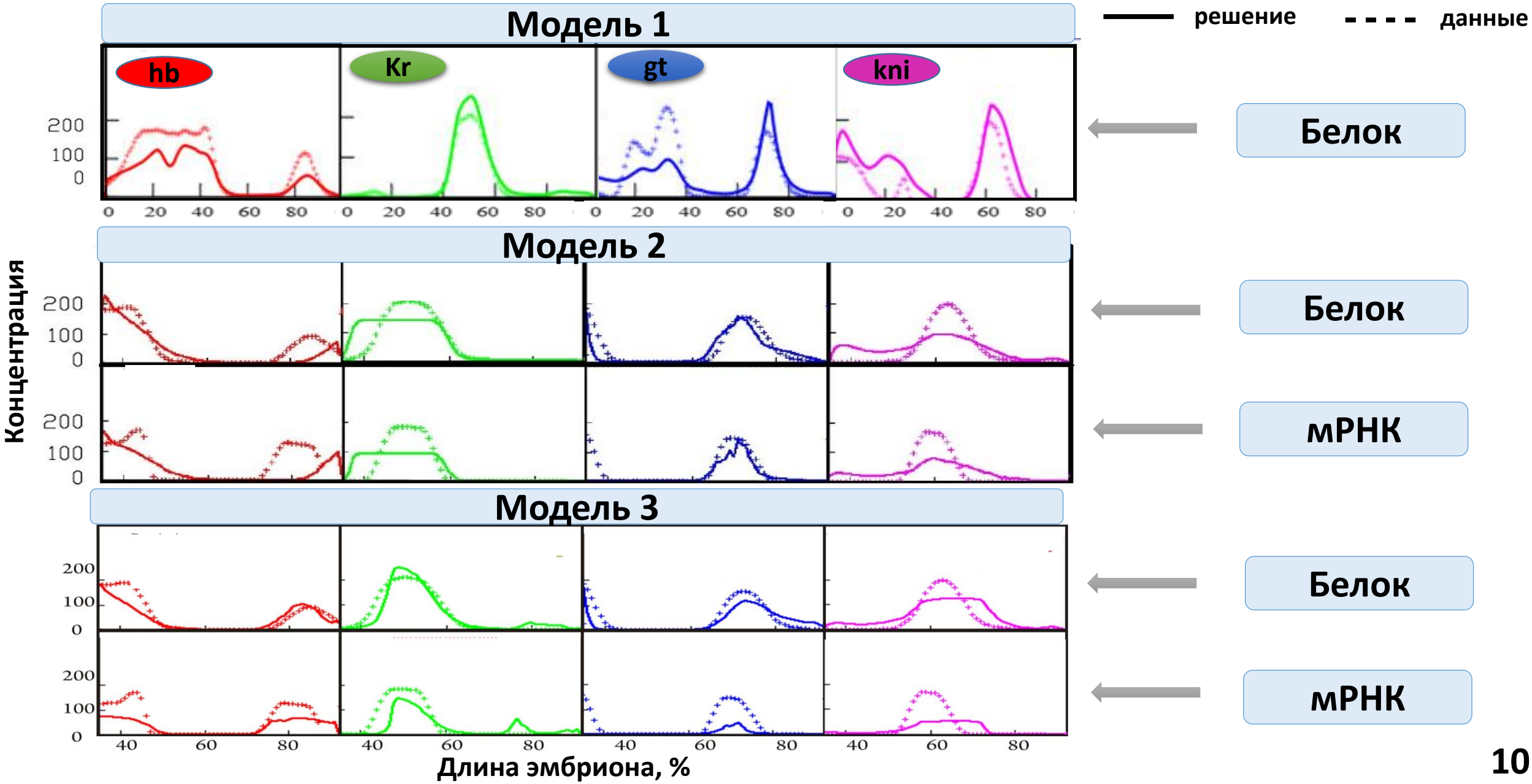
Модели регуляторной сети

Модель 1 – параметры из [1] + **расширенный список сайтов**, данные только по белку (БД FlyEx), область моделирования – 100%, концентрации ТФ - данные

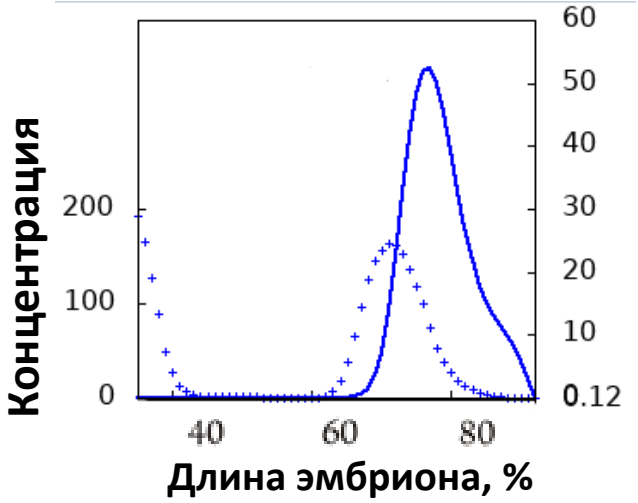
Модель 2 – **вычисленные концентрации ТФ**, **новый набор параметров**, данные по белку и **мРНК** (БД SuperFly), область моделирования 35-92%

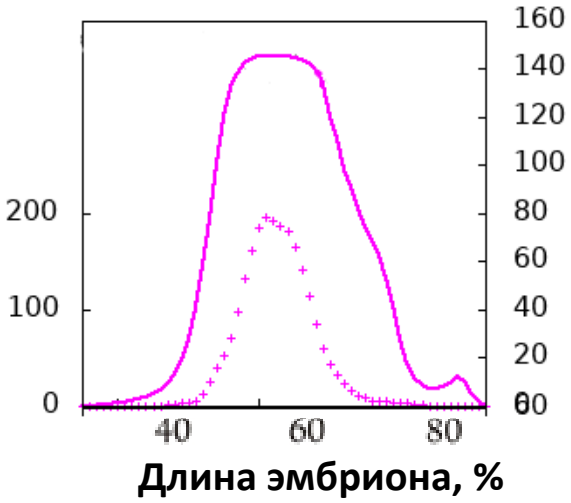
Модель 3 – **вычисленные концентрации ТФ**, **новый набор параметров**, данные по белку и **мРНК**, область моделирования 35-92%, **двоякое действие ТФ Hb на ген *Kr***

Решение модели в сравнении с данными



Независимая проверка модели





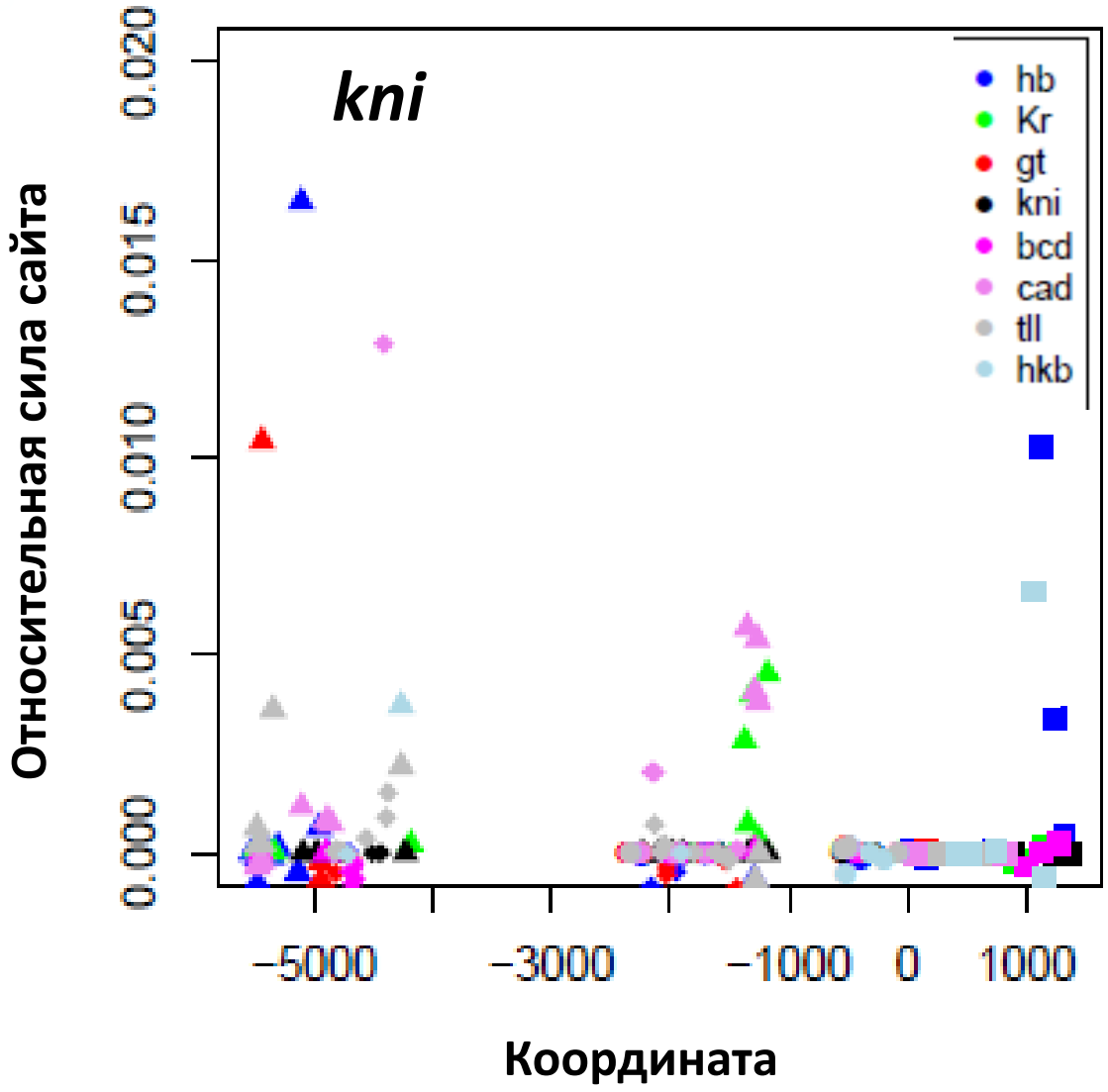
Модель предсказала экспрессию регуляторных модулей:

hb **3** (было 0) из 10;
Kr **2** (было 2) из 10;
gt **3** (было 1) из 5;
kni **8** (было 2) из 16;

—— **решение** - - - - **данные**

График экспрессии регуляторного модуля отмасштабирован для наглядности (правая шкала)

Определение значимости сайтов ТФ



Относительная сила сайта

$$w_r = (\mathrm{RSS}_{ref} - \mathrm{RSS}_{mut}) / \mathrm{RSS}_{ref}$$

mut – сайт исключен

Треугольник - сайты в регуляторном модуле
Круг - сайты в открытом хроматине
Ромб — обе группы
Квадрат – сайты, в кодирующем районе гена *kni*, *kni_+1_construct*

Распределение сильных сайтов

Модель 1: большее количество сильных сайтов в районах открытого хроматина и регуляторных модулях

Модель 2, **Модель 3**: большее количество сильных сайтов в районах открытого хроматина

Выводы

1. Разработана математическая модель регуляторной сети генов *gap.* Преимущество модели - использование вычисленных концентраций ТФ. Модель описывает основные характеристические черты картин экспрессии генов-мишеней дикого типа и некоторых регуляторных модулей;
2. Наилучший результат показала модель с учетом двоякого действия ТФ Hb на ген *Kr* (модель 3);
3. В разработанных моделях (2, 3) наибольшую значимость имеют сайты из районов открытого хроматина.

Благодарности

Выражаю благодарность:

- К.Н. Козлову, к.б.н;
- М.Г. Самсоновой, д.б.н;
- В.В. Гурскому, к.ф.-м.н..

Литература

1. Kozlov K, Gursky V, Kulakovsky I, Samsonova M.: Sequence-based model of gap gene regulatory network (2014). BMC Genomics, accepted;
2. He, X., Samee, M.A.H., Blatti, C., Sinha, S.: Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. PLoS Comput. Biol. 6(9) (2010);
3. Samee, M.A.H., Sinha, S.: Evaluating thermodynamic models of enhancer activity on cellular resolution gene expression data. Methods 62, 79–90 (2013);
4. Kozlov, K., Samsonov, A.: Deep – differential evolution entirely parallel method for gene regulatory networks. Journal of Supercomputing 57, 172–178 (2011);
5. Pisarev, A., Poustelnikova, E., Samsonova, M., Reinitz, J.: FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. Nucleic Acids Research 37, 560–566 (2008);
6. Gallo, S.M., Gerrard, D.T., Miner, D., Simich, M., Des Soye, B., Bergman, C.M., Halfon, M.S.: Redfly v3.0: Toward a comprehensive database of transcriptional regulatory elements in Drosophila. Nucleic Acids Research 10, 1–6 (2010).

Определение параметров

$$Error = RSS + 5 * 10^4 * wPGP + 0.001 * Penalty$$ → Целевая минимизируемся функция

$$RSS = \sum_{\forall g,n,t:\exists data_n^g(t)} (v_n^g(t) - data_n^g(t))^2$$ → Сумма квадратов разностей между решением модели и экспериментальными данными

$$wPGP = 0.5 + 0.5 * (penalty - reward)$$

$$penalty = \frac{\sum_i (r_{max} - r_i) * |p_i - r_i|}{\sum_i (r_{max} - r_i) * \sum_i (r_{max} - r_i)}$$

$$reward = \frac{\sum_i r_i * min(r_i, p_i)}{\sum_i r_i * r_i}$$

r_i реальная экспрессия в ядре *i*

p_i предсказанная экспрессия в ядре *i*

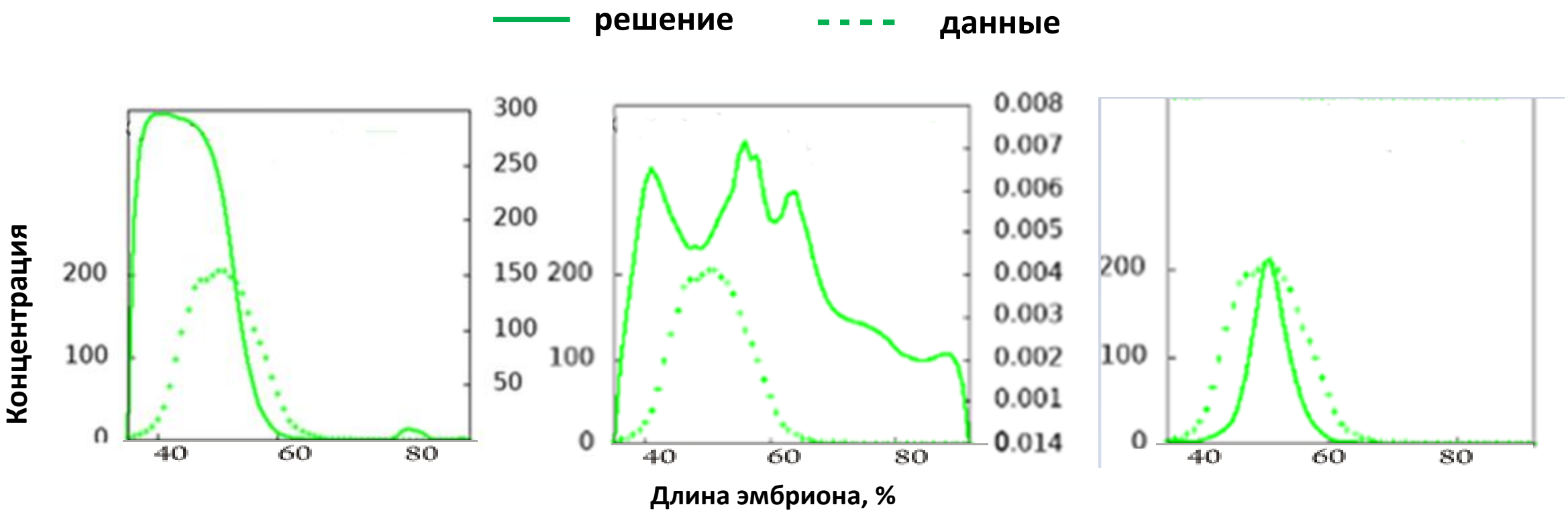
r_{max} максимальный уровень реальной экспрессии

Двоякое действия ТФ Hb на ген *Kr*

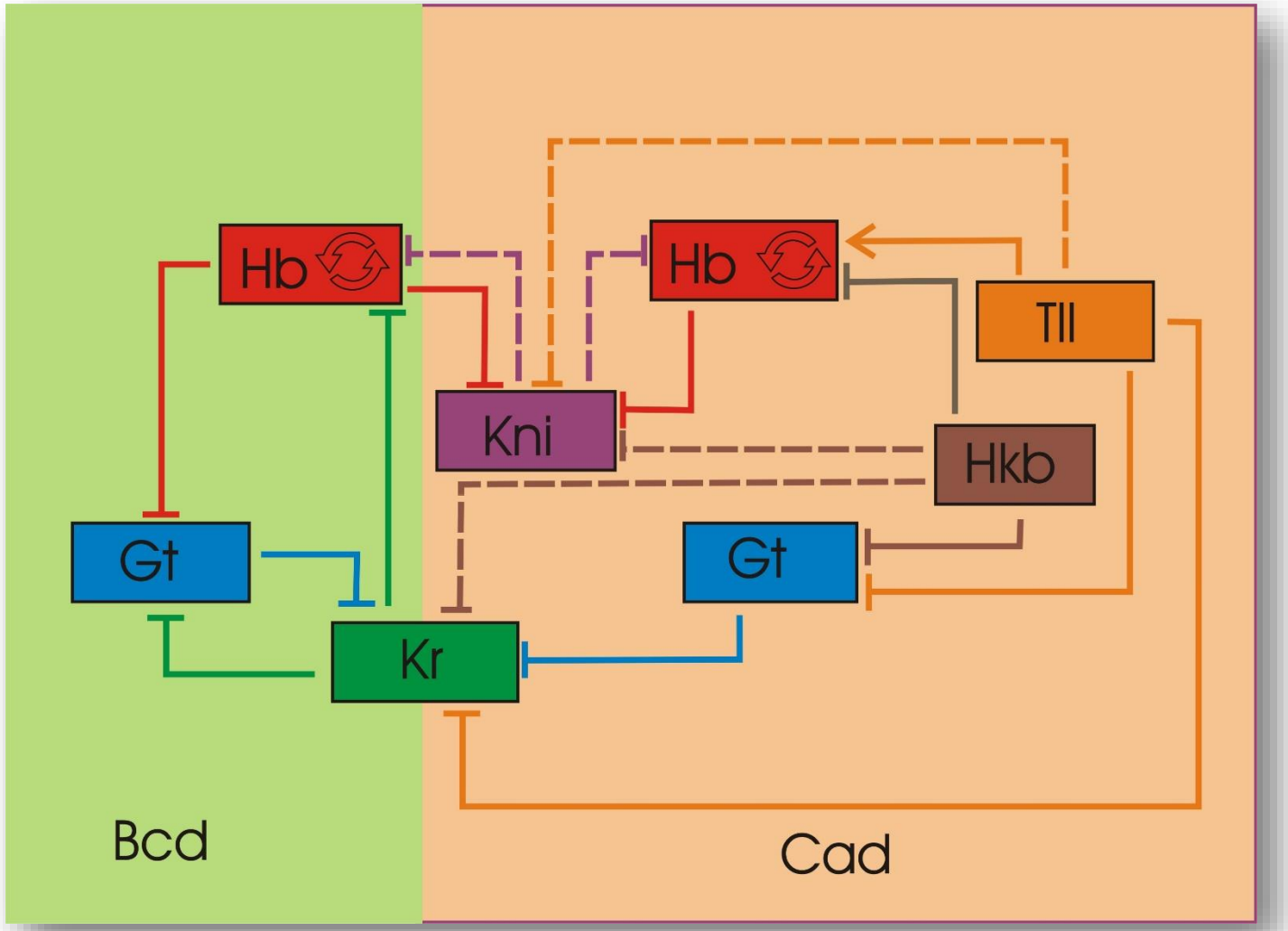
ТФ Hb – только активатор. Концентрация *Kr* - 300

ТФ Hb – только репрессор. Концентрация *Kr* - 0

ТФ Hb – порог переключения = 34.3



Регуляторная сеть генов *gap*

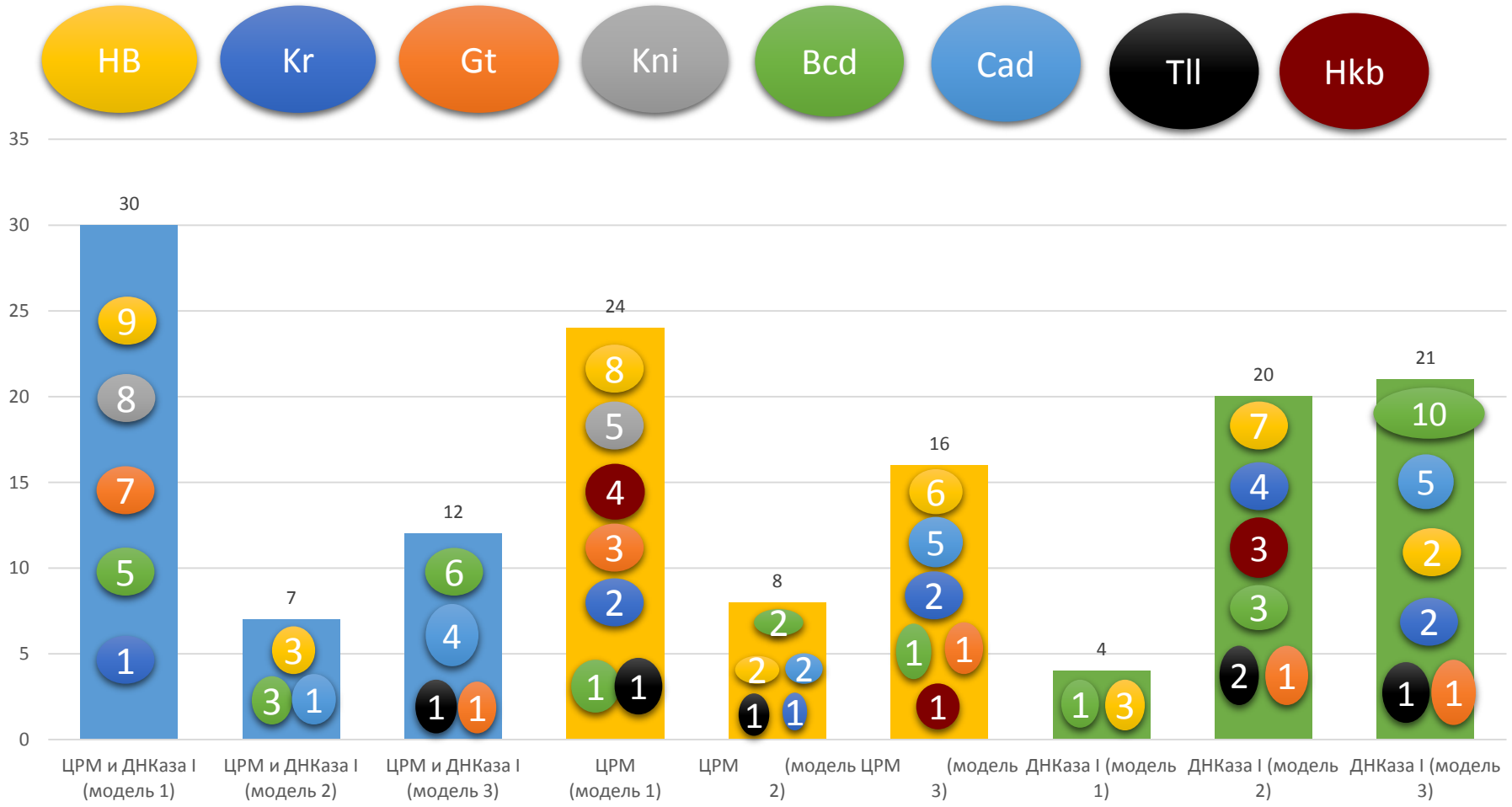
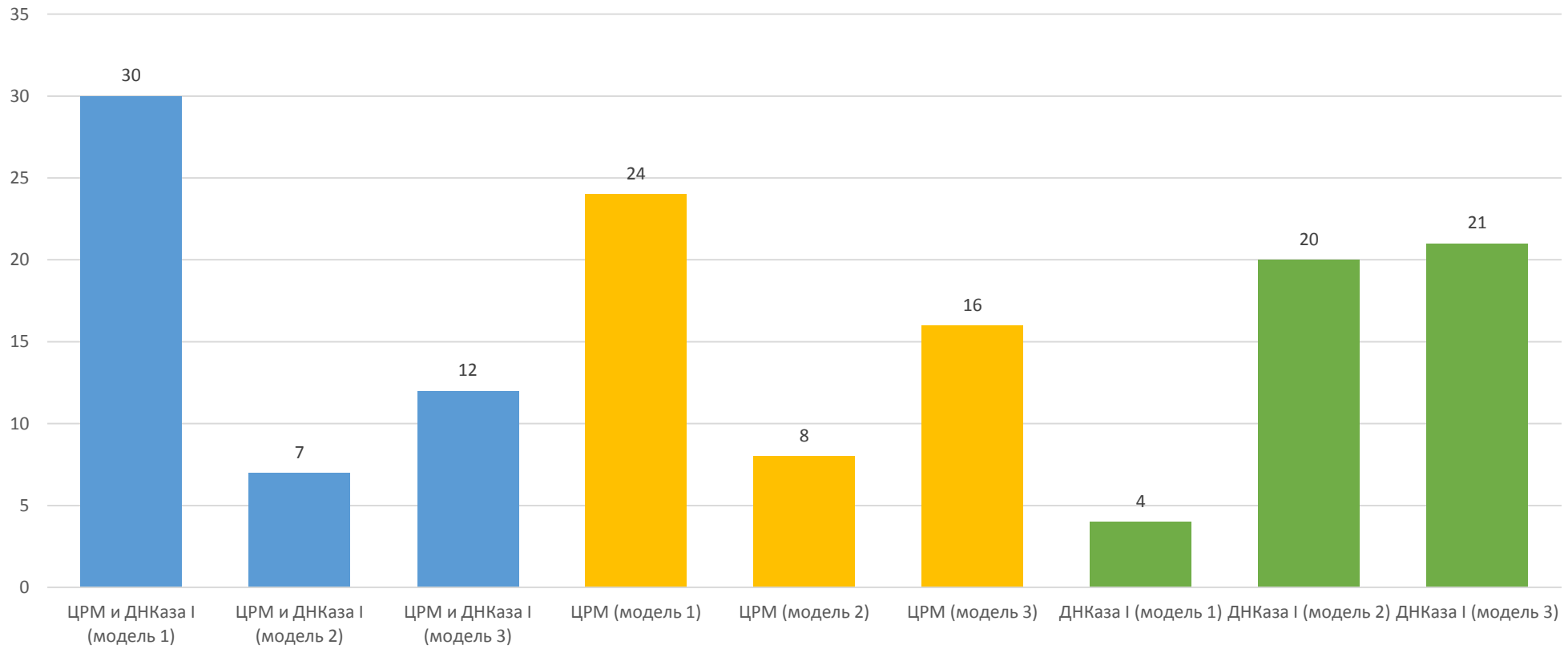


Активация во всех трех моделях

Репрессия во всех трех моделях;

Репрессия в моделях (1,3);

В моделях 2,3 большее количество сильных сайтов в районах открытого хроматина



В моделях 2,3 большее количество сильных сайто в районах открытого хроматина