666 российских геномов?

Олег Балановский

В Петербурге только что прошла конференция по проекту «Российские геномы».

На рабочем совещании присутствовали как организаторы проекта (биоинформатики Санкт-Петербургского университета и московская лаборатория С.А. Лимборской), так и В.А. Степанов (Томск), Э.К. Хуснутдинова (Уфа), М.И. Воевода (Новосибирск), О.П. Балановский (Москва) – то есть лидеры всех основных популяционно-генетических коллективов России, кроме Б.А. Малярчука (Магадан). Присутствовавшие коллективы озвучили размеры своих биобанков, варьирующие от 6 000 тысяч до 24 000 образцов представителей коренного населения. Все присутствовавшие коллективы, кроме нашего, заявили о своем решении участвовать в этом проекте. Предлагаем Вашему вниманию доклад О.П. Балановского, представленный на этой конференции. В нем, в частности, говорится, что планируемый анализ триад (отец, мать, ребенок) сокращает объем полезной геномной информации на одну треть, и поэтому вместо 1000 российских геномов фактически будет изучено 666 геномов.

В Петербурге только что прошло рабочее совещание по проекту «Российские геномы». Пригласили туда и меня, и наряду с другими приглашенными специалистами, я поделился своим мнением о проекте. Поскольку он широко обсуждается российской научной общественностью, и оценки озвучиваются диаметрально противоположные, предлагаю посетителям нашего сайта, интересующимся этим вопросом, ознакомиться с моим выступлением.

Я, наверное, не вправе пересказывать все, что говорилось на конференции – это прерогатива организаторов. Но три вещи, объявлявшиеся во всеуслышание, полагаю, сообщить все же можно.

Во-первых, перечень планируемых к изучению народов, выложенный на сайте проекта(www.genomerussia.org, http://genomerussia.bio.spbu.ru/?page_id=141) является весьма предварительным — он был обновлен в ходе совещания. В ближайшие дни организаторы планируют обновить его и на сайте.

Во-вторых, основные российские коллективы, занимающиеся генетикой народонаселения, приглашены участвовать в проекте. Коллектив С.А. Лимборской был в проекте с самого начала, потом присоединился коллектив Э.К. Хуснутдиновой, а коллективы В.А. Степанова и М.И. Воеводы приняли это приглашение в ходе совещания. Наш коллектив пока не принял решения. Еще один сильный коллектив – под руководством Б.А. Малярчука – не был представлен на совещании, и относительно его участия или неучастия мне ничего не известно. Можно полагать, что уже определившееся участие ряда основных российских игроков существенно укрепит позицию этого проекта.

В-третьих, в ходе обсуждения проекта впервые все основные коллективы озвучили размеры своих биобанков:

В.А. Степанов указал более 10 000 образцов из более 100 популяций; насколько я понял, в это число входят и образцы пациентов по более, чем 20 нозологиям;

Э.К. Хуснутдинова указала около 10 000 образцов из популяций 50 этнических групп (некоторые народы представлены несколькими популяциями) и, кроме этого, 44 415 образцов больных по 47 нозологиям. Была показана карта расположения изученных популяций.

М.И. Воевода указал 9 248 образцов (в показанной им таблице было 37 групп).

О.П. Балановский указал 24 078 образцов из 256 популяций, представляющих 90 народов. Была показана карта расположения изученных популяций.

С.А. Лимборская на этом совещании на объеме биобанка не останавливалась, но наши институты недавно обменивались этой информацией для подготовки другого проекта, и в ходе этого обмена я дал информацию, что в нашем биобанке указал 24 078 образцов, а Институт молекулярной генетики указал 6 316 образцов. Можно предполагать, что основную его часть (или весь) представлена биобанком коллектива С.А. Лимборской.

Конечно, общее число образцов мало что говорит об изученности каждого конкретного народа. Но оно показывает, сколь внушительные коллекции от коренного населения накоплены российскими учеными (как легко подсчитать, 10+10+9+24+6=59, то есть суммарно около 60 тысяч образцов). Это позволяет не только гордиться нашей наукой, которая почти без финансирования смогла провести такую работу, но и надеяться, что в наступившую «полногеномную» эру значительная часть этих образцов будет генотипирована по полногеномным панелям, сформировав огромный массив данных

о структуре генофонданародов России и сопредельных стран. Конечно, отдельный и очень важный вопрос – это тщательность отбора именно коренного, а не частично метисированного населения, но на этом вопросе докладчики, к сожалению, не останавливались.

После этих трех комментариев перехожу к публикации своего выступления.



«Глубокоуважаемые коллеги,

Я представляю здесь научный коллектив, аффилированный с Институтом общей генетики, Медико-генетическим научным центром и включающий также ученых иных организаций России и других стран. Коллектив возглавляют Елена Владимировна Балановская и я.



Некоторые научные показатели, достигнутые за почти 20 лет работы, приведены на слайде. Они стоят на трех китах:

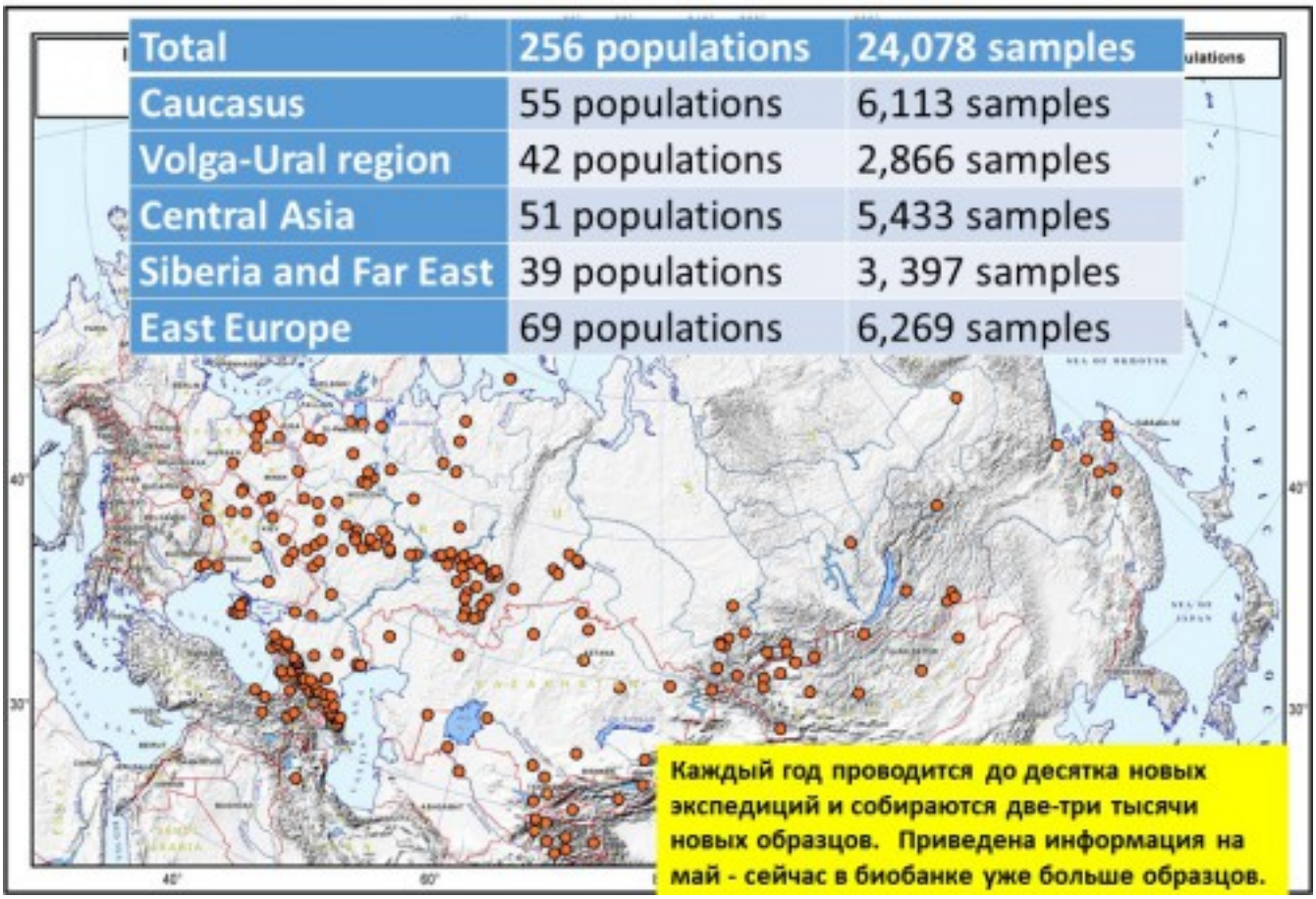
— на созданном за эти годы биобанке,

— на наших банках **данных** о генофонде мира,

— на разрабатываемых нами картографических методах изучения генофонда.

Позвольте мне кратко рассказать о нашем биобанке и о том, как он используется в геномных проектах. Возможно, эта информация будет интересна для консорциума Вашего проекта.

ОБЪЕМ И ОХВАТ БИОБАНКА



Наш биобанк включает образцы от 24 тысяч человек – представителей коренных народов Северной Евразии. Охвачены 256 популяций, представляющих 90 народов России и сопредельных стран. Каждая точка на приведенной карте – это отдельная популяция. Типичный объем выборки – от 80 до 120 образцов. Можно видеть, что очень подробно изучен и Кавказ, и Восточная Европа, и Центральная Азия, и Дальний Восток. Из Сибири хорошо покрыта только ее южная часть, но сейчас мы планируем экспедиции и в более северные регионы.

Каждый год проводится до десятка новых экспедиций и собирается две-три тысячи новых образцов. Эта информация на май, сейчас в биобанке уже больше образцов.

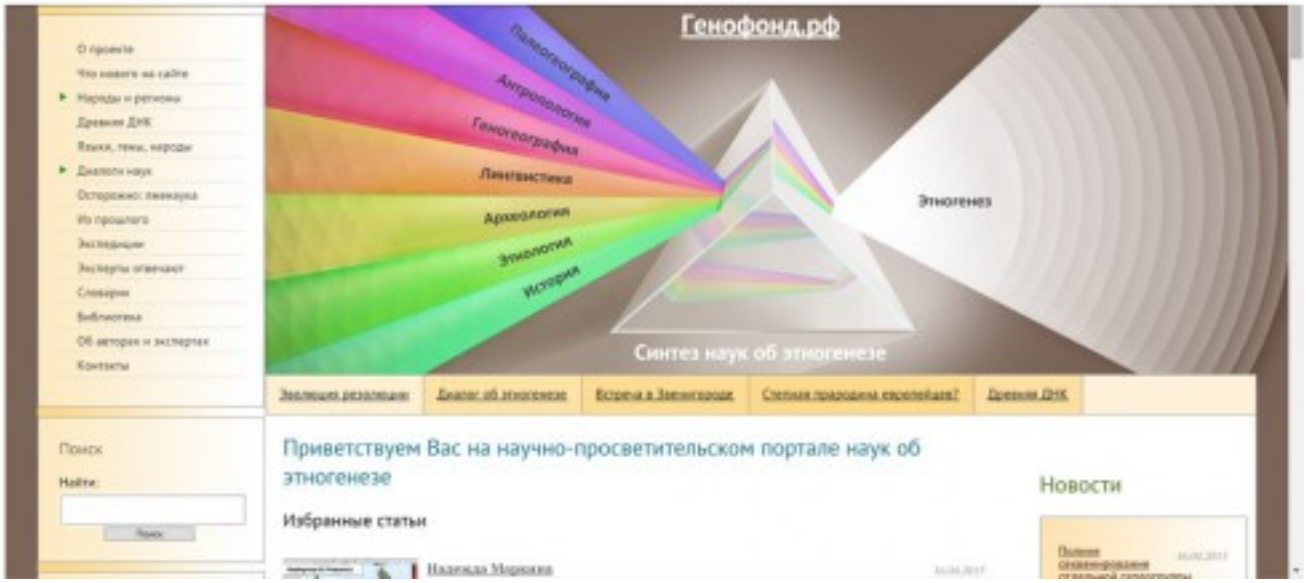


Но чтобы понять, что стоит за словами «24 тысячи образцов», надо сказать, что значит «один образец». Для каждого образца у нас есть, во-первых его генотип – то есть образец крови и выделенной из нее ДНК, с промеренной концентрацией и нередко уже генотипированный. Под фенотипом подразумевается антропологическая фотография и измерение цвета кожи, это позволит в будущем проводить исследования генетического контроля антропологических признаков. Информированное согласие обязательно имеется для каждого образца, как и анкета, указывающая места происхождения и этническую принадлежность предков до третьего поколения. В выборку включались только индивиды, все предки которых вплоть до третьего поколения относили себя к данному этносу и родились в данной популяции.

Биобанк создавался нами в ходе исследований, поддержанных многочисленными международными и российскими фондами. Особенно велика была роль The Genographic project. Но нередко образцы собирались на нашем голом энтузиазме.

Наш новый проект – сайт Генофонд.рф

Ресурс знаний о генофонде и происхождении народов.
Сайт междисциплинарный, поэтому на нем говорят языком, понятным академикам и студентам.



Биобанк предназначен для исследований генофонда, которые могут становиться междисциплинарными. И для усиления диалога представителей разных наук мы создали сайт «генофонд.рф», который является не столько научно-популярным, сколько научным – местом, где лингвисты, археологи, этнографы обсуждают с генетиками этногенез различных народов.



Проект Геногафик до сих пор остается крупнейшим в истории науки популяционно-генетическим проектом. Чтобы проследить расселение человечества по планете, было решено проанализировать мтДНК и Y-хромосому в ста тысячах образцов. На 5 континентах было создано 10 исследовательских центров, и каждому поручено по 10 тысяч образцов. Два центра – наш и китайский – план даже перевыполнили.

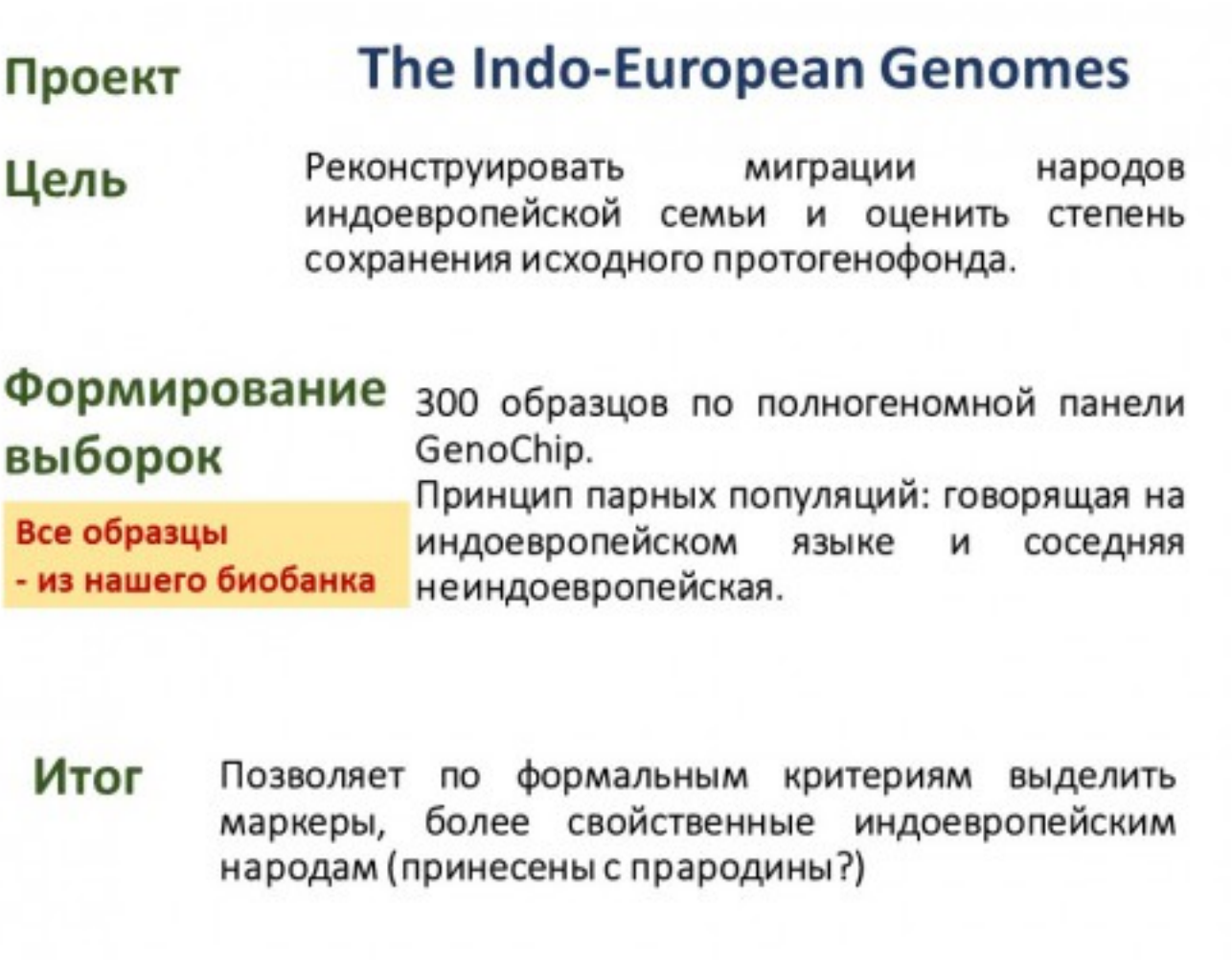
Важно, что в соответствии с целью проекта, выборки формировались по таким принципам:

— максимальное лингвистическое разнообразие – все языковые семьи и ветви,

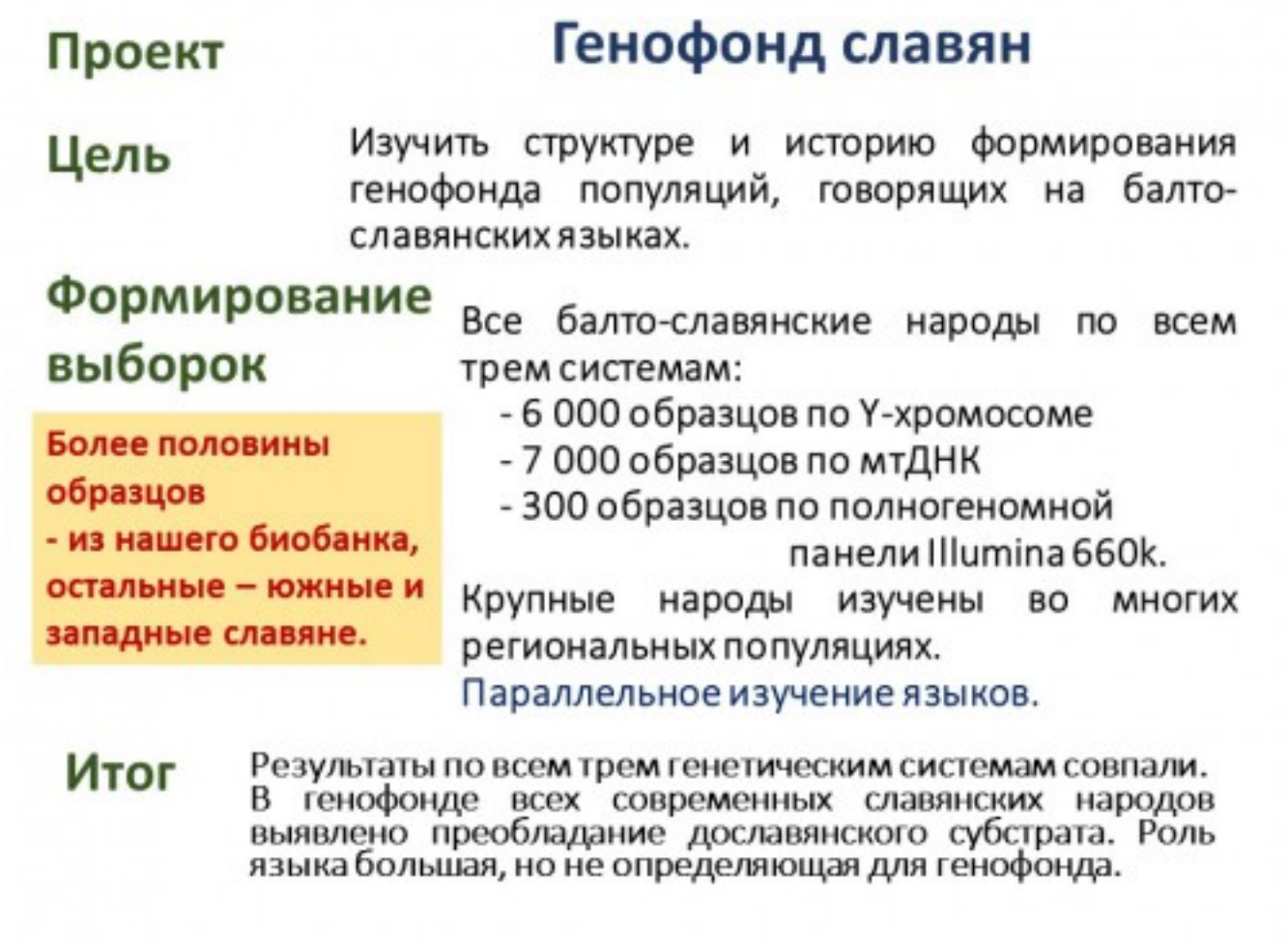
— максимальное антропологическое разнообразие – все антропологические типы,

— максимально широкий географический охват;

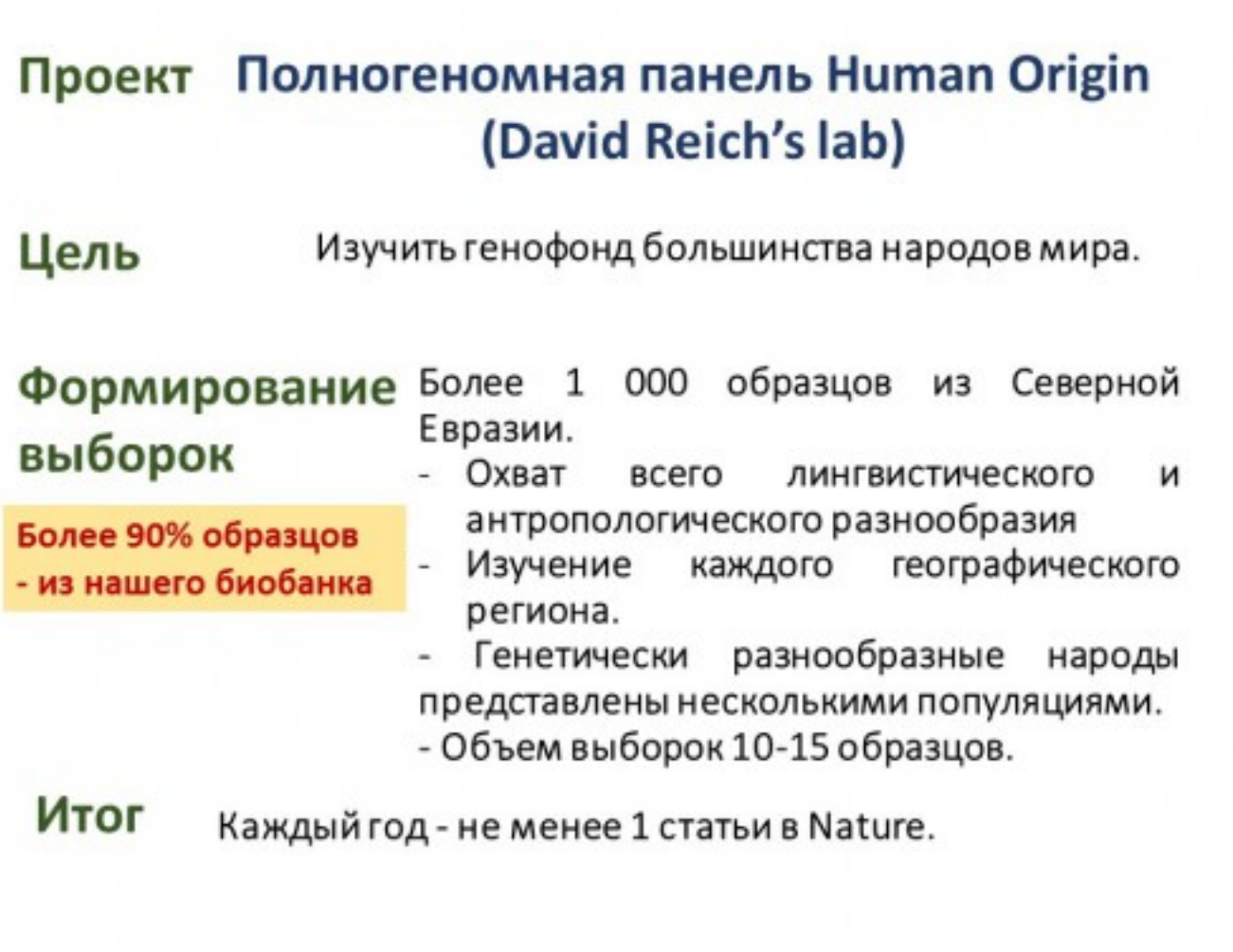
учет генетического разнообразия: все народы изучались хотя бы в двух региональных популяциях, а генетически гетерогенные народы – в пяти, десяти и даже двадцати региональных популяциях. Все эти образцы – из нашего биобанка.



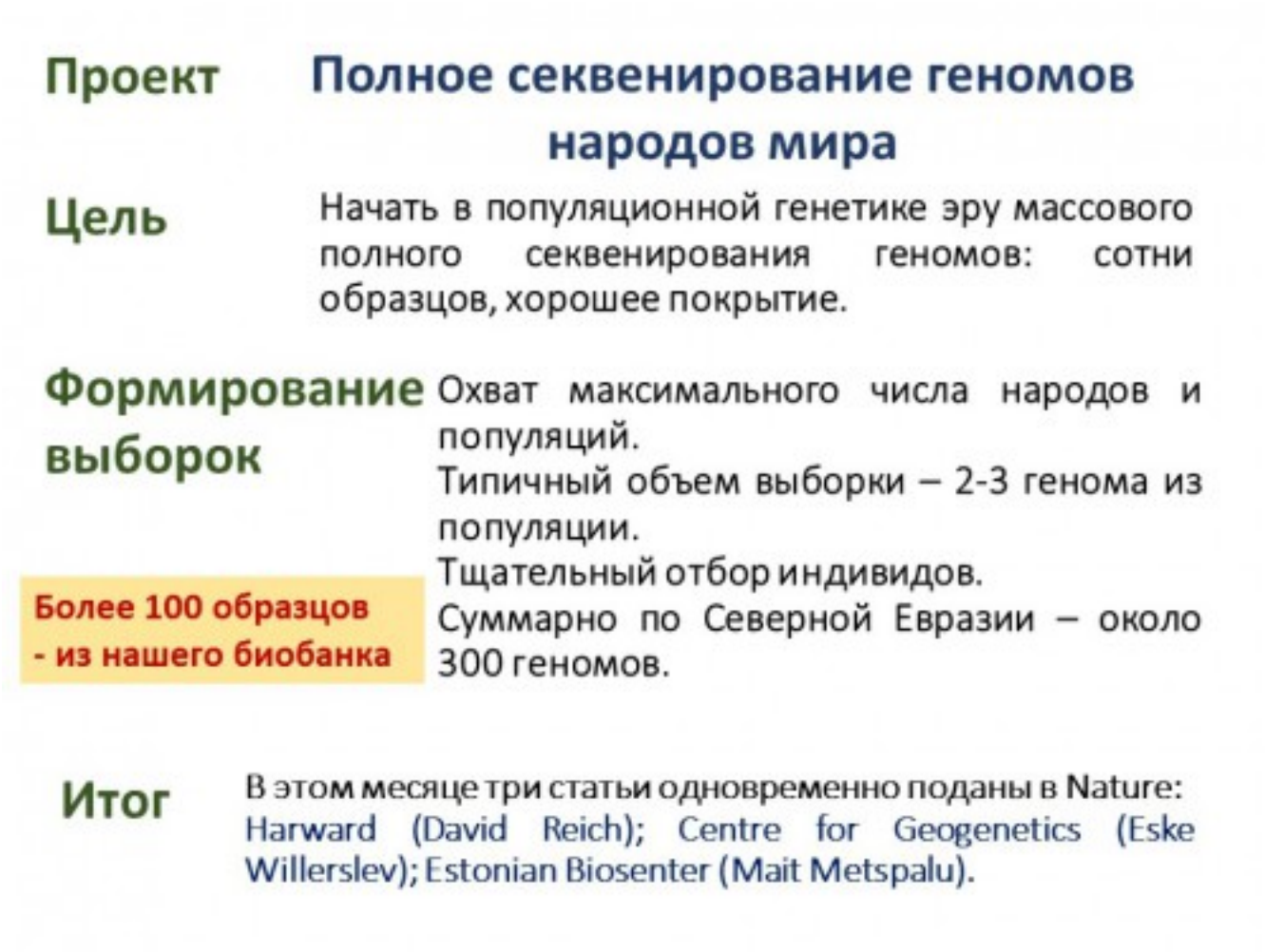
Другой проект, выполняемый под моим руководством – полногеномный анализ индоевропейских народов. Соответственно задаче проекта, сформированы пары популяций – индоевропейская и соседняя ей неиндоевропейская. Это позволяет вычленять генетический компонент, общий для индоевропейцев. Объем каждой выборки – 20 образцов, сформировано 7 пар, суммарно 300 образцов генотипировано по полногеномной панели. Все образцы — из нашего биобанка.



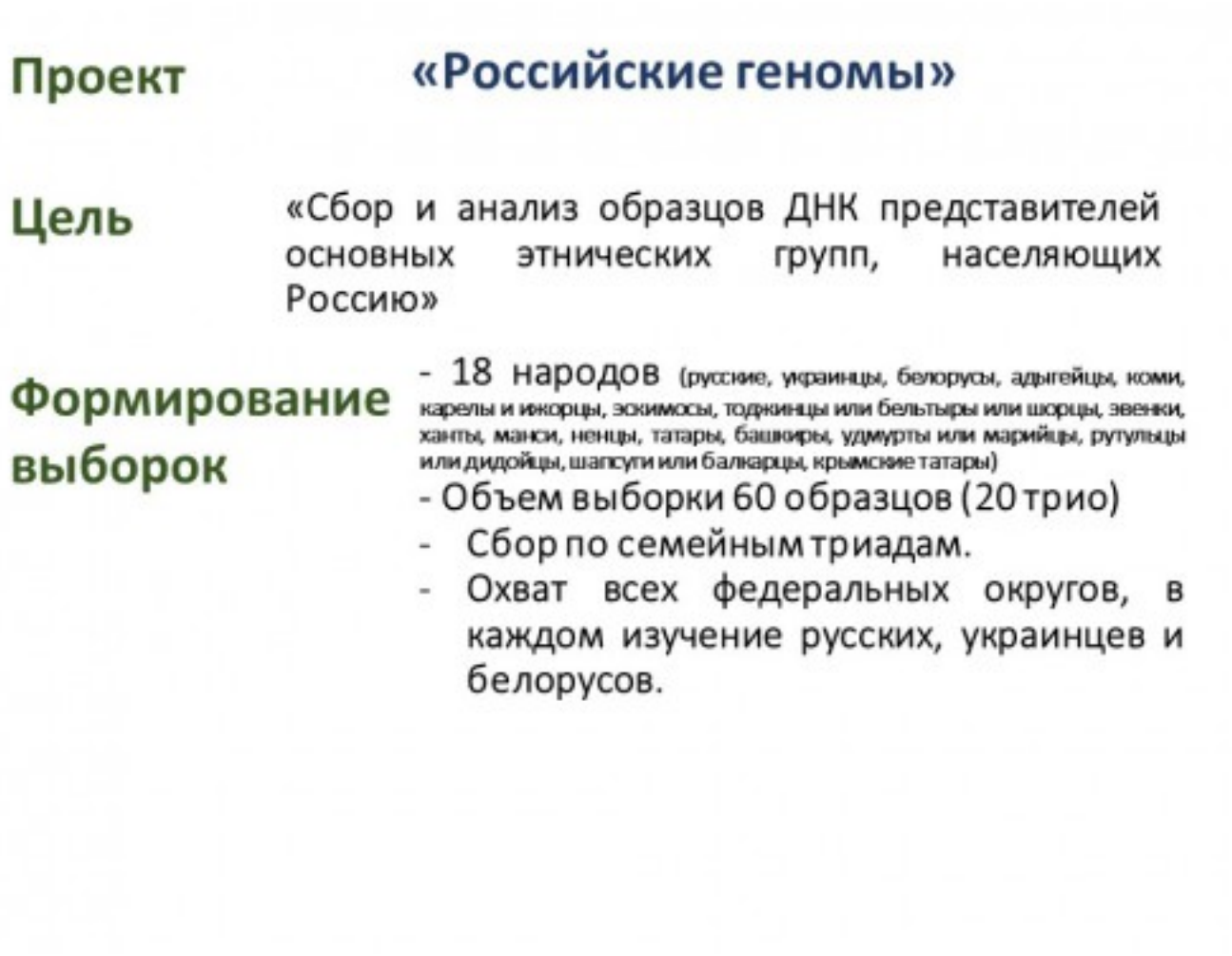
Три недели назад вышла наша обобщающая статья по генофонду славян. В ней каждый народ изучен и по Y-хромосоме (6 тысяч образцов), и по мтДНК (7 тысяч), и по полногеномным панелям (300 образцов). Важно, что изучены все славянские народы, а крупные народы представлены многими региональными популяциями. То есть опять продуманное формирование выборок в начале исследования обеспечило успех в конце его.



Лаборатория Дэвида Райха в Гарварде славится не только тем, что каждый год публикует статьи в Nature, но и тем, что им ни разу не приходилось пересматривать свои выводы. Во многом это вызвано четкой постановкой задач и оптимальным формированием выборок для их решения. Сейчас Дэвид заинтересовался Северной Евразией и генотипирует тысячу образцов. Подавляющее большинство — из нашего биобанка. Отобрали мы их очень тщательно по совокупности пяти критериев: равномерный охват лингвистического, антропологического и генетического разнообразия, охват всех географических регионов, стандартизованный объем выборки из популяции.



Это все были те исследования, которые правильнее называть «широкогеномными». Но проведено и настоящее полное секвенирование, причем не с низким покрытием, как в «1000 геномов», а в среднем с 50-кратным покрытием для сотен геномов. Это исследования трех коллективов – Дэвида Райха в Гарварде, Эске Виллерслева в Дании и Майта Метспалу в Эстонии. На днях эти три статьи одновременно поданы в Nature (мы участвуем в двух из них). И не исключено, что к Рождеству популяционные генетики получат этот роскошный подарок – новые 1000 геномов, но уже секвенированные с хорошим покрытием, равномерно представляющие весь мир и уже тщательно проанализированные ведущими биоинформатиками. **Подчеркну важнейший момент. Раньше почти по всем генетическим системам изучение Северной Евразии отставало от других регионов мира. Но на этот раз дело обстоит иначе: из новых 1000 геномов, охватывающих мир, более 300 представляют Россию и ее сопредельные страны.** В том числе около 100 полных геномов – из нашего биобанка. В ближайшее время одна из статей будет размещена на biorxiv.org, и одновременно ее разбор появится на нашем сайте генофонд.рф.



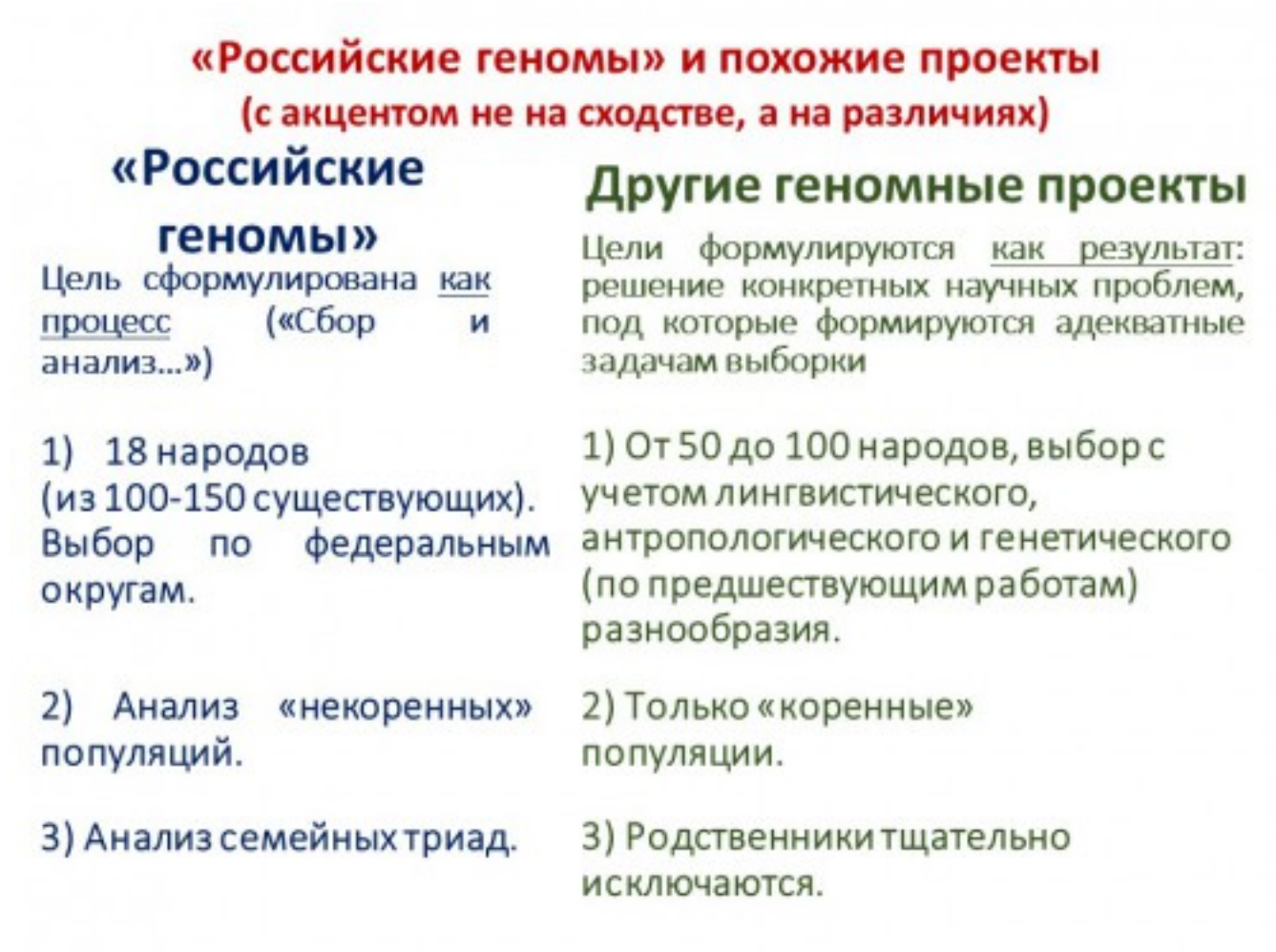
Хотя этот проект единственный, в котором не участвует наш биобанк, для сопоставимости я буду придерживаться того же плана изложения. Я опираюсь на то описание проекта, которое разослано участникам совещания ФАНО 17 июля.

Цель указана как «Сбор и анализ образцов ДНК представителей основных этнических групп, населяющих Россию».

В план включены 18 народов России, причем только для славян — русских, украинцев и белорусов — изучается несколько популяций, а для остальных народов – по одной. Объем каждой выборки – 60 образцов.

Все образцы от семейных триад – отец, мать и ребенок.

Охват идет по федеральным округам, причем в каждом округе изучаются как коренные народы, так и пришлые славянские популяции.



Понятно, что все популяционные геномные проекты похожи – во всех используются те или иные полногеномные технологии, во всех счет идет на сотни образцов. Поэтому я акцентирую внимание не на сходстве, а на отличиях.

Цель обычно формулируется как научная проблема, под оптимальное решение которой собираются образцы. В проекте «Российские геномы» иной подход – сначала сбор и анализ образцов, потом поиск конкретных проблем, которые они могут решать.

Что касается плана выборок, то проект отличается тремя аспектами.

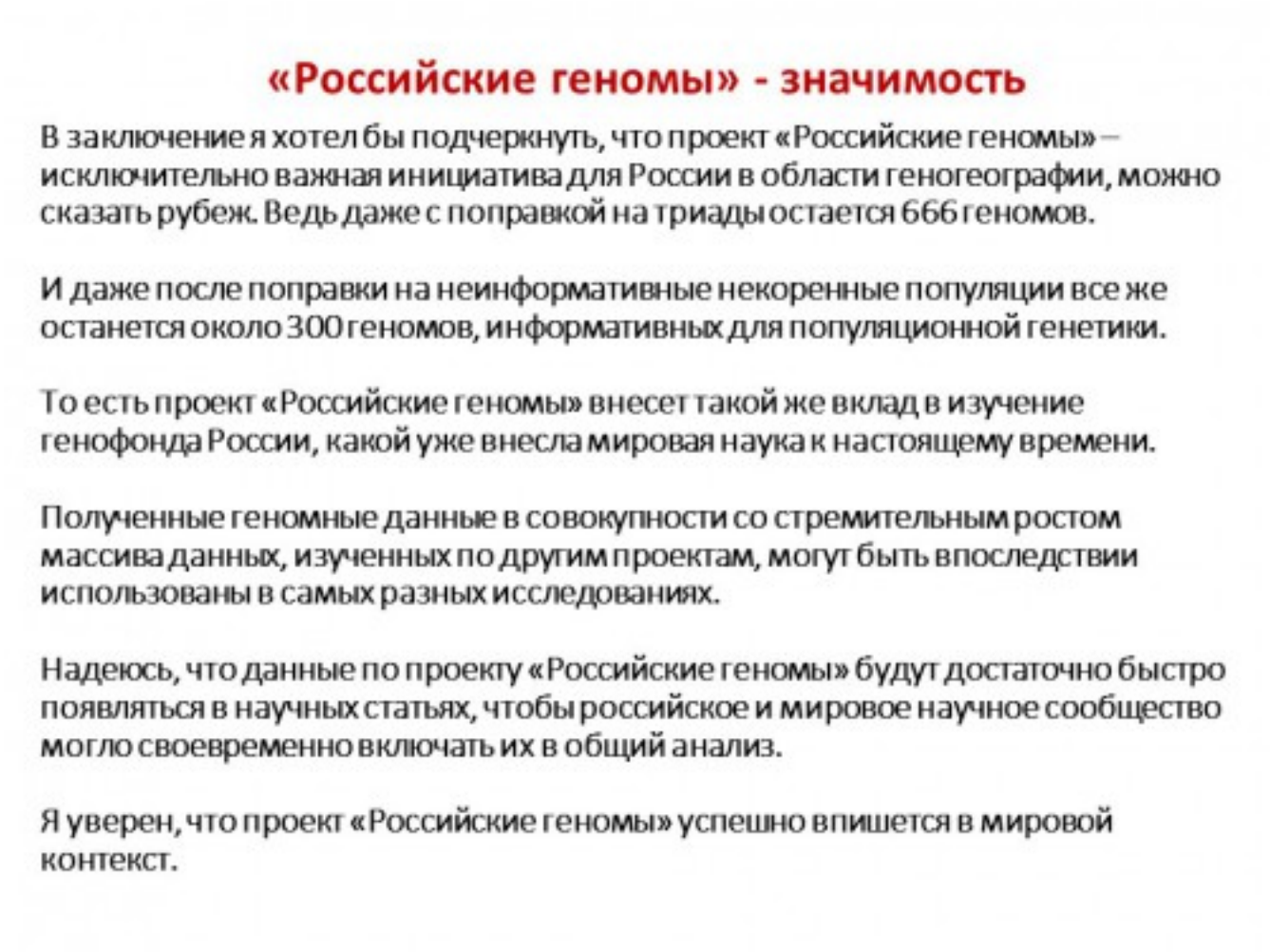
Первое – это ограниченный перечень из восемнадцати народов. Но в России более ста народов, а поскольку и древние этногенетические связи, и современные миграции тесно связывают нас с ближним зарубежьем, то народов еще больше. Другие проекты включают от 50 до 100 народов.

Второе – изучение не коренного населения. Я пользуюсь этим термином для обозначения, например, русских и украинцев Дальнего Востока. Хорошо известно, что их генофонд практически не отличается от русских и украинских популяций на их исторических ареалах в Европейской части. В работе присутствующего здесь Вадима Анатольевича Степанова, в которой и мне довелось участвовать, мы показали, что генофонды русского населения городов Сибири, Урала и Европейской части неотличимы даже по «судебным» маркерам. А в работах здесь присутствующего Михаила Ивановича Воеводы показано, что и по маркерам митохондриальной ДНК население Новосибирска почти не отличается от Европейской части. С тем же столкнулась Ирина Геннадьевна Удина, которая искала, но не нашла различия между русскими и белорусами Москвы и Минска и предложила еще больше увеличивать выборку. Конечно, увеличивая число маркеров вплоть до полных геномов и увеличивая объемы выборок, можно надеяться достичь статистической достоверности очень малых различий между генофондами городского населения. Но на это можно только надеяться, а расход значительных бюджетных средств можно гарантировать.

Третий аспект, цель которого не вполне понятна, это анализ триад. Ведь сложившаяся научная практика такова, что если анализируются гены болезней, то трио (оба родителя и ребенок) подбираются среди больных, и важна нацеленность на конкретные нозологии – а этого проект не предусматривает. Если же анализируются не болезни, а генофонды, то

родственников обязательно исключают. Ведь геном ребенка является копией геномов родителей. Получается, что из 1000 российских геномов каждый третий будет демонстрацией законов Менделя или тестом на отцовство. И на самом деле будет изучено не 1000, а только 666 геномов. Довольно мрачное число.

Это те вопросы, которые я задавал сам себе, знакомясь с проектом «Российские геномы». Те же вопросы я слышал от многих коллег, и подумал, что ими стоит поделиться.



Слайд Заключение.

В заключение я хотел бы подчеркнуть, что проект «Российские геномы» – исключительно важная инициатива для России в области геногеографии. Ведь даже с поправкой на триады остается более 666 геномов. И с поправкой на неинформативные некоренные популяции все-таки останется около 300 геномов, информативных для популяционной генетики. То есть проект «Российские геномы» внесет такой же вклад в изучение генофонда России, как внесла мировая наука к настоящему времени.

Полученные геномные данные в совокупности с растущими данными по другим проектам могут быть использованы в самых разных исследованиях. Надеюсь, что данные по проекту «Российские геномы» будут достаточно быстро появляться в научных статьях, чтобы и российское, и мировое научное сообщество могли их использовать.

Поэтому я не сомневаюсь, что Ваш проект «Российские геномы» успешно впишется в мировой контекст.

Благодарю за внимание.»