

**ГЕНОМ ЧЕЛОВЕКА:
ЧТО ЭТО ТАКОЕ В РЕАЛЬНОСТИ И НА «БУМАГЕ»**

**РУБЦОВ Николай Борисович
Институт цитологии и генетики СО РАН
Новосибирский Государственный Университет
г. Новосибирск**

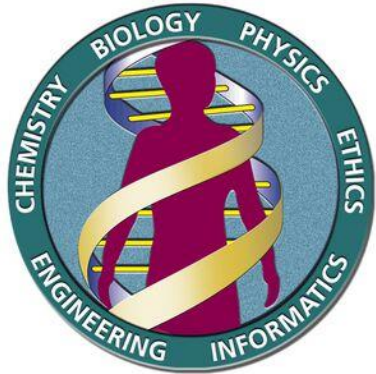
Март 2000 года: Celera объявила о секвенировании генома человека

Федеральный проект «Геном человека» договорился с Celera, и результаты обоих проектов были одновременно объявлены на пресс-конференции 26 июня 2000 года, в которой участвовали президент США Билл Клинтон, премьер-министр Великобритании Тони Блэр .

Проект «Геном человека» - «первоначальное секвенирование генома человека» ЗАВЕРШЁН!

««Стандартный», или референсный геном человека дорабатывается до сих пор. Финальная точка была очень условная. Договорились, что этот момент считать точкой, когда [Клинтон с Блэром] сделали [свое заявление]. В этот момент геном не был сделан до конца, люди потом много лет дочищали это дело».

Гельфанд Михаил



«Проект «Геном человека» (*The Human Genome Project, HGP*) — завершённый международный научно-исследовательский проект, главной целью которого было определение пар оснований, которые составляют ДНК человека, а также выявление, картирование и секвенирование всех генов человеческого генома как с физической, так и с функциональной точки зрения. ... К 2003 году было секвенировано лишь 85 % генома человека, проект был завершён в 2022 году, когда было достигнуто полное секвенирование генома человека».

https://ru.ruwiki.ru/wiki/%D0%9F%D1%80%D0%BE%D0%B5%D0%BA%D1%82_%C2%AB%D0%93%D0%B5%D0%BD%D0%BE%D0%BC_%D1%87%D0%B5%D0%BB%D0%BE%D0%B2%D0%B5%D0%BA%D0%B0%C2%BB

1986 год - начало работы над проектом, позднее названным «Геном человека».



*Александр Александрович
Баев*

28.12.1903 – 31.12.1994.

Программа «Геном человека» в СССР и РФ:

- Программа стартовала в 1987 г. Ее инициатор и безусловный лидером в те годы - академик А.А. Баев. В 1989 г. программа стала одной из ведущих Государственных научно-технических программ СССР.

Основные разделы этой программы как в России, так и во всем мире включают три главных направления научных исследований :

- 1) картирование и секвенирование генома;
- 2) структурно-функциональное изучение генома;
- 3) медицинскую генетику и генотерапию.

- В 90е гг работы по международной программе Геном Человека в СССР, а затем в РФ были сосредоточены главным образом на изучении 3-й, 13-й и 19-й хромосом. Подход к изучению этих хромосом очень различен.

1990 год: формальный запуск проекта министерством энергетики США и Национальным институтом здравоохранения (3×10^9 \$, ожидаемая длительность 15 лет). Участники международного консорциума помимо США: Китай, Франции, Германии, Японии и Великобритании, СССР.

Джеймс Уотсон возглавлял Национальный центр исследований человеческого генома в Национальной организации здравоохранения США (NIH) с 1988 года. В 1992 году ушёл в отставку из-за несогласия с позицией Бернадины Хили по вопросам патентования генов.

С апреля 1993 руководил Френсис Коллинз

В 1997 году название центра было изменено на Национальный институт исследований человеческого генома (NHGRI).



«Я не хочу положить свою жизнь на то, чтобы определить последовательность 12-й хромосомы от 100 000-й до 200 000-й пары оснований».

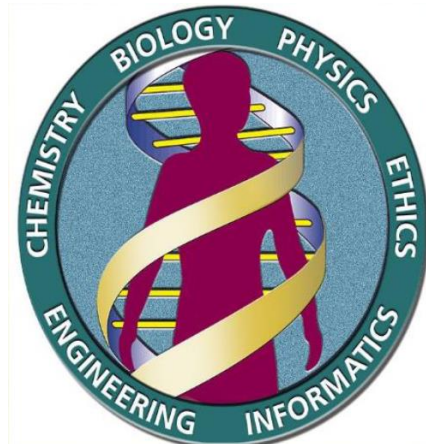
1992, Вентер Крейг организовал Institute for Genomic Research, некоммерческий научно-исследовательский институт, где в 1995 году был секвенирован первый живой организм *Haemophilus influenzae*, с использованием новой технологии полногеномного секвенирования (**shotgun technique**).

1998, Вентер Крейг организовал Celera Genomics с целью секвенирования генома человека, используя новую технологию.

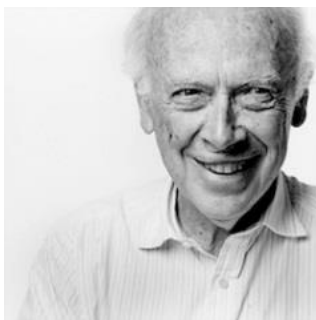
В марте 2000 Celera Genomics сделала заявление о секвенировании генома человека.

26 июня 2000 года Крейг Вентер и Фрэнсис Коллин, руководитель проекта «Геном человека» объявили «первой сборке генома человека».

В феврале 2001 года, опубликован первый предварительный драфт генома человека.



Первые персональные геномы:



Джеймс Уотсон
2007
\$1.5 миллиона



Стив Джобс



Гордон Мур



Гленн Клоуз



Оззи Осборн



Дэн Стоикеску
2008 год
\$350 000



Вентер Крейг

Секвенирование по Сэнгеру — "золотой стандарт".

Разработка методов массового параллельного секвенирования:

Секвенирование второго поколения.

Секвенирование путем гибридизации.

Секвенирование путем синтеза — Illumina.

Секвенирование путем синтеза — GenoLab M.

Секвенирование с использованием наночастиц BGI.

Секвенирование третьего поколения.

Нанопоровое секвенирование

RacBio

Основные технические проблемы сборки:

гетерозиготность

дупликация генома, районов генома, её следы

генетическое разнообразие

Technology	Reads	Highlights	Limitations
PacBio CLR	Single molecule long reads, average length ~ 20 kb	Long-read sequencing technology widely used to assemble genome in the past years	The high base error requires correction with high coverage or with short reads, also may introduce switch error into polyploid assembly
PacBio CCS (HiFi)	Single molecule long reads, circular sequencing, and consensus by multi passes, with average length ~ 15–20 kb	Generate long reads that have high base accuracy (>99%) and long length, reduced need for error correction, and lower sequencing coverage	Reads length only reach ~20 Kbps, therefore sacrificing length for accuracy
Oxford Nanopore	Single-molecule long reads, directly detected by base signal, with average length ~10 Kbps, ultra-long reads can reach 2.3 Mbps in the current record	Direct sequencing by measuring electric base signal without limited by enzyme activity	High base error and occurrence of homopolymer errors Requirement to prepare long-length DNA molecules
BioNano Genomics	Optical mapping of DNA fingerprint-labeled fluorescently in enzyme sites, with average length up to multiple hundred kilobases	Produce long-range linkage information for all chromosome	Limitations in sparse enzyme site regions and pool algorithms for integrating optical mapping and genome assembly
10X Genomics	Short reads from Illumina paired-end sequencing, but linked reads through GEMs, which provide linkage information with average length reach to 100 Kbps	Experimental methods to construct a microenvironment to sequencing a single molecule DNA fragment, capable of linked reads over long-range and ability to sequence using mainstream Illumina sequencers	Pool algorithms available for polyploid, and it is based on NGS, therefore inherits certain limitations in repetitive regions, especially for local tandem repeats where the labels are the same (i.e., unable to differentiate between tandem copies)
Hi-C	Derived short reads from Illumina paired-end sequencing can provide linkage information up to 100 Kbps for chromosomes based on proximity ligation	Methods can detect the chromosome interaction, can provide the whole-genome contact information	Requires (and sensitive to) a high-quality input haplotig set To achieve better results, chimeric contigs to be removed from the input

Сборки генома до GRCh38.p13 и T2T-CHM13

Финальная сборка генома человека T2T-CHM13?

Консорциум «От теломеры до теломеры» (T2T-Consortium, теломера — концевой участок хромосомы: 54 института и лабораторий из разных стран, включая Россию. Результат - первая действительно полная сборка генома, шесть статей в *Science*.

В T2T-CHM13 **3 054 815 472** пар нуклеотидов ядерной ДНК и **16 569** пар митохондриальной ДНК. **182 миллиона** пар нуклеотидов из новой сборки отсутствуют в предыдущих сборках.

Новая сборка генома получила название от культуры клеток CHM13, из которой была выделена ДНК (клетки культуры, полученной из опухоли, возникшей в результате пузырного заноса — оплодотворенная яйцеклетка потеряла материнские хромосомы, сохранила только отцовские).

22 аутосомы и X-хромосома (последующая диплоидизация) — геном одного из сперматозоидов, возможно, претерпевший некоторые изменения.

Решение технической проблемы гетерозиготности - различий ДНК в копиях гомологичных хромосом.

Результат:

Полная сборка без пробелов.

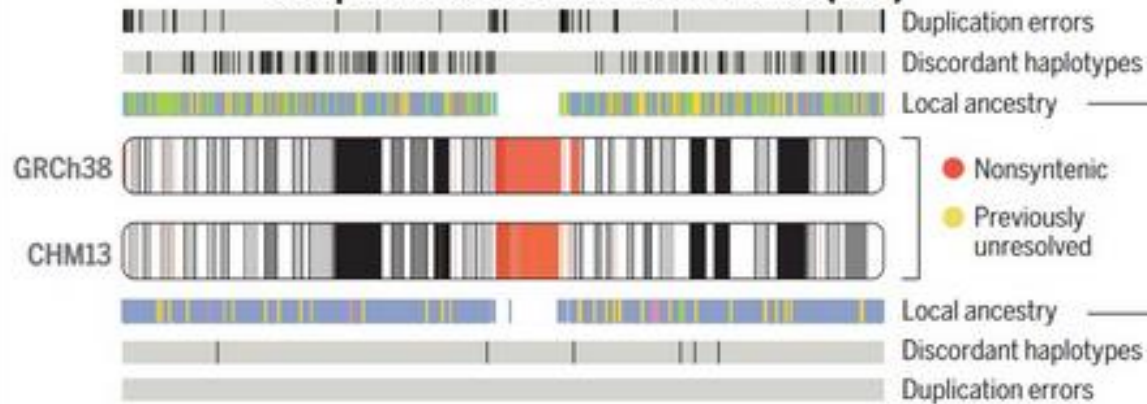
Завершен проект Геном человека.

Вопрос: может ли существовать человек с таким геномом?

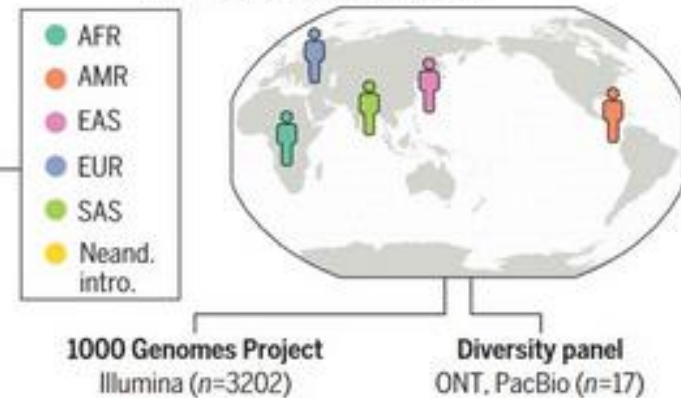
Новая сборка позволила исправить множество ошибок, присутствующих в GRCh38.p13, добавила к сиквенсу еще 200 миллионов пар оснований, содержащих 2 226 генов-паралогов, среди них 115, предположительно кодирующих белки. Основным недостатком T2T-CHM13 — это отсутствие Y-хромосомы (Y-хромосома добавлена из другой сборки).

STATISTICS	GRCH38	T2T-CHM13	DIFFERENCE (±%)
Summary			
Assembled bases (Gbp)	2.92	3.05	+4.5
Unplaced bases (Mbp)	11.42	0	-100.0
Gap bases (Mbp)	120.31	0	-100.0
Number of contigs	949	24	-97.5
Contig NG50 (Mbp)	56.41	154.26	+173.5
Number of issues	230	46	-80.0
Issues (Mbp)	230.43	8.18	-96.5
Gene annotation			
Number of genes	60,090	63,494	+5.7
Protein coding	19,890	19,969	+0.4
Number of exclusive genes	263	3,604	
Protein coding	63	140	
Number of transcripts	228,597	233,615	+2.2
Protein coding	84,277	86,245	+2.3
Number of exclusive transcripts	1,708	6,693	
Protein coding	829	2,780	

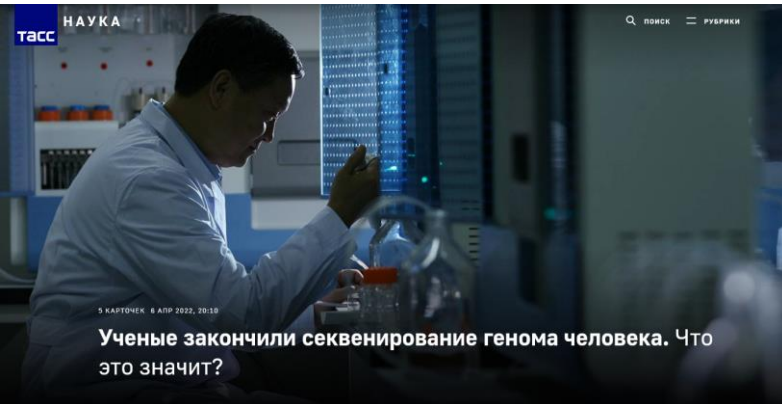
Comparison of GRCh38 and T2T-CHM13 (chr1)



Global genetic diversity



Сборка генома человека T2T-CHM13



Исследовательская группа Telomere-to-Telomere опубликовала отчет о расшифровке последних 8% человеческого генома. Хотя первый вариант полного генома был получен в 2003 году, эту работу удалось закончить только сейчас. Рассказываем, почему так получилось и что удалось узнать ученым

Это действительно сборка генома человека?

Earth's heart of iron begins to yield its secrets p. 18

Microglia in chronic pain recovery and relapse pp. 33 & 86

Particle acceleration in a nova explosion p. 77

Science

\$15
1 APRIL 2022
SPECIAL ISSUE
science.org

AAAS

FILLING THE GAPS

Closing in on a complete human genome p. 42



scientists have come to realize that a single reference genome is inadequate for many purposes.

учёные пришли к пониманию, что один референсный геном не годится для решения многих задач.

While the use of a single reference has advanced genetics immensely, it has not, as some had hoped, allowed us to find the cause of all genetic disease, a shortcoming that has prompted some commentators to call the Human Genome Project a failure.

Хотя использование одной референсной сборки гаплоидного генома значительно продвинуло генетику, оно не позволило нам, как надеялись, найти причины генетических заболеваний, этот недостаток побудил некоторых учёных назвать проект «Геном человека» провалом.

«Regardless of what methods are chosen, it is now clear that the community must move beyond reliance on a single reference genome».

Несмотря на то, какой метод (секвенирования) был выбран, становится ясно, что сообщество должно двигаться за границы опоры на один референсный геном

Что мы знаем о геноме человека?

Международный проект (HapMap) (старт в 2002 году). Целью каталогизировать индивидуальные генетические особенности людей. Публикация 2-го сентября 2010 года карты вариаций генома человека третьего поколения. Частота SNP ~ 1 на 279 bp. (2010 г.)

1000 Genomes Project (1KGP), старт в 2008 году, завершение в 2015, *Nature volume 526, pages 68–74, 01 October 2015*) реконструированы геномы 2504 человек из 26 популяций, выявлено более 88 миллионов различных вариаций (**84.7 миллионов SNPs, 3.6 миллионов indels и 60000 структурных вариантов**). Частота SNP >1% (1 на ~30 bp).

100,000 Genomes Project (старт в Великобритании в 2013 году). В Англии к 1 октября 2018 года были секвенированы 87 231 геномов.

Вопросы к качеству секвенирования и качеству сборок

ARTICLE

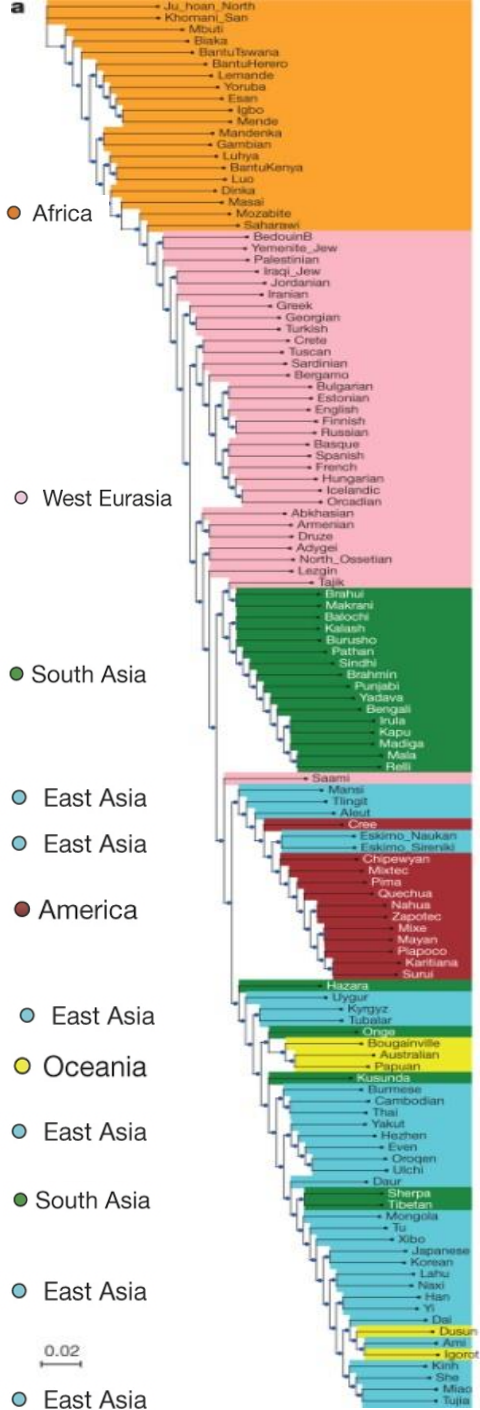
doi:10.1038/nature18964

The Simons Genome Diversity Project: 300 genomes from 142 diverse populations

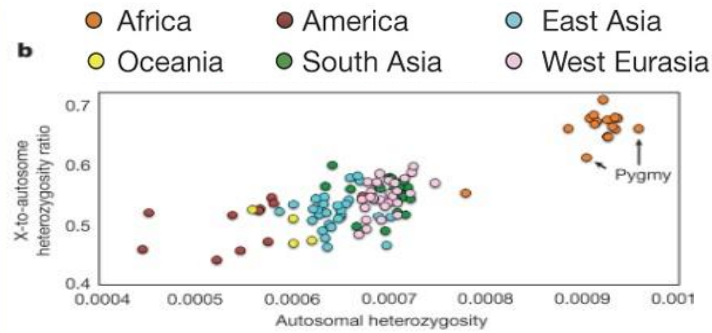
A list of authors and affiliations appears at the end of the paper.

Here we report the Simons Genome Diversity Project data set: high quality genomes from 300 individuals from 142 diverse populations. These genomes include at least 5.8 million base pairs that are not present in the human reference genome. Our analysis reveals key features of the landscape of human genome variation, including that the rate of accumulation of mutations has accelerated by about 5% in non-Africans compared to Africans since divergence. We show that the ancestors of some pairs of present-day human populations were substantially separated by 100,000 years ago, well before the archaeologically attested onset of behavioural modernity. We also demonstrate that indigenous Australians, New Guineans and Andamanese do not derive substantial ancestry from an early dispersal of modern humans; instead, their modern human ancestry is consistent with coming from the same source as that of other non-Africans.

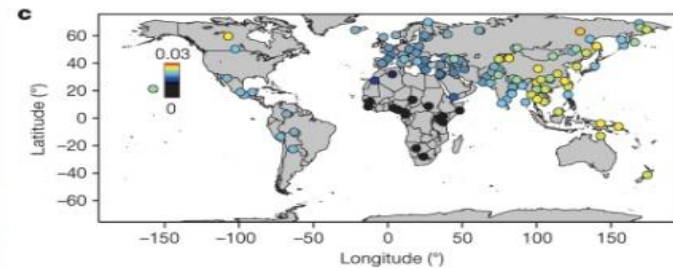
Всего 300 персональных геномов, но из 142 далёких популяций. **Новое качество секвенирования:** библиотеки приготовлены без использования ПЦР (278 из 300), в среднем 43 кратное покрытие генома (от 34 до 83). Выявлено 34.4×10^6 SNPs и 2.1×10^6 indels.



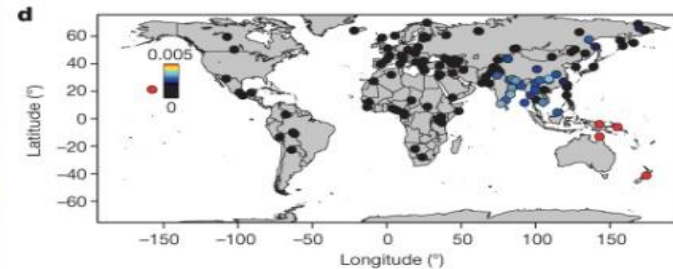
a, Дерево, построенное на анализе дивергенции при попарных сравнениях



b, Отношение аутосомной гетерозиготности к X/ аутосомной гетерозиготности показало его уменьшение у неафриканских популяций и пигмеев.



c, Оценка вклада генома неандертальцев (тепловая карта 0–3%).



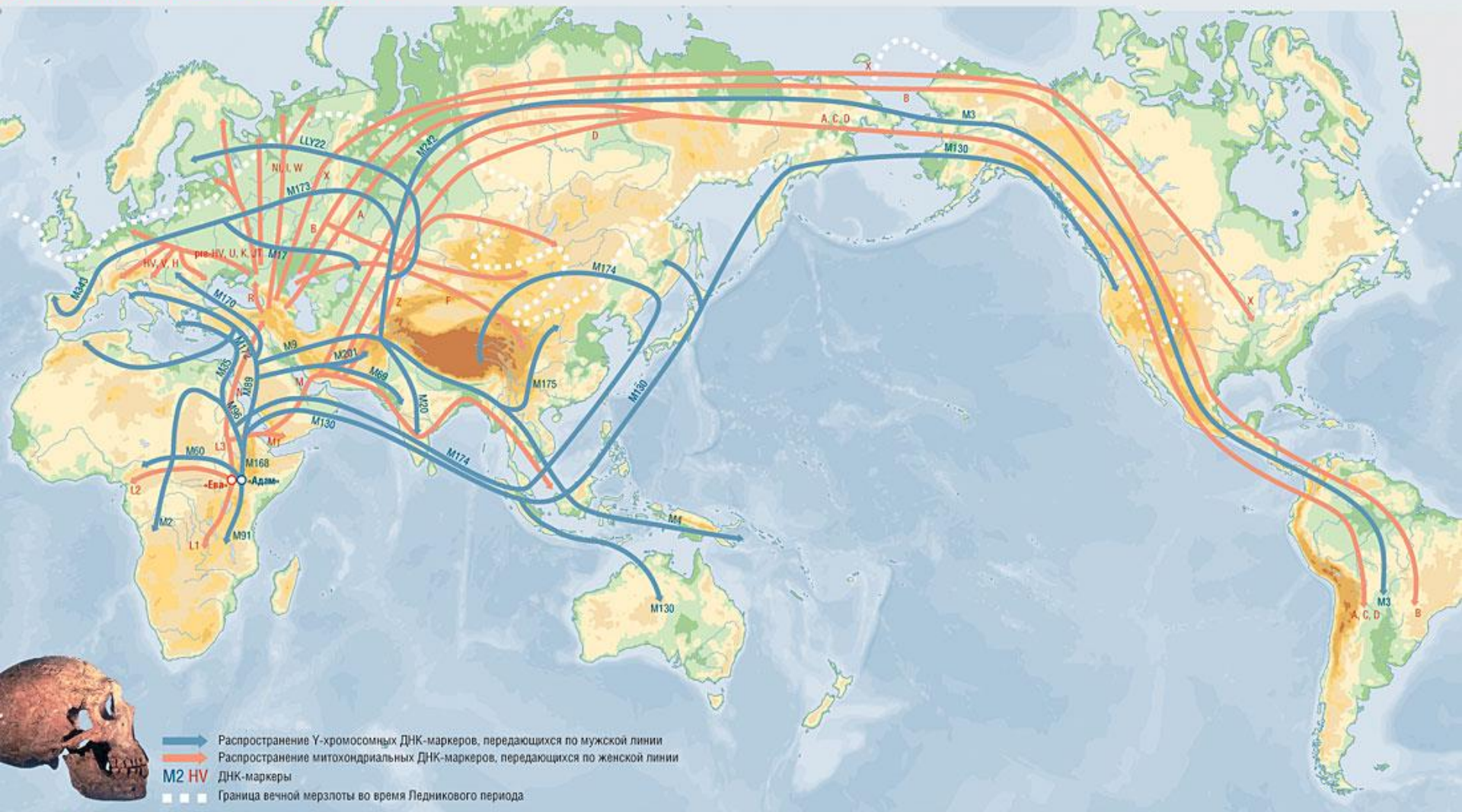
d, Оценка вклада генома денисовцев (тепловая карта 0–0,5%). В группе осеан до 5% (яркий красный).

S Mallick *et al.* *Nature* 1–6 (2016) doi:10.1038/nature18964

nature

Эволюция и генетическое разнообразие человека

Происхождение человека из Африки



Неандерталлы, Денисовцы и другие?

A Draft Sequence of the Neandertal Genome

Richard E. Green,^{1,††} Johannes Krause,^{1,†§} Adrian W. Briggs,^{1,†§} Tomislav Maricic,^{1,†§} Udo Stenzel,^{1,†§} Martin Kircher,^{1,†§} Nick Patterson,^{2,†§} Heng Li,^{2,†||} Weimei Zhai,^{3,†||} Markus Hsi-Yang Fritz,^{4,†} Nancy F. Hansen,^{5,†} Eric Y. Durand,^{3,†} Anna-Sapfo Malaspinas,^{3,†} Jeffrey D. Jensen,^{6,†} Tomas Marques-Bonet,^{7,13,†} Can Alkan,^{7,†} Kay Prüfer,^{1,†} Matthias Meyer,^{1,†} Hernán A. Burbano,^{7,†} Jeffrey M. Good,^{1,8,†} Rigo Schultz,² Aynuer Aximu-Petri,¹ Anne Butthof,¹ Barbara Höber,¹ Barbara Höffner,¹ Madlen Siegemund,¹ Antje Weihmann,¹ Chad Nusbaum,² Eric S. Lander,² Carsten Russ,² Nathaniel Novod,² Jason Affourtit,² Michael Egholm,² Christine Verha,^{2,1} Pavao Rudan,¹⁰ Dejana Brajkovic,¹¹ Zeljko Kucan,¹⁰ Ivan Gušić,¹⁰ Vladimir B. Doronichev,¹² Liubov V. Golovanova,¹² Carles Lalueza-Fox,¹³ Marco de la Rasilla,¹⁴ Javier Fortea,^{14,¶} Antonio Rosas,¹⁵ Ralf W. Schmitz,^{16,17} Philip L. F. Johnson,^{18,†} Evan E. Eichler,^{7,†} Daniel Falush,^{19,†} Ewan Birney,^{4,†} James C. Mullikin,^{2,†} Montgomery Slatkin,^{3,†} Rasmus Nielsen,^{3,†} Janet Kelso,^{3,†} Michael Lachmann,^{1,†} David Reich,^{2,20,†} Svante Pääbo^{14,†}

Neandertals, the closest evolutionary relatives of present-day humans, lived in large parts of Europe and western Asia before disappearing 30,000 years ago. We present a draft sequence of the Neandertal genome composed of more than 4 billion nucleotides from three individuals. Comparisons of the Neandertal genome to the genomes of five present-day humans from different parts of the world identify a number of genomic regions that may have been affected by positive selection in ancestral modern humans, including genes involved in metabolism and in cognitive and skeletal development. We show that Neandertals shared more genetic variants with present-day humans in Eurasia than with present-day humans in sub-Saharan Africa, suggesting that gene flow from Neandertals into the ancestors of non-Africans occurred before the divergence of Eurasian groups from each other.

The morphological features typical of Neandertals first appear in the European fossil record about 400,000 years ago (1–3). Progressively more distinctive Neandertal forms subsequently evolved until Neandertals disappeared from the fossil record about 30,000 years ago (4). During the later part of their history, Neandertals lived in Europe and Western Asia as far east as Southern Siberia (5) and as far south as the Middle East. During that time, Neandertals presumably came into contact with anatomically modern humans in the Middle East from at least 80,000 years ago (6, 7) and subsequently in Europe and Asia.

sumed ancestors of present-day Europeans. Similarly, analysis of DNA sequence data from present-day humans has been interpreted as evidence both for (12, 13) and against (14) a genetic contribution by Neandertals to present-day humans. The only part of the genome that has been examined from multiple Neandertals, the mitochondrial DNA (mtDNA) genome, consistently falls outside the variation found in present-day humans and thus provides no evidence for interbreeding (15–19). However, this observation does not preclude some amount of interbreeding (14, 19) or the possibility that Neandertals contributed other parts of their genomes to present-

changed parts of their genome with the ancestors of these groups.

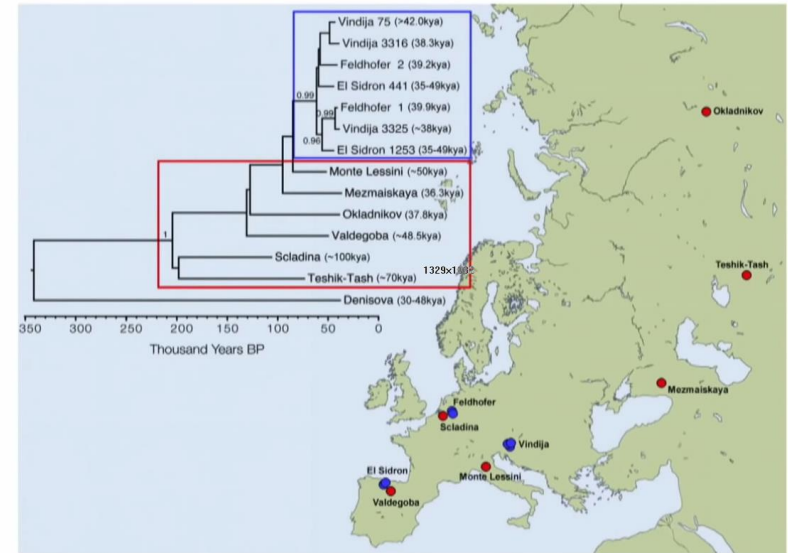
Several features of DNA extracted from Late Pleistocene remains make its study challenging. The DNA is invariably degraded to a small average size of less than 200 base pairs (bp) (21, 22), it is chemically modified (21, 23–26), and extracts almost always contain only small amounts of endogenous DNA but large amounts of DNA from microbial organisms that colonized the specimens after death. Over the past 20 years, methods for ancient DNA retrieval have been developed (21, 22), largely based on the polymerase chain reaction (PCR) (27). In the case of the nuclear genome of Neandertals, four short gene sequences have been determined by PCR: fragments of the *MC1R* gene involved in skin pigmentation (28), a segment of the *FOXP2* gene involved in speech and language (29), parts of the ABO blood group locus (30), and a taste receptor gene (31). However, although PCR of ancient DNA can be multiplexed (32), it does not allow the retrieval of a large proportion of the genome of an organism.

The development of high-throughput DNA sequencing technologies (33, 34) allows large-scale, genome-wide sequencing of random pieces of DNA extracted from ancient specimens (35–37) and has recently made it feasible to sequence ge-

¹Department of Evolutionary Genetics, Max-Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany. ²Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. ³Department of Integrative Biology, University of California, Berkeley, CA 94720, USA. ⁴European Molecular Biology Laboratory—European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK. ⁵Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA. ⁶Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01655, USA. ⁷Howard Hughes Medical Institute, Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA. ⁸Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA. ⁹454 Life Sciences, Branford, CT 06405, USA. ¹⁰Croatian Academy of Sciences and Arts, Zrinski trg 11, HR-10000 Zagreb, Croatia. ¹¹Croatian Academy of Sciences and Arts, Institute for Quaternary Paleontology and Geology, Ante Kovacic 5, HR-10000 Zagreb.

Downloaded from www.sciencemag.org on May 6, 2010

Bottleneck в Европе >48 тыс. лет назад



ЧТО МЫ ВИДИМ ГЛЯДЯ

Love Dalén et al. Mol Biol Evol 2012;29:1893-1897



- красный – азиаты
- синий – европейцы

Денисова пещера

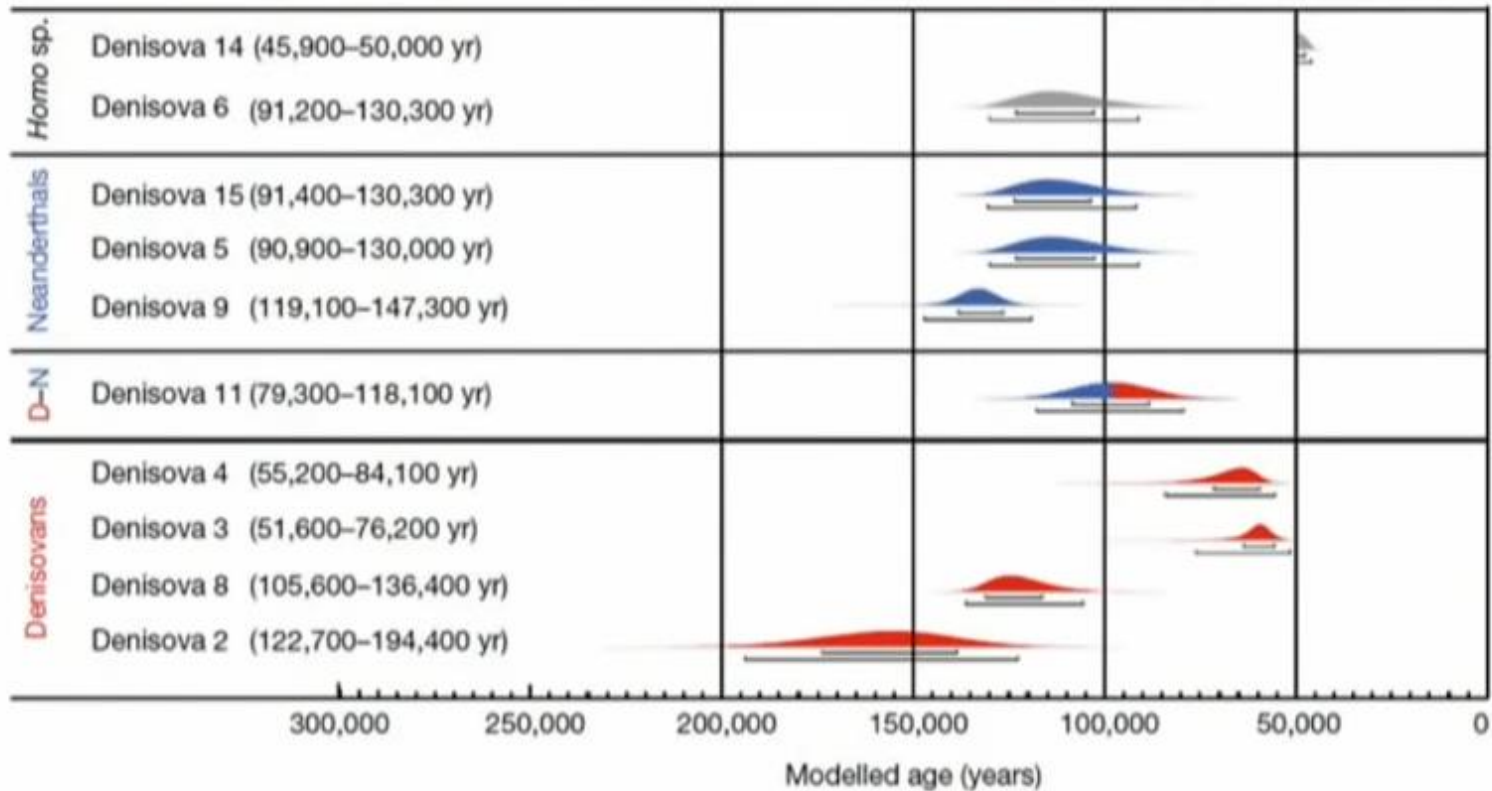


2010 год – геном денисовца
отсеквенирован со средним покрытием 31
(99,4% нуклеотидов прочтено не менее
10 раз, 92,9% — не менее 20)

Денисовец:

1984 год - зуб «Денисова 2»;
2000 год -зуб «Денисова 4»;
2010 году -зуб «Денисова 8»;
2008 год - кость последней
фаланги пальца руки
ребёнка «Денисова 3»;
2019 год – последняя 5-я
находка

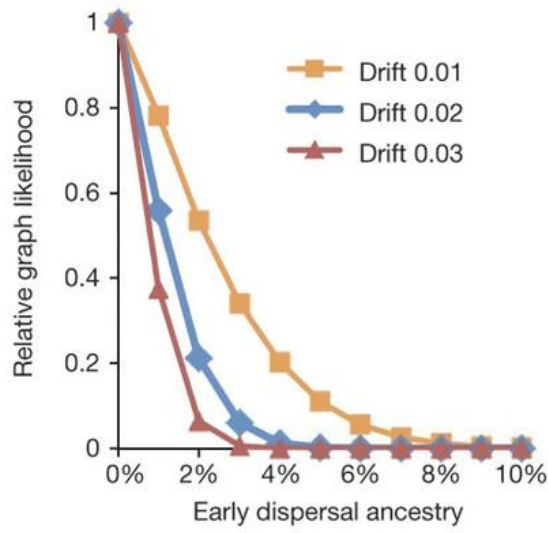
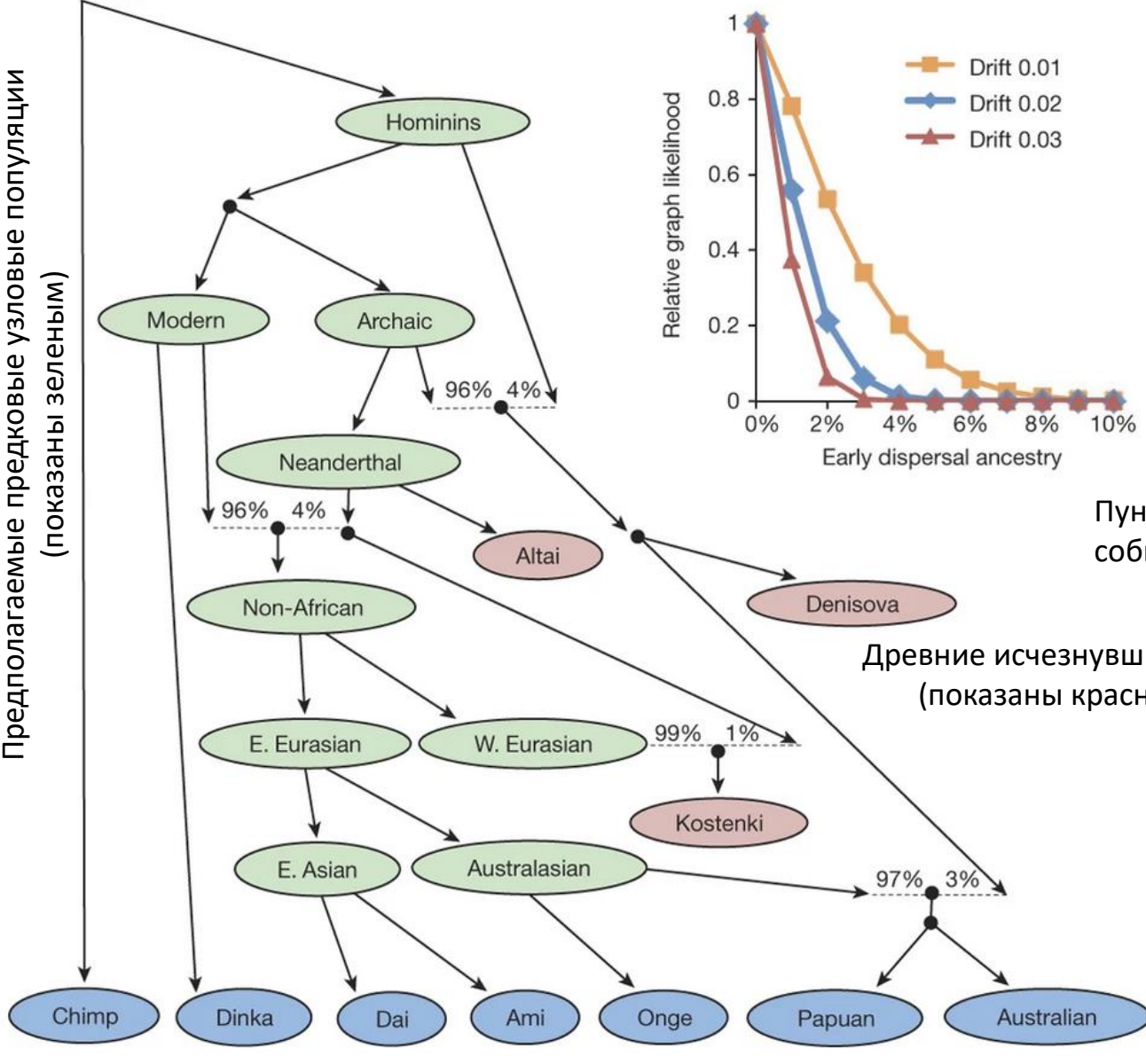
Коммунальная пещера



K.Douka et al. (2019)

nature

Предполагаемые предковые узловые популяции (показаны зеленым)



Пунктирные линии показывают события смешивания геномов

Древние исчезнувшие виды (показаны красным)

Современные популяции (показаны синим)

S Mallick *et al.* *Nature* 1–6 (2016) doi:10.1038/nature18964



Что делать с данными по секвенированию генома человека?

Sherman, R. M. et al. Assembly of a pan-genome from deep sequencing of 910 humans of African descent. *Nat. Genet.* 51, 30–35 (2019).

Более **300 Mb** новых последовательностей найдено в геномах индивидов, имеющих африканское происхождение. Значительная часть генома отсутствует в геномных сборках.

Что делать с данными по секвенированию генома человека?

Сборка гаплоидного генома или пан-генома человека

- Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. ***Nature* 538**, 201–206 (2016).
- Telenti, A. et al. Deep sequencing of 10,000 human genomes. ***Proc. Natl Acad. Sci. USA* 113**, 11901–11906 (2016).
- 1000 Genomes Project Consortium et al. A global reference for human genetic variation. ***Nature* 526**, 68–74 (2015).
- Sudmant, P. H. et al. An integrated map of structural variation in 2,504 human genomes. ***Nature* 526**, 75–81 (2015).

ISCN 2020



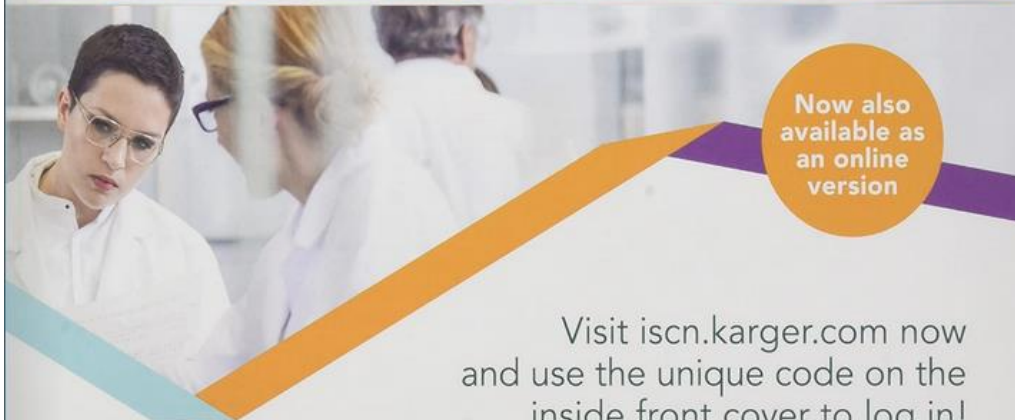
An International System for
Human Cytogenomic Nomenclature (2020)

Editors

Jean McGowan-Jordan

Ros J. Hastings



Sarah Moore



Now also
available as
an online
version

Visit iscn.karger.com now
and use the unique code on the
inside front cover to log in!

Pan-genomics in the human genome era

Rachel M. Sherman ^{1,2}✉ and Steven L. Salzberg ^{1,2,3,4}

Abstract | Since the early days of the genome era, the scientific community has relied on a single 'reference' genome for each species, which is used as the basis for a wide range of genetic analyses, including studies of variation within and across species. As sequencing costs have dropped, thousands of new genomes have been sequenced, and scientists have come to realize that a single reference genome is inadequate for many purposes. By sampling a diverse set of individuals, one can begin to assemble a pan-genome: a collection of all the DNA sequences that occur in a species. Here we review efforts to create pan-genomes for a range of species, from bacteria to humans, and we further consider the computational methods that have been proposed in order to capture, interpret and compare pan-genome data. As scientists continue to survey and catalogue the genomic variation across human populations and begin to assemble a human pan-genome, these efforts will increase our power to connect variation to human diversity, disease and beyond.

Геном человека слишком большой, слишком большое разнообразие.

Облегченные варианты пан-генома:

Genic pan-genome;

Pan-transcriptome;

Population-specific pan-genomes.

Недавно National Human Genome Research Institute запустил проект по созданию референсного пан-генома человека для 350 индивидов.

The screenshot shows the NIH website header with navigation links: ABOUT GENOMICS, RESEARCH FUNDING, RESEARCH AT NHGRI, ABOUT HEALTH, CAREERS & TRAINING, NEWS & EVENTS, and ABOUT NHGRI. Below the header is a search bar with the text "...Begin your search here". The main content area features a breadcrumb trail: Home / News & Events / News / NHGRI funds centers for advancing the reference sequence of the human genome. The headline reads "NHGRI funds centers for advancing the reference sequence of the human genome". The background image depicts a globe and silhouettes of two people's heads, one in profile and one in silhouette, with a DNA double helix overlaid on the globe.

<https://www.genome.gov/news/new-s-release/NIH-funds-centers-for-advancing-sequence-of-human-genome-reference> (2019).

The international journal of science / 11 May 2023

nature



HUMAN PANGENOME

Data from 47 individuals combine to create reference resource that reflects human diversity

Human pangenome

Nature

Volume 617 Issue 7960, 11 May 2023

The cover shows the pangenome wrapping a globe and uses a sequence tube map rendering of a pangenome graph relating ten haplotypes within the highly variable HLA-A locus on chromosome 6 created by Adam Novak at the University of California, Santa Cruz.

Cover image: Darryl Leja/NHGRI

[Read More](#)

A draft human pangenome reference

<https://doi.org/10.1038/s41586-023-05896-x>

Received: 9 July 2022

Accepted: 28 February 2023

Published online: 10 May 2023

Open access

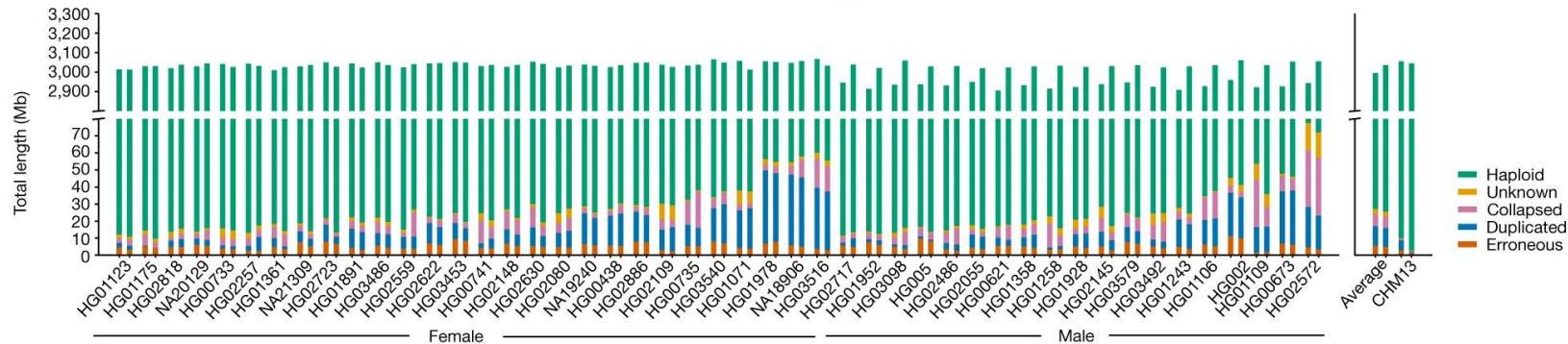


Check for updates

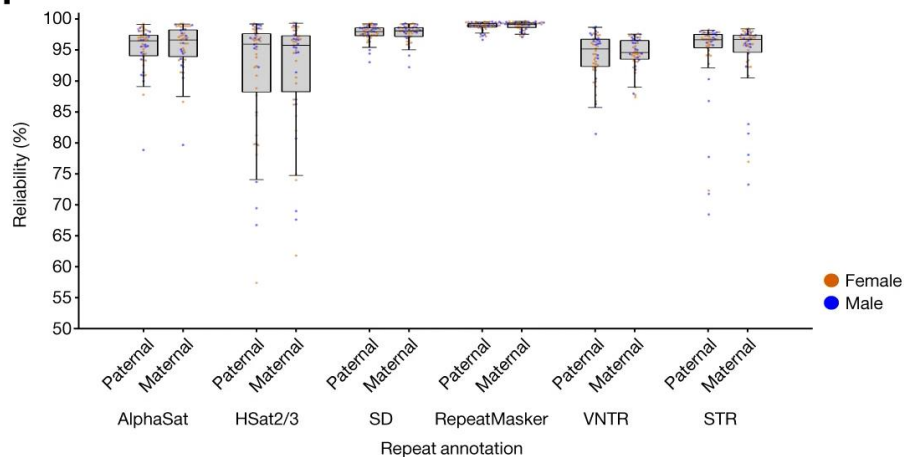
Wen-Wei Liao^{1,2,3,60}, Mobin Asri^{4,60}, Jana Ebler^{5,6,60}, Daniel Doerr^{5,6}, Marina Haukness⁴, Glenn Hickey⁴, Shuangjia Lu^{1,2}, Julian K. Lucas⁴, Jean Monlong⁴, Haley J. Abel⁷, Silvia Buonaiuto⁸, Xian H. Chang⁴, Haoyu Cheng^{9,10}, Justin Chu⁹, Vincenza Colonna^{8,11}, Jordan M. Eizenga⁴, Xiaowen Feng^{9,10}, Christian Fischer¹¹, Robert S. Fulton^{12,13}, Shilpa Garg¹⁴, Cristian Groza¹⁵, Andrea Guarracino^{11,16}, William T. Harvey¹⁷, Simon Heumos^{18,19}, Kerstin Howe²⁰, Miten Jain²¹, Tsung-Yu Lu²², Charles Markello⁴, Fergal J. Martin²³, Matthew W. Mitchell²⁴, Katherine M. Munson¹⁷, Moses Njagi Mwaniki²⁵, Adam M. Novak⁴, Hugh E. Olsen⁴, Trevor Pesout⁴, David Porubsky¹⁷, Pjotr Prins¹¹, Jonas A. Sibbesen²⁶, Jouni Sirén⁴, Chad Tomlinson¹², Flavia Villani¹¹, Mitchell R. Vollger^{17,27}, Lucinda L. Antonacci-Fulton¹², Gunjan Baid²⁸, Carl A. Baker¹⁷, Anastasiya Belyaeva²⁸, Konstantinos Billis²³, Andrew Carroll²⁸, Pi-Chuan Chang²⁸, Sarah Cody¹², Daniel E. Cook²⁸, Robert M. Cook-Deegan²⁹, Omar E. Cornejo³⁰, Mark Diekhans⁴, Peter Ebert^{5,6,31}, Susan Fairley²³, Olivier Fedrigo³², Adam L. Felsenfeld³³, Giulio Formenti³², Adam Frankish²³, Yan Gao³⁴, Nanibaa' A. Garrison^{35,36,37}, Carlos Garcia Giron²³, Richard E. Green^{38,39}, Leanne Haggerty²³, Kendra Hoekzema¹⁷, Thibaut Hourlier²³, Hanlee P. Ji⁴⁰, Eimear E. Kenny⁴¹, Barbara A. Koenig⁴², Alexey Kolesnikov²⁸, Jan O. Korbel^{23,43}, Jennifer Kordosky¹⁷, Sergey Koren⁴⁴, HoJoon Lee⁴⁰, Alexandra P. Lewis¹⁷, Hugo Magalhães^{5,6}, Santiago Marco-Sola^{45,46}, Pierre Marijon^{5,6}, Ann McCartney⁴⁴, Jennifer McDaniel⁴⁷, Jacquelyn Mountcastle³², Maria Nattestad²⁸, Sergey Nurk⁴⁴, Nathan D. Olson⁴⁷, Alice B. Popejoy⁴⁸, Daniela Puiu⁴⁹, Mikko Rautiainen⁴⁴, Allison A. Regier¹², Arang Rhie⁴⁴, Samuel Sacco³⁰, Ashley D. Sanders⁵⁰, Valerie A. Schneider⁵¹, Baergen I. Schultz³³, Kishwar Shafin²⁸, Michael W. Smith³³, Heidi J. Sofia³³, Ahmad N. Abou Tayoun^{52,53}, Françoise Thibaud-Nissen⁵¹, Francesca Floriana Tricomi²³, Justin Wagner⁴⁷, Brian Walenz⁴⁴, Jonathan M. D. Wood²⁰, Aleksey V. Zimin^{49,54}, Guillaume Bourque^{55,56,57}, Mark J. P. Chaisson²², Paul Flicek²³, Adam M. Phillippy⁴⁴, Justin M. Zook⁴⁷, Evan E. Eichler^{17,58}, David Haussler^{4,58}, Ting Wang^{12,13}, Erich D. Jarvis^{32,58,59}, Karen H. Miga⁴, Erik Garrison¹¹✉, Tobias Marschall^{5,6}✉, Ira M. Hall^{1,2}✉, Heng Li^{9,10}✉ & Benedict Paten⁴✉

The Human Pangenome Reference Consortium

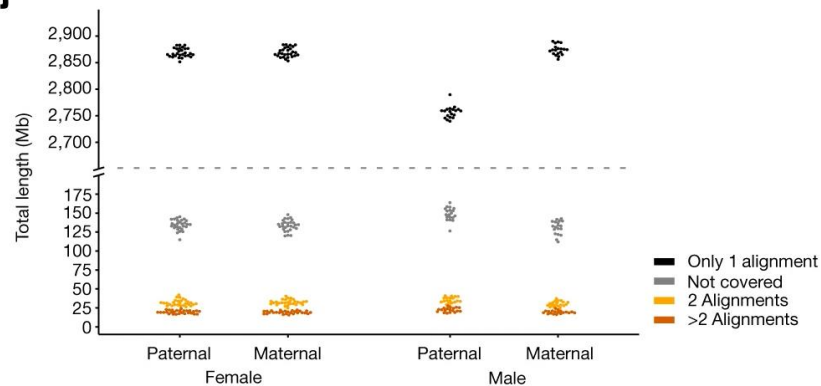
h



i



j

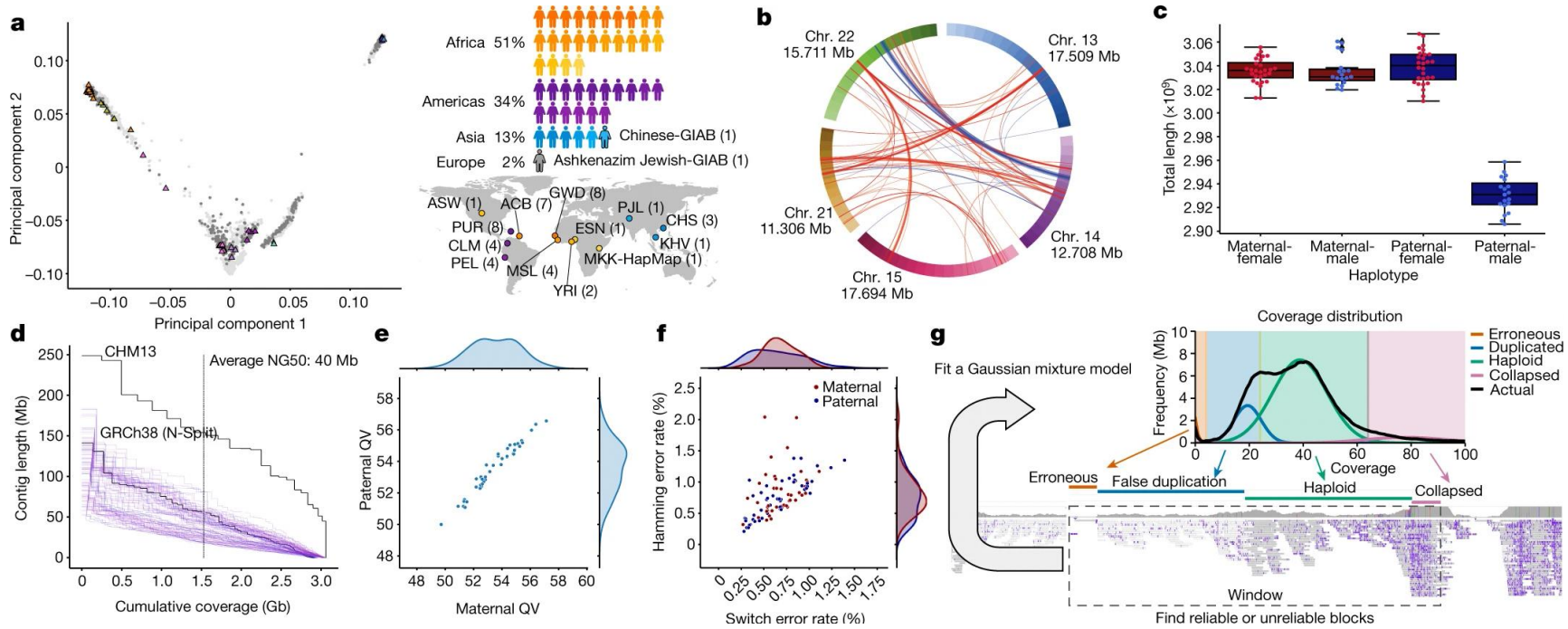


h, Качество **47 HPRC** сборок. Слева отцовский, справа материнский гаплотипы.

i, Качество сборки 6-ти типов повторов: AlphaSat, alpha satellites; HSat2/3, human satellites 2 and 3.

j, Полнота сборок относительно сборки T2T-CHM13.

The Human Pangenome Reference Consortium



a, Отбор и оценка образцов для HPRC. Слева анализ основных компонент; справа Происхождение образцов (ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest US; CHS, Han Chinese South; CLM, Colombian in Medellin, Colombia; ESN, Esan in Nigeria; GWD, Gambian in Western Division; KHV, Kinh in Ho Chi Minh City, Vietnam; MKK, Maasai in Kinyawa, Kenya; MSL, Mende in Sierra Leone; PEL, Peruvian in Lima, Peru; PJP, Punjabi in Lahore, Pakistan; PUR, Puerto Rican in Puerto Rico; YRI, Yoruba in Ibadan, Nigeria).

b, Сходные последовательности в коротких плечах акроцентриков.

c, Фазирующие сборки.

d, Сравнение с T2T-CHM13 и GRCh38 сборки.

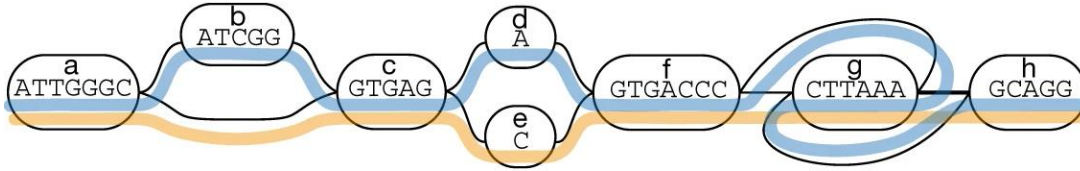
e, Оценки материнской и отцовской части сборки каждого образца.

f, Оценка точности фазирующих сборок.

g, Оценки наличия ошибочных, дублированных гаплоидных и неизвестных элементов.

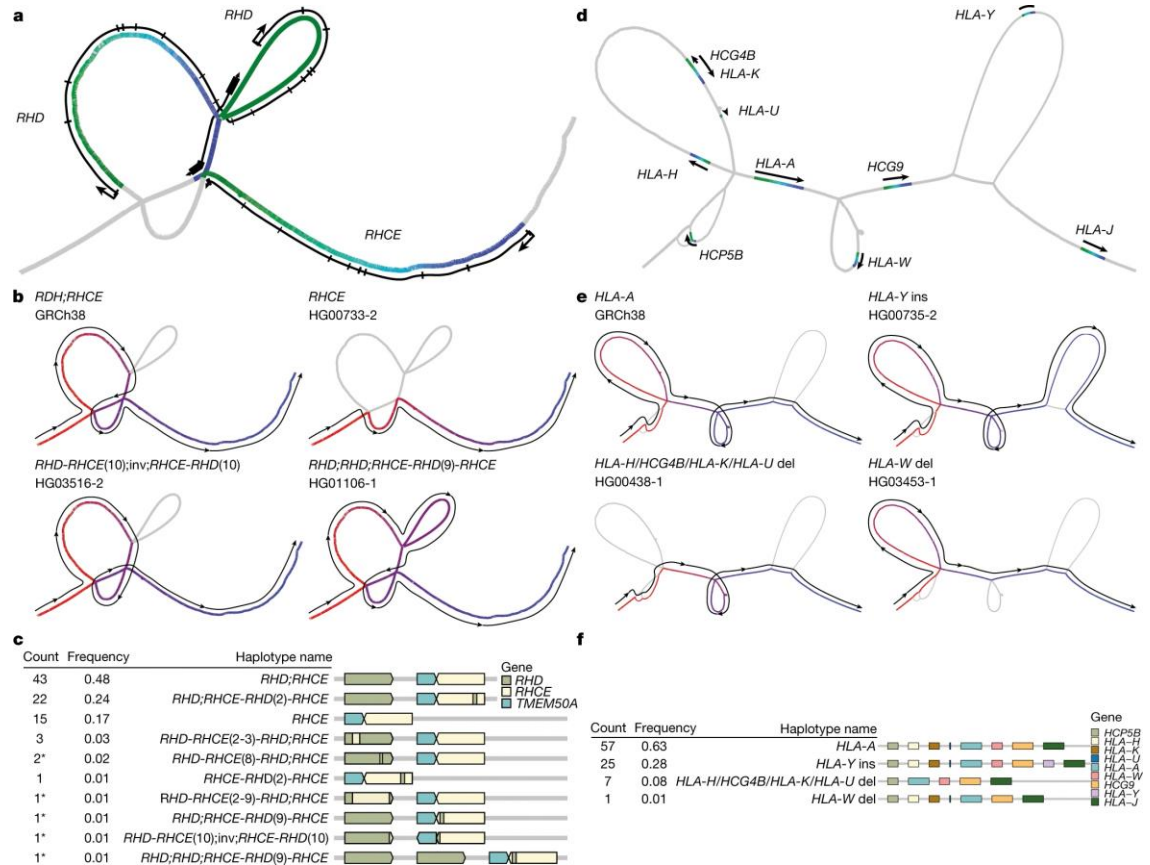
a

ATTGGGCATCGGGTGAGAGTGACCCCTTTAAGGCAGG
 ATTGGGC-----GTGAGCGTGACCCCTTAAGGCAGG

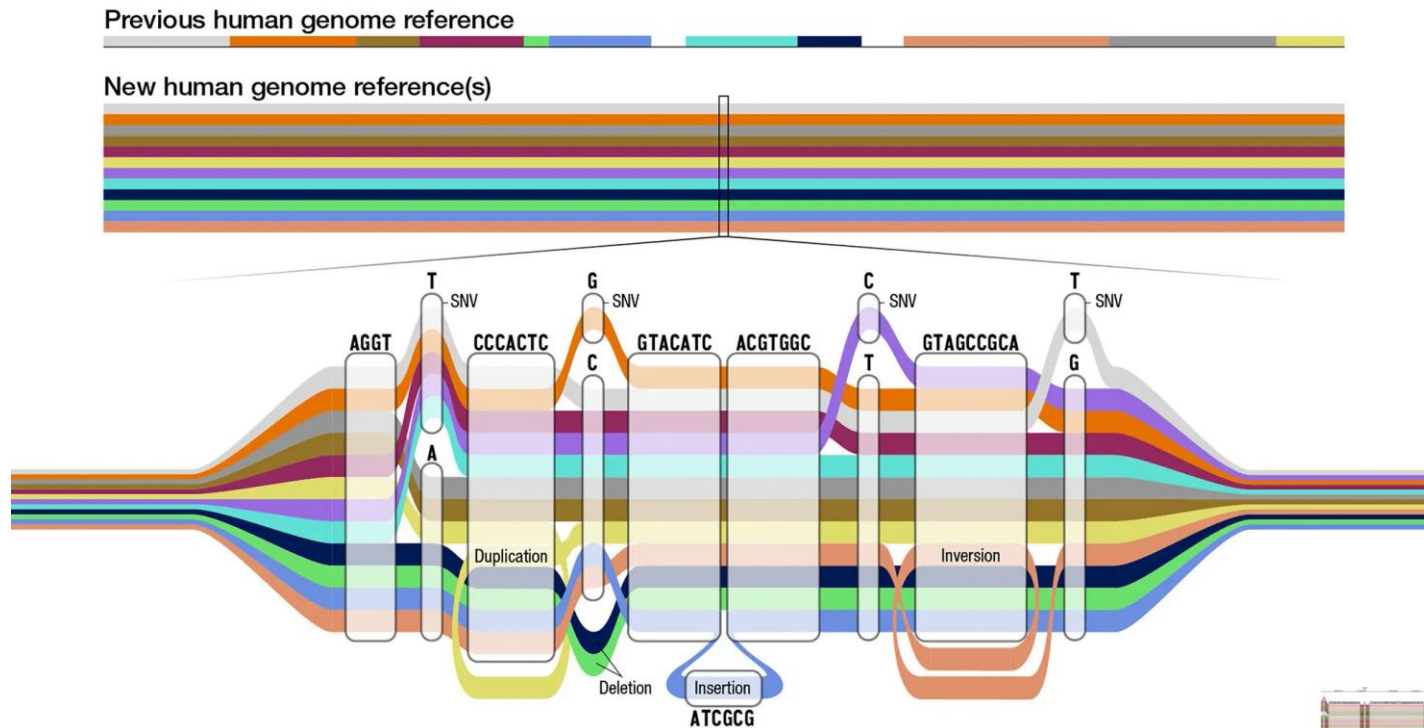


Презентация пан-генома человека

Визуализация сложных пагеномных локусов

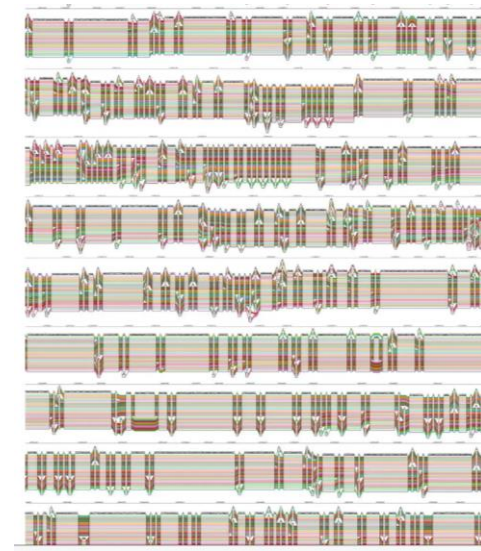


The New York Times



Scientists Unveil a More Diverse Human Genome

The “pangenome,” which collated genetic sequences from 47 people of diverse ethnic backgrounds, could greatly expand the reach of personalized medicine.



Human Pangenome Reference

Human pangenome supports analysis of complex genomic regions

Arya Massarat, Melissa Gymrek ... Hákon Jónsson
Nature | **News & Views Forum** | 10 May 2023

Increased mutation and gene conversion within human segmental duplications

Mitchell R. Vollger, Philip C. Dishuck ... Evan E. Eichler
Nature | **Article** | [Open Access](#) | 10 May 2023

'Pangenome' aims to capture the breadth of human diversity

Nick Petrić Howe & Shamini Bundell
Nature | **Nature Podcast** | 10 May 2023








Combining 47 human genomes into a single pangenome

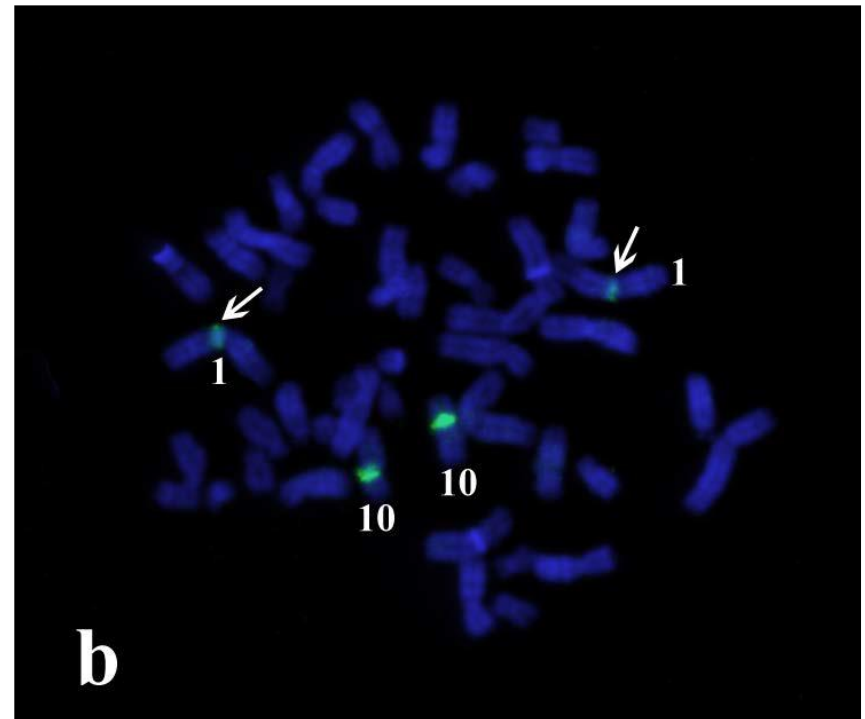
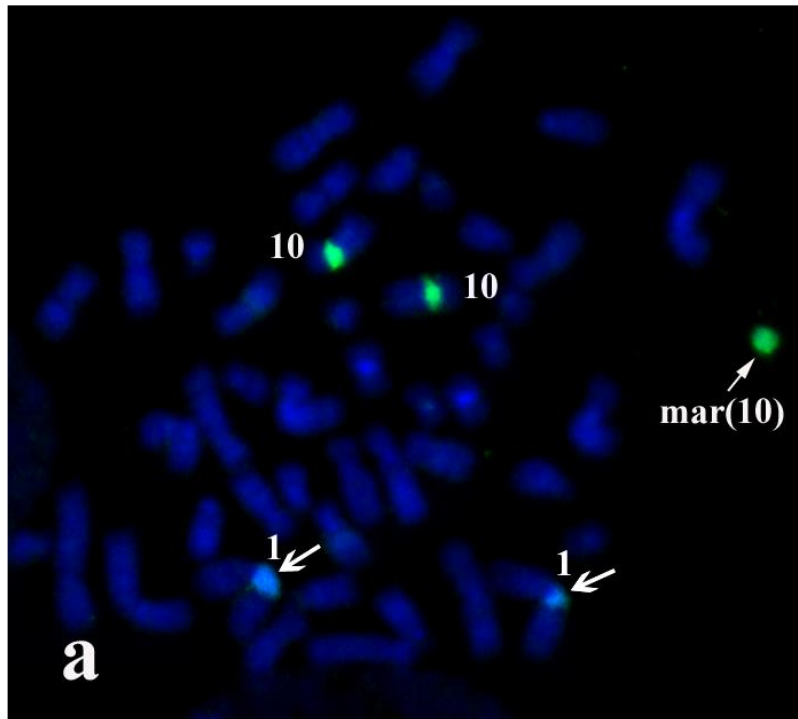
Jared B. Fudge
Nature Biotechnology | **Research Highlight** | 14 Jun 2023

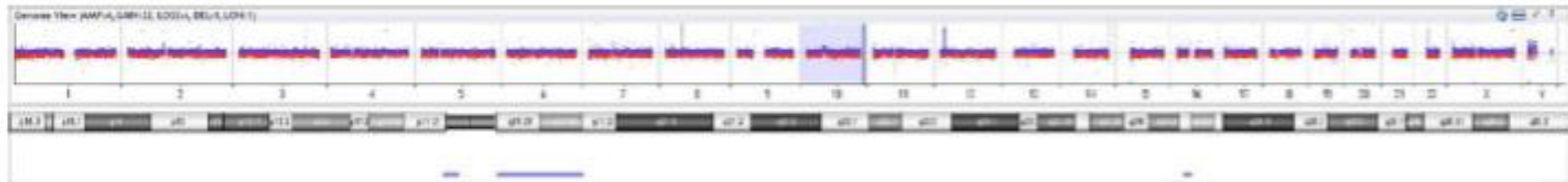
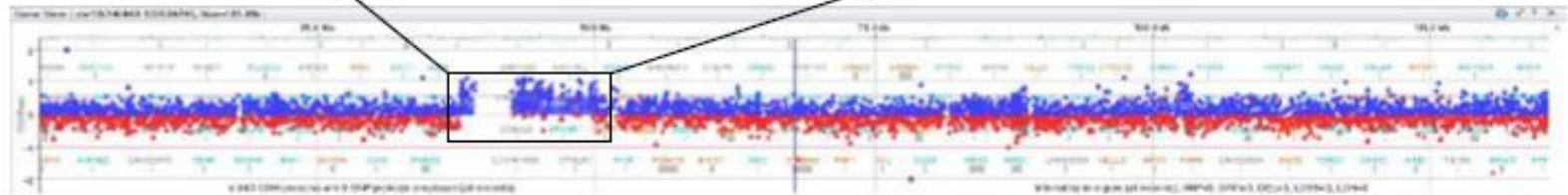
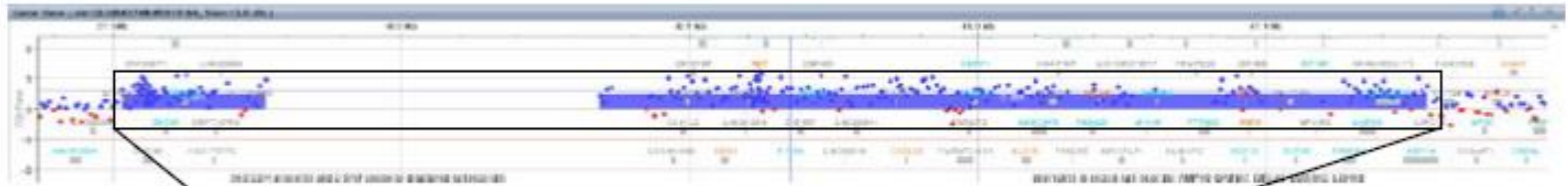
**Использование сборок генома человека при проведении
цитогеномной диагностики**

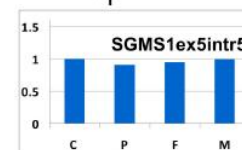
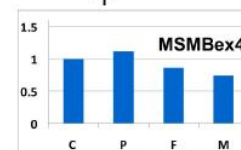
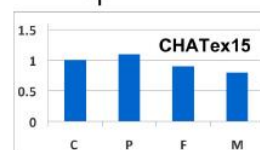
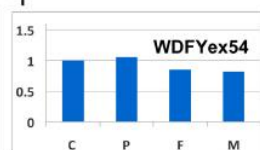
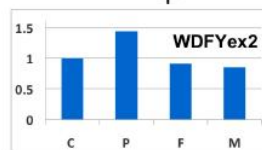
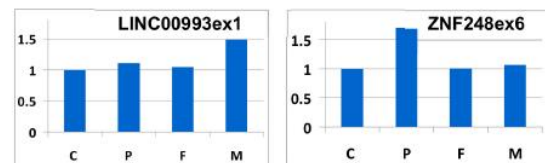
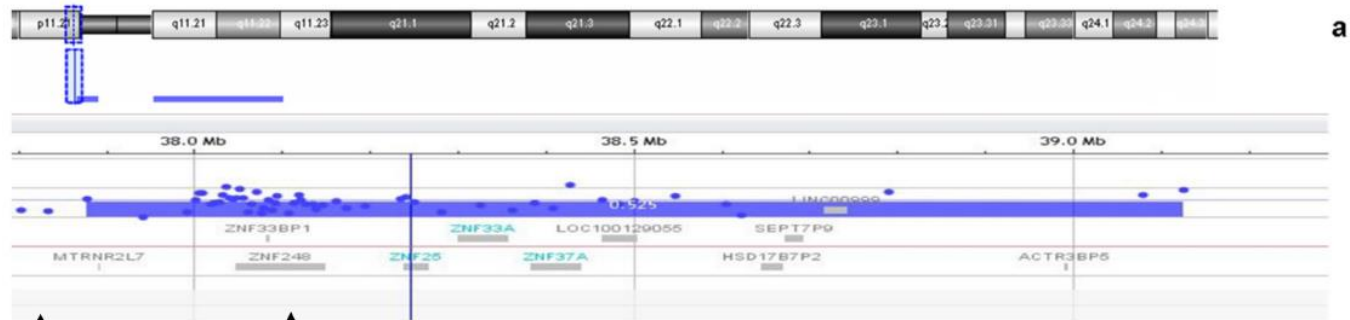
Article

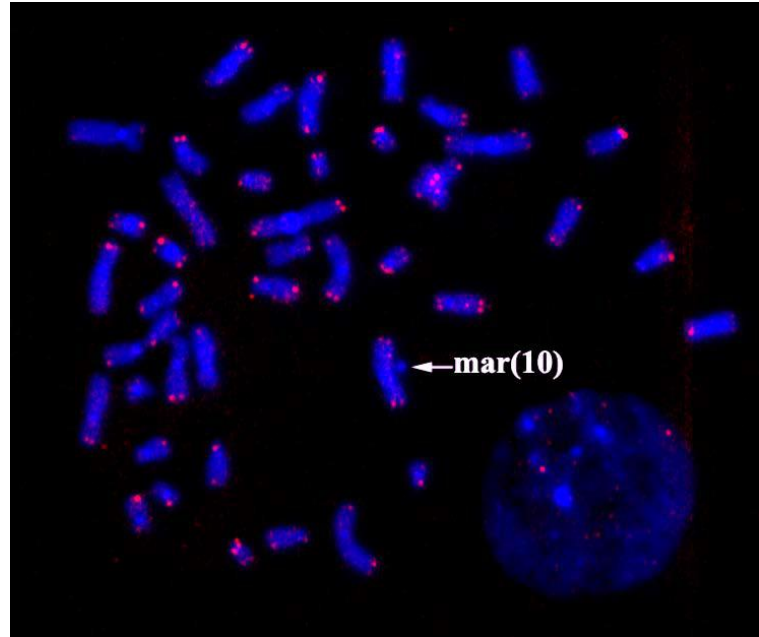
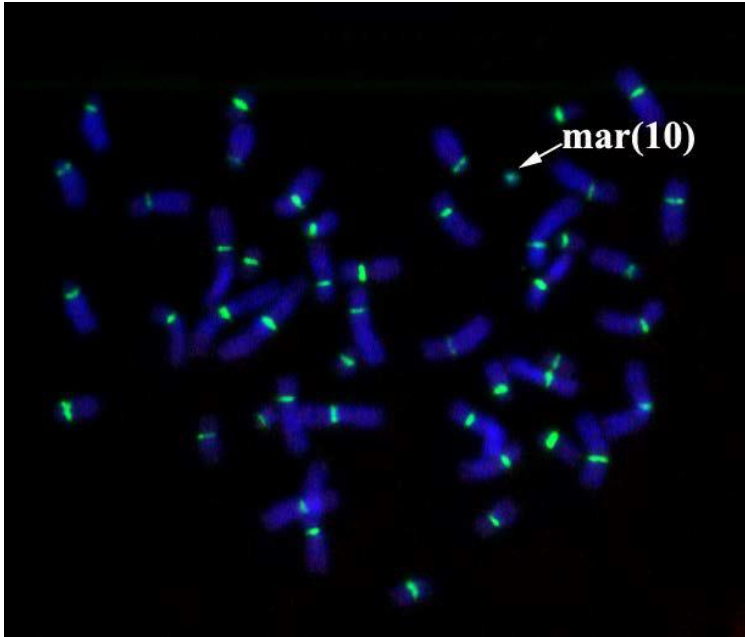
The Precise Breakpoint Mapping in Paracentric Inversion 10q22.2q23.3 by Comprehensive Cytogenomic Analysis, Multicolor Banding, and Single-Copy Chromosome Sequencing

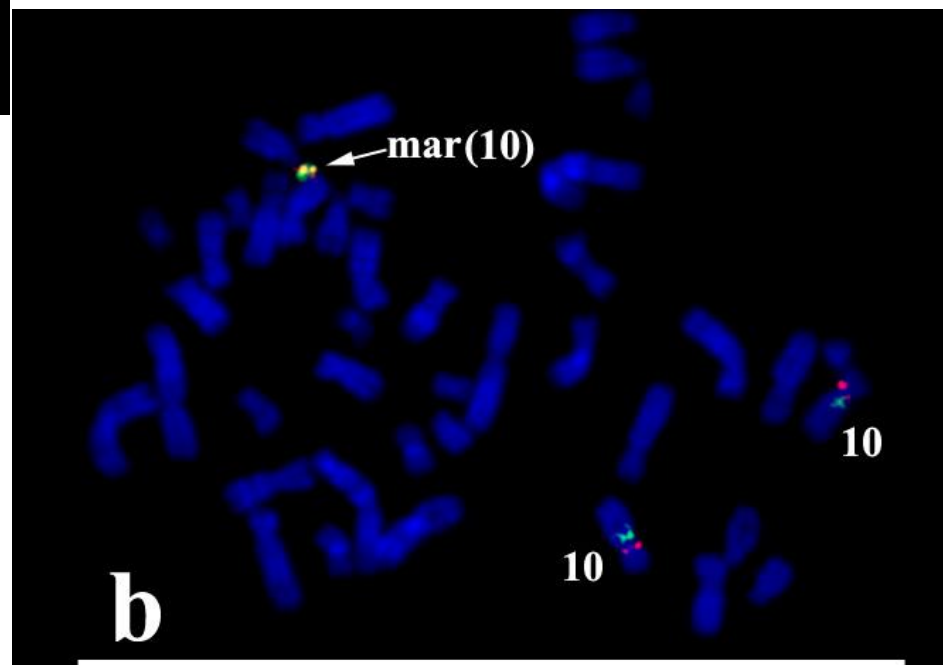
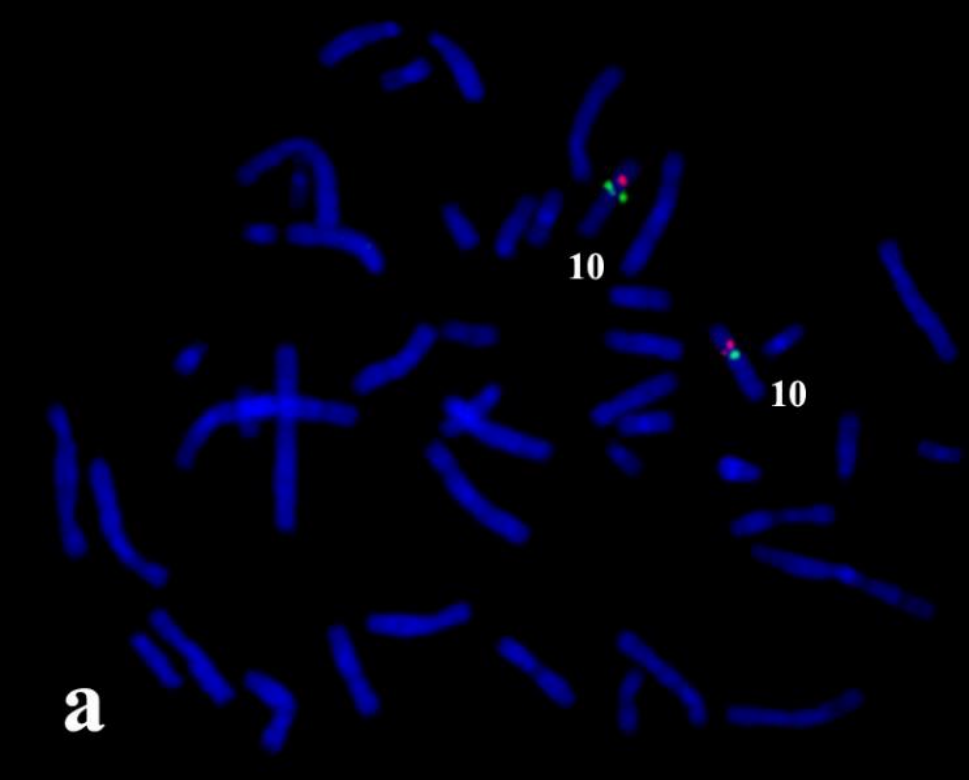
Tatyana V. Karamysheva ^{1,*}, Tatyana A. Gayner ^{2,3}, Eugeny A. Elisaphenko ¹, Vladimir A. Trifonov ^{4,5}, Elvira G. Zakirova ^{1,5}, Konstantin E. Orishchenko ^{1,5}, Mariya A. Prokhorovich ¹, Maria E. Lopatkina ⁶, Nikolay A. Skryabin ⁶, Igor N. Lebedev ⁶ and Nikolay B. Rubtsov ^{1,5}





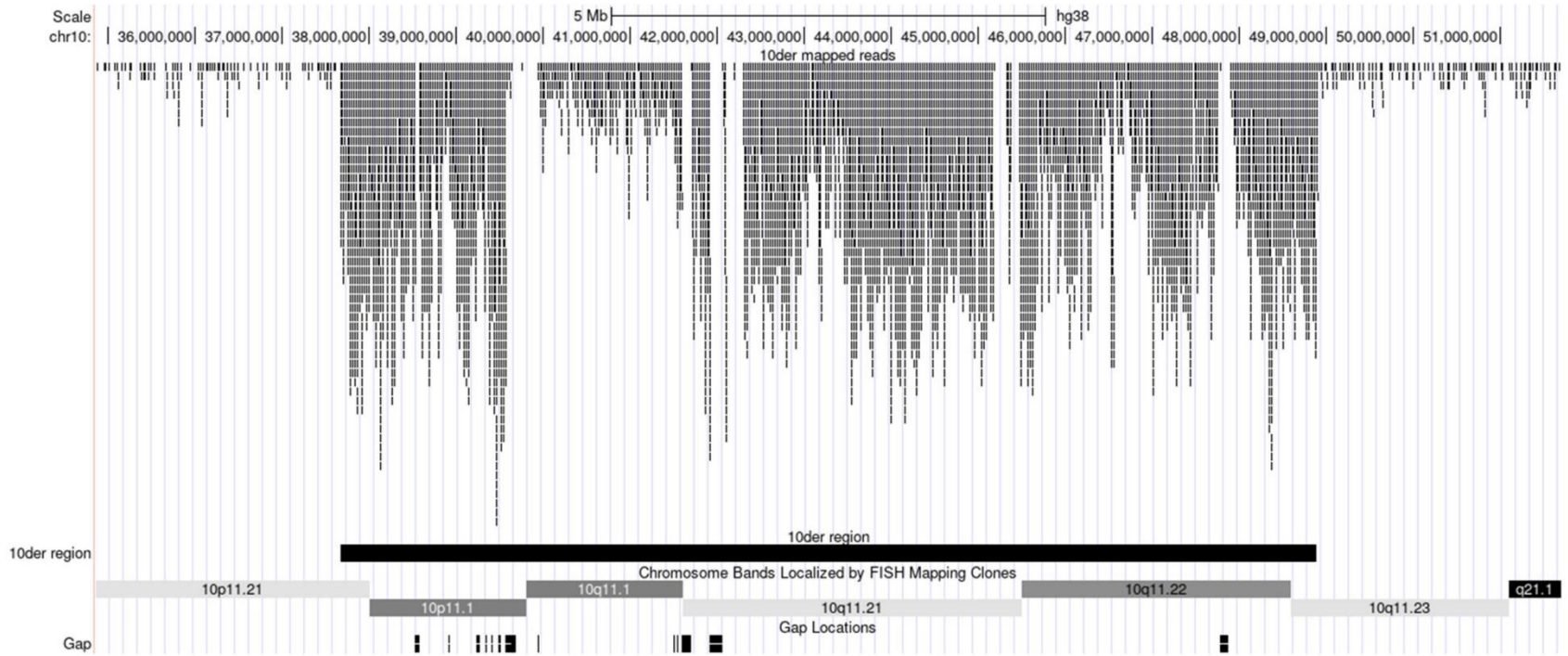


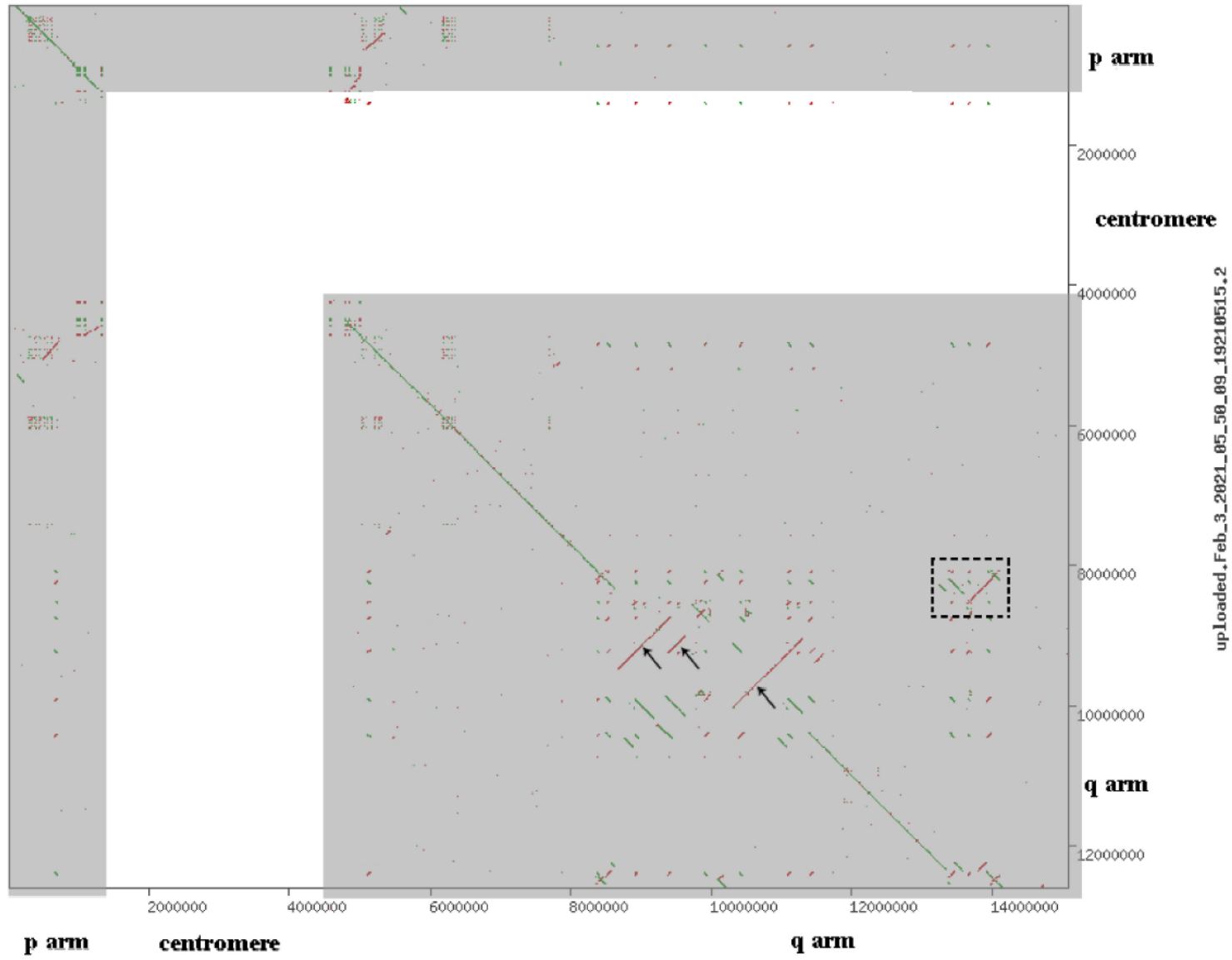




c mar(10):

A schematic diagram of a marker chromosome 10, showing a blue chromosome with a single fluorescent spot (red and green) on its short arm, representing the mar(10) configuration.



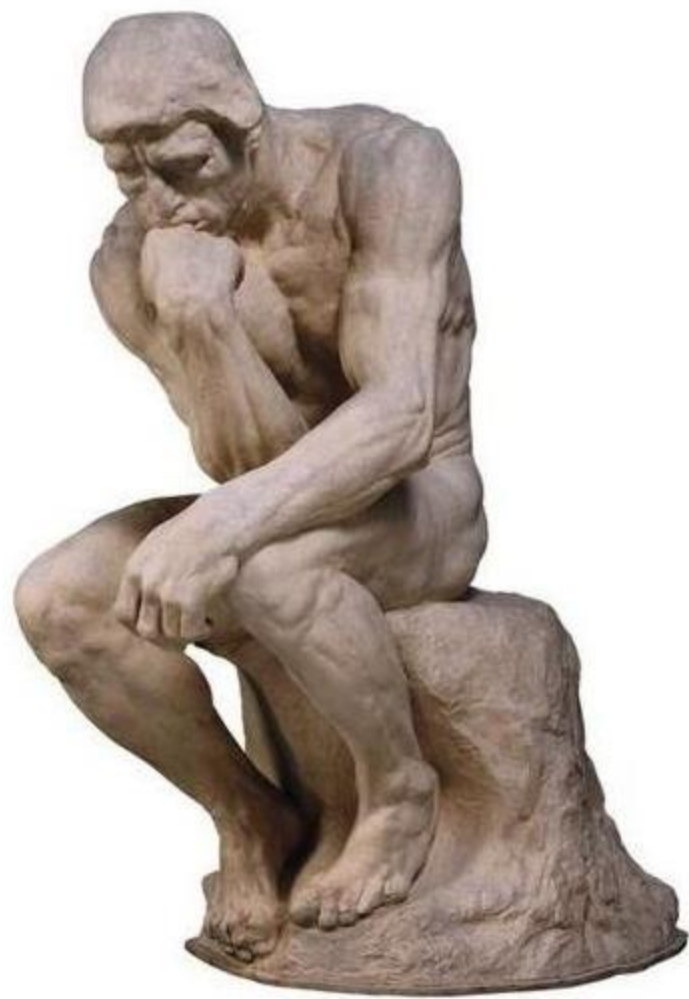


upLoaded, Feb_3_2021_05_50_09_19210515_2

Пан-геном человека

Сумма всех возможных вариантов его элементов?

Комбинаторика отличающихся элементов генома?



- Liao, Wen-Wei; Asri, Mobin; Ebler, Jana; Doerr, Daniel; Haukness, Marina; Hickey, Glenn; Lu, Shuangjia; Lucas, Julian K.; Monlong, Jean; Abel, Haley J.; Buonaiuto, Silvia; Chang, Xian H.; Cheng, Haoyu; Chu, Justin; Colonna, Vincenza (May 2023). "[A draft human pangenome reference](#)". *Nature*. **617** (7960): 312–324. [Bibcode:2023Natur.617..312L](#). [doi:10.1038/s41586-023-05896-x](#).
- Abondio, Paolo; Cilli, Elisabetta; Luiselli, Donata (June 2023). "[Human Pangenomics: Promises and Challenges of a Distributed Genomic Reference](#)". *Life*. **13** (6): 1360. [Bibcode:2023Life...13.1360A](#). [doi:10.3390/life13061360](#).
- Lee, HoJoon; Greer, Stephanie U.; Pavlichin, Dmitri S.; Zhou, Bo; Urban, Alexander E.; Weissman, Tsachy; Liao, Wen-Wei; Asri, Mobin; Ebler, Jana; Doerr, Daniel; Haukness, Marina; Hickey, Glenn; Lu, Shuangjia; Lucas, Julian K.; Monlong, Jean (2023-08-28). "[Pan-conserved segment tags identify ultra-conserved sequences across assemblies in the human pangenome](#)". *Cell Reports Methods*. **3** (8): 100543. [doi:10.1016/j.crmeth.2023.100543](#).
- Wang, T; Antonacci-Fulton, L; Howe, K; Lawson, HA; Lucas, JK; Phillippy, AM; Popejoy, AB; Asri, M; Carson, C; Chaisson, MJP; Chang, X; Cook-Deegan, R; Felsenfeld, AL; Fulton, RS; Garrison, EP; Garrison, NA; Graves-Lindsay, TA; Ji, H; Kenny, EE; Koenig, BA; Li, D; Marschall, T; McMichael, JF; Novak, AM; Purushotham, D; Schneider, VA; Schultz, BI; Smith, MW; Sofia, HJ; Weissman, T; Flicek, P; Li, H; Miga, KH; Paten, B; Jarvis, ED; Hall, IM; Eichler, EE; Haussler, D; Human Pangenome Reference, Consortium (April 2022). "[The Human Pangenome Project: a global resource to map genomic diversity](#)". *Nature*. **604** (7906): 437–446. [doi:10.1038/s41586-022-04601-8](#).
- Jarvis, Erich D.; Formenti, Giulio; Rhie, Arang; Guarracino, Andrea; Yang, Chentao; Wood, Jonathan; Tracey, Alan; Thibaud-Nissen, Françoise; Vollger, Mitchell R.; Porubsky, David; Cheng, Haoyu; Asri, Mobin; Logsdon, Glennis A.; Carnevali, Paolo; Chaisson, Mark J. P. (November 2022). "[Semi-automated assembly of high-quality diploid human reference genomes](#)". *Nature*. **611** (7936): 519–531. [Bibcode:2022Natur.611..519J](#). [doi:10.1038/s41586-022-05325-5](#).
- Li, Heng; Bloom, Jonathan M.; Farjoun, Yossi; Fleharty, Mark; Gauthier, Laura; Neale, Benjamin; MacArthur, Daniel (August 2018). "[A synthetic-diploid benchmark for accurate variant-calling evaluation](#)". *Nature Methods*. **15** (8): 595–597. [doi:10.1038/s41592-018-0054-7](#).
- Li, Heng; Feng, Xiaowen; Chu, Chong (2020-10-16). "[The design and construction of reference pangenome graphs with minigraph](#)". *Genome Biology*. **21** (1): 265. [doi:10.1186/s13059-020-02168-z](#).
- Li, Heng (2018-05-10). "[Minimap2: pairwise alignment for nucleotide sequences](#)". *Bioinformatics*. **34** (18): 3094–3100. [doi:10.1093/bioinformatics/bty191](#).
- Hickey, Glenn; Monlong, Jean; Ebler, Jana; Novak, Adam M.; Eizenga, Jordan M.; Gao, Yan; Marschall, Tobias; Li, Heng; Paten, Benedict (2023-05-10). "[Pangenome graph construction from genome alignments with Minigraph-Cactus](#)". *Nature Biotechnology*: 1–11. [doi:10.1038/s41587-023-01793-w](#).
- Armstrong, Joel; Hickey, Glenn; Diekhans, Mark; Fiddes, Ian T.; Novak, Adam M.; Deran, Alden; Fang, Qi; Xie, Duo; Feng, Shaohong; Stiller, Josefin; Genereux, Diane; Johnson, Jeremy; Marinescu, Voichita Dana; Alföldi, Jessica; Harris, Robert S. (November 2020). "[Progressive Cactus is a multiple-genome aligner for the thousand-genome era](#)". *Nature*. **587** (7833): 246–251. [Bibcode:2020Natur.587..246A](#). [doi:10.1038/s41586-020-2871-y](#).
- Singh, Vipin; Pandey, Shweta; Bhardwaj, Anshu (2022). "[From the reference human genome to human pangenome: Premise, promise and challenge](#)". *Frontiers in Genetics*. **13**. [doi:10.3389/fgene.2022.1042550](#). [ISSN 1664-8021](#). [PMC 9684177](#). [PMID 36437921](#).