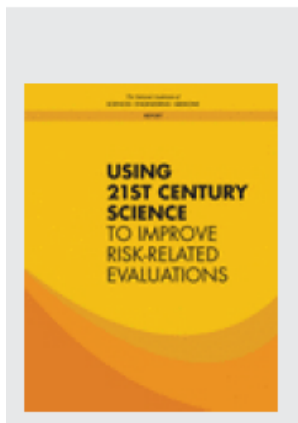


The National
Academies of

SCIENCES
ENGINEERING
MEDICINE

THE NATIONAL ACADEMIES PRESS

This PDF is available at <http://www.nap.edu/24635>



Using 21st Century Science to Improve Risk-Related Evaluations

DETAILS

200 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-45348-6 | DOI: 10.17226/24635

CONTRIBUTORS

Committee on Incorporating 21st Century Science into Risk-Based Evaluations; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Academies of Sciences, Engineering, and Medicine

Наука 21-го века: вклад в оценку токсикологических рисков

Глава 4: Достижения в Эпидемиологии

Перевод на русский язык: Баландина Юлия Геннадиевна
Редактор русского перевода: д.б.н. Суворов Александр Николаевич

Достижения в Эпидемиологии

Эпидемиология - это наука, изучающая закономерности развития и распространения заболеваний в популяциях (Gordis, 2013). Описательный элемент эпидемиологии включает слежение за показателями здоровья и факторами риска для населения (наблюдение). Этиологическая деятельность - поиск причин и детерминант заболевания - включает в первую очередь проведение контрольных и когортных исследований. Эпидемиологические исследования также включают в себя интервенционные исследования, как произвольные (при назначении профилактических мер, таких как вакцинация), так и непроизвольные.

В этой главе рассматриваются развивающиеся подходы, используемые эпидемиологами для изучения связей между факторами окружающей среды и болезнями человека и роль эпидемиологии в переходе к новой стратегии тестирования. Глава не содержит общего введения в эпидемиологию; такой материал легко доступен в учебниках и других источниках. Вместе с тем, в данной главе обсуждается роль эпидемиологии в оценке риска, направление развития эпидемиологии как науки, имеющиеся возможности по использованию баз данных и ошибки, которые следует учитывать при использовании методов, обозначенных в программах Tox21 и ES21. В данной главе основное внимание уделяется использованию -omics технологий в эпидемиологии, а также даются некоторые рекомендации по решению имеющихся проблем.

ОЦЕНКА РИСКА И ЭПИДЕМИОЛОГИЯ

Роль эпидемиологических данных многократно подчеркивалась как в парадигме оценки риска, первоначально описанной в докладе «Оценка риска в федеральном правительстве: управление процессом» («Risk Assessment in the Federal Government: Managing the Process», NRC 1983), так и в различных последующих отчетах (Samet et al., 1998). Выявление факторов риска в появлении и развитии заболеваний и причинно-следственных связей между экспозицией и болезнью с помощью эпидемиологических исследований являются важными компонентами идентификации опасности. Доказательства опасности, полученные в результате эпидемиологических исследований, имеют приоритет в рекомендациях (evidence-evaluation guidelines) по оценке фактических данных, в том числе

в Агентстве по охране окружающей среды США и Международном агентстве по исследованию рака (МАИР). Например, эпидемиологические данные для канцерогенных веществ были признаны достаточно убедительными для установления причинно-следственной связи. Однако данные о людях доступны лишь по относительно небольшому числу химических веществ, особенно по сравнению с тем их количеством, экспозицией которого подвергается население. При отсутствии естественных экспериментов эпидемиологические исследования являются единственным доступным научным и этически приемлемым подходом для изучения возможных эффектов потенциально опасных агентов непосредственно в популяциях человека.

В дополнение к предоставлению доказательств для идентификации опасности, эпидемиологические исследования могут обеспечить понимание взаимосвязи между экспозицией и ответной реакцией. Для некоторых химических веществ эффекты экспозиции изучались, прежде всего, в определенных группах работников, таких как работники асбестовых предприятий; при значениях экспозиции, как правило, гораздо более высоких, чем у населения в целом; при этом отношения «экспозиция-реакция» экстраполируются вниз, что приводит к погрешностям. В случае имеющихся необходимых данных об экспозиции населения в целом, эпидемиологические исследования могут предоставить важную информацию о концентрациях химических веществ, представляющих риск. Например, на больших когортах, включая исследование Американского общества по профилактике рака (American Cancer Society's Cancer Prevention Study «) и многочисленные исследования, проведенные в рамках Европейского исследования по воздействию загрязнения воздуха (European Study of Cohorts for Air Pollution Effects, ESCAPE 2014) были оценены уровни загрязнения воздуха. Несмотря на то, что в некоторых случаях имеют место ошибочные выводы о взаимосвязи экспозиции загрязнителями окружающей среды и экспозиции, связанной с профессиональной деятельностью с возникающими эффектами на здоровье, существует множество примеров успешного включения эпидемиологически обоснованных взаимосвязей между экспозицией и реакцией в оценки риска: ионизирующее излучение и рак, загрязнение воздуха и смертность, экспозиция мышьяком и рак, экспозиция свинцом на ранних стадиях развития и нейроразвитие. Были также разработаны методы исправления ошибок; такие поправки обычно приводят к кривым зависимости «экспозиция-ответ» с более крутыми склонами (Hart et al., 2015).

В случае доступности информации об индивидуальных особенностях участников исследования (возраст, пол или геном), эпидемиологические исследования могут также способствовать улучшению понимания взаимосвязи между экспозицией и ответной реакцией в зависимости от этих параметров. Данные, собираемые в процессе эпидемиологических исследований или в ходе эпидемиологического надзора за населением на основе вопросников, мониторинга, моделей и анализов биологических образцов, могут быть полезны для описания распределения экспозиции.

Эпидемиологические исследования могут также предоставить информацию об общем демографическом риске. Статистика по атрибутивному риску (количество смертей в результате экспозиции), первоначально разработанная для оценки рака легких, вызванного курением, дает оценку груза болезней, вызванного причинным фактором (Levin, 1953). Таким образом, эпидемиологические данные о человеческих популяциях будут способствовать улучшению всех четырех компонентов оценки риска, описанных в новой парадигме тестирования (см. гл. 1).

ЭПИДЕМИОЛОГИЯ В 21 ВЕКЕ

Развитие науки эпидемиологии

Методы эпидемиологических исследований не были статичными. Первоначально эпидемиологические исследования по этиологии неинфекционных болезней, прежде всего рака, сердечно-сосудистых заболеваний, легочных заболеваний и заболеваний, связанных с обменом веществ, были ориентированы на конкретные факторы риска; оценка экспозиции определялась в основном за счет использования анкет, замеров и оценочных методов в случае исследований, связанных с профессиональной деятельностью, и довольно грубых показателей в случае экспозиции факторами окружающей среды. В некоторых исследованиях использовались показатели из биологических образцов (концентрации свинца или кадмия), а также некоторые вычисленные показатели экспозиции в моделях на базе обширных данных. Например, при изучении жертв атомных бомбардировок Хиросимы и Нагасаки доза облучения оценивалась с помощью сложного алгоритма, который включал такую информацию, как местоположение и положение тела во время взрыва. Эпидемиологические исследования неинфекционных болезней, начатые в 1950-х годах, были сосредоточены на индивидуальных факторах риска; в некоторых

более поздних исследованиях стали учитываться детерминанты риска на более высоких уровнях социальной или организационной структуры, включая семью, места проживания и работы, а также регион и страну проживания. Предпринимались усилия по созданию исследований в концептуальных рамках, отражающих понимание структурных, социологических и культурных факторов, влияющих на состояние здоровья и риск возникновения заболевания, а в последние десятилетия все большее внимание уделяется подходам на основе жизненного цикла, где признается необходимость изучения влияния экспозиции на ранних стадиях развития, даже внутриутробные и трансгенеративные экспозиции. Кроме того, многие более поздние исследования в области окружающей среды и здоровья были разработаны для того, чтобы отразить различия в экспозиции между сообществами и внутри них.

В последнее время на эпидемиологические исследования серьезно повлияли достижения в других областях науки. Начало 21-го века характеризовалось быстрым развитием технологий, медицинских наук, биологии и генетики, имеющих отношение к эпидемиологии (Hiatt et al., 2013). Критическим для эпидемиологии стало появление усовершенствованных вычислительных моделей и создание мощных баз хранения данных. Например, важную роль в трансформации практики эпидемиологии сыграло появление исследований в области геномики (GWAS).

Необходимость в получении образцов, достаточно больших, чтобы обеспечить исследования, обладающие достаточной статистической силой, и необходимость проверки полученных данных в независимых популяциях способствовали созданию крупных групп эпидемиологических исследований, многоцентровых исследований и консорциумов, разработке мета-аналитических инструментов и культуре обмена данными. В последние десятилетия наблюдается эволюция от единичных исследовательских групп, имеющих контроль над данными и образцами к созданию исследовательских консорциумов, которые, на основе командного подхода увеличивают результативность работы за счет более широкого обмена данными, протоколами и аналитическими методами (Gutmacher Et al., 2009; Tenopir et al., 2011). Некоторые организации, финансирующие исследования, стремятся еще больше активизировать эти преобразования, поддерживая разработку и распространение валидированных, соответствующих современному состоянию науки протоколов, предназначенных для проведения широкого спектра фенотипических замеров, с тем чтобы отдельные исследовательские группы (при разработке новых

исследований) могли эффективней использовать и сравнивать данные (PhenX Toolkit NHGRI).

Тематические (case-control studies) и когортные популяционные исследования - традиционные рабочие лошадки эпидемиологии - будут по-прежнему продолжать играть значительную роль в эпидемиологических исследованиях. В частности, исследования по методу «случай-контроль» будут и далее способствовать своевременному глубокому изучению людей с конкретными эффектами на здоровье, такими как редкие виды рака, нарушение репродуктивных функций или специфические врожденные дефекты. Исследования на когортах будут и впредь помогать в определении ранних факторов, ассоциированных с развитием заболеваний, выявлении доклинических биомаркеров, факторов риска и способствовать формированию прикладных исследований и точной медицины (прим. перев.: точная медицина «precision medicine» — инновационный подход в системе здравоохранения, при котором профилактика и лечение заболеваний подбирается главным образом по генетической информации).

Когортные исследования, если они начаты в достаточно раннем возрасте, могут быть особенно полезными для определения влияния ранней экспозиции на здоровье на протяжении всего жизненного цикла. Комитет ожидает все большего числа когортных исследований, объединяющих информацию о лечении и результатах лечения из разных источников, включая информацию от организаций, оказывающих медицинскую помощь. Исследования, предоставляющие информацию об образцах из биобанков, станут главными источниками, помогающими соединить данные, полученные с помощью -omics технологий и других методов с фактами патогенеза в человеческих популяциях. Возможность включения географических сведений, документирующих физическую и социальную среду популяции в мелком масштабе, существенно обогатит базы данных и позволит внести дополнительную ясность в понимание влияния экспозиции на здоровье человека.

Итак, факторы, реконфигурирующие науку эпидемиологию в XXI веке, включают в себя: расширение междисциплинарного диапазона; возрастающую сложность научного исследования, включая многоуровневый анализ, учет этиологии и развития заболевания на протяжении всего жизненного цикла; появление новых источников и технологий для генерирования данных; прогресс в характеристике экспозиции; а также возросшие требования к интеграции новых знаний в области фундаментальных, клинических и демографических наук (Lam et al., 2013). Существует также тенденция к регистрации

прошлых и настоящих данных с тем, чтобы в случае возникновения конкретной проблемы можно было бы комбинировать новые данные с уже имеющимися. Уже существуют модели для агрегирования данных во всех исследованиях (например, National Cancer Institute Cohort Consortium и Agricultural Health cohorts). Следует подчеркнуть, что исследователи признают необходимость достижения согласованности сбора данных для облегчения агрегации набора данных в будущем (PhenX Toolkit NHGRI, Fortier et al., 2010). Они также рассматривают возможность создания глобальных биобанков (Harris et al., 2012).

Возможности новых баз данных

Эпидемиология всегда была дисциплиной, использующей большой объем информации с целью выявления факторов риска, что, в конечном счете, имеет целью снижение заболеваемости и смертности. Сегодня современные технологии, в том числе -omics платформы, а также сложные датчики и методы моделирования, облегчают создание и сбор новых типов данных. Данные могут использоваться для генерации гипотез, но они также могут использоваться для дополнения баз данных с целью увеличения достоверности результатов новых исследований (см. Вставку 4-1). Новые возможности для получения данных появились из-за изменений в медицинской практике и системах хранения и контроля данных, относящихся к вопросам здравоохранения (AACR 2015). Биобанки, созданные клиническими учреждениями¹, включают в себя хранилища биологических образцов (кровь, моча и образцы, взятые во время хирургических операций и биопсии), клиническую информацию о пациентах, которая предоставляет демографическую информацию и информацию об образе жизни, возможно, вопросник об образе жизни, экологической экспозиции и экспозиции, связанной с профессиональной деятельностью, а также о результатах оценки состояния здоровья по данным клинической практики. Таким образом, популяционные данные и биологические образцы, потенциально доступные для тестирования различными методами, включая -omics технологии, могут находиться в банках данных изначально собранных не для исследовательских целей. Однако данные, полученные при изучении тканей человека и медицинские данные, полученные из особых популяционных групп (собранные

¹ Комитет отмечает, что биобанки не являются новым изобретением. Например, в Национальном обследовании состояния здоровья и питания, которое проводилось для целей эпидемиологического надзора, были собраны и проанализированы биологические образцы; полученные данные оказались бесценными для оценки экспозиции. Было создано множество других биобанков, как правило, путем регистрации показателей, относящихся к здоровью; самые крупные из них включают Европейское перспективное исследование рака и питания (European Prospective Investigation into Cancer and Nutrition, IARC 2016) и Британский биобанк (UK Biobank 2016).

изначально не для исследовательских целей) , могут быть сложными для обобщения. Более того, такие исследования могут быть предвзятыми (также, впрочем, как и любые другие исследования), так как в них изначально не была заложена возможность избежать проблемы предвзятости с помощью хорошо продуманного дизайна эксперимента, сбора данных и протоколов для получения биологических образцов. Таким образом, новые потоки данных и технологии, хотя и многообещающие, имеют серьезные методологические проблемы и требуют разработки новых схем экспериментов и аналитических методов с учетом специфических технических требований (Khoury et al., 2013). Следует отметить, что исследователи предупреждали об увеличении вероятности появления ошибок и тупиковых моментов в процессе использования новых подходов и призывали к тщательной оценке анализируемых показателей, воспроизводимости экспериментов, безопасности, этических и юридических последствиях (Alsheikh-Ali et al., 2011; Khoury и др., 2013).

ВСТАВКА 4-1 Использование исследования «Наследие» (Legacy Studies)

В исследовании «Наследие» накоплена существенная информация о различных экспозициях, таких как употребление табака, экспозиций, связанных с профессиональной деятельностью, загрязнением воздуха, а также информация об индивидуальных факторах, включая генетические данные, события, заболевания, произошедшие за десятилетия до последующего наблюдения. Некоторые из них включают банки биологических образцов, показатели фенотипа болезни и промежуточные результаты, которые были получены при визуализации, физиологическом тестировании и других методах оценки. Некоторые образцы уже использовались для применения -омикс технологий (EXPoSOMICS 2016). Различные когорты использовались для решения проблемы ассоциации загрязнения атмосферного воздуха с заболеваемостью и смертностью путем добавления оценок загрязнения воздуха в местах проживания, полученных с помощью новых моделей экспозиции, имеющих достаточное пространственное разрешение. Объединение данных из нескольких исследований дает возможность получить статистическую мощьность и сделать результаты более точными, увеличивая разнообразие экспозиций и неоднородность участников исследования.

Цунамический масштаб данных, охватывающих все спектры геномной, молекулярной, клинической, эпидемиологической, экологической и цифровой информации, уже является реальностью эпидемиологии 21 века (Khoury et al., 2013), породившей проблемы в

использовании новых методов для систематического и эффективного анализа, обработки, интерпретации данных и поиска соответствующих сигналов в “океаническом шуме”. Для решения этих проблем правительство США в 2012 году объявило о создании проекта «Большие данные» и выделило средства для поддержки исследований в области баз данных в нескольких агентствах (Mervis 2012). Эпидемиологи готовы играть центральную роль в формировании направлений исследований и распределении инвестиций для создания надежных инфраструктур для хранения и анализа массивных и сложных баз данных. Имея опыт работы в междисциплинарных группах, эпидемиологи также хорошо экипированы для сотрудничества с экспертами в области клинических и базовых медицинских наук, биомедицинской информатики, вычислительной биологии, математики и биостатистики. Адаптация технологических достижений, таких как облачные вычисления (cloud computing), стратегическое формирование новых партнерских отношений между академическими и промышленными кругами для содействия интеграции современных компьютерных технологий в биомедицинские исследования и здравоохранение (Pechette 2012) – вот лишь некоторые из первоначальных проблем, которые должны быть решены перед тем как новые базы данных будут интегрированы в будущие эпидемиологические исследования.

Типы ошибок и проблемы, связанные с внешней валидацией

Как уже отмечалось, современная эпидемиология сталкивается с беспрецедентным распространением клинических данных и данных из области здравоохранения, -omics данных, социальных и экологических данных. Ошибки, которые в целом влияют на эпидемиологические данные, могут быть сгруппированы в три обширные категории: 1) информационная необъективность, возникающая из-за ошибки в измерениях экспозиции, вариативности результатов и ко-вариаций, 2) смещение выборки, которое возникает из-за способов отбора участников для участия в эпидемиологических исследованиях и 3) ошибок, возникающих из-за смешанных эффектов экспозиции несколькими факторами. Под внешней валидацией подразумевается обобщаемость результатов, что является ключевым фактором при оценке риска. Понимание процессов отбора, точности измерений и интерпретации анализов имеют решающее значение для использования эпидемиологических данных при оценке риска.

Множественность, разнообразие и размер источников данных вызвали широкий энтузиазм исследователей в отношении новых возможностей для анализа (Roger et al., 2015a, b).

Однако необходимо помнить, что при использовании данных могут возникнуть определенные сложности. Например, использование электронных медицинских записей в качестве единственной основы для составления когорт может усилить ошибки выборки за счет включения в статистику лишь той части популяции, которая обращается к врачу за медицинской помощью, а также способствовать неправильной классификации или неполной документации фенотипов, клинических диагнозов и процедур из-за ошибок в клинических диагнозах и методах кодирования, а также приводить к путанице из-за того, что факторы, которые необходимо принимать во внимание в качестве ко-факторов, обычно не регистрируются в медицинских отчетах. Несмотря на то, что системы электронных записей могут способствовать созданию крупных когорт для проведения исследований, наличие большого объема выборки не уменьшает вероятность ошибок и повышает вероятность статистически значимых ложноположительных результатов. Кроме того, электронные медицинские записи обычно содержат мало информации об экологических экспозициях и экспозициях, связанных с родом деятельности, связь с базами данных об экспозиции может быть проблематичной, а информация о важных потенциальных факторах, таких как употребление табака, может быть скудной и не собираться стандартизированным образом, необходимым для исследований.

При оценке рисков, создаваемых факторами окружающей среды, эпидемиологи обычно работают вместе со специалистами по оценке экспозиции, для уточнения оценок экспозиции, полученных из эпидемиологических исследований, путем увеличения разнообразия рассматриваемых экспозиций, повышения точности измерения экспозиции и понимания ошибок, которые неизбежно влияют на конечную оценку. Полный спектр достижений науки об экспозиции, описанный в отчете ES21 (NRC 2012) и второй главе настоящего доклада, применяется и в эпидемиологических исследованиях. Когда методы оценки экспозиции надлежащим образом включены в проект исследования, они облегчают исследование погрешности измерения в переменных экспозициях и ковариатах. Такие ошибки долгое время считались серьезным ограничением эпидемиологических данных в контекстах оценки риска; неслучайные ошибки могут смещать кажущиеся эффекты вверх или вниз, а случайные ошибки обычно затушевывают ассоциации и зависимости доза-реакция. Коррекция ошибок измерения может быть сделана с использованием данных валидированных исследований и статистических моделей, разработанных за последние два десятилетия и применяемых, например, в исследованиях по риску диет-ассоциированных заболеваний, зависимости развития рака от уровня радиации, а также

влияния загрязненного воздуха на здоровье (Li Et al., 2006; Freedman et al., 2015; Hart et al., 2015).

ЭПИДЕМИОЛОГИЯ И -OMICS ДАННЫЕ

На протяжении всей истории своего существования эпидемиологические исследования всегда включали появляющиеся технологии в свои исследования. Однако, несколько десятилетий назад необходимость включения новых методов ускорилась в связи с введением новой парадигмы молекулярной эпидемиологии. Новая парадигма пришла на смену эпидемиологии «черного ящика», подходу, который рассматривал ассоциации факторов риска с заболеванием, не обращая внимания на промежуточные механизмы. Молекулярно-эпидемиологическая парадигма открывает черные ящики с помощью определения и введения процедуры тестирования биомаркеров экспозиции, факторов индивидуальной восприимчивости и учета предыдущих заболеваний. В ней подчеркивается необходимость изучения путей токсичности и их пертурбаций, что стало возможным благодаря научным технологиям XXI века и конкретно – omics технологиями. Она также усиливает доказательную базу для одного из руководящих принципов Брэдфорда Хилла о причинности: понимание биологической правдоподобности. Например, считается, что канцерогенез является многофакторным процессом, в котором мутации и селективная микросреда играют критически важную роль, и ключевые этапы этого процесса могут быть изучены с помощью биомаркеров. Молекулярно-эпидемиологическая парадигма концептуально включает новые методы для анализа биомаркеров.

Как было указано, молекулярно-эпидемиологические исследования ориентированы на биологические параметры (экспозиция и патогенез заболевания), а не на эмпирические наблюдения. Таким образом, когда появились omics технологии, они были интегрированы в текущие исследования и повлияли на дизайн исследований, особенно на сбор образцов и их использование. Включение omics подходов берет начало от геномной революции и насчитывает уже около двух десятилетий. В некоторых настоящих когортных исследованиях образцы крови, которые были соответствующим образом сохранены, анализировались на однонуклеотидные полиморфизмы (SNPs) и другие маркеры для поиска генов, связанных с риском заболевания, включая те, которые изменяют риск, связанный с агентами окружающей среды.

Целесообразность привлечения в эпидемиологические исследования -omics технологий уже очевидна, о чем свидетельствуют многочисленные исследования. Одной из известных отправных точек для изучения генетической основы заболеваний стали исследования GWAS (genome-wide association studies), включающие сравнение геномных маркеров у людей, имеющих и не имеющих заболевания или состояния, представляющие интерес. Список -omics подходов, применяемых в эпидемиологических исследованиях, теперь расширился далеко за пределы геномики, включив эпигеномику, протеомику, транскриптомику и метаболомику (см. Вставку 1-1). В таблице 4-1 перечислены преимущества и недостатки их использования. Примеры их использования в конкретном контексте приводятся в приложении В, в котором излагаются смысл и ограничения подходов в рамках эпидемиологических исследований по загрязнению воздуха. Несмотря на то, что, новые методы имеют большой потенциал для получения новых данных из эпидемиологических исследований, существует и ряд проблем, касающихся их применения. В настоящий момент планируются новые исследования с целью сохранения образцов для будущего использования, как, например, в случае проектов Helix и EXPOsOMICS, финансируемых ЕС. Полученные данных популяционных исследований будут параллельны данным, которые могут быть получены из *in vitro* и *in vivo* тестов и позволят согласовать параметры экспозиции и дозы.

Иными словами, -omics подходы теперь поддерживают нецелевые исследования генов с помощью методов геномики, мРНК с помощью методов транскриптомики, протеинов с помощью методов протеомики и метаболитов с помощью методов метаболомики. За исключением геномики, измеряемые показатели обычно отражают изменения в клетках только в один или несколько моментов времени, а в качестве тканей для исследований на людях, в основном используются суррогаты, такие как кровь, моча или слюна. Однако объединение различных -omics подходов значительно расширяет возможности для изучения и понимания того, как различные химические вещества из окружающей среды взаимодействуют с внутренними молекулами - например, путем индукции мутаций - (геномика), вызывая эпигенетические изменения (эпигеномика) или изменяя внутреннюю клеточную среду более сложными способами; последние могут быть исследованы с помощью протеомики, транскриптомики или метаболомики.

Подход «промежуточного звена»

Одной из стратегий интеграции -omics технологий в эпидемиологические исследования является подход «промежуточного звена» (Vineis et al., 2013). Этот подход выявляет биологические механизмы, являющиеся отражением причинно-следственной связи. В контексте популяционного исследования этот подход обычно включает

ТАБЛИЦА 4-1 Преимущества и недостатки -омикс технологий

Преимущества	<p>Использование всего набора биологических молекул для анализа, подход, не требующий и независимый от предварительных гипотез.</p> <p>Обеспечивают лучшее понимание отношений генотип-фенотип.</p> <p>Могут помочь в понимании взаимодействий между средовыми факторами и генотипами и в понимании механизмов вовлеченных в развитие болезни.</p>
Недостатки	<p>Высокая стоимость, высокие требования к качеству биологического материала (например целостность молекул РНК), трудовые затраты.</p> <p>Технологии, находящиеся на ранних этапах разработки должны тщательно проверяться путем сопоставления с существующей биологической информацией из <i>in vitro</i> и <i>in vivo</i> тестов.</p> <p>Новые данные о промежуточных маркерах должны проходить проверку в рамках независимых исследований, использующих другие методологические подходы.</p> <p>Переход от многообещающих технологий к успешному применению биомаркеров в профессиональной медицине и в лечении болезней, вызванных воздействием загрязнителей окружающей среды, потребует стандардизации и валидации методов, а также продуманного дизайна исследований и сложной статистической интерпретации данных, особенно в случае нецелевых исследований (проблема множественных сравнений, и ложно-положительных данных).</p>

Источник: Vineis et al. 2009.

перспективный поиск промежуточных биомаркеров, связанных с основным заболеванием и присутствующих в повышенном количестве у тех, кто в конечном итоге развивает болезнь, и ретроспективный поиск, который связывает промежуточные биомаркеры с

прошлыми экспозициями химическим веществом, вызывающим озабоченность. Как показано на рисунке 4-1, этот подход можно рассматривать как три последовательно осуществляемых этапа: 1) исследование связи между экспозицией и заболеванием; 2) оценка взаимосвязи между экспозицией и биомаркерами экспозиции и ранними эффектами, 3) оценка взаимосвязи между проявлениями заболевания и промежуточными биомаркерами. Выявление причинной связи между экспозицией и болезнью усиливается, если ассоциации задокументированы для каждой из трех ключевых взаимосвязей на рисунке 4-1, соответствующих А, В и С.

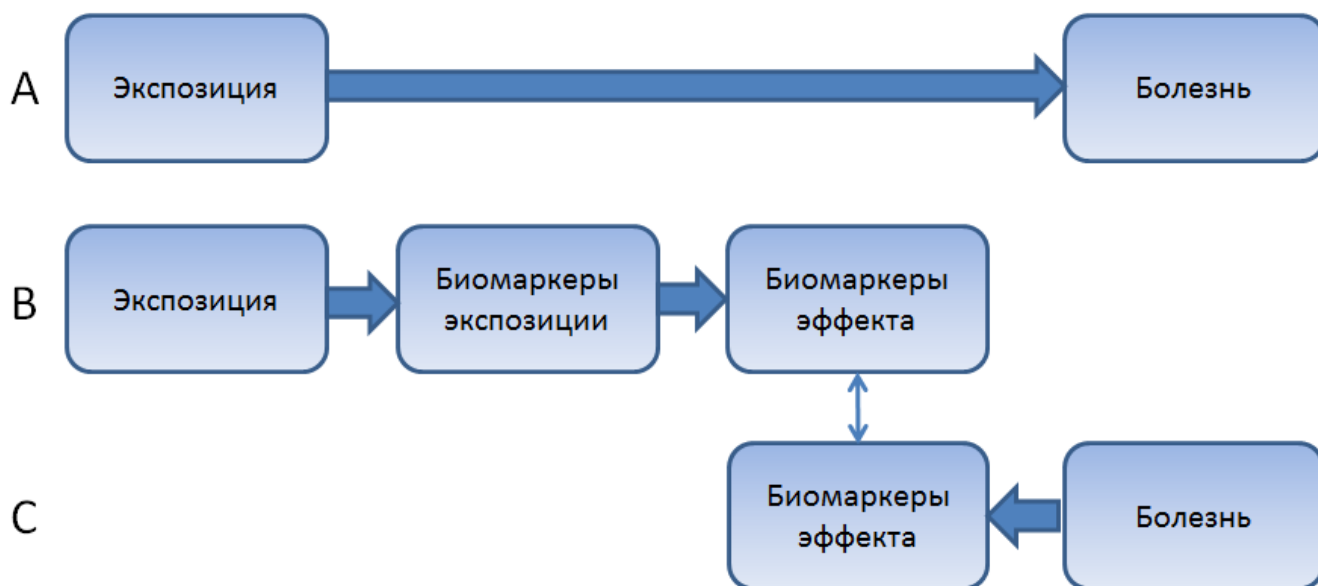


РИСУНОК 4-1 Подход, основанный на «промежуточном звене», основан на исследовании (А) связи между экспозицией и заболеванием, (В) взаимосвязи между экспозицией и биомаркерами экспозиции или эффекта и (С) взаимосвязи между заболеванием и биомаркерами экспозиции или эффекта.

Недавнее эпигенетическое исследование рака легких (Fasanelli et al., 2015) является прекрасной иллюстрацией вышеописанного подхода. В данном случае, биомаркерами являются статусы метилирования гена AHRH и гена F2RL, которые гипометилированы у курильщиков (экспозиция на Рис. 4-1В) (Vineis et al., 2013; Guida и др., 2015). Гипометилирование этих генов также связано с раком легких (болезнь на Рис. 4-1С). Вопрос в том, являются ли эти биомаркеры причинно-следственной цепью для рака легких, вызванного курением? Используя статистическую методику медиационного

анализа было показано (Fasanelli et al. 2015), что 37% легочных опухолей можно объяснить статусом метилирования этих двух генов. Таким образом, два гена представляют собой биомаркеры, которые, вероятно, находятся в причинно-следственной связи и иллюстрируют «промежуточное звено» между экспозицией и болезнью. Комитет отмечает, однако, для полной оценки причинно-следственных связей, помимо статистического анализа, потребуются другие дополнительные шаги.

Исследования совокупности экспозиций

Как было определено в главе 1, совокупность всех экспозиций на протяжении всего отрезка жизни человека принято называть экспозомом. Некоторые сомневаются в том, что экспозом возможно измерить, а следовательно, делают вывод о том, что это понятие не соответствует научным методам (Miller and Jones, 2014). В попытке определить экспозом как измеряемую величину Рапппорт и Смит (Rappaport and Smith, 2010) предложили определять начальную внутреннюю химическую среду тела и то, как организм реагирует на химические воздействия.² Они ссылались на экспозиции как на внутренний экспозом и отличали его от внешнего экспозома – внешних по отношению к телу экспозиций - и предположили, что внутренний и внешний экспозомы являются взаимодополняющими. Например, изучение внутреннего экспозома может выявлять ассоциации болезней и состояния окружающей среды (то есть порождать новые гипотезы по этиологии заболеваний), а изучение внешнего экспозома необходимо для определения источников, рассмотрения путей экспозиции и определения их пространственной и временной изменчивости (Turner et al. in press). Следовательно, оценка внешней экспозиции может произойти после того, как были сделаны гипотезы, и наступило время поиска источников внутренних изменений. Обе схемы исследования, - та, в которой исследуются внутренние изменения, начиная с измерений внешней экспозиции (оценка внешнего экспозома) и та, в которой ведется поиск источников внешней экспозиции на основе внутренних изменений (оценка внутреннего экспозома), являются взаимодополняющими и были определены как «восходящий» и «нисходящий» подходы, соответственно.

² Включение биологического ответа в концепцию помогает выйти за пределы внешней экспозиции химическими веществами и включить многие типы экспозиций (включая психологический или физический стресс, инфекции и флору кишечника), которые производят эндогенные химические вещества, такие как окислительные молекулы, и реакции, вызывающие болезни, такие как воспалительная реакция, окислительный стресс и перекисное окисление липидов..

-Omics технологии, используемые для анализа внешнего экспозома, делают возможным нецелевое тестирование, аналогичное в концепции и подходе анализам GWAS. Исследования в рамках этого проекта были названы широко-экспозомными ассоциативными исследованиями (exposome-wide association studies или EWAS).³ В частности, подход EWAS включает исследование ассоциаций большого числа малых молекул, белков или липидов с болезнью или промежуточными фенотипами для идентификации биомаркеров экспозиции или болезни. Один из общих подходов EWAS для создания новых гипотез о причинной обусловленности заболевания описан Раппортом и Смитом (Rappaport and Smith, 2010). На рис. 4-2 показан дизайн исследования, который может привести к появлению новых гипотез о химических опасностях в контексте исследования «случай-контроль». Для сравнения экспозиции в случаях с зафиксированным конкретным заболеванием с экспозицией тех индивидуумов, у которых заболевание отсутствует (группа контроля) используются целевые и нецелевые методы метаболомики. После фазы первичных находок экспериментальная схема может быть улучшена на этапе тестирования (репликации) в соответствии с проспективным контекстом (тематическое исследование в проспективной когорте). Этот подход учитывает временные факторы (для исключения или уменьшения вероятности обратной обусловленности), используя биологические образцы, собранные до проявления болезни. Затем неопределенные показатели, которые определенно связаны с представляющими интерес эффектами, химически идентифицируются с использованием методов, описанных в главе 2, например, с использованием NMR, IMS-MS/MS или хемоинформатики или путем синтеза и оценки химических стандартов для предполагаемых токсических химических веществ. На следующем этапе проверка достоверности ассоциации и установления окончательной причины нарушений в организме предпринимается путем репликации более, чем в одной когорте.

Биологическая достоверность может быть оценена с помощью целевого анализа доступных тканей человека с использованием методов протеомики, метаболомики или других методов для поиска биологических ответов, связанных с заболеванием. В качестве альтернативы для тестирования химических веществ-кандидатов и получения данных о биологических реакциях могут быть использованы новые животные модели или тесты с высокой пропускной способностью *in vitro*, описанные в главе 3. В идеале оценка

³ Комитет отмечает, что акроним EWAS был первоначально предложен Patel et al. (2010), чтобы обратиться к исследованиям ассоциаций со всеми внешними факторами, но другие, такие как Rappaport (2012), использовали EWAS для в более узком смысле, для исследования ассоциаций со всеми компонентами экспозома.



РИСУНОК 4-2 Дизайн исследования для разработки новых гипотез о причинной связи заболевания с экспозицией. Комитет отмечает, что подход и методы, используемые для исследования экспозиции и биологических путей для определения причины заболевания и профилактики необязательно отличаются от тех, которые используются для изучения биологических путей, имеющих значение для разработки лекарств. Источник: Rappaport 2012.

биологической правдоподобности также должна включать в себя уточнение характера и дозы экспозиции, а также систематическое сравнение реальной экспозиции человека с экспозицией в тестовых системах, используемых для получения подтверждающих данных о биологических реакциях. При использовании аналогичных данных и моделей, ответные реакции на экспозицию в когортах могут быть непосредственно сопоставлены с ответными реакциями в тестовых системах; в таком случае сравнительный анализ предоставит дополнительные данные биологической достоверности. Пример описанного подхода был использован для исследования рака толстой кишки. Исследование началось с

трех перекрестных тематических исследований и обнаружило связь между неизвестным аналитическим пиком (метаболитом) и раком толстой кишки (Ritchie et al., 2013). Позднее ассоциация была подтверждена на когорте Европейских перспективных исследований рака и питания (European Prospective Investigation into Cancer and Nutrition cohort), а метаболит был идентифицирован как принадлежащий к группе жирных кислот с длинной цепью (Perttula et al., 2016).

EWAS подход является довольно многообещающим при условии осторожного и грамотного применения. В данном случае, проблемы в применении методов, производящих огромные объемы данных схожи с теми, которые имеются во всех исследованиях с множественной экспозицией. Дизайн и анализ исследования должны быть тщательно продуманы и оценены с точки зрения исключения классических ошибок с целью установления причинно-следственной связи, то есть с использованием принципов, применимых к целевым проектам, фокусирующимся на единичной экспозиции и результате. Еще одна проблема состоит в определении той экспозиции, которая среди многих прочих имеет каузальную роль и отделении тех, что отражают биологические пертурбации, вызванные другими химическими веществами. EWAS подход может позволить идентифицировать полезные биомаркеры, что указывает на то, что рассматриваемые химические вещества не являются слишком короткоживущими и экспозиции не являются слишком спорадическими, чтобы быть захваченными только одним или несколькими образцами тканей, полученными в перекрестном или когортном исследовании.

Комитет отмечает, что использование ретроспективного целевого исследования для EWAS не позволяет быть уверенным, отражают ли наблюдаемые ассоциации причинно-следственную связь между экспозицией и исследованным эффектом или же ассоциация является следствием заболевания или его лечения. Как резюмирует Томас и др. (2012), метод менделианской рандомизации (Davey Smith et al., 2004) является одним из способов решения обратной обусловленности и неконтролируемых кофакторов. Ген используется в качестве инструментальной переменной (Гренландия, 2000) для оценки причинного воздействия биомаркера на риск заболевания. В подходе, который соответствует подходу «промежуточного звена», было предложено новое двухступенчатое развитие этой идеи для исследований метилирования, в котором в качестве инструментальных переменных используются два гена: один оценивает ассоциацию «экспозиция-метилирование», а другой – ассоциацию «метилирование – болезнь» (Cortessis et al., 2012; Relton and Davey

Smith, 2012). Этот подход базируется на допущении, что инструментальная переменная действительно является подходящим инструментом для оценки экспозиции.

Аналитические трудности при работе с большими данными

Процесс интеграции технологий и данных в области медицины в эпидемиологические исследования сталкивается с серьезными трудностями. Интеграция и анализ объемных данных потребует надежных аналитических инструментов. Например, статистические анализы, которые рассматривают множество одновременных экспозиции различной природы без серьезных предшествующих данных по отдельным видам экспозиции, например, в EWAS, значительно повышают риск наблюдения случайных ассоциаций (ложных срабатываний) из-за множественного тестирования. Статистические методы анализа множественных экспозиций были существенно улучшены на основании уроков, извлеченных из подхода GWAS (Shi and Weinberg 2011, Thomas и др. 2012); некоторые из них описаны ниже. Следует также отметить, что статистические методы для объемных данных должны быть адаптированы в контексте применения для долгосрочных исследований и учитывать такие вопросы как, разнообразие экспозиции и отсроченные эффекты (Buck Louis and Sundaram 2012).

Для оценки рисков для здоровья, связанных с различными типами или комбинациями экспозиций были использованы многоступенчатые аналитические подходы. Например, оценки, полученные с помощью EWAS (в случае отсутствия первичной информации) могут быть количественно дополнены при помощи классических моделей регрессии контролируя при этом процент ложных ответов, как это было сделано в GWAS (Patel et al., 2010, 2013; Vrijheid и др., 2014). Кроме того, для моделирования и описания возможных пороговых значений или отношений «экспозиция-реакция» могут использоваться гибкие методы и модели сглаживания (Slama и Werwatz, 2005).

Аналитические подходы к изучению путей токсичности все шире используются для интеграции и интерпретации данных большого объема, генерируемых с помощью -omics технологий; эти подходы позволили провести анализ взаимосвязей между множественной экспозицией и множественными последствиями для здоровья. Следует отметить, что для идентификации молекулярных реакций, вызванных экспозицией загрязнителями окружающей среды, были использованы аналитические подходы для исследования метаболитов, белков, транскриптов и продуктов метилирования ДНК в биологических

образцах (Jennen et al., 2011, Vrijheid et al., 2014). Как резюмируется Vrijheid et al. (2014), после идентификации биомаркеров, базы данных биологических путей, таких как Gene Ontology (Ashburner et al., 2000), Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto 2000), Reactome (Fabregat et al., 2016) и Comparative Toxicogenomics Database (Davis et al., 2015) могут быть использованы по отдельности или совместно для выявления релевантных биологических путей, нарушенных экспозицией. Кроме того, биологические пути могут быть сгруппированы и описаны с использованием имеющегося программного обеспечения, такого как Ingenuity Pathway Analysis (Krämer et al., 2014), Cytoscape (Saito et al., 2012) и Impala (Kamburov et al., 2011). Например, эти аналитические подходы были применены к нескольким типам -omics данных для изучения ответа на экспозицию 2,3,7,8-тетрахлордибензо-пи-диоксином и более широким набором экологических и фармакологических веществ (Jennen et al., 2011, Kamburov et al., 2011).

Другие методы также доступны для решения новых аналитических задач. Во-первых, для интеграции индивидуальных экспозиций (получаемых, например, от персональных мобильных устройств) и экспозиций на открытом воздухе (полученных, например, при мониторинге окружающей среды) использовался анализ ковариационных методов, при котором перед разбиванием экспозиций на группы или кластеры изучаются различные компоненты основных экспозиций. Во-вторых, факторный анализ и скрытый анализ классов оказались полезными для создания сокращенных списков индексов экспозиции на основе часто встречающихся, позволяя группировать людей, имеющих сходную экспозицию. В-третьих, для решения объемного характера эпигенетических данных, методы кластерного анализа, разработанные Siegmund et al. (2006) могут быть применены к EWAS исследованиям; эти методы рассматривают кластер, а не отдельные эпигенетические метки как скрытый фактор риска для заболевания (Cortessis et al., 2012). В-четвертых, подходы моделирования структурных уравнений могут использоваться для определения комбинированных переменных экспозиции на основе знаний, суммированных направленными ациклическими графами (Budtz-Jørgensen et al., 2010).

Регрессионные модели байесовского профиля могут быть использованы для идентификации групп людей, у которых есть аналогичный экспозом, но имеются заметные различия в представляющей интерес переменной эффекта на здоровье (Molitor et al., 2010; Parathomas et al., 2011; Vrijheid et al., 2014). Кластеризация на основе моделей, примененная к данным об экспозиции, позволит эффектам на здоровье влиять на членство в кластере. Байесовский метод кластеризации на основе моделей был использован, например, для

выявления кластера в группе высокого риска рака легких. Субъекты в этой группе имели такие общие характеристики как: выполнение ручного труда и проживание вблизи главной дороги с высокой экспозицией ТЧ10 (твердые частицы с аэродинамическим диаметром ≤ 10 мкм) и диоксидом азота (Parathomas et al., 2011; Vrijheid et al., 2014).

Общая потребность проявлять осторожность относительно потенциально ложноположительных ассоциаций, возникающих в результате анализа больших наборов данных, обычно признается теми, кто обрабатывает такие данные. В дополнение к аналитическим подходам, таким как коррекция значений p для множественных сравнений и использование таких параметров, как темп ложных обнаружений (FDR), комитет отмечает, что эпидемиологические данные интерпретируются целостно в контексте других соответствующих доказательств. В контексте оценки риска определение опасности редко, если вообще когда-либо, основывается на ассоциации, обнаруженной в одном эпидемиологическом исследовании при отсутствии дополнительных доказательств.

ПРОБЛЕМЫ И РЕКОМЕНДАЦИИ ПО РАЗВИТИЮ ЭПИДЕМИОЛОГИИ

С появлением подходов Tox21 и ES21 ожидается обнаружение новых связей между биомаркерами и эффектами на здоровье человека. Эпидемиологические исследования играют существенную роль в обеспечении одной из составляющих процесса оценки риска - популяционную составляющую, которая необходима для интерпретации биомаркеров, измеренных в лабораторных исследованиях. С этой целью эпидемиологи должны генерировать данные по человеческим популяциям 1) для согласования доз, используемых в тестовых системах *in vitro*, с теми, которые связаны с экспозицией, имеющей место в популяционной среде; 2) для исследования релевантности метаболических путей, идентифицированных в тестовых системах к человеческим реакциям на те же агенты и подтверждения прогностической ценности метаболических путей, обнаруженных в *in vitro* тестах для заболеваний человека; 3) для разработки и проверки моделей индивидуальной восприимчивости человека и 4) сопоставления и подтверждения зависимостей экспозиция-реакция полученных в тестах *in vitro* и в популяциях человека.

Общая цель получения новых данных путем объединения *-omics* данных, полученных в лаборатории с данными, собранными в популяции, не будет достигнута без учета необходимой исследовательской инфраструктуры и преодоления логистических барьеров

для объединения наборов данных из разрозненных источников. В заключение комитет указывает на некоторые трудности, с которыми сталкиваются эпидемиологические исследования и предлагает рекомендации по их устранению. Комитет отмечает, что ряд рекомендаций, приведенных ниже, требуют разработки или расширения баз данных. Во всех случаях при разработке и обслуживании базы данных должна проводиться рутинная оценка качества данных.

Развитие инфраструктуры и методов, необходимых для развития науки

Проблема: При использовании в эпидемиологических исследованиях (особенно с крупными когортами на основе биобанка, которые могут достигать миллиона или более участников), -омикс данные могут генерировать большие информационные базы, которые должны управляться таким образом, чтобы облегчать доступ и анализ. Дополнительные сложности возникают при анализе очень больших массивов данных с использованием подходов, основанных на гипотезах или анализе, производимом с исключительно исследовательскими целями .

Рекомендации: Следует направить ресурсы на ускоренное развитие систем управления базами данных, которые будут способны размещать чрезвычайно большие объемы данных, поддерживать инструменты анализа для разных целей, способствовать обмену данными и разработке мощных и надежных статистических методов для анализа ассоциативных связей эффектов на здоровье с данными -omics исследований и изучать такие сложные проблемы, как взаимодействие генов и окружающей среды. Такие усилия уже предпринимаются в ряде областей, например, в клинических исследованиях и медико-санитарной помощи, однако эти усилия следует распространить и на эпидемиологические исследования.

Проблема: Для описания данных, полученных и распределенных между различными дисциплинами, необходимы стандартные методы. Эта проблема была признана в геномике и привела к разработке онтологий аннотированных генов. Аналогичные подходы могут быть распространены и на другие типы -omics данных.

Рекомендации: Онтологии должны быть разработаны и расширены, чтобы данные могли быть согласованы между исследовательскими группами, как на международном уровне, так и между -omics платформами. Такие онтологии обычно не включают данные,

собранные эпидемиологами. Такие инструменты, как STROBE, следует расширить и адаптировать к новому поколению эпидемиологических исследований; ранее STROBE уже был расширен, чтобы охватить молекулярную эпидемиологию (Gallo et al., 2011). Инициативная группа Европейской семерки по координации стандартов в метаболомике (Coordination of Standards in Metabolomics COSMOS) разрабатывает «надежную инфраструктуру данных и стандарты обмена для метаболомических данных и других метаданных» (Salek et al., 2015); этот тип подхода следует распространить на другие базы – omics данных.

Обмен данными

Проблема: Совместное использование баз данных сопряжено со многими сложностями, особенно когда данные получены в результате исследований на людях. Однако, совместное использование данных могло бы быть полезным, если бы к ним можно было получить доступ, обеспечивающий единообразный анализ и интеграцию посредством иерархического анализа или мета-анализа. Обмен данными может также привести к более взвешенным оценкам опасности и отношений «экспозиция-ответ». Одним из таких примеров является объединение данных исследований по шахтерам, подвергшимся экспозиции радоном (Lubin et al., 1995).

Те же проблемы, связанные с совместным использованием данных, возникают в других областях, в которых появляются методы исследований, генерирующие большие объемы информации. Таким образом, потребуется установление общей культуры обмена данными. Что касается геномики, то публикация данных секвенирования стала нормой, при условии сохранения анонимности. Подобное совместное использование в идеале будет распространяться на другие -omics данные и приведет к развитию культуры обмена данными, разумным решениям этических проблем и стандартизации онтологий и баз данных. Комитет отмечает, что дискуссия вокруг обмена данными успешно развивается в отношении клинических испытаний; однако в отношении других эпидемиологических наблюдений все еще прилагается недостаточно усилий (Mascalzoni 2015).

Рекомендации: Необходимо принять меры для обеспечения обмена данными, относящимися к оценке риска. Как отмечено выше, для достижения этой цели необходимо разработать стандартные онтологии и использовать их для определения и кодирования ключевых переменных. Также необходимо систематически исследовать возможные

материально-технические и этические барьеры для совместного использования потенциально массивных наборов данных, полученных в популяционных исследованиях.

Сотрудничество и подготовка нового поколения ученых

Задача: Новые исследовательские модели, основанные на биобанках и крупных когортах, полученных из клинических наблюдений, станут ценным ресурсом для применения -omics подходов и поиска биомаркеров, однако использование этих методов имеет внутренние ограничения, связанные с неравномерностью данных, представленных в электронных базах данных. Другие проблемы связаны с конфиденциальностью медицинской документации и другой частной документации и обменом такого рода информацией.

Рекомендация: Разработчики биобанков и когорт должны активно взаимодействовать с эпидемиологами и учеными-исследователями для обеспечения сбора наилучших и наиболее полных данных об экспозиции и относящейся к этой теме информации. Выбор способов сбора информации об экспозиции будет особенно сложным и, вероятно, потребует сбора вспомогательных данных.

Задача: В настоящий момент собирается и отправляется на хранение широчайший спектр самых разнообразных биологических образцов для неопределенных целей, исходя из предположения, что они будут полезны в будущем, в том числе для анализа, который сегодня не востребован. Согласование процедур сбора и методов хранения является важнейшим залогом успешного использования этих данных в будущем.

Рекомендация: Эпидемиологи должны предвидеть будущее использование биологических образцов, которые собираются в ходе эпидемиологических исследований или в ходе других исследований, и обеспечить, чтобы собранные данные, их обработка и хранение были возможны для использования в будущем. Такие базы данных, ориентированные на использование в будущем, должны быть тщательно продуманы с точки зрения дизайна, с учетом рекомендаций со стороны разработчиков новых методов.

Задача: Необходимо новое поколение исследователей, которое будет обучено проводить крупномасштабные эпидемиологические исследования и интегрировать -omics подходы и другие новые технологии в популяционные исследования. Следующее поколение ученых будет также нуждаться в достаточно многодисциплинарном образовании, чтобы иметь

возможность взаимодействовать с учеными-исследователями и исследователями баз данных.

Рекомендация: Подготовка эпидемиологов должна быть дополнена более глубоким пониманием биологических механизмов, лежащих в основе заболеваний человека, и поиска биомаркеров, используемых для их исследования.

Задача: Ландшафт эпидемиологических исследований быстро меняется с переходом от когорт прошлого, таких как «Исследование здоровья медсестер» (Nurses' Health Study), к прагматично разработанным когортам, основанным на новых и практичных способах идентификации и последующего наблюдения. Также, вероятно, будут создаваться большие национальные когорты, такие как когорта, которая уже разрабатывается для Precision Medicine Initiative. Эти когорты предназначены в качестве платформ для широкого круга исследовательских вопросов; они создаются как крупные банки биологических образцов, но будут иметь определенные ограничения в отношении имеющейся информации об экспозиции.

Рекомендация: Эпидемиологам, ученым, занимающимся изучением экспозиции и научным-лабораториям следует тесно сотрудничать для обеспечения того, чтобы весь научно-технический потенциал XXI века применялся в эпидемиологических исследованиях. Необходимо стремиться к многодисциплинарности. Поскольку новые когорты находятся в стадии разработки, необходимо использовать этот момент для того, чтобы включить в их дизайн учет факторов окружающей среды.

REFERENCES

AARC (American Association for Cancer Research). 2015. AACR Cancer Progress Report 2015 [online]. Available: http://cancerprogressreport.org/2015/Documents/AACR_CPR2015.pdf [accessed July 21, 2016].

Alsheikh-Ali, A.A., W. Qureshi, M.H. Al-Mallah, and J.P. Ioannidis. 2011. Public availability of published research data in high-impact journals. PLoS One 6(9): e24357.

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock. 2000. Gene ontology: Tool for the

unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25(1): 25-29.

Buck Louis, G.M., and R. Sundaram. 2012. Exposome: Time for transformative research. *Stat. Med.* 31(22): 2569-2575.

Budtz-Jørgensen, E., F. Debes, P. Weihe, and P. Grandjean. 2010. Structural equation models for meta-analysis in environmental risk assessment. *Environmetrics* 21(5): 510-527. Cortessis, V.K., D.C. Thomas, A.J. Levine, C.V. Breton,

T.M. Mack, K.D. Siegmund, R.W. Haile, and P.W. Laird. 2012. Environmental epigenetics: Prospects for studying epigenetic mediation of exposure–response relationships. *Hum. Genet.* 131(10): 1565-1589.

Davey Smith, G., R. Harbord, and S. Ebrahim. 2004. Fibrinogen, C-reactive protein and coronary heart disease: Does Mendelian randomization suggest the associations are non-causal? *QJM* 97(3): 163-166.

Davis, A.P., C.J. Grondin, K. Lennon-Hopkins, C. Saraceni-Richards, D. Sciaky, B.L. King, T.C. Wieggers, and C.J. Mattingly. 2015. The Comparative Toxicogenomics Data-base's 10th year anniversary: Update 2015. *Nucleic Acids Res.* 43(Database issue): D914-D920.

ESCAPE (European Study of Cohorts for Air Pollution Effects). 2014. ESCAPE Project [online]. Available: <http://www.escapeproject.eu/index.php> [accessed July 21, 2016].

EXPOsOMICS. 2016. About EXPOsOMICS [online]. Available: <http://www.exposomicsproject.eu/> [accessed July 21, 2016].

Fabregat, A., K. Sidiropoulos, P. Garapati, M. Gillespie, K. Hausmann, R. Haw, B. Jassal, S. Jupe, F. Korninger, S. McKay, L. Matthews, B. May, M. Milacic, K. Rothfels, V. Shamovsky, M. Webber, J. Weiser, M. Williams, G. Wu, L. Stein, H. Hermjakob, and P. D'Eustachio. 2016. The Reactome pathway knowledgebase. *Nucleic Acids Res.* 44(D1): D481-D487.

Fasanelli, F., L. Baglietto, E. Ponzi, F. Guida, G. Campanella, M. Johansson, K. Grankvist, M. Johansson, M.B. Assumma, A. Naccarati, M. Chadeau-Hyam, U. Ala, C. Faltus, R. Kaaks, A. Risch, B. De Stavola, A. Hodge, G.G. Giles, M.C. Southey, C.L. Relton, P.C. Haycock, E. Lund, S. Polidoro, T.M. Sandanger, G. Severi, and P. Vineis. 2015. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. *Nat. Commun.* 6:10192.

Fortier, I., P.R. Burton, P.J. Robson, V. Ferretti, J. Little, F. L'Heureux, M. Deschênes, B.M. Knoppers, D. Doiron, J.C. Keers, P. Linksted, J.R. Harris, G. Lachance, C. Boileau, N.L. Pedersen, C.M. Hamilton, K. Hveem, M.J. Borugian, R.P. Gallagher, J. McLaughlin, L. Parker, J.D. Potter, J. Gallacher, R. Kaaks, B. Liu, T. Sprosen, A. Vilain, S.A. Atkinson, A. Rengifo, R. Morton, A. Metspalu, H.E. Wichmann, M. Tremblay, R.L. Chisholm, A. Garcia-Montero, H. Hillege, J.E. Litton, L.J. Palmer, M. Perola, B.H.

Wolffenbuttel, L. Peltonen, and T.J. Hudson. 2010. Quality, quantity and harmony: The DataSHaPER approach to integrating data across bioclinical studies. *Int. J. Epidemiol.* 39(5): 1383-1393.

Freedman, L.S., J.M. Commins, J.E. Moler, W. Willett, L.F. Tinker, A.F. Subar, D. Spiegelman, D. Rhodes, N. Potischman, M.L. Neuhouser, A.J. Moshfegh, V. Kipnis, L. Arab, and R.L. Prentice. 2015. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. *Am. J. Epidemiol.* 181(7): 473-487.

Gallo, V., M. Egger, V. McCormack, P.B. Farmer, J.P. Ioannidis, M. Kirsch-Volders, G. Matullo, D.H. Phillips, B. Schoket, U. Stromberg, R. Vermeulen, C. Wild, M. Porta, and P. Vineis. 2011. STrengthening the Reporting of OBServational studies in Epidemiology-Molecular Epidemiology (STROBE-ME): An extension of the STROBE statement. *PLoS Med.* 8(10): e1001117.

Gordis, L. 2013. *Epidemiology*, 5th Ed. Philadelphia: Elsevier and Saunders. 416 pp.

Greenland, S. 2000. An introduction to instrumental variables for epidemiologists. *Int. J. Epidemiol.* 29(4): 722-729.

Guida, F., T.M. Sandanger, R. Castagné, G. Campanella, S. Polidoro, D. Palli, V. Krogh, R. Tumino, C. Sacerdote, S. Panico, G. Severi, S.A. Kyrtopoulos, P. Georgiadis, R.C. Vermeulen, E. Lund, P. Vineis, and M. Chadeau-Hyam. 2015. Dynamics of smoking-induced genome-wide methylation changes with time since smoking cessation. *Hum. Mol. Genet.* 24(8): 2349-2359.

Guttmacher, A.E., E.G. Nabel, and F.S. Collins. 2009. Why data-sharing policies matter. *Proc. Natl. Acad. Sci. US* 106(40): 16894.

Harris, J.R., P. Burton, B.M. Knoppers, K. Lindpaintner, M. Bledsoe, A.J. Brookes, I. Budin-Ljøsne, R. Chisholm, D. Cox, M. Deschênes, I. Fortier, P. Hainaut, R. Hewitt, J. Kaye, J.E. Litton, A. Metspalu, B. Ollier, L.J. Palmer, A. Palotie, M. Pasterk, M. Perola, P.H. Riegman, G.J. van Ommen, M. Yuille, and K. Zatloukal. 2012. Toward a roadmap in global biobanking for health. *Eur. J. Hum. Genet.* 20(11): 1105-1111.

Hart, J.E., X. Liao, B. Hong, R.C. Pruett, J.D. Yanosky, H. Suh, M.A. Kiomourtzoglou, D. Spiegelman, and F. Laden. 2015. The association of long-term exposure to PM 2.5 on all-cause mortality in the Nurses' Health Study and the impact of measurement-error correction. *Environ. Health.* 14:38.

Hiatt, R.A., S. Sulsky, M.C. Aldrich, N. Kreiger, and R. Rothenberg. 2013. Promoting innovation and creativity in epidemiology for the 21st century. *Ann. Epidemiol.* 23(7): 452-454.

IARC (International Agency for Research on Cancer). 2016. The European Prospective Investigation into Cancer and Nutrition (EPIC) Study [online]. Available: <http://epic.iarc.fr/> [accessed July 21, 2016].

Jennen, D., A. Ruiz-Aracama, C. Magkoufopoulou, A. Peijnenburg, A. Lommen, J. van Delft, and J. Kleinjans. 2011. Integrating transcriptomics and metabolomics to unravel modes-of-action of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD) in HepG2 cells. *BMC Syst. Biol.* 5:139.

Kamburov, A., R. Cavill, T.M. Ebbels, R. Herwig, and H.C. Keun. 2011. Integrated pathway-level analysis of transcriptomics and metabolomics data with IMPaLA. *Bioinformatics* 27(20): 2917-2918.

Kanehisa, M., and S. Goto. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28(1): 27- 30.

Khoury, M.J., T.K. Lam, J.P. Ioannidis, P. Hartge, M.R. Spitz, J.E. Buring, S.J. Chanock, R.T. Croyle, K.A. Goddard, G.S. Ginsburg, Z. Herceg, R.A. Hiatt, R.N. Hoover, D.J. Hunter, B.S. Kramer, M.S. Lauer, J.A. Meyerhardt, O.I. Olopade, J.R. Palmer, T.A. Sellers, D. Seminara, D.F. Ransohoff, T.R. Rebbeck, G. Tourassi, D.M. Winn, A. Zaubler, and S.D. Schully. 2013. Transforming epidemiology for 21st century medicine and public health. *Cancer Epidemiol. Biomarkers Prev.* 22(4): 508-516.

Krämer, A., J. Green, J. Pollard, and S. Tugendreich. 2014. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 30(4): 523-530.

Lam, T.K., M. Spitz, S.D. Schully, and M.J. Khoury. 2013. "Drivers" of translational cancer epidemiology in the 21st century: Needs and opportunities. *Cancer Epidemiol. Biomarkers Prev.* 22(2): 181-188.

Levin, M.L. 1953. The occurrence of lung cancer in man. *Acta. Unio. Int. Contra. Cancrum.* 9(3): 531-541.

Li, R., E. Weller, D.W. Dockery, L.M. Neas, and D. Spiegelman. 2006. Association of indoor nitrogen dioxide with respiratory symptoms in children: Application of measurement error correction techniques to utilize data from multiple surrogates. *J. Expo. Sci. Environ. Epidemiol.* 216(4): 342-350.

Lubin, J.H., J.D. Boice, Jr., C. Edling, R.W. Hornung, G.R. Howe, E. Kunz, R.A. Kusiak, H.I. Morrison, E.P. Radford, J.M. Samet, M. Tirmarche, A. Woodward, S.X. Yao, and D.A. Pierce. 1995. Lung cancer in radon-exposed miners and estimation of risk from indoor exposure. *J. Natl. Cancer Inst.* 87(11): 817-827.

Mascalzoni, D., E.S. Dove, Y. Rubinstein, H.J.S. Dawkins, A. Kole, P. McCormack, S. Woods, O. Riess, F. Schaefer, H. Lochmüller, B.M. Knoppers, and M. Hansson. 2015. International Charter of principles for sharing bio-specimens and data. *Eur. J. Hum. Genet.* 23:721-728.

Mervis, J. 2012. US science policy. Agencies rally to tackle big data. *Science* 336(6077): 22.

Miller, G.W., and D.P. Jones. 2014. The nature of nurture: Refining the definition of the exposome. *Toxicol. Sci.* 137(1): 1-2.

Molitor, J., M. Papathomas, M. Jerrett, and S. Richardson. 2010. Bayesian pro le regression with an application to the National Survey of Children's Health. *Biostatistics* 11(3): 484-498.

NRC (National Research Council). 1983. *Risk Assessment in the Federal Government: Managing the Process*. Washington, DC: National Academy Press.

NRC (National Research Council). 2012. *Exposure Science in the 21st Century: A Vision and a Strategy*. Washington, DC: The National Academies Press.

Papathomas, M., J. Molitor, S. Richardson, E. Riboli, and P. Vineis. 2011. Examining the joint effect of multiple risk factors using exposure risk profiles: Lung cancer in non-smokers. *Environ. Health Perspect.* 119(1): 84-91.

Patel, C.J., J. Bhattacharya, and A.J. Butte. 2010. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. *PLoS One* 5(5): e10746.

Patel, C.J., D.H. Rehkopf, J.T. Leppert, W.M. Bortz, M.R. Cullen, G.M. Chertow, and J.P. Ioannidis. 2013. Systematic evaluation of environmental and behavioral factors associated with all-cause mortality in the United States National Health and Nutrition Examination Survey. *Int. J. Epidemiol.* 42(6): 1795-1810.

Pechette, J.M. 2012. Transforming health care through cloud computing. *Health Care Law Mon.* 5:2-12.

Perttula, K., W.M. Edmands, H. Grigoryan, X. Cai, A.T. Iavarone, M.J. Gunter, A. Naccarati, S. Polidoro, A. Hubbard, P. Vineis, and S. Rappaport. 2016. Evaluating ultra-long chain fatty acids as biomarkers of colorectal cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 25(8): 1216-1223.

Rappaport, S.M. 2012. Biomarkers intersect with the exposome. *Biomarkers* 17(6): 483-489.

Rappaport, S.M., and M.T. Smith. 2010. Environment and disease risks. *Science* 330(6003): 460-461.

Relton, C.L., and G. Davey Smith. 2012. Two-step epigenetic Mendelian randomization: A strategy for establishing the causal role of epigenetic processes in pathways to disease. *Int. J. Epidemiol.* 41(1): 161-176.

Ritchie, S.A., J. Tonita, R. Alvi, D. Lehotay, H. Elshoni, S. Myat, J. McHattie, and D.B. Goodenowe. 2013. Low-serum GTA-446 anti-9 inflammatory fatty acid levels as a new risk factor for colon cancer. *Int. J. Cancer.* 132(2): 355- 362.

Roger, V.L., E. Boerwinkle, J.D. Crapo, P.S. Douglas, J.A. Epstein, C.B. Granger, P. Greenland, I. Kohane, and B.M. Psaty. 2015a. Roger et al. respond to “future of population studies.” *Am. J. Epidemiol.* 181(6): 372-373.

Roger, V.L., E. Boerwinkle, J.D. Crapo, P.S. Douglas, J.A. Epstein, C.B. Granger, P. Greenland, I. Kohane and B.M. Psaty. 2015b. Strategic transformation of population studies: Recommendations of the working group on epidemiology and population sciences from the National Heart, Lung, and Blood Advisory Council and Board of External Experts. *Am. J. Epidemiol.* 181(6): 363-368.

Saito, R., M.E. Smoot, K. Ono, J. Ruschinski, P.L. Wang, S. Lotia, A.R. Pico, G.D. Bader, and T. Ideker. 2012. A travel guide to Cytoscape plugins. *Nat. Methods* 9(11): 1069- 1076.

Salek, R.M., S. Neumann, D. Schober, J. Hummel, K. Billiau, J. Kopka, E. Correa, T. Reijmers, A. Rosato, L. Tenori, P. Turano, S. Marin, C. Deborde, D. Jacob, D. Rolin, B. Dar- tiques, P. Conesa, K. Haug, P. Rocca-Serra, S. O’Hagan, J. Hao, M. van Vliet, M. Sysi-Aho, C. Ludwig, J. Bouw- man, M. Cascante, T. Ebbels, J.L. Griffin, A. Moing, M. Nikolski, M. Oresic, S.A. Sansone, M.R. Viant, R. Good- acre, U.L. Günther, T. Hankemeier, C. Luchinat, D. Wal- ther, and C. Steinbeck. 2015. COordination of Standards in MetabOmicS (COSMOS): Facilitating integrated metabolomics data access. *Metabolomics* 11(6): 1587-1597.

Samet, J.M., R. Schnatter, and H. Gibb. 1998. Invited commentary: Epidemiology and risk assessment. *Am. J. Epidemiol.* 148(10): 929-936.

Shi, M., and C.R. Weinberg. 2011. How much are we missing in SNP-by-SNP analyses of genome- wide association studies? *Epidemiology* 22(6): 845-847.

Siegmund, K.D., A.J. Levine, J. Chang, and P.W. Laird. 2006. Modeling exposures for DNA methylation pro les. *Cancer Epidemiol. Biomarkers Prev.* 15(3): 567-572.

Slama, R., and A. Werwatz. 2005. Controlling for continuous confounding factors: Non- and semiparametric approaches. *Rev. Epidemiol. Sante Publique* 53(Spec. No. 2): 2S65-2S80.

Tenopir, C., S. Allard, K. Douglass, A.U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame. 2011. Data sharing by scientists: Practices and perceptions. *PLoS One* 6(6): e21101.

Thomas, D.C., J.P. Lewinger, C.E. Murcray, and W.J. Gauderman. 2012. Invited commentary: GE- Whiz! Ratcheting gene-environment studies up to the whole genome and the whole exposome. *Am. J. Epidemiol.* 175(3): 203-207.

Turner, M.C., M. Nieuwenhuijsen, K. Anderson, D. Balshaw, Y. Cui, G. Dunton, J.A. Hoppin, P. Koutrakis, and M. Jerrett. In press. Assessing the exposome with external measures: Commentary on the State of the Science and Research Recommendations. *Annual Review of Public Health.*

UK Biobank. 2016. Biobank [online]. Available: [http:// www.ukbiobank.ac.uk/](http://www.ukbiobank.ac.uk/) [accessed July 21, 2016].

Vineis, P., A.E. Khan, J. Vlaanderen, and R. Vermeulen. 2009. The impact of new research technologies on our understanding of environmental causes of disease: The concept of clinical vulnerability. *Environ. Health* 8:54.

Vineis, P., K. van Veldhoven, M. Chadeau-Hyam, and T.J. Athersuch. 2013. Advancing the application of omics-based biomarkers in environmental epidemiology. *Environ. Mol. Mutagen.* 54(7): 461-467.

Vrijheid, M., R. Slama, O. Robinson, L. Chatzi, M. Coen, P. van den Hazel, C. Thomsen, J. Wright, T.J. Athersuch, N. Avellana, X. Basagaña, C. Brochot, L. Bucchini, M. Bustamante, A. Carracedo, M. Casas, X. Estivill, L. Fairley, D. van Gent, J.R. Gonzalez, B. Granum, R. Gražulevičienė, K.B. Gutzkow, J. Julvez, H.C. Keun, M. Kogevinas, R.R. McEachan, H.M. Meltzer, E. Sabidó, P.E. Schwarze, V. Siroux, J. Sunyer, E.J. Want, F. Zeman, and M.J. Nieuwenhuijsen. 2014. The human early-life exposome (HELIX): Project rationale and design. *Environ. Health Perspect.* 122(6): 535-544.