

УДК:475.15.275

## ПОЛНОГЕНОМНЫЙ АНАЛИЗ ПУЛИРОВАННЫХ ВЫБОРОК ДНК КОГОРТ ЧЕЛОВЕКА

© 2014 г. **В.Н. Бабенко<sup>1,2</sup>, В.Н. Максимов<sup>1,3</sup>, Е.В. Кулакова<sup>1</sup>, Н.С. Сафронова<sup>1,2</sup>,  
М.И. Воевода<sup>1-3</sup>, Е.И. Рогаев<sup>1,4</sup>**

<sup>1</sup> Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики  
Сибирского отделения Российской академии наук, Новосибирск, Россия,  
e-mail: bob@bionet.nsc.ru;

<sup>2</sup> Новосибирский национальный исследовательский государственный университет,  
Новосибирск, Россия;

<sup>3</sup> Институт терапии СО РАМН, Новосибирск, Россия;

<sup>4</sup> Институт общей генетики РАН, Москва, Россия

Поступила в редакцию 10 октября 2014 г. Принята к публикации 27 октября 2014 г.

Цель работы – разработка, валидация и применение подходов приоритетного скрининга точковых полиморфизмов в эксперименте полногеномного аллелотипирования 16 когорт человека, составленных из больных и здоровых индивидов, по схеме «случай – контроль» на основе пулированных выборок. Генотипирование произведено на платформе Illumina Omni1S, имеющей 1,2 млн маркеров на планшете, выборки в среднем составили 200 человек. При сравнении частот аллелей между когортами патологических и контрольных выборок выявляли наборы полиморфизмов, достоверно отличающиеся друг от друга по частоте. Установлено, что выборки больных индивидов показывают систематические, устойчивые отклонения от здоровых по числу достоверно отличающихся полиморфизмов, дисперсия интенсивности аллелей между повторами одной когорты меньше таковой при случайном выборе пар разных когорт.

**Ключевые слова:** аллелотипирование, пулированная выборка, полногеномный анализ, частота аллеля, точковый полиморфизм.

### ВВЕДЕНИЕ

Пулированная, или смешанная, ДНК представляет собой смесь образцов геномной ДНК нескольких индивидов. Она может служить основой для выявления изменения частот точковых полиморфизмов в данных по сравнительному анализу выборок больных и здоровых индивидов, а именно: при достоверном различии частоты точкового полиморфизма между нормой и патологией мы классифицируем данный полиморфизм как кандидатный для дальнейшего рассмотрения.

Применяя технологию анализа пулированной выборки, заметим, что генотипирование невозможно для пулированного материала, поскольку индивидуальные генотипы в смешанном образце не различимы. Поэтому цель такого анализа – оценка частот аллелей по интенсив-

ности сигнала каждого из аллелей в выборке (аллелотипирование). Для этого вполне подходит методология генотипирования на чипе. В настоящее время аллелотипирование пулированных выборок на генотипирующих чипах Illumina широко распространено (Wilkening *et al.*, 2007; Iliadis *et al.*, 2012; Berglund *et al.*, 2013; Gai *et al.*, 2013; Ozerov *et al.*, 2013; Teumer *et al.*, 2013) и доказало свою эффективность в вопросе выявления кандидатных полиморфизмов по схеме «случай – контроль».

С другой стороны, фирма Illumina прогрессивно увеличивает плотность точковых полиморфизмов на планшете, вводя технологически новые платформы и используя новые полиморфизмы, выявленные в проекте «1000 геномов» (1000GP; <http://www.1000genomes.org>). В частности, массив, использованный в нашей рабо-

те, содержит 1,2 млн маркеров ([http://support.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet\\_human\\_omni1s.pdf](http://support.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_human_omni1s.pdf)). Такое число полиморфизмов является представительной выборкой относительно состояния dbSNP к настоящему времени, что предполагает извлечение новой информации из ранее не известных полиморфных сайтов.

В настоящем исследовании метод пулированного аллелотипирования применен к ряду когорт индивидов Новосибирской области для выявления кандидатных полиморфизмов. Шестнадцать изученных выборок представляли образцы крови больных и здоровых когорт индивидов проекта НАРЕЕ, а также образцы, собранные в Институте терапии РАМН г. Новосибирска (Malyutina *et al.*, 2002; Nicholson *et al.*, 2008; Bobak *et al.*, 2009; Pajak *et al.*, 2013).

В указанных работах дан анализ биометрических, патофизиологических и биохимических медицинских показателей. В предлагаемой работе впервые проведен анализ полногеномного аллелотипирования этих данных, проанализированы их качество и статистические

свойства для оценки их пригодности в целях дальнейшего рассмотрения.

## МАТЕРИАЛЫ И МЕТОДЫ

### Описание выборок

Основная часть выборок собрана в рамках проекта НАPIEE (Health, Alcohol and Psychosocial factors In Eastern Europe; [http://www.ucl.ac.uk/easteurope/hapiee\\_open.htm](http://www.ucl.ac.uk/easteurope/hapiee_open.htm)) в г. Новосибирске. Методы построения выборок вкуче с анализом морфометрических, биохимических и психосоциальных данных по данному проекту неоднократно опубликованы (Malyutina *et al.*, 2002; Nicholson *et al.*, 2008; Bobak *et al.*, 2009; Vikhireva *et al.*, 2010; Pajak *et al.*, 2013). Анализ проведен на 16 когортах (табл. 1), краткая аннотация выборок и способ выбора представлены в приложении.

Геномную ДНК выделяли из венозной крови методом фенол-хлороформной экстракции. Концентрацию исходных образцов оценивали на спектрофотометре Nanodrop 1000 (ThermoFisher

Таблица 1

Генотипированные когорты

Когорта	Кол-во человек	Кол-во репликаций
Мальчики	200	2
Девочки	200	2
Мужчины	200	2
Женщины	200	2
Инфаркт миокарда (НАPIEE), мужчины	200	2
Инфаркт миокарда (Куимов), мужчины	200	2
Внезапная смерть, мужчины	200	2
Артериальная гипертензия с метаболическим синдромом		
мужчины	200	2
женщины	200	2
Артериальная гипертензия без метаболического синдрома		
мужчины	200	2
женщины	200	2
Долгожители, женщины	130	2
Инсульт инцидентный ишемический, мужчины	60	2
Глаукома, мужчины и женщины	149	2
Пилоты, мужчины	202	2
Синдром Вольфа – Паркинсона – Уайта, мужчины и женщины	81	2

Scientific Inc., USA). Затем оцененные концентрации ДНК были выравнены до концентрации 50 нг/мкл. Качество ДНК-образцов проверено с помощью ПЦР. После этого 10 мкл каждого образца были взяты для пулирования. Для анализа на чипе отобрано 1 мкг пулированной ДНК с концентрацией 50 нг/мкл. Две реплики для каждой когорты были собраны и пулированы (объединены) независимым образом. Препарат приготовлен в соответствии с протоколом изготовителя ([www.illumina.com](http://www.illumina.com)).

Генотипирование проводили в центре «Биоинженерия» РАН (<http://www.biengi.ac.ru/>) на платформе Illumina Omni1S-8v1H12 ([www.illumina.com](http://www.illumina.com)), содержащей 8 планшетов по 1,2 млн маркеров. Были использованы 4 чипа, таким образом, проанализировано 32 образца пулированной ДНК. Полученные предварительные данные были процессированы с помощью программы Illumina GenomeStudio 2011.1 (<http://support.illumina.com/downloads/ilmn>) и размещены в СУБД Mysql v.6.3. ([www.mysql.com](http://www.mysql.com)).

Из 32 экспериментов были извлечены выборки полиморфизмов с частотами в диапазоне [0,01 ... 0,99], средняя численность таких полиморфизмов на планшет составила 830 тыс. Затем были скомпилированы целевые выборки 16 когорт, в которых в качестве частоты аллеля взята средняя двух реплик. Средняя численность этих полиморфизмов на когорту 560 тысяч. (Обе частоты не ниже 0,01.) После этого произведено попарное сравнение выборок когорт по частотам полиморфизмов.

Достоверность попарных различий частот оценивали с помощью критерия  $\chi^2$  для четырехпольной таблицы, составленной как

$$(n_{11} = N_1 p_1, n_{12} = N_1(1 - p_1), n_{21} = N_2 p_2, n_{22} = N_2(1 - p_2)):$$

$$\chi^2 = \frac{N(n_{11}n_{22} - n_{21}n_{12})^2}{(n_{11} + n_{12})(n_{11} + n_{21})(n_{12} + n_{22})(n_{21} + n_{22})}$$

где  $p_1, p_2$  – частоты  $B$  аллелей полиморфизма в выборке 1 и 2,  $N_1, N_2$  – объемы выборок 1 и 2.

В качестве внешних контрольных частот взяты частоты соответствующих полиморфизмов выборки полиморфизмов 360 европейских представителей проекта «1000 геномов» в популяциях GBR, TSI, FIN, CEU ([www.1000genomes.org](http://www.1000genomes.org)). Для внутреннего контроля взяты когорты девочек, мальчиков, мужчин и женщин. Данные

об аннотированных полиморфизмах в базах данных OMIM, GENEREVIEWs взяты из dbSNP v.137 с указанием соответствующей аннотации в системе фильтров dbSNP (<http://www.ncbi.nlm.nih.gov/snp/limits>), насчитывающих 18 291 и 34 373 полиморфизмов соответственно. Таблица о полиморфизмах, найденных в проектах полногеномного ассоциативного анализа (Genome Wide Association Studies, GWAS), взята на сайте базы данных Университета Калифорнии Санта-Круз ([www.genome.ucsc.edu](http://www.genome.ucsc.edu)).

Оценка LD парного сцепления по полиморфизмам осуществлена программой plink (Purcell *et al.*, 2007) на европейской популяции (см. выше) с параметром  $r^2 > 0,8$ . Половые хромосомы в случае мужчин обработаны по отдельному сценарию.

## РЕЗУЛЬТАТЫ

### Валидация исходных данных

#### *Внутри- и межпопуляционная дисперсия интенсивности аллелей*

Результатом генотипирования по отдельной точке на платформе Illumina является значение двумерного вектора  $(X, Y)$  ([www.illumina.com/genotyping/ilmn](http://www.illumina.com/genotyping/ilmn)). Значения вектора  $(N, 0)$  и  $(0, N)$  соответствуют гомозиготным генотипам, а  $(N/2, N/2)$  – гетерозиготе, где  $N$  – значение сигнала на красной или зеленой «бусинах», соответствующих аллелям. В общем случае сигнал, называемый интенсивностью, имеет целые значения  $(N_1, N_2)$ . В целом на анализируемой платформе Illumina значения интенсивности  $N$  менялось от 0 до 62 000. Значение  $X$  было, как правило, выше значения  $Y$ , видимо, в силу выбора  $b$  аллеля на платформе в сторону минимального.

Мы применили дисперсионный F-тест для анализа внутри- и межпопуляционных различий для оценки конкордантности реплик по отношению к случайным данным (пары разных когорт). В качестве внутривнутрипопуляционных данных взяты соответствующие двумерные Raw координаты выявленных гетерозиготных генотипов каждой из когорт по каждому маркеру. Для межпопуляционных различий были случайно выбраны 16 пар реплик без совпадения по когорте. По каждой паре выборок вычислено

суммарное отличие всех общих полиморфизмов отдельно по осям  $x$  и  $y$  (табл. 2). В результате мы получили достоверное различие ( $p < 0,015$ ) по абсциссе ( $x$ ) и недостоверное, хотя систематическое, отличие ( $p < 0,12$ ) по ординате ( $y$ ) между указанными парами, подтверждающее систематическое отличие в меньшую сторону внутривнутрипопуляционных различий (табл. 2).

Другим критерием валидации было попарное различие мужчин и женщин одной когорты, где были представлены и мужчины, и женщины, и случайный выбор мужчин и женщин разных когорт. F-критерий показал достоверное отличие от случайного ( $p < 0,014$ ) и здесь.

### **Попарное сравнение когорт по частотам аллелей**

После компилирования усредненных значений частот аллелей мы провели попарное сравнение между когортами. Для этого определили матрицу попарных различий как число достоверно различающихся полиморфизмов (см. Материалы и методы) относительно контрольной (мужчины или женщины) когорты. В качестве иллюстрации на рис. 1 даны числа статистически отличных по частоте (СОПЧ) полиморфизмов относительно соответствующих контрольных групп (мужчины и женщины) для 14 когорт. Видно, что значительное число СОПЧ-полиморфизмов относительно контрольных когорт наблюдается (в порядке убывания) у когорт внезапной смерти (ВС), глаукомы и синдрома Вольфа – Паркинсона – Уайта (Wolff – Parkinson – White, WPW), при этом когорта девочек также характеризуется

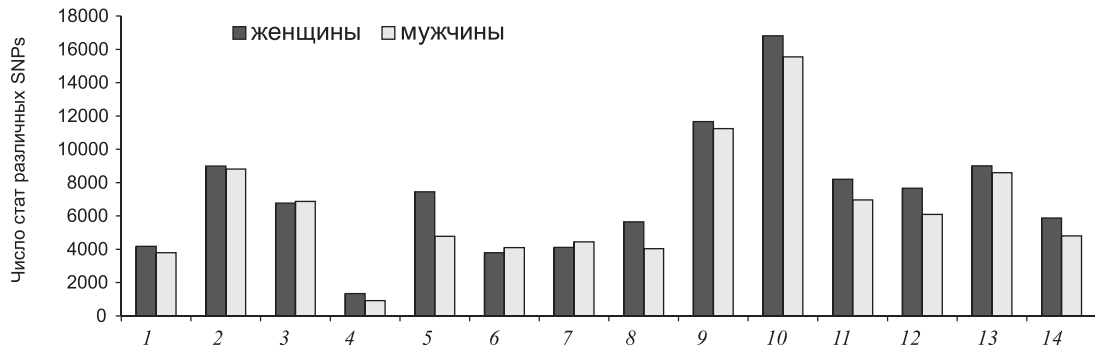
высокими отличиями от женской и мужской контрольных когорт по числу СОПЧ-полиморфизмов. Характерную особенность составляет систематическое, хотя и небольшое, превышение числа различий СОПЧ-полиморфизмов при сравнении с контрольной когортой женщин практически у всех когорт относительно сравнения с контрольной когортой мужчин (столбик «женщины» в гистограмме систематически выше), что, видимо, обусловлено наличием двух X-хромосом при оценке полиморфизма у женщин. Малое число СОПЧ-полиморфизмов в когорте инсульт инцидентный ишемический обусловлено сравнительно малым размером выборки (см. табл. 1). Рисунок 1 показывает систематические отличия, по крайней мере по количеству найденных кандидатных СНП, при сравнении с независимыми выборками контроля.

Для визуализации распределения 16 когорт использована матрица попарных отличий по числу СОПЧ-полиморфизмов для построения графика многомерного шкалирования (МШ) (рис. 2). Видно, что контрольные когорты оказались в левой половине графика (девочки, мальчики, мужчины, женщины). Положение когорты инцидентного инсульта, также попавшей в левую половину, можно считать незначимой вследствие значительно меньшего объема выборки и, следовательно, минимального числа СОПЧ-полиморфизмов (см. рис. 1), делающей данную выборку «аутлайером». Следует заметить, что девочки и женщины расположены на графике достаточно далеко, в то время как мужчины и мальчики расположены относитель-

**Таблица 2**

F-тест по сравнению внутри- и межпопуляционных различий сигнала Raw по осям  $x$  и  $y$

Показатель	Raw $x$ различия		Raw $y$ различия	
	наблюдаемые	ожидаемые	наблюдаемые	ожидаемые
Среднее	651,8	1 102,3	634,4	803,1
Дисперсия	44 737,6	143 406,7	63 672,7	11 8123,2
Наблюдения	16	16	16	16
Df	15	15	15	15
F	0,3119		0,539	
P ( $F \leq f$ ) одностороннее	0,015		0,12	
F критическое одностороннее	0,41		0,41	



**Рис. 1.** Количество статистически достоверных отличий полиморфизмов относительно контрольной выборки мужчин и женщин в 14 когортах.

1 – мальчики; 2 – девочки; 3 – долгожители, женщины; 4 – инсульт инцидентный ишемический, мужчины; 5 – артериальная гипертензия (АГ) с метаболическим синдромом (МС), мужчины; 6 – АГ с МС, женщины; 7 – АГ без МС, женщины; 8 – АГ без МС, мужчины; 9 – глаукома, мужчины и женщины; 10 – внезапная смерть, мужчины; 11 – инфаркт миокарда (Куимов), мужчины; 12 – инфаркт миокарда (НАПЕЕ), мужчины; 13 – синдром Вольфа – Паркинсона – Уайта, мужчины и женщины; 14 – пилоты, мужчины.

но близко, что, по всей видимости, свидетельствует о смещении выборки девочек, вероятно, в силу некоторого отличия концентрации ДНК. Наибольшее отличие от контрольных групп по числу СОПЧ-полиморфизмов, как и на рис. 1, имеют: а) ВС, мужчины, б) синдром WPW и в) глаукома.

Рисунки 1 и 2 иллюстрируют результат выполнения первой фазы выявления кандидатных полиморфизмов визуализацией распределения СОПЧ-полиморфизмов; можно сделать выводы о том, что в среднем несколько тысяч полиморфизмов в каждой выборке достоверно отличаются от контроля. Различия количества СОПЧ-полиморфизмов при смене контрольной существенны. Мы решили также проверить, насколько часто они повторены в обеих контрольных выборках.

Для пяти патологий с наиболее выраженным числом отличий (см. рис. 1) мы провели валидацию оценки числа общих при разных контрольных выборках (мужчины, женщины) СОПЧ-полиморфизмов (табл. 3). Видно, что в среднем 25–38 % отличий систематически повторены относительно обоих контролей (мужчин и женщин). Ожидаемая вероятность случайного совпадения одного маркера при выборе 9 тыс. из 500 тыс. маркеров ( $p = 0,018$ ) равна  $p = 0,0003$ . Систематическое отклонение оценено как число совпадающих у 14 когорт

СОПЧ-полиморфизмов по отношению к двум контрольным когортам (мужчин и женщин), таких полиморфизмов найдено не было. При числе достоверных различий между мужчинами и женщинами 6 700 из 500 000, и, следовательно, 30 % (около 3 000) совпадений, данной вероятностью можно пренебречь.

### ОБСУЖДЕНИЕ

В работах по анализу данных при пулированном способе нанесения материала ставятся две задачи: 1) валидация и адекватная оценка частот аллелей по данным эксперимента и 2) поиск наиболее вероятных кандидатных полиморфизмов, в нашем случае по схеме «случай – контроль». При валидации полиморфизмов одним из вариантов является проверка гаплотипического конформизма, использованная, в частности, Gaj и соавт. (2012). При такой валидации одним из критериев к кандидатным полиморфизмам было отклонение от контроля не менее двух сцепленных полиморфизмов, включая кандидатный. В этой работе контрольные гаплотипы и частоты восстановлены по выборке нативной (польской) популяции, на основе индивидуального полногеномного генотипирования (секвенирования). К сожалению, индивидуальных полногеномных данных по российской (сибирской) популяции достаточного объема у нас не было, поэтому ис-

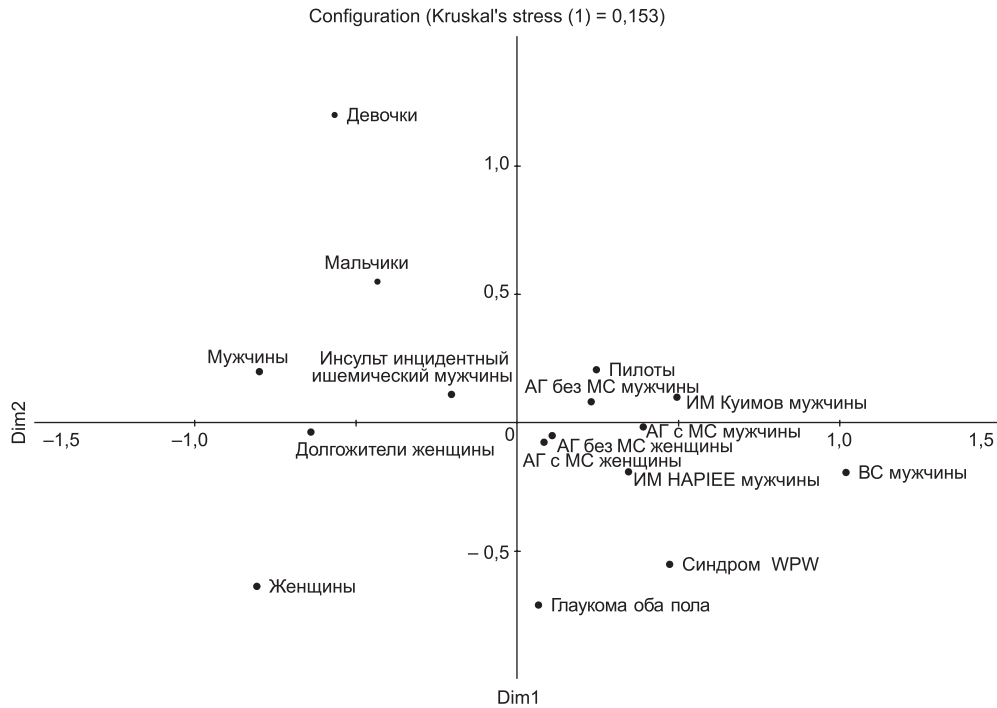


Рис. 2. График многомерного шкалирования матрицы попарных различий частот аллелей когорт.

Таблица 3

Распределение СОПЧ-полиморфизмов по выборкам пяти патологических когорт относительно контрольной выборки женщин (КВЖ) и мужчин (КВМ)

Кол-во СОПЧ-полиморфизмов	Всего	Общие, в КВЖ и КВМ*
Внезапная смерть		
КВЖ	16 820	5 473 (33 %)
КВМ	15 550	
Глаукома		
КВЖ	11 672	3 564 (32 %)
КВМ	11 244	
Инфаркт миокарда (НАРІЕЕ)		
КВЖ	7 663	1 519 (25 %)
КВМ	6 089	
Синдром WPW		
КВЖ	9 009	3 254 (38 %)
КВМ	8 596	

\* В процентах по отношению к предыдущему столбцу.

пользована выборка европейцев для выявления сцепленных по  $r^2$  полиморфизмов, содержащих кандидатные. Мы идентифицировали их с помощью программы plink (Purcell *et al.*, 2007) по европейской популяции из проекта «1000 геномов» (Abecasis, Auton, 2012). Установлено,

что примерно 10 % кандидатных полиморфизмов значимо сцеплены между собой ( $r^2 > 0,8$ ), при этом их частоты имеют достоверно меньшую дисперсию, чем при отсутствии сцепления, что подтверждает состоятельность оценки частот для пулированных образцов на чипе. Рассматривая

обогащение различных геномных сегментов полиморфизмами из GWAS, стоит отметить, что, как и ожидалось, наибольшее количество ассоциированных с признаком полиморфизмов расположено в экзонных областях (плотность превышена в 2,6 раза по сравнению с контролем). При анализе взаимодействия SNP-eQTL или влияния полиморфизма на экспрессию генов стоит отметить, что она практически полностью соответствует распределению GWAS-полиморфизмов: наибольшую роль играют экзонные состояния, и потом все остальные.

В ряде случаев мы наблюдали совместный, противоположный эффект меж- (внутригеномных) полиморфизмов на несколько соседних генов, что, по-видимому, обусловлено модулированием хроматинового состояния доменов ДНК. Доля СОПЧ-полиморфизмов, имеющих известную eQTL-ассоциацию по внешним данным (2,1 %), недостоверно превышает долю таковых в общем пуле полиморфизмов (1,1 %). Тем не менее мы планируем использовать эту информацию при дальнейшей селекции кандидатных полиморфизмов.

## ЗАКЛЮЧЕНИЕ

Стратегия выявления кандидатных полиморфизмов по пулированным данным при наличии множества когорт показывает систематичность отличий когорт больных индивидов от когорт здоровых по числу СОПЧ-полиморфизмов. Безусловно, этот вывод основан на обобщенных характеристиках когорт и условном показателе соотношений СОПЧ-полиморфизмов. Он не исключает как возможные систематические смещения по точному соотношению ДНК пулированных индивидов, вариации концентрации ДНК в различных выборках, так и возможное выпадение качества генотипирования на отдельных полиморфизмах. Работа по оценке указанных факторов продолжается.

## БЛАГОДАРНОСТИ

Работа поддержана бюджетным проектом ИЦиГ СО РАН VI.61.1.2, грантом Правительства Российской Федерации 14.B25.31.0033 и грантом РФФИ 14-04-01906. Вычисления

производились в суперкомпьютерном центре ИВМиМГ СО РАН ([www.sccc.ru](http://www.sccc.ru)).

Авторы благодарны Елене Пивоваровой и Татьяне Колесниковой, а также рецензентам за ценные замечания.

## ЛИТЕРАТУРА

- Abecasis G.R., Auton A. *et al.* An integrated map of genetic variation from 1,092 human genomes. 1000 Genomes Project Consortium // *Nature*. 2012. V. 491. No. 7422. P. 56–65.
- Berglund E.C., Lindqvist C.M., Hayat S. *et al.* Accurate detection of subclonal single nucleotide variants in whole genome amplified and pooled cancer samples using HaloPlex target enrichment // *BMC Genomics*. 2013. No. 14. P. 856.
- Bobak M., Richards M., Malyutina S. *et al.* Association between year of birth and cognitive functions in Russia and the Czech Republic: cross-sectional results of the HAPIEE study // *Neuroepidemiology*. 2009. V. 33. No. 3. P. 231–239.
- Gaj P., Maryan N., Hennig E.E. *et al.* Pooled sample-based GWAS: a cost-effective alternative for identifying colorectal and prostate cancer risk variants in the Polish population // *PLoS One*. 2012. V. 7. No. 4. P. e35307.
- Iliadis A., Anastassiou D., Wang X. Fast and accurate haplotype frequency estimation for large haplotype vectors from pooled DNA data // *BMC Genet*. 2012. V. 13. P. 94.
- Malyutina S., Bobak M., Kurilovitch S. *et al.* Relation between heavy and binge drinking and all-cause and cardiovascular mortality in Novosibirsk, Russia: a prospective cohort study // *Lancet*. 2002. V. 360. No. 9344. P. 1448–1454.
- Nicholson A., Pikhart H., Pajak A. *et al.* Socio-economic status over the life-course and depressive symptoms in men and women in Eastern Europe // *J. Affect. Disord*. 2008. V. 105. No. (1–3). P. 125–136.
- Ozerov M., Vasemägi A., Wennevik V. *et al.* Cost-effective genome-wide estimation of allele frequencies from pooled DNA in Atlantic salmon (*Salmo salar* L.) // *BMC Genomics*. 2013. V. 14. P. 12.
- Pajak A., Szafraniec K., Kubinova R. *et al.* Binge drinking and blood pressure: cross-sectional results of the HAPIEE study // *PLoS One*. 2013. V. 8. No. 6. P. e65856.
- Purcell S., Neale B., Todd-Brown K. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis // *American J. Human Genetics*. 2007. V. 81. No. 3. P. 559–575.
- Teumer A., Ernst F.D., Wiechert A. *et al.* Comparison of genotyping using pooled DNA samples (allelotyping) and individual genotyping using the affymetrix genome-wide human SNP array 6.0 // *BMC Genomics*. 2013. V. 14. P. 506.
- Vikhireva O., Pikhart H., Pajak A. *et al.* Non-fatal injuries in three Central and Eastern European urban population samples: the HAPIEE study // *Eur. J. Public Health*. 2010. V. 20. No. 6. P. 695–701.
- Wilkening S., Chen B., Wirtenberger M. *et al.* Allelotyping of pooled DNA with 250 K SNP microarrays // *BMC Genomics*. 2007. V. 8. P. 77.

## ПРИЛОЖЕНИЕ

## Краткая аннотация выборок

Мальчики Девочки	14–17 лет, учащиеся 9–11 классов общеобразовательных школ Октябрьского района (обследовано 10 % от всех учащихся этих классов в этом районе)
Мужчины Женщины	45–69 лет (НАРИЕЕ)
Инфаркт миокарда (НАРИЕЕ), мужчины	Группа больных инфарктом миокарда была сформирована на основе популяционной выборки 45–69-летних жителей Октябрьского и Кировского районов г. Новосибирска (9 400 человек), которая была собрана НИИ терапии СО РАМН в ходе работы по международному проекту НАРИЕЕ (Health, Alcohol and Psychosocial factors In Eastern Europe). Программа исследования включала: измерение артериального давления, антропометрия (рост, вес, объем талии, бедер), социально-демографические характеристики, опрос о курении, потреблении алкоголя (частота и типичная доза), уровне физической активности, оценку липидного профиля (общий холестерин; триглицериды, холестерин липопротеидов высокой плотности), опрос на выявление стенокардии напряжения (Rose), ЭКГ покоя в 12 отведениях
Инфаркт миокарда (Куимов), мужчины	В исследование включены больные ОКС, поступившие в блок интенсивной терапии городской клинической больницы № 1 Новосибирска с 01 апреля 2009 по 30 марта 2010 г. (средний возраст $59,1 \pm 6,1$ года), в том числе с ОКС с подъемом сегмента ST 180 человек (117 мужчин) и 58 человек с ОКС без подъема сегмента ST. Средний возраст мужчин $56,2 \pm 5,2$ года. Диагноз ОКС установлен по совокупности критериев, разработанных Европейским обществом кардиологов и Американской коллегией кардиологов (2000), включающих: а) типичный болевой приступ, б) изменения ЭКГ в двух и более последовательных отведениях (высокоамплитудный Т, отрицательный Т, подъем сегмента ST, патологический Q, депрессия сегмента ST, наличие QR), в) динамические изменения в уровне ферментов (КФК, КК-МВ, ТнТ, ТнI)
Внезапная смерть, мужчины	Набор аутопсийного материала проводился у мужчин, умерших внезапно в возрасте 25–64 лет (жителей Октябрьского района Новосибирска), подвергнутых судебно-медицинскому исследованию. Средний возраст умерших составил $53,6 \pm 7,9$ года. При секционном исследовании производили забор образцов ткани печени или миокарда в количестве 5–10 г. Мужчины, умершие внезапно (без морфологических изменений характерных для инфаркта миокарда, дилатационной и гипертрофической кардиомиопатий и др.). С учетом ограниченной информации о времени развития фатального события в исследуемую группу включены случаи смерти, развившейся в течение 1 ч или при отсутствии свидетелей смерти в течение не более 24 ч и расцененных по данным аутопсии как смерть сердечного генеза
АГ с МС, мужчины	45–69 лет НАРИЕЕ
АГ без МС, мужчины	45–69 лет НАРИЕЕ
АГ с МС, женщины	45–69 лет НАРИЕЕ
АГ без МС, женщины	45–69 лет НАРИЕЕ
Геронты, женщины	90 лет и старше
Инсульт инцидентный ишемический, мужчины	Инсульт в НАРИЕЕ 45–69 лет (инсульт, случившийся после прохождения скрининга). Данные предоставлены регистром инсульта
Глаукома, мужчины и женщины (ПОУГ)	79 женщин, 70 мужчин, 41–85 лет, средний возраст 67,2 года
Пилоты, мужчины	Действующие пилоты гражданской авиации, 35–55 лет



Синдром Вольфа – Паркинсона – Уайта, мужчины и женщины	Группа представлена пациентами с синдромом WPW в количестве 81 человек, среднего возраст $32,0 \pm 1,6$ года, из которых 43 были мужчинами (53,0 %) и 38 женщинами (47,0 %). Они были выбраны из группы пациентов с пароксизмальными реципрокными атриовентрикулярными тахикардиями, пролеченных в отделении сердечно-сосудистой хирургии ГУЗ дорожной клинической больницы на ст. Новосибирск-главный, отделениях нарушений ритма Областного кардиодиспансера и НИИ патологии кровообращения им. Е.Н. Мешалкина г. Новосибирска в течение 2000–2003 гг. Критериями включения были: отсутствие ИБС, врожденных и приобретенных пороков сердца, перенесенных острых мио- и перикардитов различной этиологии в течение не менее одного месяца до обследования, нарушения функции щитовидной железы, онкологических заболеваний и беременности
--	---

## GENOME-WIDE SNP ALLELOTYPING OF HUMAN COHORTS BY POOLED DNA SAMPLES

V.N. Babenko<sup>1</sup>, V.N. Maximov<sup>1,3</sup>, E.V. Kulakova, E.S. Safronova<sup>1,2</sup>, M.I. Voevoda<sup>1-3</sup>, E.I. Rogaev<sup>1,4</sup>

<sup>1</sup> Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,  
e-mail: bob@bionet.nsc.ru;

<sup>2</sup> Novosibirsk National Research State University, Novosibirsk, Russia;

<sup>3</sup> Institute of Therapy SB RAMS, Novosibirsk, Russia;

<sup>4</sup> Institute of Genetics RAS, Moscow, Russia

### Summary

The research concerns the task of identification of contrasted single nucleotide polymorphisms (SNPs) obtained in genome-wide pooled allelotyping of 16 human cohorts, comprising healthy and ill persons, by the nested case-control approach. The genotyping platform was the Illumina Omni1S chip with 1.2 million markers. The mean pooled sample size was about 200 individuals. The candidate selection was based on statistical comparison of allele frequencies in a “case-control” study. Samples of ill patients show significant deviations from healthy persons in the numbers of significantly differing polymorphisms. The variance of allele frequencies among repeats in a single cohort was less than that in random choice of pairs from different cohorts.

**Key words:** genome-wide genotyping, pooled sample, allele frequency, SNP.