

[Переведено с английского, оригинал доступен по адресу:
http://www.ibmcm.sk.ru/content/intelligence/RHUPRO_Roadmap.pdf]

Министерство образования и науки РФ

Рабочая группа по международному Проекту «Протеом человека»
Российское протеомное общество (RHUPRO)

Дорожная карта

**Протеом 18-ой хромосомы человека: геноцентричный подход к
идентификации транскриптов, белков и пептидов**

(31 мая 2010 г.)

Москва, 2010

(с) 2009-2010, Рабочая группа по проекту «Протеом человека»

- Минобрнауки России;
- Минздравсоцразвития России;
- Департамент науки, технологии и образования при правительстве РФ;
- Российская академия наук;
- Российская академия медицинских наук;
- Российское Протеомное Общество (RHUPO)

www.proteome.ru

www.ibmc.msk.ru

(с) Арчаков А.¹, Быков В.², Говорун В.³, Иванов В.³, Хлунов А.⁴, Лисица А.¹,
Макаров А.⁵, Сагдеев Р.⁶, Скрыбин К.⁷

1 - Институт биомедицинской химии им. В.Н. Ореховича (Москва)

2 - ЗАО НТ-МДТ (Зеленоград)

3 - Институт биоорганической химии им. М.М. Шемякина и Ю.А.Овчинникова
(Москва)

4 - Департамент науки, технологии и образования при правительстве РФ

5 - Институт молекулярной биологии им. В.А.Энгельгарда (Москва)

6 - Международный центр томографии (Новосибирск)

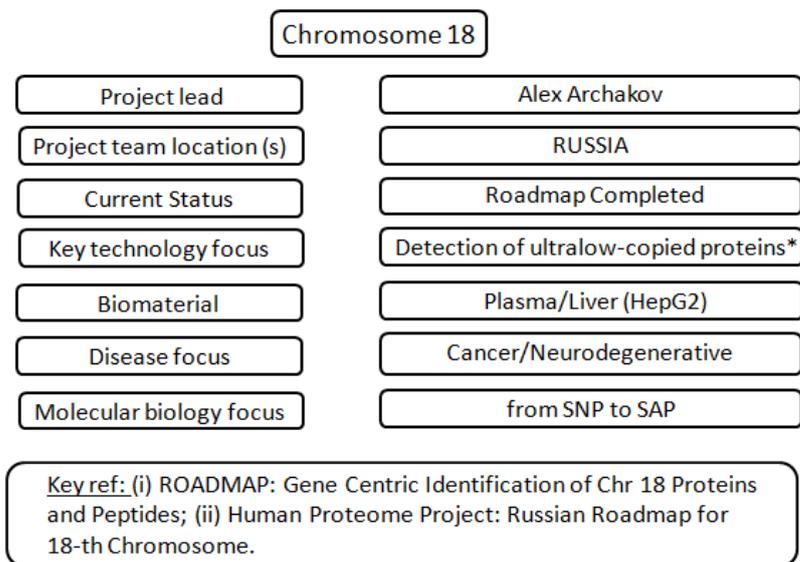
7 - Центр «Биоинженерия» (Москва)

Аннотация

Проект «Протеом человека» (ППЧ) направлен на инвентаризацию всех белков человека и выявление взаимодействий между ними. По своему масштабу, он превосходит проект «Геном человека», формально заверченный в 2001 г. Основные проблемы, стоящие перед ППЧ: невозможность детектирования единичных молекул в биологическом материале и ситуационный характер протеома, т.е. зависимость белкового состава от времени, типа ткани и клеток. Благодаря наличию единичных замен аминокислотных остатков (SAP), альтернативного сплайсинга (AS) и посттрансляционных модификаций (PTM), в составе полного генома можно предполагать присутствие нескольких миллионов различных белков.

На данный момент создана дорожная карта по проекту в виде Российского национального геноцентричного подхода. На предварительном этапе работ по проекту будут выявлены транскрипты 18-ой хромосомы в клетках печени и 286 белков в клетках печени и в плазме крови, а также будут разработаны новые технологии, такие, как технология на базе сопряжения АСМ и МС, аналитическое устройство с использованием нанопроводного детектора, информационно-аналитическая система распределенного (облачного) типа на платформе персональных супер-ЭВМ со специализированным программным обеспечением.

Текущая стадия экспериментальной реализации проекта включает идентификацию 220 транскриптов, кодирующих белки 18-ой хромосомы в HepG2, 90 белков 18-ой хромосомы, обнаруженных методом ионной ловушки 3D-LC-MS/MS в HepG2, а также выявление 36 белков в плазме крови. С помощью биоинформационных методов характеризуются белки 18-ой хромосомы, в том числе, оценивается медицинская значимость белков и потенциальные белки-партнеры.



* Down to concentration 10^{-16} M in plasma and ~ 10 copies per liver (HepG2) cell

Оглавление

1. Введение	5
2. Геноцентричный подход.....	6
3. График реализации проекта	8
4. Выбор хромосомы	9
5. Хромосома №18.....	11
6. Биоматериал: ткань печени и плазма крови	12
7. Стандартизация и протеотипирование	16
8. Анализ транскриптов и высоко/среднекопийных белков	20
9. Конструирование лекарств на базе пептидома.....	22
10. Детектирование низко/ультранизкокопийных белков	23
11. Интерактомика.....	25
12.База знаний.....	27
13. Краткосрочные перспективы (5-ти летний план).....	28
14.Долгосрочные перспективы (15-ти летний план)	29
15. Заключение.....	30
Список литературы.....	33

1. Введение

По аналогии с проектом «Геном человека», который позволил расшифровать гены в составе генома, конечная цель проекта «Протеом человека» (ППЧ) - осуществить инвентаризацию всех белков с помощью этих генов. Однако, даже самые убежденные сторонники протеомики сомневаются, что задача инвентаризации всех белков может быть выполнена в ближайшее время. Во-первых, существующие в протеомике технологии не позволяют выявлять единичные копии белковых молекул в биологическом материале. В геномике для таких случаев применяется амплификация путем полимеразной цепной реакции (ПЦР) [1]. Во-вторых, если геном, по большей части, одинаковый в каждой клетке организма, то белковый состав существенно меняется в зависимости от биологической жидкости, типов клеток и тканей. В третьих, существует огромное количество как трансляционных, так и посттрансляционных модификаций белковых молекул, то есть на 1 ген приходится широкое разнообразие различных белковых форм.

Кроме того, геном является константным параметром организма; он, как правило, постоянен, не зависимо от физиологических или патологических состояний и его последовательность не зависит от времени. Протеом же, наоборот, зависит от времени, физиологических стимулов и поэтому является ситуационным [2]. Состав протеома крайне изменчив и, в зависимости от времени, содержание белков и их модифицированных вариантов резко меняется. Правомочен вопрос, являются ли вышеуказанные проблемы непреодолимым препятствием для начала работ по проекту «Протеом человека»?

Для ответа на этот вопрос целесообразно провести аналогию с проектом «Геном человека», который формально был завершен в 2001 году [3]. Это был усредненный геном 6-ти индивидуумов, в составе которого были выявлены некоторые из экспрессированных участков, за исключением регулирующих элементов [4]. Однако с тех пор произошли большие изменения: многие геномы расшифрованы и общедоступны и в ближайшее время появятся тысячи других (геномов) [5]. Проведен анализ индивидуальных особенностей геномов у больных и здоровых людей, с целью выявления сплайс-вариантов и однонуклеотидных полиморфизмов. Однако, работа по генотипированию далеко не завершена, поскольку полная расшифровка целого генома человека не вполне пригодна для массового применения. Несмотря на окончание проекта, знания о геноме в основном исчерпываются экспрессируемой частью генома, то есть той частью генома, которая реализуется в виде экспрессии белков через стадию транскрипции, а это всего не более 1,5% от полного генома человека. Свойства и назначение остальных 98% генома остаются неизученными.

На самом деле, изложенная выше неоднозначная ситуация с «полностью расшифрованным» геномом человека [6] позволяет ответить на вопрос, нужно ли приступать к исследованию протеома. Очевидно, что даже задача изучения

полного генома человека, являясь несравнимо более легкой, чем инвентаризация протеома, до сих пор остается нерешенной в полной мере. Однако, сегодня можно говорить о ее решении только благодаря программе «Геном человека», стартовавшей около 20 лет назад. Именно вследствие необходимости создания аналогичного долгосрочного проекта, специалисты призывают начать работу по проекту «Протеом человека», несмотря на ясное понимание того, что задача не будет решена в ближайшее время.

Начало работы над проектом «Протеом человека» позволит участникам определить задачи и разработать более или менее общие подходы для реализации этих задач. В настоящее время нет единого методического подхода к идентификации полного протеома. Наиболее перспективным является так называемый «геноцентричный подход» [7]: белки идентифицируются путем проецирования результатов масс-спектрометрического анализа биоматериала на весь геном. Геноцентричный подход позволяет сконцентрировать внимание каждого участника проекта на определенной части целого генома, хромосоме, чтобы провести детальный анализ белков, кодируемых генами выбранной хромосомы.

2. Геноцентричный подход

На самом деле, геноцентричный подход, направленный на исследование белков с проекцией на гены, используется повсеместно. Классическая протеомика основана на знаниях о геномах, поэтому переход от анализа всех генов к анализу генов отдельной хромосомы является простым сужением задачи по сравнению с доминирующей сегодня геномо-центричной парадигмой. Сужение научной задачи за счет геноцентричности делает ее более реалистичной по сравнению с альтернативным подходом тканеспецифичной протеомики [8], состоящем в анализе белков отдельных органов и тканей на основе информации о целом геноме. По сравнению с геноцентричным подходом, инвентаризация всего протеома в различных органах и тканях — более амбициозная задача, которая кажется не реальной в ближайшем будущем.

Применение геноцентричного подхода позволит сконцентрировать усилия на решении важных технологических задач. Прежде всего, необходимо решить проблему концентрационной чувствительности аналитических методов протеомики, снять ограничение на выявление низко- и ультранизкокопийных¹ белков в биологическом материале [9]. В отсутствие аналога ПЦР для белков,

¹ Дорожная карта устанавливает диапазоны концентрации белков в биоматериале, которые соответствуют пределам чувствительности аналитических методов, используемых в настоящее время [10]. Это: высококопийные ($\geq 10^{-6}$ M), среднекопийные (10^{-7} -- 10^{-10} M), низкокопийные (10^{-11} -- 10^{-14}) и ультранизкокопийные ($< 10^{-14}$ M). «Copied» используется вместо «abundance», чтобы подчеркнуть технологическую направленность дорожной карты на детектирование единичных молекул.

концентрационный предел детектирования (DL) становится реальной проблемой [10]. В результате, лишь небольшая часть имеющихся белков, представленных в высоко/среднекопийном диапазоне доступна для протеомной инвентаризации в настоящее время (Рис. 1). Вместе с тем, логично ожидать, что биомаркеры ранних стадий заболеваний будут присутствовать в низко- и ультранизкокопийных диапазонах.

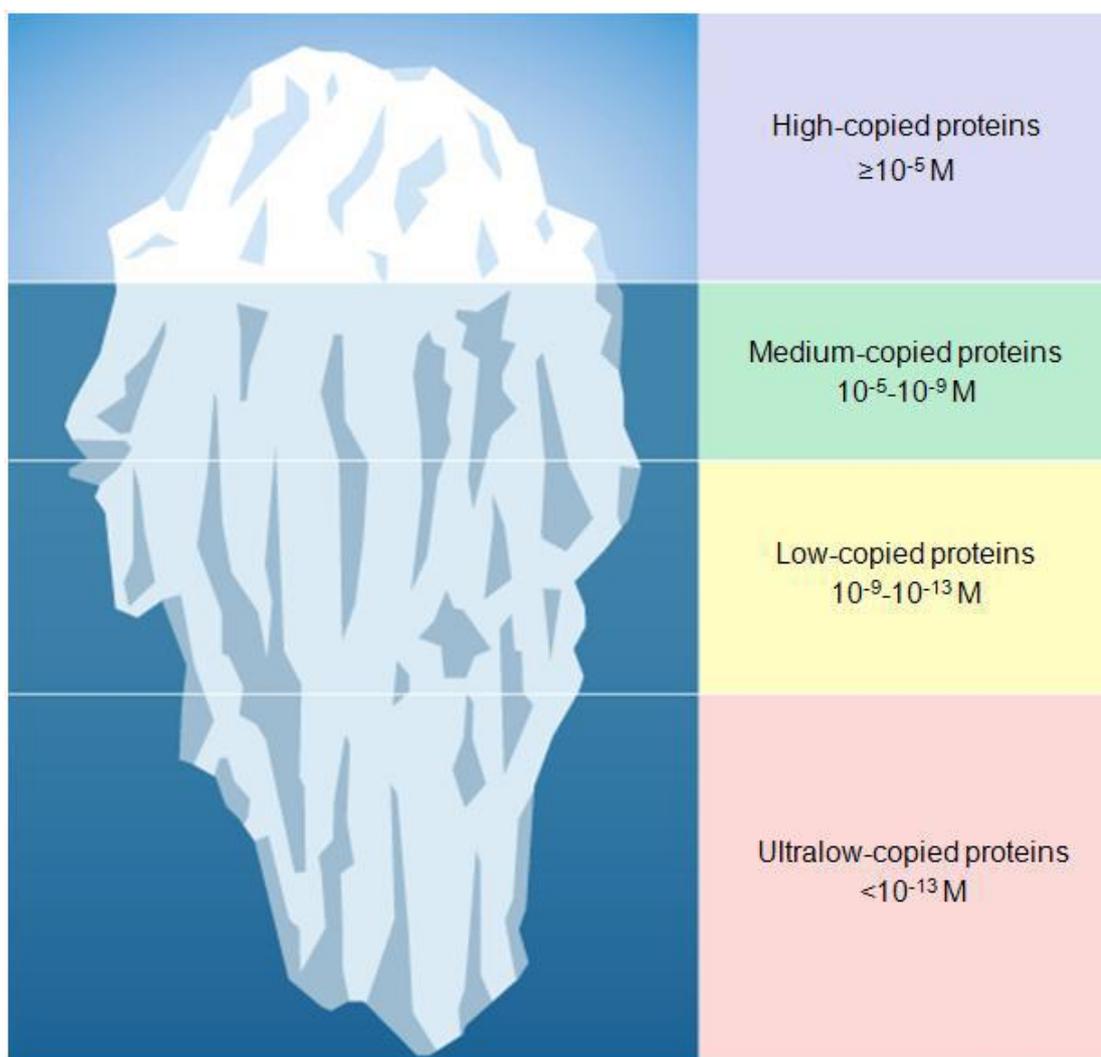


Рис. 1. С использованием существующих протеомных методов может быть исследована лишь небольшая «верхушка» протеомного «айсберга». Большая часть белков, включая биомаркеры конкретных заболеваний, скорее всего, расположены в нижней части, в ультранизких концентрациях.

Настоящая дорожная карта основана на концепции использования нанотехнологии для получения ультранизкокопийных белков, находящихся «на дне» протеомного айсберга. Нанотехнологии активно проявляют себя в геномике, обеспечивая возможности для работы с единичными молекулами нуклеиновых кислот [11]. Если аналогичный подход можно использовать для белков, то проблема концентрационной чувствительности будет решена и в протеомике [10]. Не исключено, что одновременно удастся решить смежную проблему, связанную с

ситуационной изменчивостью протеома. Благодаря возможности анализировать клетки и биологические жидкости для детектирования нескольких десятков белковых копий, динамика протеома будет выражаться в виде зависимости количества копий от времени, и, скорее всего, будет способствовать появлению новых белков и/или их модификаций.

Геноцентричный подход позволяет выполнить систематизированное крупномасштабное сравнение транскриптома и протеома. С учетом вариативности SAP и AS, анализ транскриптов одной хромосомы представляется более реальной задачей, чем исследование продуктов трансляции в широком масштабе генома. Геноцентричное сопоставление транскриптов и белков позволит определить корреляцию между уровнем экспрессии на mRNA и на белковом уровне.

Ограничив разнообразие объектов исследования генами одной хромосомы, можно приступить к решению задачи, связанной с определением взаимодействий и взаимосвязей белков, то есть к исследованию интерактома. Это значит, что будет проведен анализ взаимодействия белков выбранной хромосомы со всеми остальными белками, кодируемыми геномом человека.

Таким образом, преимущества геноцентричного подхода очевидны: он позволяет, во-первых, использовать традиционные протеомные методы, основанные на результатах анализа полного генома, и, во-вторых, сконцентрировать усилия отдельных групп исследователей на конкретной задаче исследования генных продуктов отдельной хромосомы.

3. График реализации проекта

Дорожная карта предусматривает осуществление ППЧ для одной хромосомы в течение 8-10 лет. График выполнения проекта включает предварительный этап (3-4 года) и основной этап (5-7 лет). На предварительном этапе будет решаться задача идентификации, минимум, одного белка для каждого гена определенной хромосомы человека, с приблизительной оценкой уровня его экспрессии и преобладающих модификаций. В результате выполнения этой работы, будет создана база знаний о распределении белков по выбранным тканям на уровне чувствительности до десятков молекул на клетку [12]. Предварительный этап работ в рамках предлагаемой дорожной карты состоит из геномных и протеомных задач.

Первые включают:

- установление критерия для выбора хромосомы;
- глубокое секвенирование выбранной хромосомы, конкретизация открытых рамок считывания;
- экстракция данных об известных транскриптах для генов выбранной хромосомы;
- экспериментальная схема идентификации транскриптов в гепатоцитах и клеточной линии HepG2.

Протеомные задачи предварительного этапа проекта включают:

- экстракцию данных о белках, экспрессируемых генами выбранной хромосомой, выбор ткани печени и плазмы крови в качестве исследуемого биоматериала;
- стандартизация транскриптомных и протеомных технологий, стандартизация вариабельности протеома плазмы путем исследования образцов от здоровых добровольцев;
- идентификация высоко/среднекопийных белков с использованием метода MS/MS, идентификация низкокопийных белков с использованием технологии MRM, идентификация пептидома;
- протеотипирование и протеогеномное профилирование белков выбранной хромосомы в ткани печени, клеточной линии HepG2 и в плазме крови;
- определение интерактомного компонента выбранной хромосомы;
- создание геноцентричной базы знаний по белкам выбранной хромосомы.

Основной этап реализации дорожной карты направлен на существенное повышение чувствительности протеомных технологий и, за счет этого, на увеличение объема информации о белковых молекулах, кодируемых выбранной хромосомой. Перед основным этапом работы необходимо выполнить две ключевые задачи:

- создать гибридную технологию на базе сопряжения атомно-силовой микроскопии и масс-спектрометрии для идентификации низко- и ультранизкокопийных белков;
- идентифицировать белки в диапазоне концентраций от 10^{-13} до 10^{-18} М в плазме крови, и на уровне ~ 10 копий на клетку печени.

4. Выбор хромосомы

В связи с геноцентричной концепцией, выбор хромосомы становится важным этапом региональных проектов, выполняемых в рамках «Протеома человека». Исходя из общих соображений, для выбора хромосомы для Российской части проекта были предложены следующие критерии:

- количество белок-кодирующих генов;
- наличие генов, которые в соответствии с имеющимися литературными данными, связаны с заболеваниями;
- отсутствие генов, кодирующих иммуноглобулины.

По данным ресурса по иммуногенетике IMGT [13] всего в составе генома выявлено 430 генов, кодирующих иммуноглобулины; примерно половина этих генов является псевдогенами. Гены иммуноглобулинов неравномерно распределены по хромосомам. Большая их часть находится на хромосомах №2, №14 и №22. Хромосомы №№3-7, 10-13, 17, 19-21 вообще не содержат иммуноглобулины; по 1-2 гена выявлены в хромосомах №№1,8,9 и №18, причем в хромосомах №№1, 9 и 18 иммуноглобулины представлены в виде псевдогенов.

В среднем в геноме человека на 1 хромосому приходится 921 ± 443 ген, кодирующий белки. Существенно меньшим количеством генов по отношению к

среднему отличаются хромосомы №№13, 18 и 21, содержащие по 353, 286 и 253 гена, соответственно. Именно эти хромосомы и стали самыми важными объектами для обсуждения на конференциях по ППЧ, которые организует HURO. В частности, хромосома №13 была выбрана корейскими учеными, а №21 – предложена учеными США и Канады (см. Рис. 2). Учитывая предыдущий критерий, следует отметить, что в хромосомах №13 и №21 гены иммуноглобулинов отсутствуют, а в хромосоме №18 единственный иммуноглобулин IG-lambda кодируется псевдогеном.

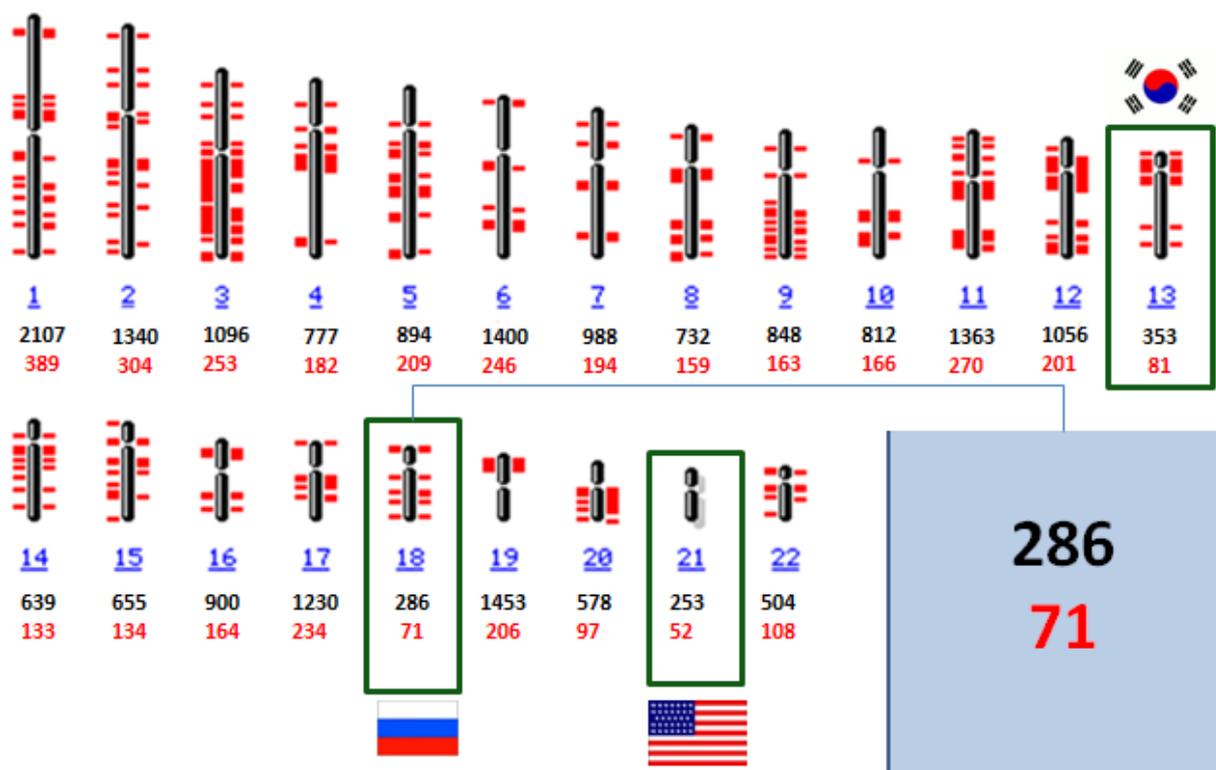


Рис. 2. Хромосомы в составе генома человека. Номера аутосом подчеркнуты, для каждой хромосомы указано количество белок-кодирующих генов и количество генов, связанных с заболеваниями.

С учетом объемов работы, хромосома №21, содержащая всего 253 гена, представляется наиболее подходящим объектом для исследований. Однако, эта хромосома уступает по медицинской значимости, которую можно оценить как отношение количества генов, ассоциированных с каким-либо заболеванием, к общему количеству генов в хромосоме. Ассоциацию генов и их соответствующих продуктов с заболеваниями установили на основе данных ресурса GeneCards [14]. В таблице 1 приведены соотношения генов, ассоциированных с заболеваниями, в некоторых хромосомах. В среднем, около 20% белок-кодирующих генов, так или иначе связаны с заболеваниями. Хромосома №18 занимает первое место в рейтинге, тогда как хромосомы №13 и №21 уступают 18-й хромосоме по данному критерию около 2% и 4%, соответственно.

Таблица 1. Общее количество белок кодирующих генов (БКГ), и процентное содержание (%) генов, ассоциированных с заболеваниями. Приведены данные для 10 хромосом с максимальным процентным содержанием. По данным ресурса GeneCards на 15 февраля 2010 г.

№хр.	18	4	5	3	13	2	8	22	14	21
кол-во БКГ	286	777	894	1096	353	1340	732	504	639	253
%	24,8	23,4	23,4	23,1	22,9	22,7	21,7	21,4	20,8	20,6

Несмотря на значительные (более чем в 5 раз; сравните хромосомы №2 и №21 в табл. 1) различия по количеству генов, содержащихся в хромосомах человека, процентное содержание генов, ассоциированных с заболеваниями, существенно не различается (~3-4%). Это наблюдение справедливо и для остальных хромосом, не вошедших в таблицу 1. Равное значение хромосом для целей геноцентричного подхода к ППЧ стало основанием для выбора для Российской части ППЧ хромосомы №18 и разработки настоящей дорожной карты.

5. Хромосома №18

С точки зрения медицины, важной особенностью 18-ой хромосомы является высокая частота трисомии, описанной как синдром Эдвардса [15]. Дети с трисомией 18-й хромосомы рождаются с частотой примерно 1 раз на 6000 родов и умирают преимущественно в возрасте до 1 года вследствие многочисленных пороков развития [16]. Можно предположить, что гиперэкспрессия белков, кодируемых 18-ой хромосомой, приводит к тяжелым нарушениям в организме человека.

Генетические исследования показали, что 18-я хромосома связана с целым рядом заболеваний (Рис. 3). Важной хромосомной мутацией является транслокация между 14-й и 18-й хромосомами, ведущая к фолликулярной лимфоме, с гиперэкспрессией онкогена *bcl2* [17]. Онкологическая направленность вообще характерна для выбранной хромосомы. На рисунке 3 обращают на себя внимание такие заболевания, как ретинобластома, колоректальный рак, рак поджелудочной железы, синовиальная саркома, лейкемия и сквамозно-клеточная карцинома. Другой социально значимый аспект 18-й хромосомы становится очевидным в связи с заболеваниями центральной нервной системы, в том числе, амилоидная нейропатия, паркинсонизм, шизофрения.

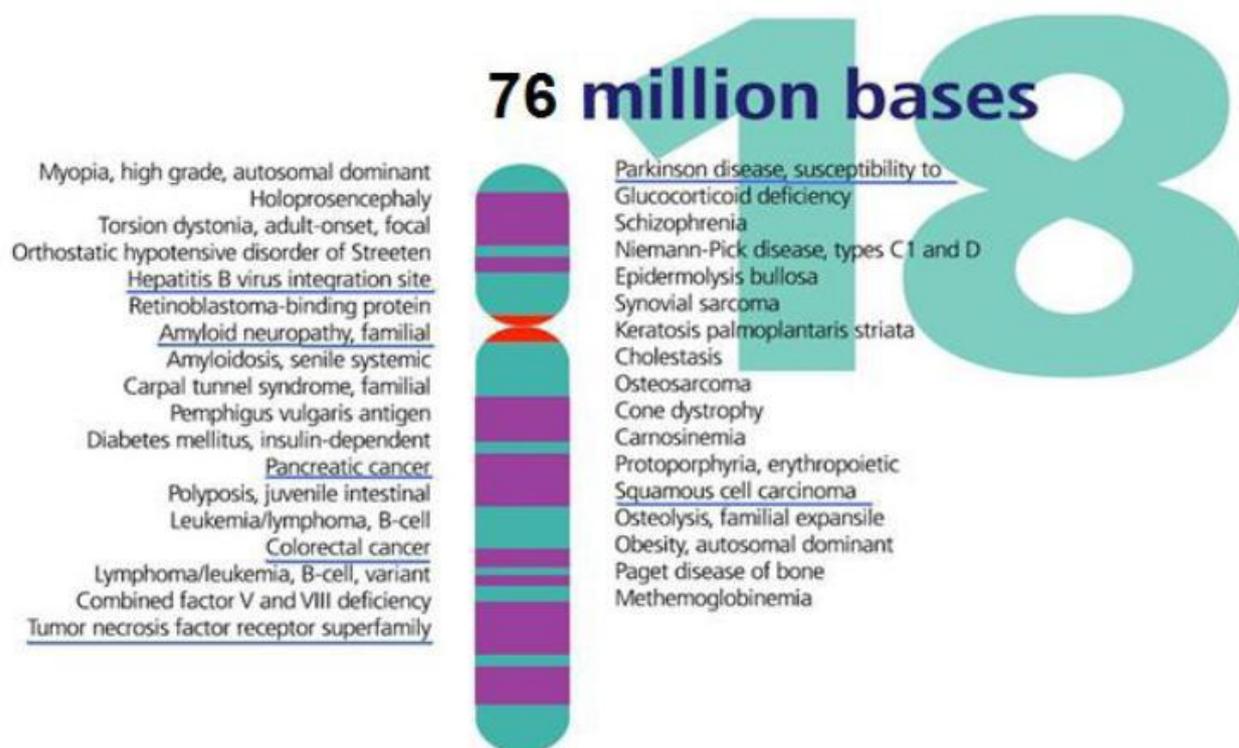


Рис. 3. Заболевания, ассоциированные с генетическими аномалиями в хромосоме №18. Иллюстрация с официального сайта проекта «Геном человека» [18].

Хромосома №18 содержит 76 млн. пар оснований [19]. В среднем на каждый ген этой хромосомы приходится по 3,1 вариантов альтернативной трансляции кДНК. Общее количество генов составляет 480 (RefSeq, Feb 2010); однако, как уже указывалось, только 286 из них (ProteinAtlas, Feb 2010) кодируют потенциально экспрессируемые белки, а остальные либо являются псевдогенами, либо кодируют молекулы РНК. Согласно данным UniProt [20] трансляция была подтверждена присутствием кДНК для 89% генов 18-й хромосомы. Из 286 белков 18-й хромосомы 104 (белки) были выявлены иммуногистохимическими методами в образцах тканей в ходе выполнения проекта ProteinAtlas [21]. Протеомными методами были идентифицированы 226 белков исследуемой хромосомы: соответствующие масс-спектры доступны в базе данных PRIDE [22]. Из всех белков кодируемых 18-й хромосомой, 90 белков выявлены одновременно иммуногистохимическими и масс-спектрометрическими методами, тогда как 136 белков содержатся только в базе данных PRIDE, а 14 - только в ProteinAtlas. Для 46 белков данные об идентификации в настоящее время отсутствуют.

6. Биоматериал: ткань печени и плазма крови

Выбор биоматериала необходим для определения общих основных направлений по координированию действий в рамках дорожной карты, несмотря на то, что геноцентричный подход, в принципе, не зависит от такого выбора. Поскольку в отличие от генома, протеом человека является тканеспецифичным, то прежде всего необходимо определиться с биоматериалом.

Дорожная карта, направленная на исследование ткани или органа, обеспечивает связь между геноцентричным подходом ППЧ и уже существующими проектами HUPO. Среди проектов HUPO, наибольшее развитие получили проект протеома печени (HLPP, [23]) и проект протеома плазмы (HPPP, [24]). Данные, накопленные в ходе осуществления этих проектов, если их использовать для геноцентричного подхода к ППЧ, могут стать базой для предварительного этапа работ по ППЧ. Другие ткани могут быть использованы как дополнительный материал для исследований на основном этапе работ, если некоторые хромосомо-специфичные генные продукты не будут выявлены в плазме и печени, даже при использовании высокочувствительных методов.

В связи с затрудненным доступом к биоматериалу печени, в протеомике используют клеточные линии, происходящие из гепатоцеллюлярной карциномы. Некоторые из них обладают признаками гепатоцитов, в частности, линия HepG2 является неагрессивной клеточной линией, не вызывающей опухоли у безтимусных мышей. Данная линия используется в качестве модели гепатоцитов для исследования их цитофизиологических свойств [25]. Протеомный анализ HepG2 выявил ее существенное сходство с первичной культурой гепатоцитов [26]. Поэтому, линия HepG2 может служить исходным биоматериалом для исследования протеома печени.

Работы по идентификации протеома печени активно проводятся в России на протяжении последних 5-ти лет. Разработаны технологические подходы к идентификации мембранных белков микросомальной фракции печени [27]. Проведены тщательные исследования микрогетерогенности цитохромов P450 в ткани печени с использованием SDS-PAGE [28, 29]. Были установлены профили дифференциально экспрессируемых белков, индуцированные ксенобиотиками, в результате серии экспериментов на базе 2DE [30]. Работы, связанные с исследованием протеома показали, что сотрудничество, существующее между клиническими и научными центрами, должно стать неотъемлемой составной частью выполнения работ в рамках дорожной карты.

Любая ткань, включая ткань печени, сформирована клетками нескольких типов. Протеом образца ткани представляет собой комплексную смесь протеомов клеточных типов, что усложняет интерпретацию результатов идентификации белков. Современные тенденции в протеомике тканей человека требуют применения лазерной микродиссекции (Laser Capture Microdissection – LCM) для выделения однородной клеточной популяции. Использование данного метода в протеомике находится на стадии становления из-за потребности в большом количестве белка для протеомного анализа, в отличие от анализа нуклеиновых кислот. В последнее время, усовершенствование методов микродиссекции позволяет собирать 2x10000 клеток в образце (1-4 мкг белка) [31], что совместимо с идентификацией белков методом LC-MS/MS. Метод LCM использовали для выделения гепатоцитов в протеомных исследованиях печени [32, 33], хотя широкое применение LCM в этой области серьезно затрудняют условия предварительного

протокола. В контексте дорожной карты, ориентированной на исследование печени, предполагается использование LCM подходов.

На белковом уровне, печень как орган тесно связана с плазмой крови. Существует высокая корреляция между протеомом печени и протеомом плазмы крови [34]. PRIDE сообщала, что практически все белки 18-й хромосомы обнаруженные в печени, были идентифицированы и в плазме крови. Только 3 белка из кодируемых 18-ой хромосомой белков выявлены исключительно в ткани печени; в плазме крови и в других органах – они отсутствуют. И наоборот, 216 белков 18-ой хромосомы были идентифицированы в плазме крови; из них 110 также присутствовали и в ткани печени.

Кровь – своего рода коллектор, содержащий все белки человеческого организма. Протеом плазмы, по сути, содержит в качестве подмножеств протеомы других тканей. Плазма крови исторически является одним из первых направлений в протеомике человека [35]. Основная цель исследования плазмы – применение результатов в медицинской диагностике. Плазма крови в наибольшей степени отвечает требованиям, предъявляемым к объекту клинических молекулярных исследований. Она является наиболее доступной для малоинвазивного отбора тканей человека. По сравнению с более доступными биологическими жидкостями (слюна, моча) плазма, как составная часть крови, в большей степени гомеостатична, т. е. она характеризуется, в основном, постоянным соотношением высококопийных белков.

На предварительном этапе реализации проекта Human Plasma Proteome Project (HPPP), в десятках лабораторий разными протеомными методами анализировали стандартные образцы плазмы крови человека [36]. Основной проблемой для протеомного анализа плазмы крови является огромный динамический диапазон концентраций белков (~10-кратное различие концентраций). Большинство способов анализа нефракционированной плазмы позволяют выявить только несколько мажорных белков, оставляя вне рассмотрения потенциально важные для медицинской диагностики низкокопийные белки [37].

На рисунке 4 приведена диаграмма, отражающая современные представления о структуре протеома плазмы крови. Видно, что почти 60% белков, идентифицированных в плазме крови, функционируют в клетках, тогда как «классические» белки плазмы составляют только 13%, а на долю иммуноглобулинов приходится 8%. Среди клеточных белков, 53% локализованы в цитоплазме, 9% - в ядре и 19% - в мембране, тогда как фракция секретируемых белков занимает менее 14%. Эти данные показывают, что в составе протеома плазмы содержатся белки, отражающие молекулярные процессы, происходящие в клетках, и, следовательно, эти белки могут быть отобраны в качестве кандидатов на биомаркеры заболеваний.

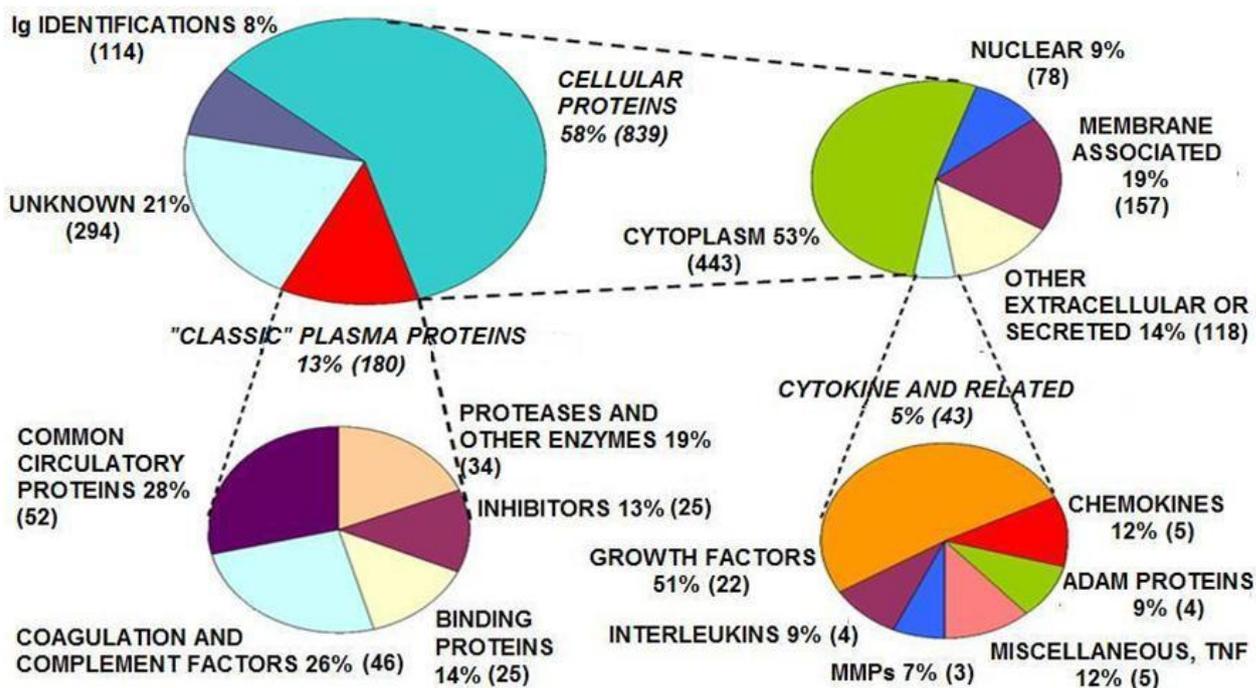


Рис. 4 «Нет плазмы – нет денег» - С. Ханаш, Конференция по HPPP HUPO, 2009 г. Распределение идентифицированных белков плазмы крови согласно [38], [39].

Несмотря на технические сложности, нет сомнений, что плазма крови является биоматериалом выбора в геноцентричном подходе к реализации проекта «Протеом человека». Не стоит недоумевать по поводу такой трансформации геноцентричной парадигмы в отношении конкретного типа биоматериала. Следует подчеркнуть, что протеом плазмы крови может содержать субпротеомы различных органов и тканей, обнаружить которые в настоящее время не удастся только в связи с низкой чувствительностью существующих протеомных технологий.

Для того, чтобы определить «точку отсчета» в анализе протеома плазмы, мы обратились к списку из 3020 белков, полученных по итогам предварительного этапа HPPP [24] (идентификация 1175 белков из этого списка была подтверждена независимо четырьмя аналитическими методами [40]). В HPPP белок интерпретировали как продукт одного гена, без учета модификаций и процессинга.

Геноцентричный подход подразумевает распределение идентифицированных белков соответственно хромосоме, которой они кодируются. Идентифицированные в плазме белки равномерно распределены по хромосомам, т.е. чем больше генов в составе хромосомы, тем большее количество кодируемых этими генами белков обнаруживается в плазме крови (Рис. 5).

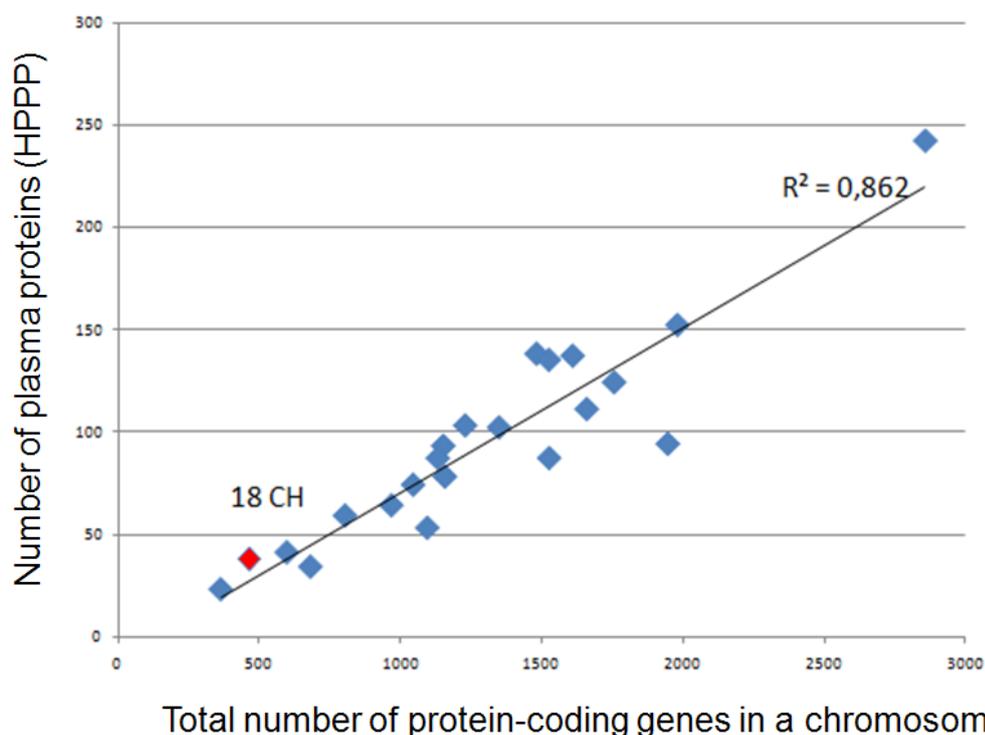


Рис. 5. Зависимость количества идентифицированных в плазме крови белков от общего количества белок-кодирующих генов (БКГ) в хромосоме. Точка, соответствующая 18-й хромосоме, отмечена красным цветом. Белки плазмы взяты из [41].

На рисунке 5 приведены данные, полученные в результате нанесения на хромосомы человека информации об идентифицированных в плазме крови белках. Наблюдается высокая степень корреляции ($R^2 \sim 0,9$) между количеством белок-кодирующих генов в хромосоме и количеством идентифицированных в плазме белков; корреляция показывает, что каждая хромосома кодирует белки плазмы крови пропорционально количеству содержащихся на ней генов, т.е. плазма и в отношении хромосом может быть охарактеризована как своеобразный «коллектор».

Принимая во внимание изложенные выше факты, можно предположить, что плазма как биоматериал может играть такую же решающую роль в проекте «Протеом человека», как клеточные элементы крови в проекте «Геном человека». Для этого, протеомике необходимы новые технологии для идентификации низко- и ультранизкокопийных белков. В отношении же высоко- и среднекопийных белков актуальность приобретают вопросы стандартизации высокопроизводительного эксперимента.

7. Стандартизация и протеотипирование

Используемые в протеомике методы идентификации высоко/среднекопийных белков основаны на жидкостной хроматографии, сопряженной с тандемной масс-спектрометрией (LC-MS/MS). Вопрос меж-

лабораторной воспроизводимости данного метода остается пока открытым, поскольку результаты идентификации сильно зависят от более чем 46-ти технических параметров [42]. Избежать разнообразия результатов при анализе сходных образцов в различных лабораториях можно при помощи стандартизации. С этой целью в разных научных центрах были проведены независимые исследования: одни и те же стандартные образцы, как искусственно полученных смесей рекомбинантных белков [43], так и смесей белков лизата дрожжей [44], были охарактеризованы методом LC-MS/MS. Различия в наборе белков одного тестового образца, идентифицированных разными исследователями, основывались не только на причинах, связанных с техническими процедурами, но в большой степени, с алгоритмами анализа масс-спектральной информации [44]. Из последних публикаций на эту тему становится ясно, что для целей проекта «Протеом человека» самые современные масс-спектрометрические платформы должны пройти рутинный тест с использованием стандартных образцов.

Одним из подходящих стандартов для многоцентровых исследований является комплексный протеомный стандарт, разработанный компанией Agilent (США) с участием лабораторий-лидеров в области протеомики. Набор, специально предназначенный для тестирования систем LC-MS/MS, содержит экстракт белков стандартного штамма бактерии *Pyrococcus furiosus* (Pfu). Число белков, содержащееся в данном реагенте, превышает 1500, что соответствует по сложности образцам, происходящим из организма человека. Более того, преимуществом стандарта Pfu является низкая степень сходства последовательностей белков микроорганизма и человека. Реагент Pfu может служить стандартом для сравнения и контроля качества технологии LC-MS/MS, применяемой лабораториями-участницами российской части проекта «Протеом человека».

Стандартные пептиды используются для количественного анализа белков посредством масс-спектрометрии путем мониторинга множественных реакций (MRM). Приготовленные искусственно пептидные стандарты смешивают с биологическим образцом или вводят отдельно в масс-спектрометр для его настройки [45]. Использование стандартов в процессе анализа существенно снижает вариабельность его результатов. Так, в многоцентровом исследовании одних и тех же образцов была показана высокая степень достоверности результатов MRM-анализа [46]. Поэтому, для использования MRM-анализа в рамках проекта, следует придерживаться стандартных протоколов. В процессе реализации дорожной карты, искусственно созданные стандартизованные белковые смеси предлагается использовать для оценки количественных распределений в межлабораторных условиях.

Особенностью биологических исследований, в частности, анализа протеома, является то, что погрешность измерения вызвана не только приборными и технологическими аспектами, но и индивидуальной вариабельностью исследуемого объекта. Выявление такого рода вариабельности составляет сущность задачи протеотипирования, которой в последнее время уделяется все

большее внимание [47]. Предлагаемая дорожная карта проекта «Протеом человека» предусматривает определение индивидуальной вариабельности протеомов плазмы крови и печени.

Выявление статистических диапазонов колебаний уровня белков плазмы крови является существенным условием для определения биомаркеров заболеваний [48]. Определение вариабельности протеома плазмы потребует анализа индивидуальных различий у здоровых людей. К использованию в экспериментах будут привлечены добровольцы, состояние здоровья которых соответствует существующим нормам, что подтверждено врачебной комиссией. В рамках дорожной карты будет проведен статистический анализ концентрационной изменчивости белков плазмы, кодируемых 18-й хромосомой.

Целесообразно также определить сравнительные значения уровней экспрессии белков в печени, хотя особый формат соответствующих экспериментов определяется в основной стадии дорожной карты. Условно нормальные образцы печени чаще всего получают в результате хирургического вмешательства при резекции холецистэктомии и гемангиомы.

Таким образом, в проекте «Протеом человека» стандартизация должна быть направлена, с одной стороны, на обеспечение эталонных режимов эксплуатации высокотехнологичного оборудования, а с другой – на определение межиндивидуальных различий протеома.

Еще одной проблемой, связанной со стандартизацией исследований протеома, является протеотипирование. В этом случае, стандарты следуют применять к определению структуры белка, как продукта одноаминокислотного полиморфизма (SAPs, [49]), альтернативного сплайсинга (AS, [50]) и посттрансляционных модификаций. Для решения проблемы структурной микрогетерогенности белков предлагается два подхода. Первый подход «top-down» заключается в высокоточном измерении массы цельных белков [49], [51], при этом смещение измеренной массы относительно теоретической - сопоставляется со структурой одноаминокислотных полиморфизмов и посттрансляционных модификаций. Для регистрации таких характерных сдвигов в масс-спектрах необходима высокая точность, которую предоставляют современные приборы типа Orbitrap или FT-ICR. Однако, при регистрации целых белковых молекул с высокой молекулярной массой, уровень чувствительности невысок: около $10^{-5} - 10^{-6}$ М [52]. Низкая чувствительность не позволяет применять подход «top-down» для достижения целей, заявленных в дорожной карте.

Чувствительность протеотипирования можно повысить за счет подхода «bottom-up», используя стандартный масс-спектрометрический анализ протеолитических пептидов. При этом, чувствительность повышается на 4-5 порядков и достигает уровня $10^{-8} - 10^{-9}$ М или даже 10^{-12} М в режиме MRM. В ходе выполнения предварительного этапа работ по дорожной карте ППЧ, планируется использовать итеративный алгоритм идентификации протеотипических пептидов [53]. В соответствии с этим алгоритмом, сначала проводится идентификация

белков в сравнении с полногеномной базой данных, затем, в состав базы данных включаются варианты последовательностей, содержащие SAP, AS и вероятные PTM для идентифицированных белков. После расширения базы данных, проводится повторная идентификация, результатом которой является информация о наличии в образце микрогетерогенных вариантов белков. Далее следует идентификация протеотипического пептида, содержащего SAP, AS или PTM в режиме MRM (см. ниже).

С учетом информации, существующей в современных базах данных, можно приблизительно оценить количество микрогетерогенных вариантов белков, относящихся к 18-й хромосоме. Для генов 18-й хромосомы известно 650220 вариантов нуклеотидного полиморфизма (по данным National Center for Biotechnology Information [54]), из которых 2386 являются несинонимичными (nsSNP) и потенциально могут рассматриваться как SAPs.

Для белков 18-й хромосомы, в базе данных UniProt содержится информация о 796 одноаминокислотных полиморфизмах. В подавляющем большинстве случаев это предположительные SAP, наличие которых можно ожидать, исходя из данных о SNP в соответствующих генах. В среднем, на один белок можно ожидать по 4 SAP. Среди всех белок-кодирующих генов 18-й хромосомы наибольшее количество – 168 SAP - приходится на ген из семейства Нимана-Пика типа С (NPC). Существует 168 участков одноаминокислотного полиморфизма, что дает увеличение комбинаторных вариантов до $2^{168} \sim 3,74 \cdot 10^{50}$ первичной структуры в популяции. Функции NPC связаны с внутриклеточным транспортом холестерина; дефекты этого гена способствуют развитию сфингомиелиноза (болезнь Нимана-Пика). Следует отметить, что в наших экспериментах ген NPC был идентифицирован путем анализа протеома клеточной линии НерG2.

Данные по AS и PTM также оценивали с использованием информации, доступной в ресурсе UniProt. Оказалось, что для 120 белков 18-й хромосомы известна, минимум, 1 сплайс - изоформа. В соответствии с данными из UniProt, при альтернативном сплайсинге генов этой хромосомы образуется, в среднем, 2,81 белковых продуктов (что не отличается от среднестатистических показателей по целому геному), а наибольшее разнообразие изоформ наблюдается для гена NFATC1 (9 изоформ). Посттрансляционной модификации подвергаются, примерно, 60% белков, кодируемых 18-й хромосомой. Как обычно, наиболее часто наблюдается фосфорилирование (112 белков), гликозилирование (53 белка) и ацетилирование (36 белков). В базе данных UniProt указано, в среднем, по 4 модификации на белок, при этом среди белков с известной функцией наибольшее количество модификаций (23, в т.ч. 21 сайт фосфорилирования серина и треонина, и 2 сайта ацетилирования по лизину и метионину) известно для регуляторного белка с идентификатором Q9Y2J2, участвующего в патогенезе менингиомы.

Из приведенных выше средних величин можно дать минимальную оценку количества изоформ белков, происходящих от 18-й хромосомы. Допуская, что на каждый из 286 белок-кодирующих генов, в среднем, приходится 4

одноаминокислотных полиморфизма, 2,81 вариантов альтернативного сплайсинга и по 4 пост-трансляционных модификаций (не принимая во внимание комбинаторные варианты), мы получаем $N=286 \times (4+2,81+4) \sim 3046$ структурно различающихся изоформ белков. Таков, по самым скромным подсчетам, объем работ, который предстоит выполнить в ходе реализации дорожной карты по 18-ой хромосоме.

8. Анализ транскриптов и высоко/среднекопийных белков

С целью проиллюстрировать возможности и недостатки существующих подходов к выполнению ППЧ, был проведен анализ несинхронизированной клеточной линии гепатоцитов HepG2, образцов ткани печени здоровых доноров и образцов плазмы крови. Анализ клеток и ткани проводился на уровне транскриптома с применением полногеномных микрочипов фирмы Agilent Inc. На уровне протеома биоматериал анализировали с применением многомерной хроматографии в сочетании с MS/MS идентификацией белков.

С помощью полногеномных транскрипционных микрочипов в клеточной линии выявили экспрессию 10 865 генов (из этого количества, 347 генов относилось 18-й хромосоме). В базе данных Ensembl [55], идентификаторам этих генов соответствует только 179 белков, тогда как остальные имеют временные экспериментальные названия фрагментов неизвестных генов.

Анализ по базе GeneOntology показал, что транскрипты 18-ой хромосомы участвуют в 30-ти различных биохимических процессах. Соответствующие гены, в большинстве случаев, способствуют регуляции транскрипции (24 белка) и кодируют транспортные белки (19 белков). Локализация продуктов транскрибируемых генов весьма разнообразна: мембранные белки составляют около трети экспрессированных транскриптов (53), ядерные - представлены 43 генами и в цитоплазме располагаются 37 белков. По данным Uniprot, только 22 белка из всех рассматриваемых связаны с развитием какого-либо заболевания. Часть этих заболеваний обусловлена nsSNP, которые могут транслироваться в виде SAP (см. далее).

Сопоставление транскриптома клеточной линии с транскриптомом ткани печени здоровых людей, полученной при холецистэктомии, показало, что более 73% функционирующих генов совпадают. Это говорит о том, что клеточная линия является адекватной моделью, отражающей основные физиологические процессы печеночной ткани. Если рассматривать транскрипты только 18-й хромосомы, то уровень совпадения будет несколько ниже – около 62%.

К достоинствам клеточной линии можно, безусловно, отнести простоту получения материала в нужных количествах и высокую степень стандартизации условий эксперимента. С другой стороны, эти клетки являются раковыми, с частично утраченными в процессе культивирования функциональными особенностями. Явным недостатком в работе с образцами печени является

сложность получения нужного количества исходного материала, необходимость выделения гомогенной популяции клеток из ткани, а также индивидуальные различия между пробами.

Протеомный анализ вышеупомянутых биологических образцов был выполнен с использованием методики 3-х мерного разделения [27]. Белки разделяли на обращеннофазовой колонке на 5 фракций, каждую из которых затем подвергали гидролизу и анализировали по технологии MudPIT [56]. С использованием описанного метода было идентифицировано более 4100 белков HerG2, из которых 91 относятся к 18-й хромосоме. Объединив эти данные по идентификации с разделом “Natural Variants” из UniProt, мы обнаружили, что 17 белков несут мутации, ассоциированные с болезнями, включая такие как кардиомиопатия, катаракта, рак и ранее упомянутая болезнь Ньюмена-Пика.

Необходимо обратить внимание, что некоторые из идентифицированных белков не были выявлены на предыдущей стадии полногеномного транскриптомного исследования. Транскрипты 17 генов не регистрировались выше порога интенсивности на микрочипе, тем не менее, соответствующие им белковые продукты были выявлены в протеоме с высокой степенью достоверности. Подобное несоответствие между транскриптомными и протеомными данными наблюдается довольно часто [57], поэтому, чтобы подтвердить данное наблюдение, мы проверили различия в копиях белков, используя спектральный подход без меток [58] сопряженный с тандемной масс спектрометрией (MS/MS).

На рис. 6 показана копияность белков в зависимости от соответствующей интенсивности мРНК-экспрессии. Копийности белков (количество пептидов, участвовавших в идентификации белков) не согласовывались с абсолютной интенсивностью флуоресценции их мРНК ($R^2 = 0,02$). Это может указывать на низкую стабильность мРНК по сравнению с молекулами белков, но нельзя исключать и технические причины: зависимость уровня сигнала от количества и специфичности зондов, размещенных на чипе для определенного гена [59].

С использованием платформы MudPIT, был также проведен анализ белкового состава плазмы крови. Среди 2000 идентифицированных белковых продуктов было обнаружено 66 белков, которые кодируются генами 18-й хромосомы.

Анализ транскриптов 18-й хромосомы и кодируемых ими белков, показал, что существенное пересечение данных по мРНК и белкам обеспечивает возможности для реализации геноцентрического подхода. Несовпадения, наблюдаемые на рис. 6, указывают на метрологическую некорректность сравнения интенсивности сигнала на микрочипе с индексом копияности пептида. Дорожная карта должна решать такого рода погрешности, выражая содержание в виде количества копий биомолекул. Подсчитывание отдельных молекул вместо измерения их концентраций обосновано концепцией обратного числа Авогадро [10]. Сегодня мы являемся свидетелями того, как данная концепция находит все большее применение в технологии секвенирования единичных молекул ДНК [60].

В рамках дорожной карты, концепция обратного числа Авогадро используется для соединения обычной протеомики с нанопроотеомикой [9],[61].

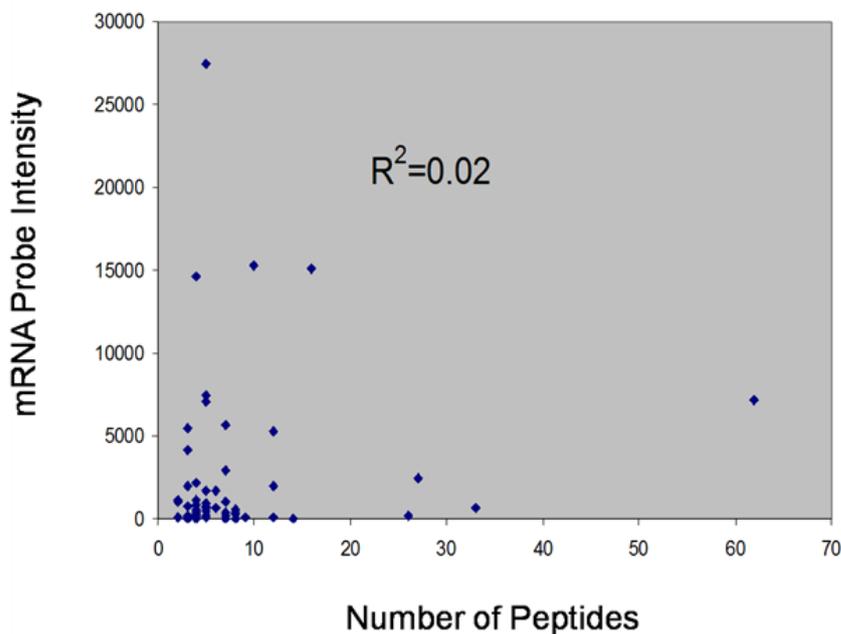


Рис. 6. Диаграмма разброса данных о количестве пептидов в зависимости от уровня экспрессии соответствующей мРНК для транскриптов/белков 18-ой хромосомы.

Протеогеномное профилирование. Сопоставление транскриптомных и протеомных результатов позволяет выполнить форматирование генома по принципу обратной связи. Обратная геномика (протеогеномное типирование) – это анализ сплайс-вариантов на уровне белков для расшифровки реальных границ кодирующей рамки в геноме и выявление неаннотированных генов [62]. Для протеогеномного типирования важна не только информация о полных белках, но и об их низкомолекулярных фрагментах – пептидах.

9. Конструирование лекарств на базе пептидома

При аннотации генов, локализация и идентификация возможных «пептидных» генных продуктов (белков, имеющих массу менее 10 кДа), а также семейств пептидов, являющихся продуктами нативной и патологической деградации белков протеолитическими ферментами, является, фактически, «темным пятном». Большое количество продуктов деградации могут обладать физиологической активностью.

Особый интерес представляет идентификация пептидов, являющихся продуктами деградации белков, и выявление их специфической биологической и маркерной активности. Дорожная карта рассматривает пептидом как источник дополнительной информации, которая собирается в соответствии с

геноцентричным принципом. Биологическая активность пептидов рассматривается в качестве отправной точки для создания прототипов лекарственных средств.

Для систематизированной характеристики пептидома необходимо разработать универсальные методы экстракции и десорбции пептидов с мажорных тканевых и плазменных белков. Методы экстракции должны обеспечивать воспроизводимость результатов структурно-функционального анализа с помощью методов высокоточной масс-спектрометрии. Несмотря на то, что в подобного рода экспериментах интерес представляют любые пептиды, обнаруживаемые в сыворотке или плазме крови, существует возможность их селективной идентификации и идентификации белков-предшественников, кодируемых 18-ой хромосомой человека. Методы селективной масс-спектрометрической идентификации позволяют осуществлять целенаправленный поиск заданных пептидных фрагментов в низкокопийном диапазоне концентраций.

10. Детектирование низко/ультранизкокопийных белков

Основной проблемой при анализе биологического материала является низкая концентрационная чувствительность, характерная для современных протеомных технологий. Даже самые сложные аффинные методы идентификации белков едва позволяют достичь уровня чувствительности ниже 10^{-12} М. Как следствие, большинство способов протеомного анализа биоматериала позволяют выявить лишь несколько сотен высоко/среднекопийных белков, тогда как низкокопийные белки остаются недоступными для анализа (Рис. 1). Проблема идентификации низкокопийных белков, по-видимому, является основной технической сложностью, которую необходимо преодолеть в рамках проекта «Протеом человека».

Мониторинг множественных реакций (MRM, [45]) является масс-спектрометрическим методом, отличающимся высокой надежностью обнаружения белков и пептидов. По сравнению с другими современными методами в масс-спектрометрии, MRM обеспечивает высокую чувствительность и широкий динамический диапазон концентраций белков. Согласно литературным данным, чувствительность данного метода находится в пределах 10^{-12} - 10^{-14} М для дрожжей [12] (что составляет несколько сотен копий белков на клетку).

Для использования метода MRM необходимо определить набор протеотипических пептидов, которые будут являться объектом мониторинга. Протеотипический пептид по своим физико-химическим свойствам должен легко подвергаться ионизации и давать характерный масс-спектр с явно выраженными пиками. Для идентификации определенного белка необходимо, чтобы протеотипический пептид не только хорошо ионизировался, но и чтобы он был уникален, т.е. не встречался в белках, кодируемых другими генами.

Можно ожидать, что применение MRM в сочетании с современными методами фракционирования белков, позволит достичь предела детектирования

низко-копийных белков, т.е. достигнуть уровня концентраций 10^{-15} М. Дальнейшее увеличение чувствительности может быть основано на использовании метода необратимого фишинга белков из биологического материала на поверхность химически активированного биочипа. Регистрация белков «выловленных» на поверхность будет осуществляться с помощью молекулярного детектора, например, атомно-силового микроскопа, и, одновременно, будет проводиться масс-спектрометрическая идентификация детектируемых молекул [64].

Необратимый фишинг позволяет концентрировать низко/ультранизкокопийные белки на подложке с химически активированными группами, через которые происходит ковалентная иммобилизация белков. Теоретические оценки показывают, что концентрирование белка на поверхности чипа с помощью необратимого фишинга, может достигать 8 порядков [65]. Это означает, что низкокопийные белки, выявленные благодаря фишингу, будут переводиться в категорию среднекопийных, а ультранизкокопийные – в категорию низкокопийных.

Использование атомно-силового микроскопа в качестве молекулярного детектора позволяет подсчитывать количество белковых молекул, иммобилизовавшихся на биочипе в результате фишинга [66], [67]. Поскольку чувствительность АСМ находится на уровне единичных молекул, то процесс регистрации молекул перестает быть лимитирующим звеном в технологии анализа протеома [65]. Комбинирование АСМ с масс-спектрометрией позволит не только визуализировать и подсчитывать молекулы белков, но и осуществлять их идентификацию.

Несмотря на то, что фишинг в комбинации с молекулярным детектированием является перспективным направлением технологического развития протеомики, нельзя не указать на важное ограничение атомно-силовой микроскопии (АСМ) – низкая скорость сканирования поверхности. Скорость сканирования самых скоростных коммерческих моделей составляет 1 – 10 Герц; например, сканирование небольшой площадки в 400 мкм^2 занимает около 3-х часов. Успехи в этом направлении связаны с увеличением скорости сканирования. Одним из достижений можно считать увеличение скорости в 1000 раз, до 1 КГц [68], тогда скорость сканирования стандартной площадки в 400 мкм^2 уменьшится до 10 сек. Кроме этого, достигнутая скорость сканирования дает возможность наблюдать движение молекул в реальном времени. Это значит, что АСМ также будет обладать возможностью идентифицировать белки, главным образом, по динамическим изменениям их конформации. В будущем, АСМ в реальном времени будет способна фиксировать ситуационные изменения протеома, путем мониторинга формирования и распада функциональных комплексов.

11. Интерактомика

Исследования интерактома человека обеспечивают новый уровень понимания молекулярных механизмов и машин в норме и патологии, оказывая большое влияние на лечение заболеваний. Ниже представлена некая концепция о хромосомо-центричном подходе к объяснению интерактома.

На предварительном этапе реализации дорожной карты будет выполнен поиск по интерактомным базам данных для выявления белок-белковых взаимодействий с участием белков 18-ой хромосомы. Данные о белок-белковых взаимодействиях будут извлечены из ресурса APID2NET [69], объединяющего информацию из DIP, MINT, BioGRID, IntAct, HRPD и BIND. Машинный поиск информации о метаболических взаимодействиях белков будет вестись при помощи KEGG [70]. Белковые сети будут распределены соответственно их связи с заболеваниями (см. раздел «18-я хромосома»). Отобранные пары взаимодействующих белков будут для определения достоверных идентификаций. В результате такой инвентаризации будет получен список протеотипических пептидов для белков, кодируемых на других хромосомах, но взаимодействующих внутри путей и комплексов, связанных с 18-ой хромосомой. Такие протеотипические пептиды будут включены в MRM процедуру.

Гены 18-ой хромосомы имеют варибельность согласно базе данных IneterPro. В базе данных IneterPro определен всего 231 белок; из них 33 белка, содержащие C2P2-домены типа «цинковые пальцы» и 51 — кадгериновые домены и кадгерин-подобные домены, 11 белков - протеазные ингибиторы, 14 — домены серпина и 136 — домены других типов. Эти данные подтверждают, что большинство белков 18-ой хромосомы имеют сходную архитектуру доменов. Данная информация может быть использована для фильтрации известных и предсказанных взаимодействий, например, белки с одинаковыми аннотациями GO и/или с одинаковыми доменами, экспрессирующимися одними и теми же тканями могут иметь одинаковых партнеров взаимодействий.

KEGG содержит 60 белков 18-ой хромосомы, ассоциированных с сигнальными и метаболическими путями. Было отмечено, что не более 3-4 белков 18-ой хромосомы участвуют в одном сигнальном пути и, что белки 18-ой хромосомы распределены по различным клеточным процессам. На рис. 7 показан фрагмент сигнального пути TGF. Представлены 4 белка, кодируемые генами 18-ой хромосомы, связанными с этим путем, а именно Smad4, Smad2, Smad7 и ROCK. Белок Smad7 ингибитор Smad2 и Smad4, а Smad2 входит в состав транскрипционного комплекса, который активирует пути клеточного роста, миграции, пути апоптоза [71, 72]. Данный сигнальный путь представляет наибольший интерес для применения в медицине при лечении раков легкого, молочной железы и простаты [73].

PRIDE содержит информацию об идентификации 82 белков описанного сигнального пути. Эти белки были идентифицированы в 5-10 экспериментах с плазмой крови, однако, в большинстве экспериментов сообщается только об одном

протеотипическом пептиде, а какие-либо подробности масс-спектров отсутствуют. Кроме продуктов 18-ой хромосомы, TGF-путь указывает на белки-партнеры в составе другой хромосомы, например, SMURF2 (Q9HAU4, хромосома №17) и SMURF1(Q9HCE7, хромосома №7), которые также были обнаружены в PRIDE. Однако, для некоторых белков TGF-пути (например, Q96S42 и Q6GTN3) отсутствуют МС-данные, поэтому такие белки должны идентифицироваться в ходе интерактивных исследований в рамках дорожной карты.

Основной этап работы по дорожной карте включает экспериментальную схему для расшифровки новых взаимодействий, связанных с белками, транслированными из 18-ой хромосомы. Первое, *in silico* выявление возможных белок-белковых взаимодействий будет выполнено путем анализа данных и алгоритмов предсказания типа локального сходства сиквенсов известных партнеров взаимодействия [74] и сравнительной геномики филогенетических профилей белков [75,76].

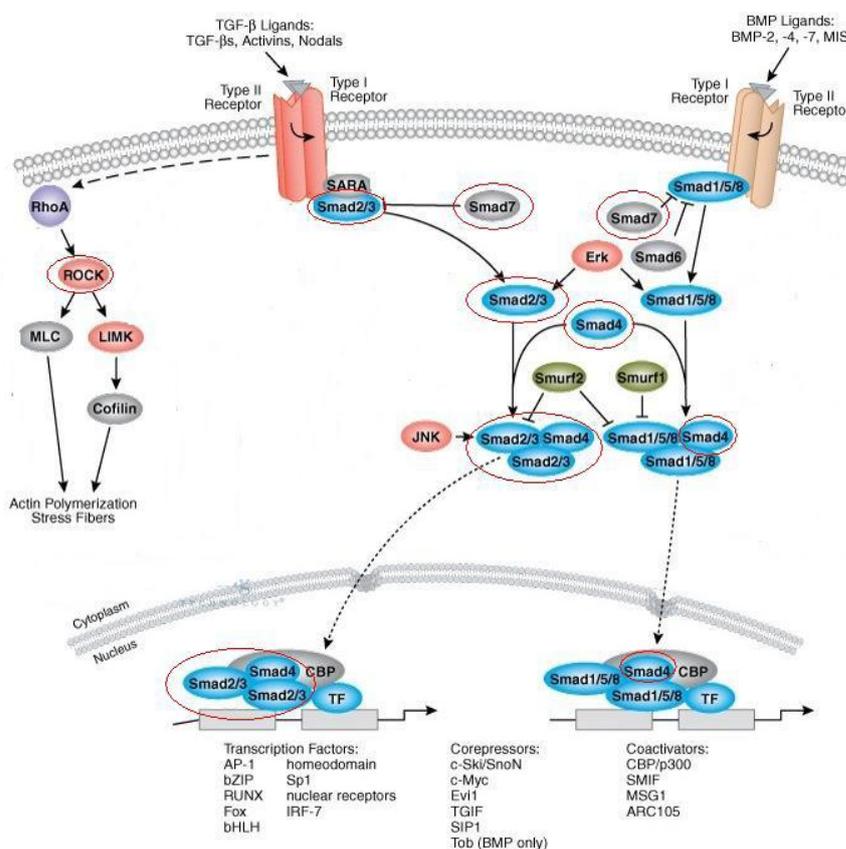


Рис. 7. Фрагмент TGF-β сигнального пути, включающего белки 18-ой хромосомы (показаны красными овалами).

Сочетание методов *in silico* и экспериментальных методов позволяет оценить известные взаимодействия в различных экспериментальных условиях и предсказанных *in silico* новых белок-белковых взаимодействиях для построения надежных белок-белковых сетей белков 18-ой хромосомы.

Биосенсор на основе метода плазмонного резонанса в сочетании с масс-спектрометрической идентификацией белков обеспечивает высокое качество измерений; протокол делится на 3 главных этапа [77], [78]: (1) экстракция белково-мишеней в комплексе с белками-партнерами на чипе оптического биосенсора после введения клеточного лизата (или его хроматографической фракции) — молекулярный фишинг, (2) элюция экстрагированных белков с оптического чипа; (3) масс-спектрометрический анализ и идентификация белков. Совокупность экспериментальных данных о взаимодействии и идентификации белков, влияние различных условий экспериментов на их состав и патологические процессы составят информацию для базы знаний.

12.База знаний

Базы данных использовались в качестве средств информационного обеспечения при реализации проекта «Геном человека» [79]. Большая сложность протеома по сравнению с геномом вынуждает предусматривать программные средства не только для хранения информации, но и для ее аналитической обработки.

В рамках российской части проекта «Протеом человека» предлагается создать геноцентричную базу знаний по белкам 18-й хромосомы. В случае если предлагаемый подход будет успешным, его можно будет использовать и для исследований других хромосом.

Как показано на рис. 8, база знаний будет действовать как посредник между дорожной картой и доступными ресурсами, необходимыми для выполнения экспериментов в области молекулярной биологии. Пользователи базы знаний делают запрос в доступных ресурсах через интерфейс, который аккумулирует знания [80]. Аккумулированные знания доступны всему консорциуму ППЧ, чтобы выявлять интеллектуальные решения [81].

В первую очередь база знаний рассчитывает индексы, которые обеспечивают мониторинг состояния работ выполняемых по дорожной карте. Первый индекс оценивает медицинскую значимость выполняемого исследования; он уже обсуждался при выборе хромосомы. С помощью базы знаний, информация, полученная в результате экстракции данных, сравнивается с данными о вновь идентифицированных белках и пептидах. Значение индекса зависит также полноты покрытия последовательности белков, ассоциированных с заболеванием, от количества протеотипических SAP-содержащих пептидов, от диагностических средств и т.д.

Второй индекс приблизительно отражает общий успех научных экспериментов в виде процентного соотношения белок-кодирующих генов к соответствующим данным MS-анализа. Этот индекс зависит от точности собранных данных масс-спектров, а также от количества идентифицированных белков в разных независимых экспериментах.

Третья часть базы знаний касается технологии исследований в рамках дорожной карты: оценка пропорции ультранизкокопийных белков исследуемой хромосомы.

База знаний будет создаваться на основе алгоритмов экстракции текстов научных статей, многие из которых активно используются в биомедицинских исследованиях. Стержень базы знаний, предположительно, составят технологии по обработке информации на базе MeSH [82],[83] анализа и библиографического картирования белков [84]. Чтобы достигнуть подробной базы знаний будут необходимы персональные суперкомпьютеры [85], подобно кластерным вычислениям, которые обеспечил ПГЧ (Проект Генома Человека).

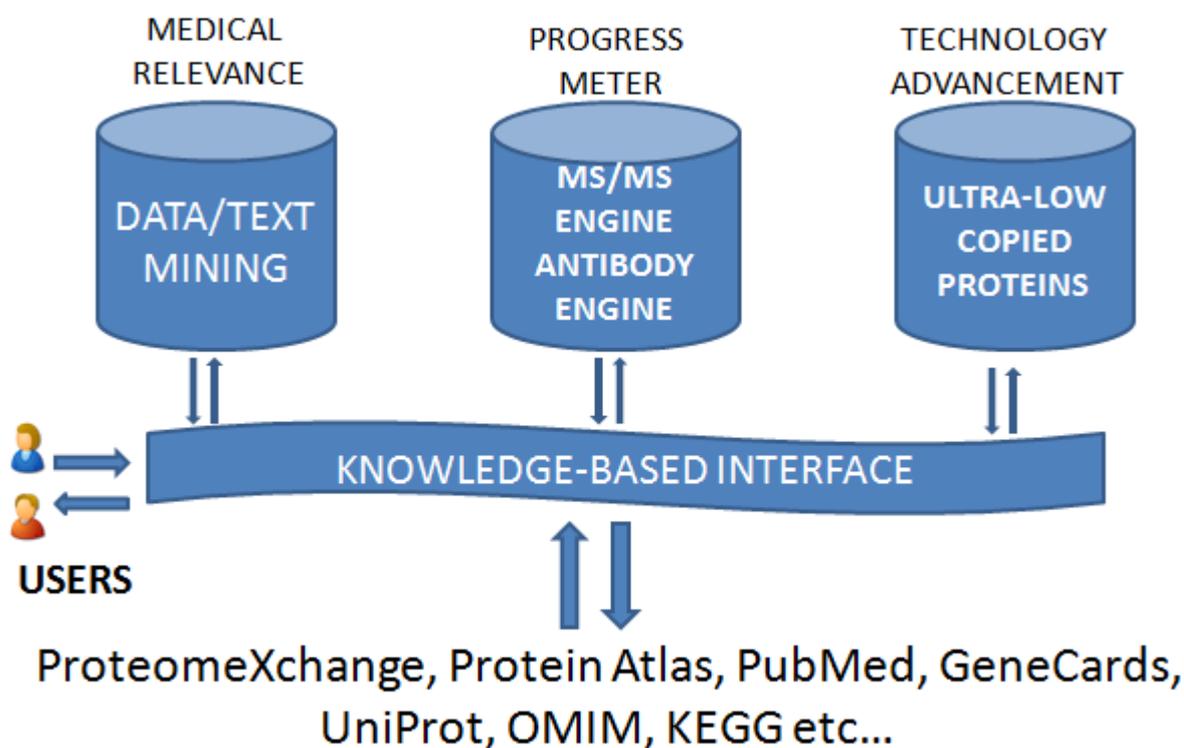


Рис. 8. Концепция базы знаний для реализации ПГЧ

13. Краткосрочные перспективы (5-ти летний план)

Предварительный этап исполнения работ в рамках дорожной карты состоит в инвентаризации высоко/средне-копийных белков и пептидов 18-й хромосомы и рассчитан на 3 года. Инвентаризация позволит получить:

- (i) минимум один белковый продукт для каждого гена 18-ой хромосомы минимум с 1 протеотипическим пептидом;
- (ii) список часто наблюдаемых модификаций, сплайсовые варианты и одно-аминокислотные полиморфизмы;

- (iii) распределение белков 18-ой хромосомы согласно их медицинской значимости (биомаркеры и/или лекарства мишени) и частоте идентификации в плазме крови и клетках ткани печени (HerG2 клетки);
- (iv) исходные диапазоны концентраций содержания высоко/среднекопийных белков в норме;
- (v) протоколы стандартизации для протеомных процедур;
- (vi) распределение взаимодействий белков в соответствии с их медицинским назначением и доступностью для идентификаций методами масс-спектрометрии;
- (vii) протоколы проведения экспериментов хромосомо-центричной интерактомики.

14. Долгосрочные перспективы (15-ти летний план)

В результате решения проблемы детектирования низко/ультранизкокопийных белков будет создан принципиально новый способ регистрации биомолекул с использованием нанотехнологии. Будут выпущены промышленные комплексы на базе гибридной технологии, сочетающие возможности масс-спектрометрии и атомно-силовой микроскопии. Объективно можно прогнозировать достижение чувствительности на уровне 10^{-18} М в плазме крови, и на уровне 10-100 копий белковой молекулы на 1 клетку ткани печени. Применение гибридных комплексов в научных целях позволит выявить биомаркеры ранних стадий заболеваний, непосредственно связанные с развитием патологических процессов.

На этапе внедрения в практику появятся прототипы молекулярных сенсоров на базе наноразмерных полупроводниковых элементов [86]. Данные устройства приведут к резкому удешевлению и масштабному распространению многопараметрических диагностических тестов. По уровню своей надежности, данные устройства будут пригодны для применения в режиме «у постели больного». Технологические достижения основной фазы дорожной карты позволят перейти к мониторинговой диагностике изменений молекулярных систем организма, что создаст научную основу для персонифицированной медицины.

На рис. 9 представлены основные возможности дорожной карты. В исходной точке — два, уже существующих лабораторных прототипа: атомно-силовой микроскоп, сопряженный с МС и нанопроводным сенсором. От лабораторных прототипов стрелка указывает на клиническое применение за счет открытия новых биомаркеров. Далее, в этом направлении, представлены устройства для применения непосредственно «у постели больного», что создает технологические условия для персонифицированной и децентрализованной медицины.

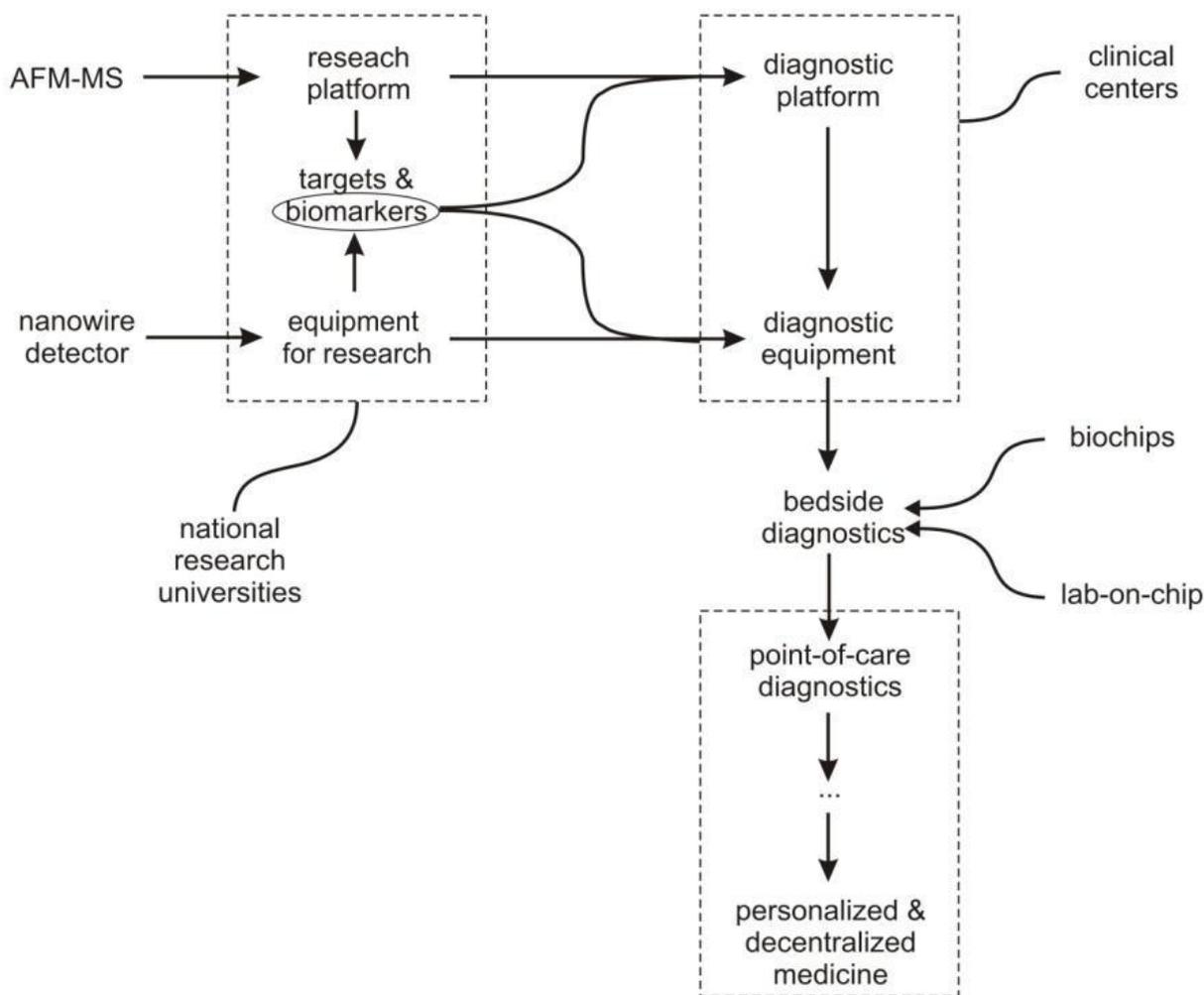


Рис. 9. Долгосрочные прогнозы на базе геноцентричного подхода к реализации Проекта Протеома Человека.

15. Заключение

Весь объем работ по проекту «Протеом человека» необходимо распределить между его участниками. Также как в проекте «Геном человека», предлагается распределить работу, основываясь на геноцентричном подходе [7], т.е. идентифицировать белковые продукты генов в соответствии с их распределением по хромосомам.

Геноцентричный подход, в отличие от прежних подходов на базе исследования ткани и изучения заболеваний, обеспечивает выполнение протеомных исследований. Отличие дорожной карты от научно-исследовательского плана состоит в том, что научные проблемы по дорожной карте решаются с использованием достижений в аналитической технологии, которые, в свою очередь, создают основу для разработки трансляционной медицины.

Дорожная карта, разработанная российскими учеными, основана на составлении списка белковых частей, кодируемых генами 18-ой хромосомы.

Данная хромосома сравнительно небольшая и примерно 80% кодируемых ею белков уже идентифицированы.

Предлагаемая дорожная карта включает предварительный и основной этапы научных исследований. В течение трехлетнего предварительного периода планируется идентифицировать белки 18-ой хромосомы в трех видах биологического материала: плазме крови, клеточных культурах HepG2 и ткани печени. Цель предварительного этапа исследований — идентифицировать, минимум, один белок для каждого гена, чтобы определить уровень его экспрессии и преобладающие модификации. В результате этой работы будут получены данные об индивидуальной вариабельности протеома плазмы крови и ткани печени.

Основной этап работ в рамках дорожной карты будет выполнен в 5-ти летний срок; он предусматривает выявление модификаций всех белков 18-ой хромосомы в ходе экспериментов. Модификации включают одноаминокислотные полиморфизмы (SAP), продукты альтернативного сплайсинга и пост-трансляционные модификации.

Дорожная карта ставит несколько задач, которые охватывают оба этапа исследований, направленных на реализацию проекта «Протеом человека»:

- анализ генома/транскриптома с использованием одно-молекулярных детекторов ДНК/РНК [60], чтобы получить более специфичный генокодирующий сиквенс 18-ой хромосомы и выявить транскрипты альтернативного сплайсинга;
- детектирование средне- и низко-копийных белков по технологии многомерного разделения MRM; протеотипирование и протеогеномное профилирование; расшифровка хромосома-центрированной части интерактома;
- усовершенствованная технология детектирования ультранизко-копийных белков путем интеграции молекулярных детекторов с помощью масс-спектрометрии [64],[78].

Усовершенствование аналитических технологий является основным стержнем дорожной карты, разработанной Россией. На наш взгляд, главной целью в этом направлении является использование подходов на базе нанотехнологии. Для участия в проекте ПЧ, Российские ученые планируют разработать новое оборудование для протеомных исследований, сочетающее возможности масс-спектрометрии и атомно-силовой микроскопии. С помощью таких комплексов АСМ-МС можно достичь чувствительности на уровне 10^{-18} М в плазме крови, и на уровне 10-100 копий белковой молекулы на 1 клетку ткани печени.

Дорожная карта перекрывается с трансляционной медициной, включая план по созданию молекулярных сенсоров на базе наноразмерных полупроводниковых элементов. В обозримом будущем, эти устройства обеспечат доступные средства для проведения комплексных, многопараметрических молекулярных анализов. Соответствующая надежность таких устройств позволит использовать их в режиме «у постели больного», а также в домашних условиях. Выполнение основного этапа работ по дорожной карте откроет дорогу для персонифицированной медицины

посредством предварительного создания научной основы для развития индивидуальной профилактики и лечения социально-значимых заболеваний.

Список литературы

- [1] R.K. Saiki, D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich, "Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.," *Science (New York, N.Y.)*, vol. 239, 1988, pp. 487-91.
- [2] M. Frenkel-Morgenstern, A.A. Cohen, N. Geva-Zatorsky, E. Eden, J. Prilusky, I. Issaeva, A. Sigal, C. Cohen-Saidon, Y. Liron, L. Cohen, T. Danon, N. Perzov, and U. Alon, "Dynamic Proteomics: a database for dynamics and localizations of endogenous fluorescently-tagged proteins in living human cells.," *Nucleic acids research*, vol. 38, 2010, pp. D508-12.
- [3] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al., "Initial sequencing and analysis of the human genome.," *Nature*, vol. 409, 2001, pp. 860-921.
- [4] I. Consortium, "Finishing the euchromatic sequence of the human genome.," *Nature*, vol. 431, 2004, pp. 931-45.
- [5] K.U. Mir, "Sequencing genomes: from individuals to populations.," *Briefings in functional genomics & proteomics*, vol. 8, 2009, pp. 367-78.
- [6] M. Snyder, J. Du, and M. Gerstein, "Personal genome sequencing: current approaches and challenges.," *Genes & development*, vol. 24, 2010, pp. 423-31.
- [7] HUPO, "A gene-centric human proteome project: HUPO--the Human Proteome organization.," *Molecular & cellular proteomics : MCP*, vol. 9, 2010, pp. 427-9.
- [8] L.A. Liotta and E.F. Petricoin, "Beyond the genome to tissue proteomics.," *Breast cancer research : BCR*, vol. 2, 2000, pp. 13-4.
- [9] Y.D. Ivanov, V.M. Govorun, V.A. Bykov, and A.I. Archakov, "Nanotechnologies in proteomics.," *Proteomics*, vol. 6, 2006, pp. 1399-414.
- [10] A.I. Archakov, Y.D. Ivanov, A.V. Lisitsa, and V.G. Zgoda, "AFM fishing nanotechnology is the way to reverse the Avogadro number in proteomics.," *Proteomics*, vol. 7, 2007, pp. 4-9.

- [11] J. Hahm and C.M. Lieber, "Direct Ultrasensitive Electrical Detection of DNA and DNA Sequence Variations Using Nanowire Nanosensors," *Nano Letters*, vol. 4, 2004, pp. 51-54.
- [12] P. Picotti, B. Bodenmiller, L.N. Mueller, B. Domon, and R. Aebersold, "Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics.," *Cell*, vol. 138, 2009, pp. 795-806.
- [13] J. Robinson, M.J. Waller, S.C. Fail, H. McWilliam, R. Lopez, P. Parham, and S.G. Marsh, "The IMGT/HLA database.," *Nucleic acids research*, vol. 37, 2009, pp. D1013-7.
- [14] M. Rebhan, V. Chalifa-Caspi, J. Prilusky, and D. Lancet, "GeneCards: integrating information about genes, proteins and diseases.," *Trends in genetics : TIG*, vol. 13, 1997, p. 163.
- [15] J.H. EDWARDS, D.G. HARNDEN, A.H. CAMERON, V.M. CROSSE, and O.H. WOLFF, "A new trisomic syndrome.," *Lancet*, vol. 1, 1960, pp. 787-90.
- [16] B. Goc, Z. Walencka, A. Włoch, E. Wojciechowska, D. Wiecek-Włodarska, J. Krzystolik-Ładzińska, K. Bober, and J. Swietliński, "Trisomy 18 in neonates: prenatal diagnosis, clinical features, therapeutic dilemmas and outcome.," *Journal of applied genetics*, vol. 47, 2006, pp. 165-70.
- [17] K. Gu, W.C. Chan, and R.C. Hawley, "Practical detection of t(14;18)(IgH/BCL2) in follicular lymphoma.," *Archives of pathology & laboratory medicine*, vol. 132, 2008, pp. 1355-61.
- [18] "Human Genome Project Information
http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml."
- [19] C. Nusbaum, M.C. Zody, M.L. Borowsky, M. Kamal, C.D. Kodira, T.D. Taylor, C.A. Whittaker, J.L. Chang, C.A. Cuomo, K. Dewar, M.G. FitzGerald, X. Yang, A. Abouelleil, N.R. Allen, S. Anderson, T. Bloom, B. Bugalter, J. Butler, A. Cook, D. DeCaprio, R. Engels, M. Garber, A. Gnirke, N. Hafez, J.L. Hall, C.H. Norman, T. Itoh, D.B. Jaffe, Y. Kuroki, J. Lehoczky, A. Lui, P. Macdonald, E. Mauceli, T.S. Mikkelsen, J.W. Naylor, R. Nicol, C. Nguyen, H. Noguchi, S.B. O'Leary, K. O'Neill, B. Piquani, C.L. Smith, J.A. Talamas, K. Topham, Y. Totoki, A. Toyoda, H.M. Wain, S.K. Young, Q. Zeng, A.R. Zimmer, A. Fujiyama, M. Hattori, B.W. Birren, Y. Sakaki, and E.S. Lander, "DNA sequence and analysis of human chromosome 18.," *Nature*, vol. 437, 2005, pp. 551-5.
- [20] "UniProt <http://www.uniprot.org/>."
- [21] "Human Protein Atlas <http://www.proteinatlas.org/>."
- [22] "PRoteomics IDEntifications database (PRIDE) <http://www.ebi.ac.uk/pride/>."
- [23] F. He, "Human liver proteome project: plan, progress, and perspectives.," *Molecular & cellular proteomics : MCP*, vol. 4, 2005, pp. 1841-8.
- [24] G.S. Omenn, D.J. States, M. Adamski, T.W. Blackwell, R. Menon, H. Hermjakob, R. Apweiler, B.B. Haab, R.J. Simpson, J.S. Eddes, E.A. Kapp, R.L. Moritz, D.W. Chan, A.J. Rai, A. Admon, R. Aebersold, J. Eng, W.S. Hancock, S.A. Hefta, H. Meyer, Y. Paik, J.

Yoo, P. Ping, J. Pounds, J. Adkins, X. Qian, R. Wang, V. Wasinger, C.Y. Wu, X. Zhao, R. Zeng, A. Archakov, A. Tsugita, I. Beer, A. Pandey, M. Pisano, P. Andrews, H. Tammen, D.W. Speicher, and S.M. Hanash, "Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database.," *Proteomics*, vol. 5, 2005, pp. 3226-45.

[25] C.J. Mee, H.J. Harris, M.J. Farquhar, G. Wilson, G. Reynolds, C. Davis, S.C. van IJzendoorn, P. Balfe, and J.A. McKeating, "Polarization restricts hepatitis C virus entry into HepG2 hepatoma cells.," *Journal of virology*, vol. 83, 2009, pp. 6211-21.

[26] A. Slany, V.J. Haudek, H. Zwickl, N.C. Gundacker, M. Grusch, T.S. Weiss, K. Seir, C. Rodgarkia-Dara, C. Hellerbrand, and C. Gerner, "Cell characterization by proteome profiling applied to primary hepatocytes and hepatocyte cell lines Hep-G2 and Hep-3B.," *Journal of proteome research*, vol. 9, 2010, pp. 6-21.

[27] V.G. Zgoda, S.A. Moshkovskii, E.A. Ponomarenko, T.V. Andreewski, A.T. Kopylov, O.V. Tikhonova, S.A. Melnik, A.V. Lisitsa, and A.I. Archakov, "Proteomics of mouse liver microsomes: performance of different protein separation workflows for LC-MS/MS.," *Proteomics*, vol. 9, 2009, pp. 4102-5.

[28] N.A. Petushkova, I.P. Kanaeva, A.V. Lisitsa, G.F. Sheremetyeva, V.G. Zgoda, N.F. Samenkova, I.I. Karuzina, and A.I. Archakov, "Characterization of human liver cytochromes P450 by combining the biochemical and proteomic approaches.," *Toxicology in vitro : an international journal published in association with BIBRA*, vol. 20, 2006, pp. 966-74.

[29] A.V. Lisitsa, N.A. Petushkova, H. Thiele, S.A. Moshkovskii, V.G. Zgoda, I.I. Karuzina, A.L. Chernobrovkin, O.G. Skipenko, and A.I. Archakov, "Application of slicing of one-dimensional gels with subsequent slice-by-slice mass spectrometry for the proteomic profiling of human liver cytochromes P450.," *Journal of proteome research*, vol. 9, 2010, pp. 95-103.

[30] V. Zgoda, O. Tikhonova, A. Viglinskaya, M. Serebriakova, A. Lisitsa, and A. Archakov, "Proteomic profiles of induced hepatotoxicity at the subcellular level.," *Proteomics*, vol. 6, 2006, pp. 4662-70.

[31] L. Zang, D. Palmer Toy, W.S. Hancock, D.C. Sgroi, and B.L. Karger, "Proteomic analysis of ductal carcinoma of the breast using laser capture microdissection, LC-MS, and ¹⁶O/¹⁸O isotopic labeling.," *Journal of proteome research*, vol. 3, pp. 604-12.

[32] D.J. Johann, J. Rodriguez-Canales, S. Mukherjee, D.A. Prieto, J.C. Hanson, M. Emmert-Buck, and J. Blonder, "Approaching solid tumor heterogeneity on a cellular basis by tissue proteomics using laser capture microdissection and biological mass spectrometry.," *Journal of proteome research*, vol. 8, 2009, pp. 2310-8.

[33] G. Marko-Varga, M. Berglund, J. Malmström, H. Lindberg, and T.E. Fehniger, "Targeting hepatocytes from liver tissue by laser capture microdissection and proteomics expression profiling.," *Electrophoresis*, vol. 24, 2003, pp. 3800-5. 34

- [34] L. Beretta, "Comparative analysis of the liver and plasma proteomes as a novel and powerful strategy for hepatocellular carcinoma biomarker discovery.," *Cancer letters*, vol. 286, 2009, pp. 134-9.
- [35] L. Anderson and N.G. Anderson, "High resolution two-dimensional electrophoresis of human plasma proteins.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 74, 1977, pp. 5421-5.
- [36] A.J. Rai, C.A. Gelfand, B.C. Haywood, D.J. Warunek, J. Yi, M.D. Schuchard, R.J. Mehigh, S.L. Cockrill, G.B. Scott, H. Tammen, P. Schulz-Knappe, D.W. Speicher, F. Vitzthum, B.B. Haab, G. Siest, and D.W. Chan, "HUPO Plasma Proteome Project specimen collection and handling: towards the standardization of parameters for plasma proteome samples.," *Proteomics*, vol. 5, 2005, pp. 3262-77.
- [37] M.A. Karpova, S.A. Moshkovskii, I.Y. Toropygin, and A.I. Archakov, "Cancer-specific MALDI-TOF profiles of blood serum and plasma: biological meaning and perspectives.," *Journal of proteomics*, vol. 73, 2010, pp. 537-51.
- [38] Y. Shen, J. Kim, E.F. Strittmatter, J.M. Jacobs, D.G. Camp, R. Fang, N. Tolié, R.J. Moore, and R.D. Smith, "Characterization of the human blood plasma proteome.," *Proteomics*, vol. 5, 2005, pp. 4034-45.
- [39] "Proteomics National Center for Research Resource
<http://www.pnl.gov/biology/programs/msd/ncrr.stm>."
- [40] N.L. Anderson, M. Polanski, R. Pieper, T. Gatlin, R.S. Tirumalai, T.P. Conrads, T.D. Veenstra, J.N. Adkins, J.G. Pounds, R. Fagan, and A. Lobley, "The human plasma proteome: a nonredundant list developed by combination of four separate sources.," *Molecular & cellular proteomics : MCP*, vol. 3, 2004, pp. 311-26.
- [41] "HUPO Plasma Proteome Project <http://www.bioinformatics.med.umich.edu/hupo/ppp>."
- [42] P.A. Rudnick, K.R. Clauser, L.E. Kilpatrick, D.V. Tchekhovskoi, P. Neta, N. Blonder, D.D. Billheimer, R.K. Blackman, D.M. Bunk, H.L. Cardasis, A.L. Ham, J.D. Jaffe, C.R. Kinsinger, M. Mesri, T.A. Neubert, B. Schilling, D.L. Tabb, T.J. Tegeler, L. Vega-Montoto, A.M. Variyath, M. Wang, P. Wang, J.R. Whiteaker, L.J. Zimmerman, S.A. Carr, S.J. Fisher, B.W. Gibson, A.G. Paulovich, F.E. Regnier, H. Rodriguez, C. Spiegelman, P. Tempst, D.C. Liebler, and S.E. Stein, "Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses.," *Molecular & cellular proteomics : MCP*, vol. 9, 2010, pp. 225-41.
- [43] A.W. Bell, E.W. Deutsch, C.E. Au, R.E. Kearney, R. Beavis, S. Sechi, T. Nilsson, and J.J. Bergeron, "A HUPO test sample study reveals common problems in mass spectrometry-based proteomics.," *Nature methods*, vol. 6, 2009, pp. 423-30.
- [44] D.L. Tabb, L. Vega-Montoto, P.A. Rudnick, A.M. Variyath, A.L. Ham, D.M. Bunk, L.E. Kilpatrick, D.D. Billheimer, R.K. Blackman, H.L. Cardasis, S.A. Carr, K.R. Clauser, J.D. Jaffe, K.A. Kowalski, T.A. Neubert, F.E. Regnier, B. Schilling, T.J. Tegeler, M. Wang, P. Wang, J.R. Whiteaker, L.J. Zimmerman, S.J. Fisher, B.W. Gibson, C.R. Kinsinger, M. Mesri, H. Rodriguez, S.E. Stein, P. Tempst, A.G. Paulovich, D.C. Liebler, and C. Spiegelman, "Repeatability and reproducibility in proteomic identifications by liquid

chromatography-tandem mass spectrometry.," *Journal of proteome research*, vol. 9, 2010, pp. 761-76.

[45] M.A. Kuzyk, D. Smith, J. Yang, T.J. Cross, A.M. Jackson, D.B. Hardie, N.L. Anderson, and C.H. Borchers, "Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma.," *Molecular & cellular proteomics : MCP*, vol. 8, 2009, pp. 1860-77.

[46] T.A. Addona, S.E. Abbatiello, B. Schilling, S.J. Skates, D.R. Mani, D.M. Bunk, C.H. Spiegelman, L.J. Zimmerman, A.L. Ham, H. Keshishian, S.C. Hall, S. Allen, R.K. Blackman, C.H. Borchers, C. Buck, H.L. Cardasis, M.P. Cusack, N.G. Dodder, B.W. Gibson, J.M. Held, T. Hiltke, A. Jackson, E.B. Johansen, C.R. Kinsinger, J. Li, M. Mesri, T.A. Neubert, R.K. Niles, T.C. Pulsipher, D. Ransohoff, H. Rodriguez, P.A. Rudnick, D. Smith, D.L. Tabb, T.J. Tegeler, A.M. Variyath, L.J. Vega-Montoto, A. Wahlander, S. Waldemarson, M. Wang, J.R. Whiteaker, L. Zhao, N.L. Anderson, S.J. Fisher, D.C. Liebler, A.G. Paulovich, F.E. Regnier, P. Tempst, and S.A. Carr, "Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma.," *Nature biotechnology*, vol. 27, 2009, pp. 633-41.

[47] M.J. Roth, B.A. Parks, J.T. Ferguson, M.T. Boyne, and N.L. Kelleher, "'Proteotyping': population proteomics of human leukocytes using top down mass spectrometry.," *Analytical chemistry*, vol. 80, 2008, pp. 2857-66.

[48] T.H. Corzett, I.K. Fodor, M.W. Choi, V.L. Walsworth, K.W. Turteltaub, S.L. McCutchen-Maloney, and B.A. Chromy, "Statistical analysis of variation in the human plasma proteome.," *Journal of biomedicine & biotechnology*, vol. 2010, 2010, p. 258494.

[49] D. Nedelkov, "Population proteomics: investigation of protein diversity in human populations.," *Proteomics*, vol. 8, 2008, pp. 779-86.

[50] J. Godovac-Zimmermann, O. Kleiner, L.R. Brown, and A.K. Drukier, "Perspectives in spicing up proteomics with splicing.," *Proteomics*, vol. 5, 2005, pp. 699-709.

[51] B.A. Parks, L. Jiang, P.M. Thomas, C.D. Wenger, M.J. Roth, M.T. Boyne, P.V. Burke, K.E. Kwast, and N.L. Kelleher, "Top-down proteomics on a chromatographic time scale using linear ion trap fourier transform hybrid mass spectrometers.," *Analytical chemistry*, vol. 79, 2007, pp. 7984-91.

[52] N.M. Karabacak, L. Li, A. Tiwari, L.J. Hayward, P. Hong, M.L. Easterling, and J.N. Agar, "Sensitive and specific identification of wild type and variant proteins from 8 to 669 kDa using top-down mass spectrometry.," *Molecular & cellular proteomics : MCP*, vol. 8, 2009, pp. 846-56.

[53] G. Alves, A.Y. Ogurtsov, and Y. Yu, "RAId_DbS: mass-spectrometry based peptide identification web server with knowledge integration.," *BMC genomics*, vol. 9, 2008, p. 505.

[54] "National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/>."

[55] "Ensembl www.ensembl.org."

- [56] C.M. Delahunty and J.R. Yates, "MudPIT: multidimensional protein identification technology.," *BioTechniques*, vol. 43, 2007, p. 563, 565, 567 passim.
- [57] G. Chen, T.G. Gharib, C. Huang, J.M. Taylor, D.E. Misek, S.L. Kardia, T.J. Giordano, M.D. Iannettoni, M.B. Orringer, S.M. Hanash, and D.G. Beer, "Discordant protein and mRNA expression in lung adenocarcinomas.," *Molecular & cellular proteomics : MCP*, vol. 1, 2002, pp. 304-13.
- [58] A.T. Kopylov, V.G. Zgoda, and A.I. Archakov, "[Label-free quantitative analysis of proteins using mass-spectrometry]," *Biomeditsinskaiā khimiā*, vol. 55, pp. 125-39.
- [59] R. Kakuhata, M. Watanabe, T. Yamamoto, E. Obana, N. Yamazaki, M. Kataoka, T. Ooie, Y. Baba, T. Hori, and Y. Shinohara, "Importance of probe location for quantitative comparison of signal intensities among genes in microarray analysis.," *Journal of biochemical and biophysical methods*, vol. 70, 2008, pp. 926-31.
- [60] M. Tsutsui, M. Taniguchi, K. Yokota, and T. Kawai, "Identifying single nucleotides by tunnelling current.," *Nature nanotechnology*, vol. 5, 2010, pp. 286-90.
- [61] A. Archakov, "Introducing Nanoproteomics, a new section in PROTEOMICS," *PROTEOMICS*, vol. 7, 2007, pp. 4409-4412.
- [62] J. Armengaud, "Proteogenomics and systems biology: quest for the ultimate missing parts.," *Expert review of proteomics*, vol. 7, 2010, pp. 65-77.
- [63] V.T. Ivanov and O.N. Yatskin, "Peptidomics: a logical sequel to proteomics.," *Expert review of proteomics*, vol. 2, 2005, pp. 463-73.
- [64] A.L. Kaysheva, Y.D. Ivanov, V.G. Zgoda, P.A. Frantsuzov, T.O. Pleshakova, N.V. Krokhin, V.S. Ziborov, and A.I. Archakov, "Visualization and identification of hepatitis C viral particles by atomic force microscopy combined with MS/MS analysis," *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry*, vol. 4, 2010, pp. 15-24.
- [65] A. Archakov, Y. Ivanov, A. Lisitsa, and V. Zgoda, "Biospecific irreversible fishing coupled with atomic force microscopy for detection of extremely low-abundant proteins.," *Proteomics*, vol. 9, 2009, pp. 1326-43.
- [66] V.Y. Kuznetsov, Y.D. Ivanov, V.A. Bykov, S.A. Saunin, I.A. Fedorov, S.V. Lemeshko, H.B. Hoa, and A.I. Archakov, "Atomic force microscopy detection of molecular complexes in multiprotein P450cam containing monooxygenase system.," *Proteomics*, vol. 2, 2002, pp. 1699-705.
- [67] V.Y. Kuznetsov, Y.D. Ivanov, and A.I. Archakov, "Atomic force microscopy revelation of molecular complexes in the multiprotein cytochrome P450 2B4-containing system.," *Proteomics*, vol. 4, 2004, pp. 2390-6.
- [68] M. Shibata, H. Yamashita, T. Uchihashi, H. Kandori, and T. Ando, "High-speed atomic force microscopy shows dynamic molecular processes in photoactivated bacteriorhodopsin.," *Nature nanotechnology*, vol. 5, 2010, pp. 208-12. 37

- [69] J. Hernandez-Toro, C. Prieto, and J. De Las Rivas, "APID2NET: unified interactome graphic analyzer.," *Bioinformatics (Oxford, England)*, vol. 23, 2007, pp. 2495-7.
- [70] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa, "KEGG for representation and analysis of molecular networks involving diseases and drugs.," *Nucleic acids research*, vol. 38, 2010, pp. D355-60.
- [71] A. Villanueva, C. García, A.B. Paules, M. Vicente, M. Megías, G. Reyes, P. de Villalonga, N. Agell, F. Lluís, O. Bachs, and G. Capellá, "Disruption of the antiproliferative TGF-beta signaling pathways in human pancreatic cancer cells.," *Oncogene*, vol. 17, 1998, pp. 1969-78.
- [72] Y. Shi and J. Massagué, "Mechanisms of TGF-beta signaling from cell membrane to the nucleus.," *Cell*, vol. 113, 2003, pp. 685-700.
- [73] G. Yang and X. Yang, "Smad4-mediated TGF-beta signaling in tumorigenesis.," *International journal of biological sciences*, vol. 6, 2010, pp. 1-8.
- [74] B. Sobolev, D. Filimonov, A. Lagunin, A. Zakharov, O. Koborova, A. Kel, and V. Poroikov, "Functional classification of proteins based on projection of amino acid sequences: application for prediction of protein kinase substrates.," *BMC bioinformatics*, vol. 11, 2010, p. 313.
- [75] M. Pellegrini, E.M. Marcotte, M.J. Thompson, D. Eisenberg, and T.O. Yeates, "Assigning protein functions by comparative genome analysis: protein phylogenetic profiles.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, 1999, pp. 4285-8.
- [76] M.A. Piatnitskiĭ, A.V. Lisitsa, and A.I. Archakov, "[Prediction of functionally related proteins by comparative genomics in silico]," *Biomeditsinskaiâ khimiâ*, vol. 55, pp. 230-46.
- [77] O. Buneeva, O. Gnedenko, V. Zgoda, A. Kopylov, V. Glover, A. Ivanov, A. Medvedev, and A. Archakov, "Isatin-binding proteins of rat and mouse brain: proteomic identification and optical biosensor validation.," *Proteomics*, vol. 10, 2010, pp. 23-37.
- [78] D. Nedelkov, "Integration of SPR biosensors with mass spectrometry (SPR-MS).," *Methods in molecular biology (Clifton, N.J.)*, vol. 627, 2010, pp. 261-8.
- [79] B.J. Strasser, "Genetics. GenBank--Natural history in the 21st Century?," *Science (New York, N.Y.)*, vol. 322, 2008, pp. 537-8.
- [80] O. Kuchar, J. Reyes-Spindola, and M. Benaroch, "Cognitive Bioinformatics: Computational Cognitive Model for Dynamic Problem Solving," *Proceedings of the Third IEEE International Conference on Cognitive Informatics*, 2004, pp. 84-92.
- [81] D. Hull, S.R. Pettifer, and D.B. Kell, "Defrosting the digital library: bibliographic tools for the next generation web.," *PLoS computational biology*, vol. 4, 2008, p. e1000204.
- [82] "Collexis <http://collexis.com/>."
- [83] "Neosemantic Framework (ProContent) <http://www.neosemanticsoft.com/>."

- [84] Z. Hu, I. Mani, V. Hermoso, H. Liu, and C.H. Wu, "iProLINK: an integrated protein resource for literature mining.," *Computational biology and chemistry*, vol. 28, 2004, pp. 409-16.
- [85] R. Hussong, B. Gregorius, A. Tholey, and A. Hildebrandt, "Highly accelerated feature detection in proteomics data sets using modern graphics processing units.," *Bioinformatics (Oxford, England)*, vol. 25, 2009, pp. 1937-43.
- [86] M. Lee, D. Lee, S. Jung, K. Lee, Y.S. Park, and W. Seong, "Measurements of serum C-reactive protein levels in patients with gastric cancer and quantification using silicon nanowire arrays.," *Nanomedicine : nanotechnology, biology, and medicine*, vol. 6, 2010, pp. 78-83.