

**МЕДИЦИНА И ТЕОРИЯ СЛОЖНЫХ СЕТЕЙ ЗАБОЛЕВАНИЙ ЧЕЛОВЕКА**

Евин И.А., Розин А.В. Мелерзанов А.В., Чернышова Д.И.

Московский физико-технический институт (государственный университет)

**Аннотация.** В статье дан краткий обзор основных понятий новой области медицины, изучающей взаимосвязи между болезнями человека методами теории сложных сетей. Узлами в сетях заболеваний являются различные заболевания человека, и два узла соединены связью, если соответствующим этим двум узлам заболеваниям имеются общие причины. Дано краткое описание метаболической и фенотипической сетей заболеваний. Рассматриваются понятия интерактом, модуль заболевания, дисизом. В заключение обсуждается вопрос приложения теории управления сетями к сетевой медицине.

**Ключевые слова:** нозология, сопутствующие заболевания, безмасштабная сеть, ведущие узлы

**Введение**

Важнейший фактор развития медицины – это постоянное совершенствование методов диагностики и терапии заболеваний человека. Нозология (раздел медицины о классификации болезней, условиях, причинах и закономерностях возникновения заболеваний и патологических процессов) до недавнего времени основывалась главным образом на фенотипах болезней – их симптомах. Однако, врачам известно, что у двух разных болезней могут быть очень схожие симптомы. [2,3,13,16] С другой стороны, бывает, что считавшиеся различными заболеваниями, при более тщательном исследовании могут иметь одну первопричину на молекулярном уровне.

Выход методов медико-биологических исследований на молекулярно-генетический уровень существенно изменил ситуацию в диагностике. Появилась возможность ставить диагнозы более точно, учитывать специфичность многих заболеваний. Возникла необходимость

пересмотреть всю существующую классификацию болезней, взяв за основу не их симптомы или вызванные ими физиологические изменения, а молекулярные причины их происхождения.

Фенотип болезни - ее симптом или набор признаков (синдром) редко является следствием мутаций одного или нескольких генов, а чаще всего отражает набор отклонений, которые имеют место в сложной сети молекулярных взаимодействий компонентов живой клетки. Такая сеть получила название интерактом (от английского глагола to interact – взаимодействовать и по аналогии с названием сети генов – геном). Узлами в интерактоме являются отдельные субъединицы молекулярных размеров: гены, белки и ферменты, субстраты метаболических реакций. Два узла связаны между собой, если соответствующие им молекулярные субъединицы взаимодействуют между собой. В таблице 1 даны приблизительные величины некоторых параметров известного в настоящее время интерактома человека [3,11].

Таблица 1.

Число узлов	13460
Число связей	141296
Среднее число связей у узла	21
Среднее значение коэффициента кластеризации	0.17
Среднее расстояние между узлами	3.6
Диаметр сети	12

Функция распределения узлов по числу связей в данном интерактоме описывается степенным законом (Рисунок 1).

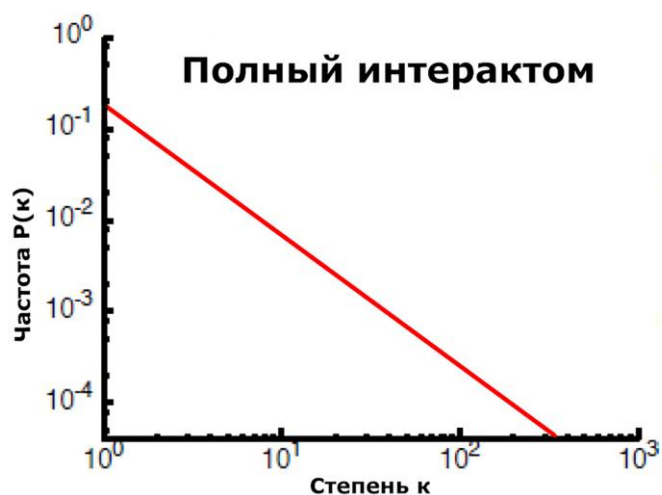


Рисунок 1. Распределение узлов по узлов по числу связей в интерактоме подчиняется степенному закону. Обе координаты на графике представлены в логарифмическом масштабе. Такие сети называются безмасштабными сетями [1].

Взаимозависимость между молекулярными субстратами клетки приводит к глубоким причинно-следственным связям между кажущимися различными фенотипами. Например, мутации в сходном наборе генов приводит у мальчиков к мышечной дистрофии Дюшенна и к сердечно-сосудистым заболеваниям у взрослых мужчин. Новая классификация заболеваний может помочь лучше понять природу болезни, ее взаимосвязь с другими болезнями, а также создать новые или подобрать среди имеющихся более эффективные лекарства. Существует множество препаратов для поддержания нормальной работы сердца, но лечение такого пациента лекарствами от мышечной дистрофии, может оказаться более адекватным и эффективным методом лечения сердечных приступов [13].

### **Теория сложных сетей и медицина.**

С возникновением в конце прошлого века теории сложных сетей как эффективного инструмента изучения сложных систем различной природы, в том числе биологии и медицины, а также развитием более точных и объективных методов диагностики заболеваний возникли условия для того, чтобы начать создавать полную карту взаимосвязи

различных заболеваний между собой. Появились первые сети заболеваний, в которых связи между различными заболеваниями (узлами сети) устанавливаются в том случае, если соответствующие заболевания имеют одинаковую причину на молекулярном уровне (например, мутацию одних и тех же генов) или нарушения в метаболических реакциях. Возникли сетевая медицина и сетевая фармакология. Взгляд на медицину с точки зрения теории сложных сетей облегчает выявление крупномасштабных последствий от различных локальных изменений, например, мутаций, и помогает устанавливать, иногда неожиданные причинно-следственные связи.

Сетевая медицина опирается на ряд фундаментальных принципов теории сложных сетей, которая стала важной составной частью набора инструментов исследования свойств сложной биологической природы. В частности, было показано, что биологические сети, как и многие другие реальные сети (технологические, социальные, когнитивные) не являются случайными сетями [1]. Распределения узлов по числу связей в таких сетях подчиняются степенным законам. Это означает, что в отличие от случайных сетей, в реальных сетях имеются узлы с большим числом связей - хабы (от английского hub – концентратор, накопитель). В сетях белковых взаимодействий и метаболических реакций живых клеток также имеются узлы с большим количеством связей. Белки и метаболические субстраты, соответствующие этим хамам (а их всегда небольшое число) играют особую биологическую роль. Имеющиеся данные свидетельствуют, что гены, кодирующие белки-хабы, являются наиболее старыми, и они эволюционируют более медленно, чем гены, кодирующие белки с небольшим числом связей.

Первым шагом в изучении взаимосвязей между сетями молекулярных взаимодействий и заболеваниями человека, а также связями заболеваний между собой, должна быть оценка, насколько полными и точными являются нынешние молекулярные и фенотипические сети человека. В последние годы мы являемся свидетелями больших усилий в расширении

диапазона исследований в целях повышения точности и объема охвата сети интерактома человека, в том числе для оценки ее размера. При этом важно иметь в виду, что большая часть текущей работы в изучении сетей организма человека сфокусирована на внутриклеточных сетях, игнорируя, в связи с отсутствием систематических данных, молекулярные сети, которые соединяют между собой клетки, ткани и органы.

### Модули заболеваний человека

Наблюдения показывают, что, если выявлены несколько компонентов вызывающих болезнь (белки, ферменты, и т.п.) существует большая вероятность, что в ближайшей окрестности интерактома можно найти и другие компоненты также относящиеся к данной болезни. То есть, можно ожидать, что каждое заболевание связано с четко определенной окрестностью интерактома, называемой "модуль заболевания". Выявление модуля заболевания есть первый шаг в понимании молекулярных механизмов болезни, а возможное перекрытие между двумя такими модулями объясняет существование связи между соответствующими заболеваниями (сопутствующие заболевания).

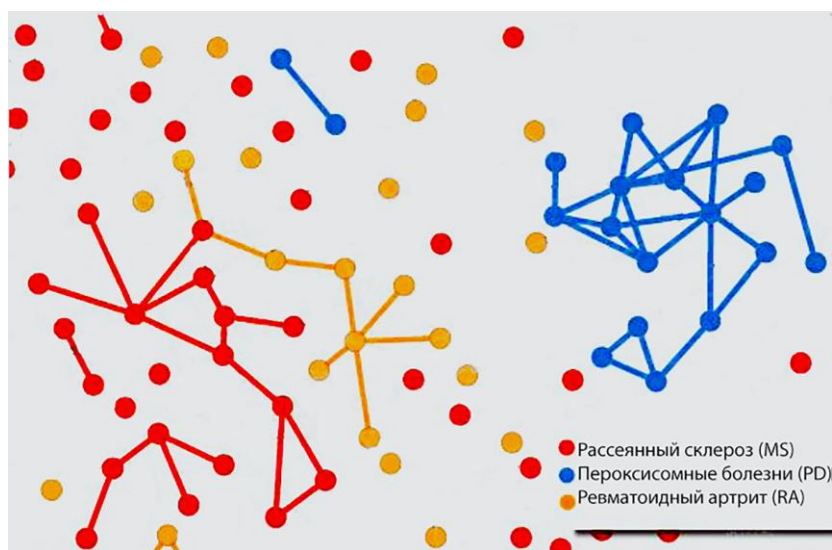


Рисунок 2. Схематичное представление модулей трех заболеваний: рассеянного склероза, ревматоидного артрита и пероксисомных болезни. Белки, относящиеся к одному определенному заболеванию, образуют связанный подграф, называемый модулем

заболевания. Пары заболеваний с перекрывающимися модулями заболеваний (рассеянный склероз и ревматоидный артрит) имеют большое число сходных симптомов (синдромов) [11]. Такие заболевания называют сопутствующими заболеваниями.

Goh et al [5] в 2007 году построили двудольный граф состоящий из двух до того времени несвязанных множеств: 1284 заболеваний генетической природы и 1777 генов заболеваний (disease genes) в геноме человека. Перечень заболеваний генетического происхождения, гены заболеваний и связи между ними были взяты в базе данных Online Mendelian Inheritance in Man (OMIM) [7,12]. На основе этого двудольного графа были построены две сетевых проекции: сеть заболеваний человека (human disease network – HDS) или дисизом, (от английского disease – болезнь, заболевание), в которой узлами являются болезни и два узла связаны связью если у них есть по крайней мере один общий ген мутации которого связаны с обоими заболеваниями. Другой проекцией двудольного графа является сеть генов заболеваний (disease gene network – DGN). Узлами этой сети являются гены заболеваний и два гена связаны если они вызывают одну и ту же болезнь.

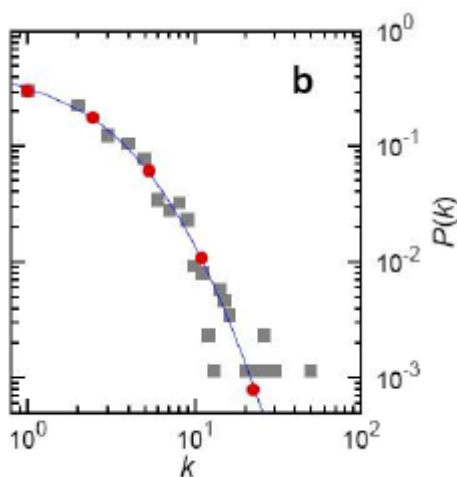


Рисунок 3. Распределение числа заболеваний с которыми данное заболевание имеет общий мутантный ген в сети заболеваний человека [5].

На рисунке 3 представлено распределение по числу связей заболеваний, означающее, что большинство болезней вызвано небольшим числом генов. Однако, ряд заболеваний, такие как глухота, имеют десятки связей (глухота – 41, лейкемия – 37, рак толстой кишки – 34). Распределение по числу связей заболеваний показывает, что большинство болезней связаны с небольшим числом других болезней, но есть болезни-хабы, такие как рак толстой кишки (связан с 50 другими болезнями) и рак молочной железы (также 50). Преобладание рака среди хабов сети заболеваний человека объясняется в частности тем, что многие клинически различимые виды рака тесно связаны друг с другом через общие подавляющие опухоли гены (tumor repressor genes) такие как *TP53* *PTEN*.

Из 299 исследованных заболеваний 20 связаны генетически, а 226 имеют более широкие молекулярные связи в интерактоме. Общие гены не дают полной информации о связях между болезнями. Необходимо исследовать и находить такие связи в более широком плане – на молекулярном уровне. Выяснилось, например, что бронхиальная астма и заболевание тонкой кишки (целиакия), респираторное и аутоиммунное заболевания имеют общие причины на молекулярном уровне, несмотря на довольно различные симптомы [5]. Такие связи между болезнями обнаружены в значительной мере благодаря методам теории сложных сетей.

### **Сети заболеваний на основе метаболических реакций**

Метаболические заболевания могут быть связаны через метаболиты и общие метаболические реакции. Чтобы исследовать эту гипотезу, Lee *et al* сначала построили сеть метаболических реакций на основе информации, взятой из двух баз данных: the Kyoto Encyclopedia of Genes Genomes (KEGG) [8] и Biochemically Genetically and Genomically (BiGG), содержащих хорошо известные метаболические реакции с описанием участвующих в них метаболитов [4]. В дополнении к этому была использована база данных OMIM [12] содержащая связи «ген-болезнь». Заключительным этапом этого исследования было

конструирование сети метаболических заболеваний, в которой две болезни соединились связью, если связанные с этими болезнями гены были связаны в сети метаболических реакций общим метаболитом или метаболиты участвовали в общей реакции. Таким образом, болезни, связанные с обменом веществ (метаболическими реакциями) напрямую не вызваны мутациями в генах. Природа таких болезней связана с нарушениями в сложной сети биохимических реакций - основного источника клеточной активности. Исследования показали, что эти нарушения, скорее всего, связаны с ферментами – катализаторами биохимических реакций и что именно дефектный фермент отвечает за соответствующее определенное заболевание. Ферментативный дефект, который влияет на течение одной реакции, может отрицательно повлиять на течение всех последующих реакций в той же цепи. Два заболевания связаны, если связанные с этими заболеваниями гены связаны в метаболической сети общим метаболитом или несколькими метаболитами.

Сеть метаболических заболеваний (Metabolic Disease Network – MDN) содержит гигантскую связанную компоненту из 197 заболеваний состоящую из нескольких кластеров (модулей) таких как сахарный диабет, гипертония, ожирение и заболеваний коронарной артерии [9].

Статистические исследования сети заболеваний на основе метаболических реакций показывают, что в среднем - одно заболевание имеет примерно пять связей с другими заболеваниями, а распределение узлов по числу связей с другими узлами значительно шире чем у случайной сети с таким же числом узлов и связей. Это указывает на то, что существует различие между разными заболеваниями на основе метаболизма. Например, гипертония является хабом с числом связей равным 27, однако большое число болезней связано с другими болезнями небольшим числом связей.

В результате масштабных исследований более 13 миллионов пожилых пациентов американских госпиталей было показано, что около 17% всех заболеваний вызванных нарушениями метаболизма имеют сопутствующие заболевания [9].



## Фенотипическая Сеть Болезней (Phenotypic Disease Network (PDN))

Немало полезной информации можно получить, применяя сетевые методы при анализе фактических данных по заболеваниям на фенотипическом уровне, в виде истории болезней пациентов. Эти данные содержат информацию о связях различных заболеваний между собой и о прогрессировании этих заболеваний.

Чтобы понять связи болезней, каждую из них определяют как специфический набор фенотипов, влияющих на одну или несколько функциональных систем. При этом подразумевается, что живой организм может быть описан в сложном фенотипическом пространстве как связанная из отдельных элементов дискретная система. Для этого в 2009 году была построена Фенотипическая Сеть Болезней (Phenotypic Disease Network (PDN)) на основе изучения корреляций историй болезней более тридцати миллионов пациентов [6.14].

Было установлено, что болезни прогрессируют и развиваются преимущественно вдоль связей этой сети и, более того, было выявлено, что такое развитие различно для пациентов разного пола и различных рас. Было установлено также, что пациенты, страдающие заболеванием, имеющими большое число связей в Фенотипической Сети Болезней умирают раньше, чем пациенты с заболеванием с небольшим числом связей. В этом исследовании была создана онлайн-запрашиваемая база данных из 18 различных наборов данных, полученных из истории болезней от более 31 миллиона пациентов. Связи (сообщества) заболеваний могут быть изучены в режиме онлайн или загружены в полном объеме.

Обозначим  $RR_{ij} = \frac{C_{ij}N}{P_iP_j}$  - (Relative Risk of disease pairs  $i$  and  $j$ ) относительный риск

заболеть двумя заболеваниями для одного и того же пациента,

где  $C_{ij}$  - число пациентов, страдающих обоими заболеваниями,  $N$  - полное число пациентов,  $P_i$  и  $P_j$  число пациентов с заболеваниями соответственно  $i$  и  $j$ , то есть распространенность заболеваний  $i$  и  $j$ .

Корреляция Пирсона  $\Phi_{ij}$  как мера связи между двумя бинарными переменными

$$\Phi_{ij} = \frac{C_{ij}N - P_i P_j}{\sqrt{P_i P_j (N - P_i)(N - P_j)}}.$$

Эти две меры различия между сопутствующими заболеваниями (comorbidity) не являются абсолютно независимыми. Например, они обе растут с ростом числа пациентов, страдающих обоими заболеваниями.

При этом сеть, построенная на основе  $RR_{ij}$  связей, переоценивает расстояния между редкими заболеваниями, в то время как  $\Phi_{ij}$  недооценивает расстояния между парами редких заболеваний.

### **Динамические свойства фенотипической сети заболеваний**

PDN сеть представляет собой своеобразную «карту» заболеваний, по которой можно изучать их развитие. Это развитие можно рассматривать как динамический сетевой процесс, когда пациенты «перепрыгивают» от одного заболевания к другому вдоль связей PDN сети.

Основываясь на имеющихся статистических данных о влиянии половых и расовых различий на RR связи PDN сети, была построена сетевая структура, демонстрирующая различие в связях между заболеваниями у белых и афроамериканских мужчин [9,10].

В сети PDN узлами являются фенотипы заболеваний, описываемые уникальными ICD9 кодами, состоящим из 5 цифр. Первые три цифры обозначают основную категорию заболевания, а оставшиеся две цифры дают дополнительную информацию о заболевании. В совокупности, ICD9 классификация состоит из 657 различных категорий на уровне 3 цифр и 16459 категорий на уровне 5 цифр.

Используя только наиболее сильные и важные связи между узлами в сети PDN, предлагается два упрощенных варианта этой сети: первый на основе RR связей и второй на основе корреляционных  $\Phi$  связей.

У этих двух версий есть некоторые сходства, например, в обоих выявляется связь между гипертонией и почечными заболеваниями. При этом в первой версии мы видим сравнительно редко встречающиеся заболевания, и в ней присутствуют четко различающиеся сообщества (модули), в то время как во втором варианте PDN сети больше широко распространенных заболеваний и связей между разными категориями ICD9 кода.

Идеи и методы этого исследования были развиты для построения сети метастаз и создания сетевой модели развития ракового заболевания [6,14].

### **Выявление заболеваний из неполного интерактома**

Предсказательные возможности современных подходов, основанных на теории сложных сетей, ограничены множеством причин. Во-первых, сегодня нам известно лишь около 20% общей структуры интерактома, и поэтому в поисках механизмов заболеваний мы вынуждены опираться на пока еще очень неполные данные о сети молекулярных взаимодействий. Во-вторых, как уже упоминалось выше, продукты мутаций генов – соответствующие белки – рассеяны по интерактому не случайным образом, а имеют тенденцию взаимодействовать друг с другом, формируя один из нескольких подграфов (кластеров), называемых модулями заболевания.

В работе [11] даны примеры взаимодействий модулей заболеваний для рассеянного склероза (multiple sclerosis (MS)), пароксизмальных расстройств (peroxisomal disorders(PD)) и ревматоидного артрита (rheumatoid arthritis (RA)). Заболевания с перекрывающимися модулями (MS и RA) имеют фенотипические сходства и высокую вероятность быть

сопутствующими заболеваниями. Непересекающиеся модули заболеваний, такие как MS и PD, не имеют выявленных общих клинических симптомов.

Однако, несмотря на частое использование или просто упоминания в научных дискуссиях и докладах, гипотеза о модулях заболевания не имеет конкретных алгоритмов и правил их выявления и построения, а также сколько-нибудь надежного обоснования с точки зрения современной теории сложных сетей.

### **Фрагментация модулей заболеваний**

Построение интерактома было начато с составления сети из 141296 физических взаимодействий между 13460 белками, экспериментально задокументированных в клетках человека. Сюда вошли белок-белковые, регуляторные взаимодействия и метаболические реакции. Также был составлен список 299 заболеваний из медицинского рубрикатора заболеваний (Medical Subject Headings (MeSH)), содержащий не меньше 20 связанных генов на основе данных, содержащихся в Mendelian Inheritance in Man (OMIM) и баз данных genome-wide association study (GWAS), включающих в себя 2436 белков, связанных с болезнями. Несмотря на некоторый прогресс, значительная часть интерактома из списка болезней, связанных с заболеваниями генов, остается несвязанной.

Например, для рассеянного склероза из 69 белков всего 11 образуют связанный подграф, остальные же случайным образом рассеяны по интерактому.

Эта картина имеет место для всех 299 заболеваний. Их выявленные модули составляют около 20% соответствующих генов болезни.

Существует несколько причин фрагментации интерактома как сетевой структуры. Основной причиной является недостаточность данных: отсутствующие связи оставляют многие связанные с заболеваниями белки изолированными в их модулях заболеваний. Методы теории сложных сетей позволяют вычислить предельный порог неполноты данных интерактома, начиная с которого появляется возможность идентифицировать модуль

заболевания. Вычисления показали, что, согласно теории перколяции, для заболевания, причиной которого было меньше 25 генов, его модуль слишком фрагментирован, чтобы быть наблюдаемым в интерактоме.

Для количественной оценки степени перекрытия двух модулей заболевания вводится понятие расстояния между модулями

$$s_{AB} \equiv \langle d_{AB} \rangle - \frac{\langle d_{AA} \rangle + \langle d_{BB} \rangle}{2}$$

$s_{AB}$  сравнивает кратчайшие расстояния между белками внутри каждого модуля заболевания,  $d_{AA}$  и  $d_{BB}$  с кратчайшими расстояниями  $d_{AB}$  между  $A-B$  белковыми парами. Белки связанные как с  $A$  так и с  $B$  имеют  $d_{AB} = 0$ . Если  $s_{AB} < 0$ , то соответствующие модули заболеваний перекрываются. Если  $s_{AB} > 0$ , то модули расположены в интерактоме отдельно.

В таблице 2 даны примеры количественных характеристик для нескольких пар сопутствующих заболеваний, то есть болезней с перекрывающимися модулями заболеваний.

Таблица 2. Основные показатели для двух заболеваний с перекрывающимися модулями внутри интерактома.  $N_A$  и  $N_B$  число генов, связанных с болезнями  $A$  и  $B$  соответственно,  $N_A \cap N_B$ , число общих генов генов.

Болезнь $A$	Болезнь $B$	$N_A$	$N_B$	$N_A \cap N_B$	Общие гены	$s_{AB}$
Бронхиальная	Целиакия	37	37	3	HLA-	-0.13

<b>астма</b>					<b>DQA1, IL1RL1, IL18R1</b>	
<b>Гломерулонефрит</b>	<b>Биллиарный цирроз печени.</b>	<b>18</b>	<b>23</b>	<b>2</b>	<b>HLA- DQB1, HLA- DQA2</b>	<b>-0.05</b>
<b>Псориаз</b>	<b>Васкулит</b>	<b>54</b>	<b>15</b>	<b>9</b>	<b>IER3, DDR1, MICA, IL23R, HL- B5, HLA-C, PSORS1C2, C6orf15, CCHCR1</b>	<b>-0.35</b>
<b>Болезнь Грейвса</b>	<b>Васкулит</b>	<b>13</b>	<b>15</b>	<b>2</b>	<b>C6orf15, MICA</b>	<b>-0.3</b>

### **Сетевые модели отдельных заболеваний**

В статье [8] на основе многочисленных биологических баз данных построены сетевые структуры белок-белковых взаимодействий для клеток поджелудочной железы при сахарном диабете второго типа, проанализированы свойства состояний в норме и патологии и сделаны сравнения со свойствами случайных сетей Эрдеша-Реньи, сетей «тесного мира» Уоттса-Строгатца, а также сетей, построенных по конфигурационной модели, в которых сохраняется последовательность степеней узлов, следовательно, сохраняется и степенной закон сетей (scale free networks).

Распределение числа узлов по числу связей в сетях белок-белковых взаимодействий в клетках поджелудочной железы в норме и патологии является степенным. Такие сети со степенным законом распределения узлов по числу связей получили название безмасштабных сетей.

Таблица 1. Основные характеристики двух сетевых структур белок-белковых взаимодействий в норме ( $N_1$  и  $N_2$ ) и патологии ( $D_1$  и  $D_2$ ) при заболевании сахарным диабетом второго типа.  $N$  - общее число узлов (белков) в сети,  $N_c$  - общее число связей в сети,  $\langle k \rangle$  - среднее число связей узла сети,  $D$  - диаметр сети,  $\langle CC \rangle$  - среднее значение кластеризации сети.

Сеть	Число узлов	Число связей	Среднее число связей узла	Диаметр сети	Коэффициент кластеризации
$N_1$	2083	11017	10	17	0,35
$N_2$	1707	9888	11	12	0,33
$D_1$	656	3628	10	10	0,26
$D_2$	384	1882	10	9	0,46

Анализ свойств сетей белковых взаимодействий в норме и патологии показывает, что сети здоровых клеток поджелудочной железы являются более случайными, чем сети больных клеток и такое отсутствие достаточно высокой случайности сети, по –видимому и ведет к заболеванию.

Для подтверждения этой гипотезы целесообразно сравнить энтропии сетей белковых взаимодействий здоровых и больных клеток.

. В работе [ ] проанализированы некоторые методологические проблемы приложения сетевой идеологии в изучении раковых заболеваний и построена сеть метастаз рака толстой кишки . Узлам сети соответствуют анатомические участки с метастазами организма человека и два узла связаны, если наблюдается совместное появление метастаз в

соответствующих участках. Показано, что сетевые модели некоторых видов рака дает более точные предсказания, чем традиционные методы математической статистики.

В работе [17] на основе баз данных GEO и Array Express, полученных с помощью ДНК - микрочипов, была построена сеть ко-экспрессии генов для исследования транскриптома глиальной опухоли мозга человека. База данных GEO (Gene Expression Omnibus) создана в рамках американского биотехнологического центра National Centre for Biotechnology ([www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/)), а база данных Array Express при Европейском биоинформационном институте ([www.ebi.ac.uk/microarray-us/aer/](http://www.ebi.ac.uk/microarray-us/aer/)).

Сети ко-экспрессии генов строятся на основе профилей экспрессии генов представленных в виде таблиц, в которых генам соответствуют строки, а образцам – столбцы. Производится оценка сходства профилей на основе коэффициента корреляции Пирсона. Результатом такой оценки и является сеть коэкспрессии, в которой узлами являются гены и гены связаны между собой, если значение соответствующего коэффициента Пирсона выше некоторого порогового положительного значения, но меньше единицы. Это значение берется за вес связи.

Для верификации предсказаний функций генов в работе используется база данных Human Protein Atlas ([www.proteinatlas.org](http://www.proteinatlas.org)), в которой содержатся описания локализации белков человека в различных тканях и типах клеток. В качестве объекта для исследования была выбрана клеточная органелла ресничка.

### **Заключение. Сетевая медицина и управление в сетях.**

В последние годы в рамках общей теории сложных сетей развивается новое научное направление – управление в сетях [10].

Согласно теории управления, система управляема, если она управляема из любого начального состояния до любого желаемого конечного состояния по возможным переменным входа. Здесь мы рассмотрим линейные системы и предположим, что индивиды



в системах не имеют никакой внутренней динамики. Тогда линейные системы управления описываются следующим уравнением.

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t) \quad (1)$$

где  $\dot{\mathbf{x}}(t) = \{\dot{x}_1(t), \dot{x}_2(t), \dots, \dot{x}_N(t)\}^T$  есть состояние системы  $N$  узлов в момент времени  $t$ .

Матрица  $A$ , размером  $N \times N$ , есть матрица смежности сети, представляющей систему. В этой матрице элементами являются сильные взаимодействия между отдельными узлами. Матрица  $B$ , размером  $N \times M$ , есть матрица ввода, которая описывает узлы, где входящий сигнал является наложенным. Вектор входного сигнала  $\mathbf{u}(t) = \{u_1(t), u_2(t), \dots, u_M(t)\}^T$ , как можем видеть, зависит от времени. Состояние каждого узла в каждый момент времени контролируется линейной комбинацией элементов вектора входа. Система, определяемая формулой (1) называется управляемой, тогда и только тогда, когда  $N \times NM$  управляемая матрица:

$$C = (B, AB, A^2B, \dots, A^{N-1}B) \quad (2)$$

имеет полный ранг, который равен

$$\text{rank}(C) = N \quad (3)$$

Это условие управляемости сети предложил Калман [7].

Математическое условие управляемости системы состоит в том, что для того, чтобы система была управляема, нужно выбрать такую матрицу  $B$  и такой вектор  $\mathbf{u}(t)$ , чтобы выполнялось условие  $\text{rank}(C) = N$ . Возникает вопрос: как именно определить минимальное количество  $M$  ( $M < N$ ) различных ведущих (управляющих) узлов, воздействия на которые дают нам полный контроль над сетью. Критерий Калмана эффективен для небольших сетей, в то время как для реальных сетей весьма трудоемко высчитать ранг  $C$ , так как матрица  $C$  обычно довольно большая. Лиу и др. предложили алгоритм, с помощью которого можно определять все ведущие узлы сети [10].

В недавней работе [15] по изучению управляемости ориентированной сети белковых взаимодействий у человека было обнаружено, что белки, соответствующие определенным узлам, в наибольшей степени подвержены мутациям, вызывающих заболевания и воздействию вирусов находящихся в человеке, а при лечении от заболеваний являются потенциальными мишенями для лекарств. Предложен алгоритм выявления таких узлов в ориентированных сетях. Важной задачей является обобщение такого алгоритма на неориентированные сети и выявление таких узлов в сетях заболеваний.

### Литература

1. Barabasi A.-L. Network Science. <http://barabasi.com/networksciencebook/>
2. Barabasi A.-L., Gulbahce N., Loscalzo J. Network medicine: a network based approach to human disease. Nature Review. Genetics. vol.12, January 2011, 57
3. Barabasi, A.-L. Network medicine – from obesity to the “diseaseome”. N.Engl.J.Med. 257:4, July 2007
4. Duarte, N.D., Becker S.A., Jamshidi N., Thiele I., Mo M.L., Vo T.D., Srivas R., Palsson B. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. PNAS 104: 1777- 1782
5. Goh K., Cusick M.E., Valle D., Childs B., Vidal M., Barabasi A.-L. Human disease network. PNAS. May 22, 2007, vol. 104, no. 21, 8687
6. Hidalgo C. A., Blumm, N., Barabasi A.L.& Christakis, N. A dynamic network approach for the study of human phenotypes. PLoS. Comput.Biol. 5, e1000353 (2009)

7. Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., McKusick, V.A. (2005) *Nucleic Acid Res.* 33:D514-D517
8. Kanehisa M., Goto S., Hattori M., Aoki-Kinoshita K.F., Itoh M., Kawashima S., Katayama T., Araki M., Hirakawa M. (2006) From genomics to chemical genomics. *Nucleic Acid Res.* 34:D3354-D3357
9. Lee, D.-S., Park, J., Kay, K.A., Christakis, N.A., Olivai, Z.N., and Barabasi, A.-L. The implication of human metabolic network topology for disease comorbidity. *PNAS.* 105, no29, 9880-9885 (2008)
10. Liu, Y.-Y., Slotine, J.J., and Barabasi, A.L. Controllability of Complex Networks. *Nature*, 473, 167-173. (2011)
11. Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., Barabási, A.-L. Uncovering disease-disease relationships through the incomplete interactome. *Science.* 2015 • v. 347 ISSUE 6224, pp.841-848
12. Online Mendelian in heritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acid Res.* 33:D514-D517
13. Pollak, A. Mapping the Human 'Diseaseome'. *The New York Times.* May, 6, 2008
14. Rzhetsky, A., Wajngurt, D., Park, N & Zheng, T. Probing genetic overlap among complex human phenotypes. *PNAS.* **104**, 11694-11699 (2007).
15. Vinayagam A., Gibson T., Lee H.J., Yimase B., Roesel Ch., Hu Ya., Kwon Yo., Liu Y., Perrimon N., and Barabasi A.L. Controllability analysis of the direct human protein interaction network identifies disease genes and drug targets. *PNAS Early Edition*, 2016
16. Zhou X. Z., Mencht J., Barabasi A.L., Sharma A.. Human symptoms-disease network. *Nature Communications.* 2014, 5-4212.DOI: 10.1038  
, scale-free network, driver nodes
17. Ivliev A.E. P.A. t Hoen, M.G. 2011, *Sergeeva Cancer Research.* **70**, 10060-70

44. E-mail: [chernyshova@phystech.edu](mailto:chernyshova@phystech.edu)