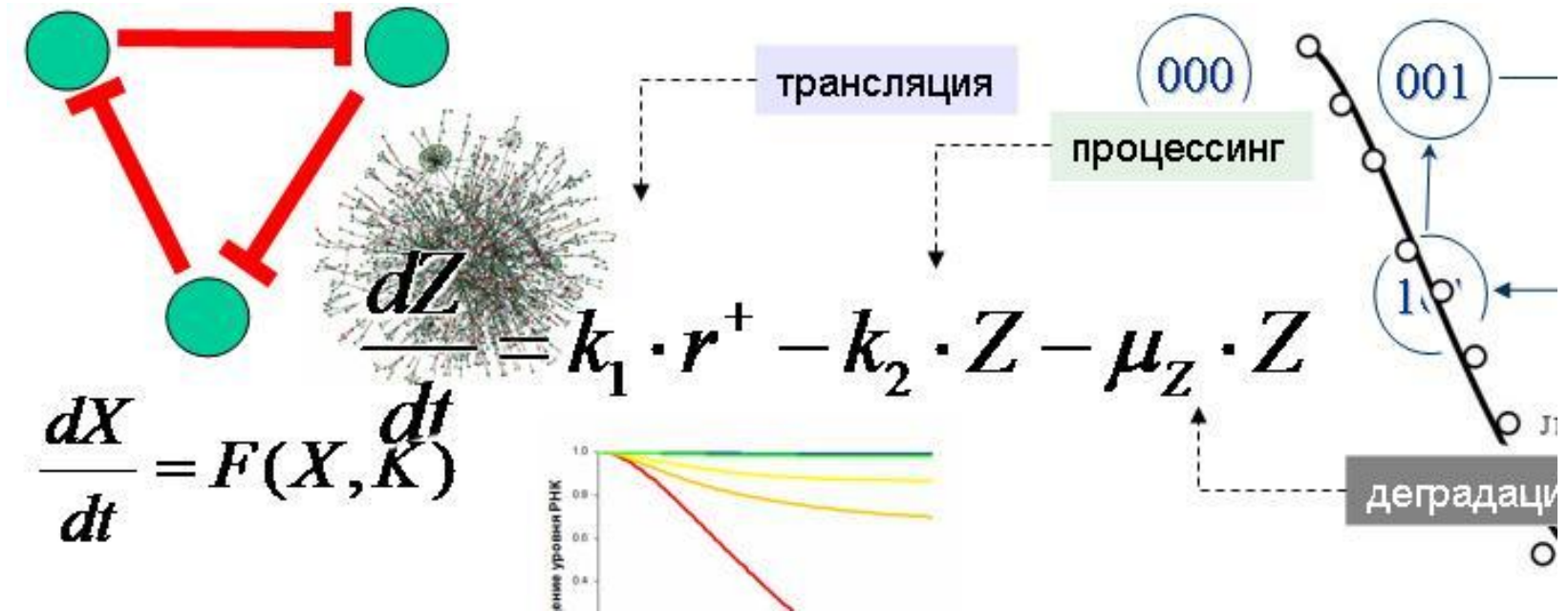


# Введение в системную биологию



Соколик Анатолий Иосифович,  
доцент каф. клеточной биологии и  
биоинженерии растений

# ВВЕДЕНИЕ

*Понятие «системная биология», различные его трактовки и содержание, место среди других приложений к биологии математики, информационных технологий и компьютерной техники.*

Биоинформатика, компьютерная геномика, компьютерная биология, математическая биология. Системная биология. История.

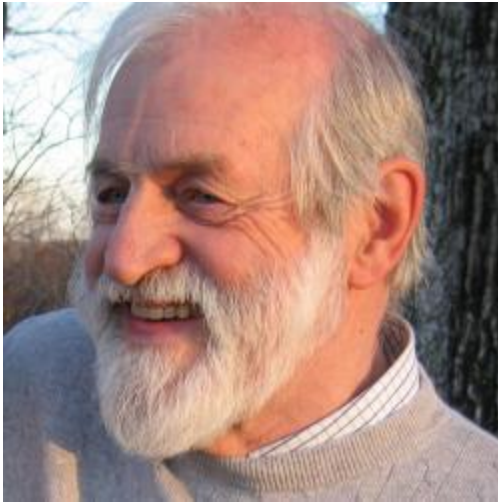
Моделирование систем – основной подход системной биологии. Анализ сложных систем с большими массивами данных. Основа системной биологии – математика.

**Системная биология** — развивающаяся междисциплинарная область биологии, которая анализирует сложные биологические системы разного уровня исходя из их многокомпонентности, наличия прямых и обратных связей, разнородности экспериментальных данных, характеризующих системы.

**Предмет исследований** - биологические системы от субклеточного и клеточного уровней – например, система регуляции генов, метаболизм, клеточная динамика, взаимодействия в клеточной популяции – до уровня популяций организмов и целых экосистем.

**Методологическая основа** системной биологии – математика

# Дж. Мюррей (James Murray) – математик:



«Чтобы обеспечить дальнейшее процветание своей науки, математикам придется заниматься биологией. Памятуя о том, сколь полезной для математики оказалась физика и как повлияла на нее математика, становится ясным, что, если математики не «влезут» в биологические науки, они просто-напросто останутся в стороне от научных открытий, которые обещают стать самыми важными и волнительными за всю историю науки»



# Основные варианты применения математики в биологии

## Статистика

**Биоинформатика** (объединение биологии, математики и информатики для решения задач молекулярной биологии, биохимии, генетики, клеточной биологии, фармакологии, здравоохранения и т.д. синоним **вычислительной молекулярной биологии**).

## Включает:

- *Биоинформатика последовательностей.*
- *Структурная биоинформатика.*
- *Компьютерная геномика*
- *Применение известных методов анализа для получения новых биологических знаний.*
- *Разработка новых методов анализа биологических данных*
- *Разработка новых баз данных*

# Биоинформатика последовательностей

В базе данных EMBL (Европейской молекулярно-биологической лаборатории) на 1.09.2015 г. хранится **13 634 705** документов с описанием **14 579 744 964** нуклеотидных последовательностей, содержащих в целом столько символов (нуклеотидов), что соответствует примерно библиотеке в  $10^5$  толстых томов с убористым шрифтом

трансляции с использованием известного генетического кода можно получить. аминокислотные (белковые) последовательности.

Из известных сегодня 5 миллионов белков 95 % последовательностей – это такие гипотетические трансляты, и больше о них ничего не известно

# Структурная биоинформатика

Структурная биоинформатика занимается анализом пространственных структур молекул. Известно всего около 100 000 структур из нескольких миллионов последовательностей.

Молекулярный **ДОКИНГ** (молекулярная стыковка) — метод моделирования, который позволяет предсказать наиболее выгодную для образования устойчивого комплекса ориентацию и положение одной молекулы по отношению к другой.



# Программы для молекулярной

## СТЫКОВКИ

AutoDock (<http://autodock.scripps.edu>)

FlexX (<http://www.biosolveit.de/FlexX/>)

Dock (<http://dock.compbio.ucsf.edu>)

Surflex (<http://www.biopharmics.com>, [www.tripos.com](http://www.tripos.com))

Fred (<http://www.eyesopen.com/products/applications/fred.html>)

Gold ([http://www.ccdc.cam.ac.uk/products/life\\_sciences/gold/](http://www.ccdc.cam.ac.uk/products/life_sciences/gold/))

PLANTS (<http://www.tcd.uni-konstanz.de/research/plants.php>)

3DPL (<http://www.chemnavigator.com/cnc/products/3dpl.asp>)

Lead Finder (<http://www.moltech.ru>)

Molegro Virtual Docker (<http://www.molegro.com>)

ICM Pro ([http://www.molsoft.com/icm\\_pro.html](http://www.molsoft.com/icm_pro.html))

Ligand fit, Libdock and CDocker

(<http://accelrys.com/services/training/life-science/StructureBasedDesignDescription.html>)

DockSearch (<http://www.ibmcm.sk.ru>)

eHiTS (<http://www.simbiosys.ca/ehits/index.html>)

Glide (<http://www.schrodinger.com/productpage/14/5/>)

DockingShop ([http://vis.lbl.gov/~scrivelli/Public/silvia\\_page/DockingShop.html](http://vis.lbl.gov/~scrivelli/Public/silvia_page/DockingShop.html))

HADDOCK (<http://www.nmr.chem.uu.nl/haddock/>)



## геномика

Сегодня определены полные или почти полные последовательности геномов многих организмов, но это не самоцель, а первый шаг для исследования того, как функционирует та или иная клетка

Изучение геномов позволяет найти новые метаболические пути или ферменты, которые будут применены в биотехнологическом производстве (например, витаминов и других биологически активных веществ )

Компьютерный анализ позволяет с известной степенью точности охарактеризовать несколько тысяч генов силами небольшой группы примерно за неделю, тогда как Экспериментальное определение функции только одного гена требует интенсивной работы одной лаборатории как минимум в течение нескольких

# Биоинформатика

**Применение известных методов анализа для получения новых биологических знаний**

Существует множество методов и инструментов для компьютерного анализа биологических данных, представленных в виде программ в Интернете и имеющих удобный пользовательский интерфейс .

На неправильный вопрос компьютер всегда дает неправильный ответ. Нужно учитывать границы применимости тех или иных методов.

компьютерный анализ биологических данных является экспериментом (только сделанным не в пробирке) и к нему предъявляются те же требования – четкость постановки, контроли

# **Биоинформатика**

**Разработка новых методов анализа биологических данных**

**Разработка новых баз данных**

# Математическая биология

Математическая биология относится к прикладной математике и использует её методы.

В математической биологии исследуются биологические задачи и проблемы методами современной математики, а результаты имеют биологическую интерпретацию

Пример - закон Харди-Вайнберга (для идеальной популяции),

$$p^2 + 2pq + q^2 = 1$$

где  $p$  и  $q$  – частоты аллелей гена

# Компьютерная

# биология

Частично перекрывается с биоинформатикой

Область науки о компьютерном анализе генетических текстов, аминокислотных последовательностей, пространственной структуры и динамики белков,

Этот анализ лежит в основе определения макромолекул-мишеней, и поиска низкомолекулярных комплексов с целью создания новых лекарств,

Компьютерная биология превратилась в быстроразвивающееся направление биомедицины

# Компьютерная

# биология

Процесс создания нового лекарственного соединения можно разделить на следующие этапы:

- (1) поиск мишени (например, белка) действия нового лекарства;
- (2) поиск низкомолекулярного соединения, обладающего нужным фармакологическим действием;
- (3) изучение этого соединения в эксперименте;
- (4) проведение испытаний в клинике.

Уже первый этап поиска подходящего кандидата на лекарство – перебор сотен миллионов вариантов из соответствующей базы данных низкомолекулярных соединений

## Оценки вычислительных потребностей для полного расчета энергии связывания всех низкомолекулярных соединений, входящих в различные базы данных

<b>Уровень сложности моделирования</b>	<b>Метод</b>	<b>Размер базы</b>	<b>Время расчета</b>
Молекулярная механика	SPECTTOPE	140000	1 час
Жесткие лиганд/мишень	LUDI	30000	1-4 часов
Молекулярная механика	Hammerhead	80000	3-4 дня
Частично деформируемый лиганд	DOCK	17000	3-4 дня
Жесткая мишень	DOCK	53000	14 дней
Молекулярная механика	ICM	50000	21 день
Молекулярная механика	AMBER, CHARMM	1	несколько дней
Квантомеханический активный сайт	Gaussian, Q-Chem	1	несколько недель

# Производительность суперкомпьютеров

Название	год	флопсы
флопс	1941	$10^0$
<u>килофлопс</u>	1949	$10^3$
<u>мегафлопс</u>	1964	$10^6$
<u>гигафлопс</u>	1987	$10^9$
<u>терафлопс</u>	1997	$10^{12}$
<u>петафлопс</u>	2008	$10^{15}$
<u>эксафлопс</u>	2019 или позже	$10^{18}$
<u>зеттафлопс</u>	не ранее 2030	$10^{21}$
<u>йоттафлопс</u>	-	$10^{24}$
<u>ксерофлопс</u>	-	$10^{27}$



# Самый мощный сегодня суперкомпьютер в мире

## Tianhe-2 (Млечный путь 2)

2013 году. 200-300

миллионов

долларов.

1300 ученых и

инженеров

трудились над

созданием

Tianhe-2,

«Млечный путь-2».

Стойки: 125

Cores: 3120000

Производительность : 33862.7

TFlop/s

ТFlop/s

Мощность: 17808.0

0 kW

Память: 1024000

GB



# Системная биология (systems biology)

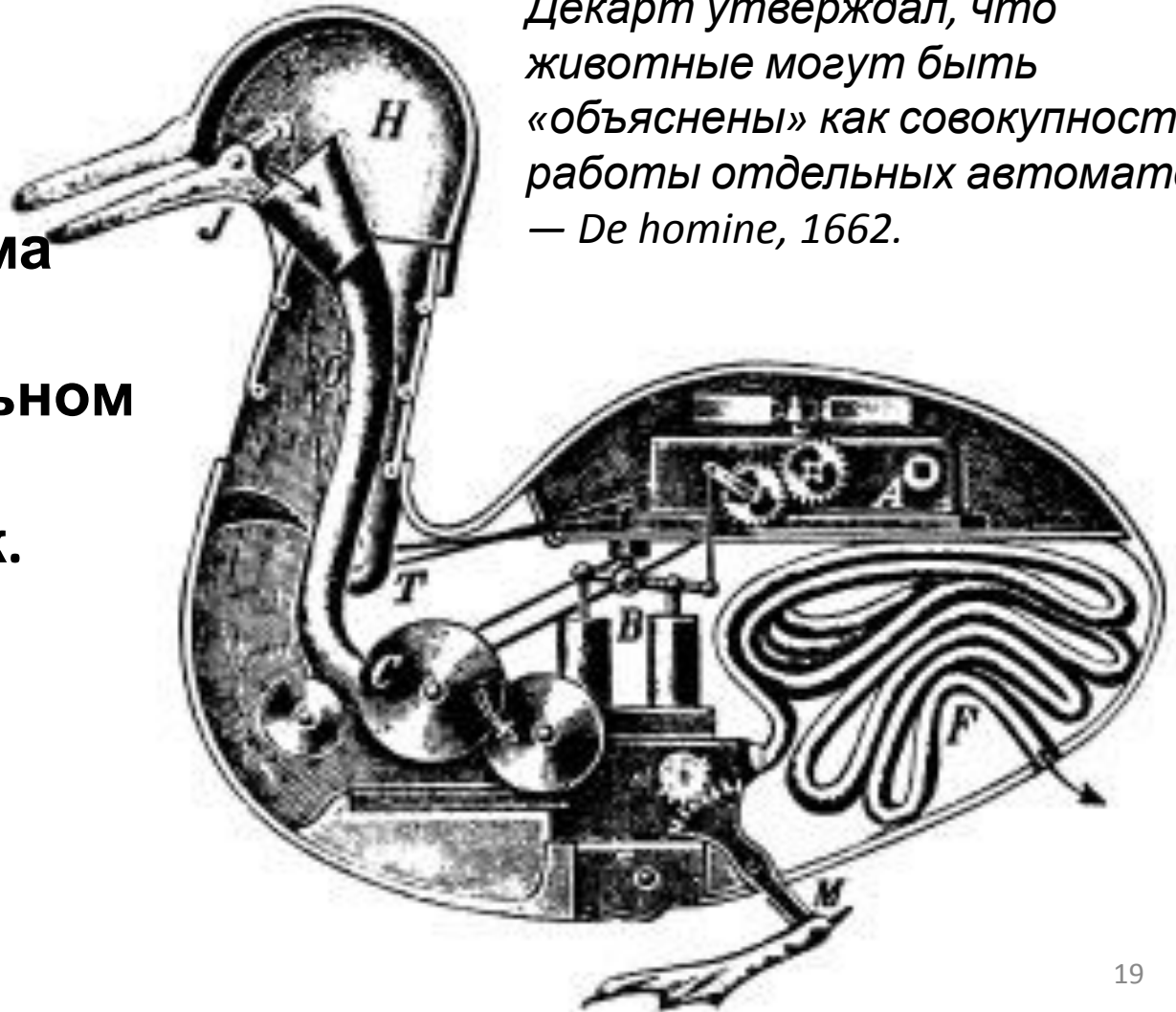
Системная биология — активно развивающаяся междисциплинарная область науки, которая анализирует сложные биологические системы с учетом их многокомпонентности, наличия прямых и обратных связей, а также разнородности и большого количества экспериментальных данных. Предметом исследований в этой области может являться система регуляции генов, метаболизм, а также клеточная динамика и взаимодействия в клеточной популяции

*(Биохимик может определить ферменты и продукты цикла Кребса, но рассчитать динамику изменения их концентрации может только системный биолог.)*

Важнейшим принципом для системной биологии является **«ХОЛИЗМ»**, который должен заменить **«редукционизм»**.

**Редукционистский подход** предполагает, что свойства сложной многокомпонентной системы **МОЖНО** получить только при рассмотрении ее отдельных компонентов.

Например, физиологические функции организма станут понятны только при детальном знании его отдельных клеток.



*Декарт утверждал, что животные могут быть «объяснены» как совокупность работы отдельных автоматов — De homine, 1662.*

# Холистический подход

предполагает, что свойства сложной многокомпонентной системы **НЕВОЗМОЖНО** представить как сумму свойств ее отдельных компонент.

Например, физиологические функции организма «не обнаружимы» при рассмотрении его отдельных клеток.

# Системная биология (systems biology)

Основная задача системной биологии, которая не пересекается с биоинформатикой это – **моделирование свойств динамических биосистем с дискретным (имеющим рамки) и непрерывным временем (большая часть био-систем).**

В целом биологические системы неравновесны (открыты, они постоянно обмениваются со средой энергией и веществом) и нелинейны (изменения их состояния не полностью определяется предшествующим).

Поэтому для них используются специальные методы анализа и описания (нелинейная динамика).

# Системная биология (systems biology)

Предпосылками появления системной биологии являются:

- Количественное моделирование ферментативной кинетики — направление, формировавшееся между 1900 и 1970 годами,
- Математическое моделирование роста популяций,
- Моделирование в нейрофизиологии,
- Теория динамических систем и кибернетика.

# Развитие системной биологии:

## Организационная и теория и теория систем

Богданова – как не странно *беларуский* ученый и революционер из Гродно - *Александр Малиновский* (псевдоним *Богданов* – один из создателей и лидеров РСДРП, совместно с Лениным). Выдающийся философ, написавший несколько больших трудов о Тектологии, введенной им науки, раскрывающий единый принцип устройства, организации и управления биологических и небиологических систем. Именно он ввел понятия

открытости биологической системы, её саморегуляции, самоорганизации, «самоусложнения», возможности убывания энтропии,

благодаря которым многие такие системы обладают холистическими свойствами. Малиновский/Богданов - признанный создатель основ системной биологии, биоинформатики и кибернетики



# Карл Людвиг фон Берталанфи

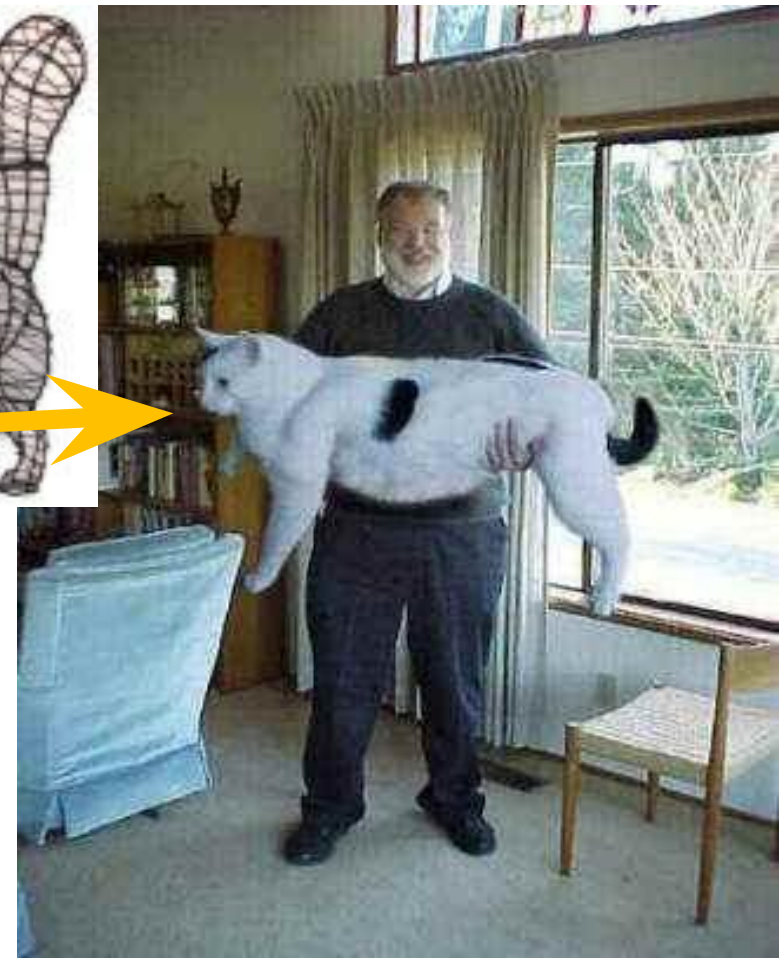
основной популяризатор **теории систем** в США. Главным образом, заимствовал и развивал идеи в математике систем. Широко известен как «отец» общей теории систем.



Теоретически обосновал, что термодинамические классические законы (сохранения энергии и массы и возрастания энтропии) **«не работают»** при рассмотрении биологических систем



Открытые системы по [Берталанфи](#) – могут принимать больше энергии, чем отдавать. Они усовершенствуют себя сами, по заложенному в них принципу организации, саморегуляции и самоуправления. В случае биологии – на основе генетического кода и его реализации (фенома) в пределах, задаваемых данными условиями существования



# Этапы развития системной биологии:

## Модель биологического роста Берталанфи

Самое простое дифференциальное уравнение (уравнение для описания динамических процессов – подставляются известные параметры и их соотношения, т.е. коэффициенты, что позволяет найти неизвестные интересующие нас параметры, а также построить график и по нему установить неизвестные параметры).

Уравнение изменения длины (любых размеров) со временем:

$$L'(t) = r_B (L_{\infty} - L(t))$$

$L$  – длина,  $t$  – время

$r_B$  – скорость роста по Берталанфи  $L_{\infty}$  – предельная длина организма.

Дополнительные коэффициенты (не указаны выше) – доступности пищи, уровня метаболизма, фазы онтогенеза и т.п.. Они помогут более точно рассчитать изменение роста во времени. Модель используется и сейчас.

**Одна из первых моделей**, в которой решалась задача физиологии, была модель распространения нервного импульса (потенциала действия), созданна **А. Ходжкиным и Э. Хаксли** для аксона кальмара (1952 г.)

**В 1960 Денис Нобл** создал первую модель клеток-пейсмейкеров в сердце – математическую модель сердечного ритма.

**Официальное признание современной Системной биологии** как отдельной науки относят к международному симпозиуму, проходившему в Кливленде в 1966 г., под названием "Systems Theory and Biology» - *Системная теория и биология*.

**В 1960-70 годах** развивались первые метаболические модели – модели сетей ферментов и их активности. Появились теории Метаболического контроля, обратной негативной и позитивной связи для регуляции, появились первые доступные расчетные модели для структур белка.

**1980-е годы:** во время бурного развития молекулярной биологии о моделировании забыли, тем более, что у биологов развился скептицизм к всемогуществу математики и физики. Компьютеры были маломощные и не позволяли производить необходимые биологам расчеты.

**С начала 90-х годов,** так называемой эры геномики, когда появились первые огромные массивы нуклеотидных и аминокислотных последовательностей, потребности в их анализе привели к новому бурному витку развития системной биологии.

**Прорыв в быстродействии и доступности компьютерных технологий (1990-2000 гг.)** вылился в привлечение всё большего числа программистов, математиков и физиков-теоретиков в биологию.

**После 2000** появились **-омиксы** – семейство наук, создавших необходимость обработки огромных массивов биологических данных.

# Смежные дисциплины (и их объекты), из которых, главным образом, берутся и анализируются данные в биоинформатике и системной биологии:

- **Феномика:** вариации в фенотипе и ее изменение в течение жизненного цикла.
- **Геномика:** ДНК последовательности организмов или клеток. Аннотация, картирование и анализ генов, экзонов (кодир.) и интронов (некодир.), других участков.
- **Эпигеномика / Эпигенетика:** транскриптомная регуляция, не кодирующиеся геномом, например, ДНК метилирование или ацетилирование гистонов.
- **Транскриптомика:** измерение изменения экспрессии отдельных генов при помощи «DNA microarrays» (ДНК-чипов).
- **Интерферомика:** знание о механизмах и многообразии систем «корректировки» транскриптов, например, РНК-интерференция.

# Смежные дисциплины (и их объекты), из которых, главным образом, берутся и анализируются данные в биоинформатике и системной биологии:

- **Протеомика** (транслятомика – более редкое название): измерения белков и пептидов при помощи двух-мерного гель-электрофореза в комбинации с масс-спектрометрией, HPLC и других детекторов. Подразделяется на фосфопротеомикс, гликопротеомикс, мембранный и эндомембранный протеомикс и др. типы.
- **Метаболомика**: измерение соотношения, разнообразия и распределения, а также связи с функциями организма небольших молекул (т.н. метаболитов), не относящихся к биополимерам.
- **Гликомика**: измерение соотношения, разнообразия и распределения, а также связи с функциями организма углеводов.

# Смежные дисциплины (и их объекты), из которых, главным образом, берутся и анализируются данные в биоинформатике и системной биологии:

- **Липидомика:** измерение соотношения, разнообразия и распределения, а также связи с функциями организма липидов.
- **Интерактомика:** измерение и анализ взаимодействий между молекулами, химических реакций. Например, белок-белковые взаимодействия.
- **Нейроэлектродинамика:** анализ организация и функция нейронов как динамической системы, способной обрабатывать информацию при помощи электрических сигналов.
- **Иономика и флаксомика:** области, изучающие активности и распределение ионов и их потоков, соответственно.
- **Биомика:** системный анализ биома (проявлений жизни – явлений присущих только живым системам).

# Инструменты системной биологии

Исследования в области системной биологии чаще всего заключаются в разработке модели сложной биологической системы, то есть модели, сконструированной на основе количественных данных об элементарных процессах, составляющих систему.

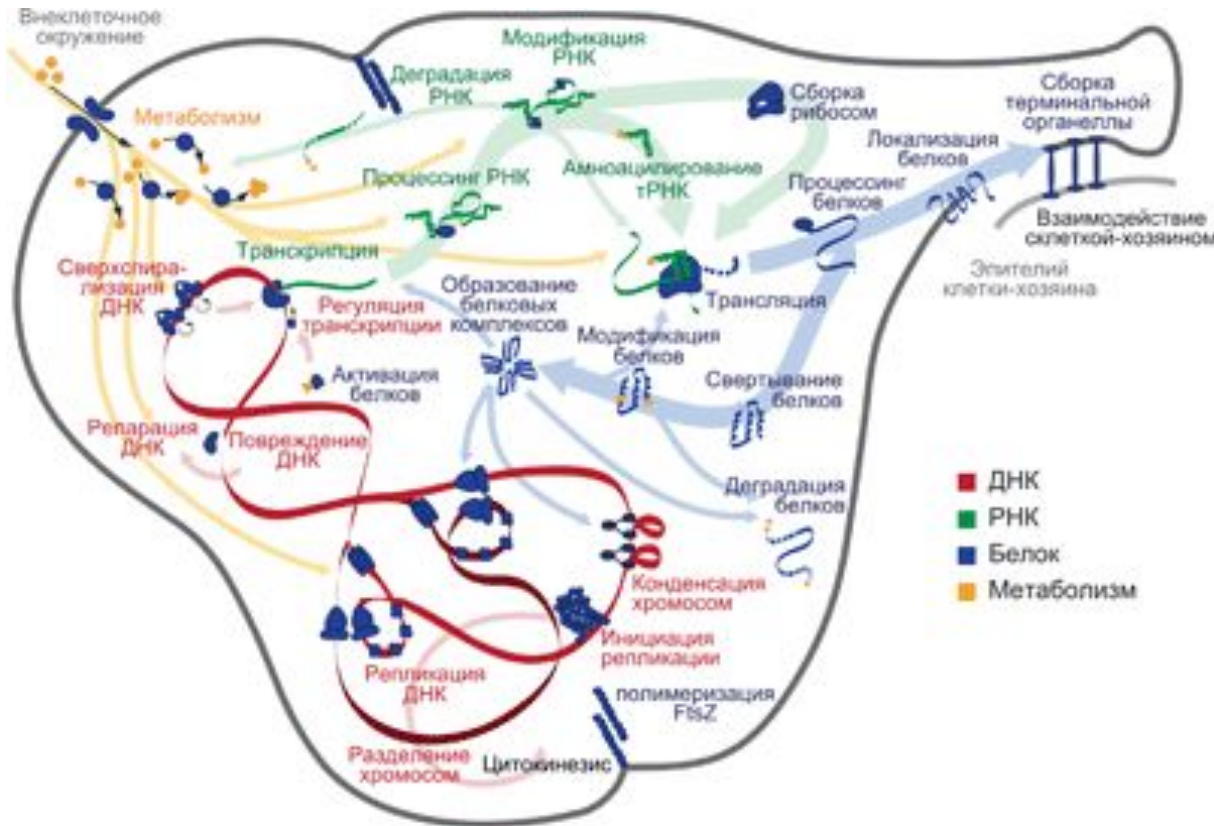
Для анализа полученных систем могут применяться математические методы нелинейной динамики, теории случайных процессов, либо использоваться теория управления.

Из-за сложности объекта изучения, большого количества параметров, переменных и уравнений, описывающих биологическую систему, современная системная биология немыслима без использования компьютерных технологий



**National Science Foundation (NSF) – своего рода  
фонд фундаментальных исследований США**

**среди задач биологии 21-го века поставил и  
серьезный вызов для системной биологии –  
построение модели функционирования целой  
клетки. Эта задача уже в какой-то степени  
решена**



Karr J.R., Sanghvi J.C., Macklin D.N., Gutschow M.V., Jacobs J.M., Bolival B., Assad-Garcia N., Glass J.I., Covert M.W. (2012).

A Whole-Cell Computational Model Predicts Phenotype from Genotype.

Cell 150, 389–401;

Модель клетки *Mycoplasma genitalium* как целого, которая состоит из 28 субмоделей различных клеточных процессов. Субмодели сгруппированы по категориям: ДНК, РНК, белки и метаболизм. Субмодели связаны друг с другом через общие метаболиты, РНК, белки и хромосомальную ДНК, что показано стрелками соответствующих цветов.

# **Системная биология (systems biology)**