



Состояние исследований генома человека

Чемерис А.В.

доктор биологических наук, профессор,

Институт биохимии и генетики

Уфимского федерального исследовательского центра

Российской академии наук

Вологда - 2020



Johann Friedrich Miescher (1844–1895)

Некоторые важные вехи в исследовании ДНК и геномов

1869: Friedrich Miescher впервые выделил ДНК (нуклеин)

1871: Первые публикации, описывающие ДНК (нуклеин), написанные Friedrich Miescher, Felix Hoppe-Seyler, P. Plosz и Nikolai Lubavin

1889: Richard Altmann переименовал «нуклеин» в «нуклеиновую кислоту»

1909: Wilhelm Johannsen использовал слово «ген» для обозначения единиц наследственности

1920: Hans Winkler Н. для гаплоидного набора хромосом предложил термин «геном»

1929: Phoebus Levene вместе с Ефимом Семеновичем Лондоном установил, что в состав ДНК входит пятиатомный сахар дезоксирибозы

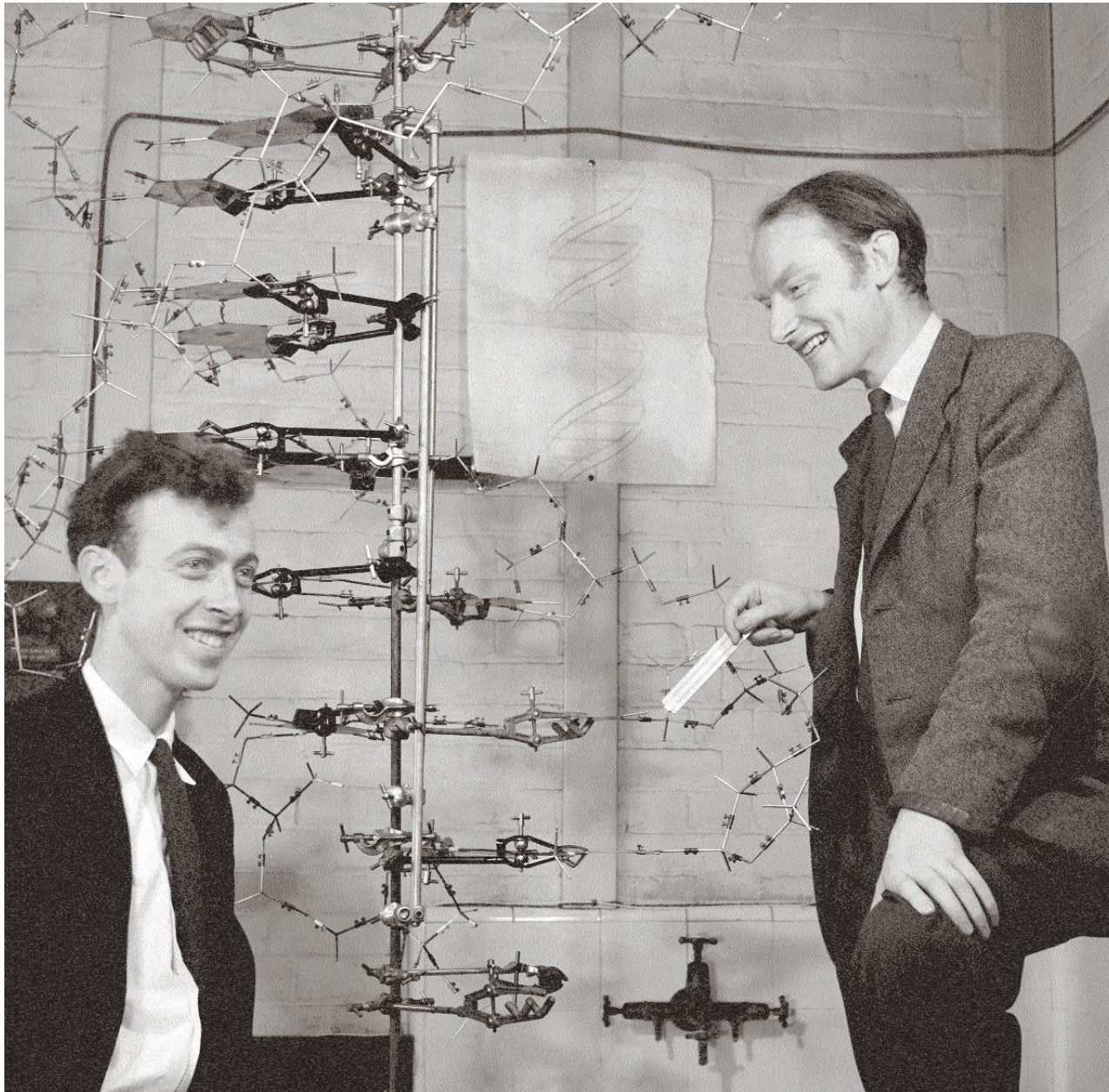
1934: Андрей Николаевич Белозерский впервые выделил ДНК из растений

1944: Oswald T. Avery, Colin MacLeod и Maclyn McCarty показали, что ДНК является генетическим материалом

1949–1950: Erwin Chargaff обнаружил фиксированные соотношения между азотистыми основаниями в ДНК, известными как «правила Чаргаффа».

1951-1952: Rosalind Franklin и Maurice Wilkins получили высококачественные рентгенограммы ДНК

1953: James Watson и Francis Crick открыли «двойную спираль ДНК»



Anthony Barrington Brown's photograph of Watson and Crick with their model of DNA at the Cavendish Laboratory in Cambridge, 21 May 1953.

Некоторые важные вехи в исследовании ДНК и геномов (продолжение)

1977: Frederick Sanger, Allan Maxam и Walter Gilbert разработали быстрые методы секвенирования ДНК

1983-1987: Kary Mullis придумал ПЦР как метод, амплифицирующий ДНК *in vitro*

1985: Alec Jeffreys предложил использовать ДНК в криминалистике

1990: Начало секвенирования генома человека

1995: Первый геном свободноживущей бактерии *Haemophilus influenzae*

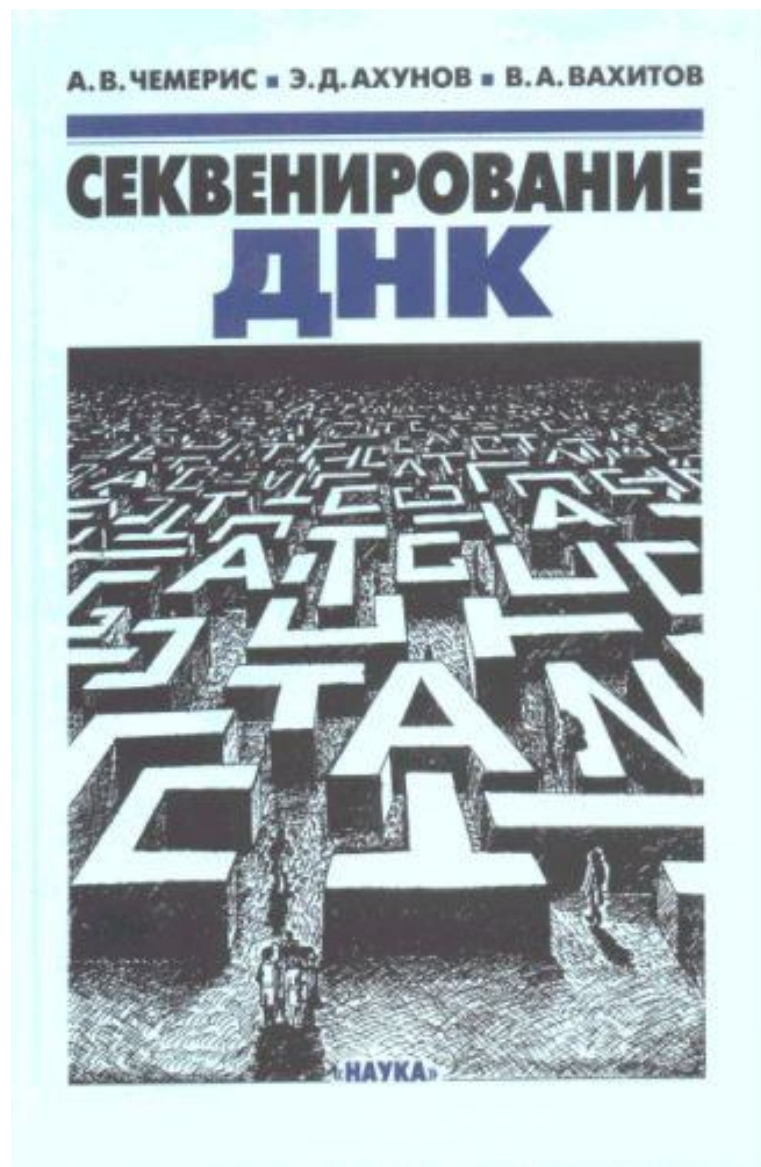
1996: Первый геном эукариотического организма - дрожжи *Saccharomyces cerevisiae*

1998: Первый геном многоклеточного организма - нематода *Caenorhabditis elegans*

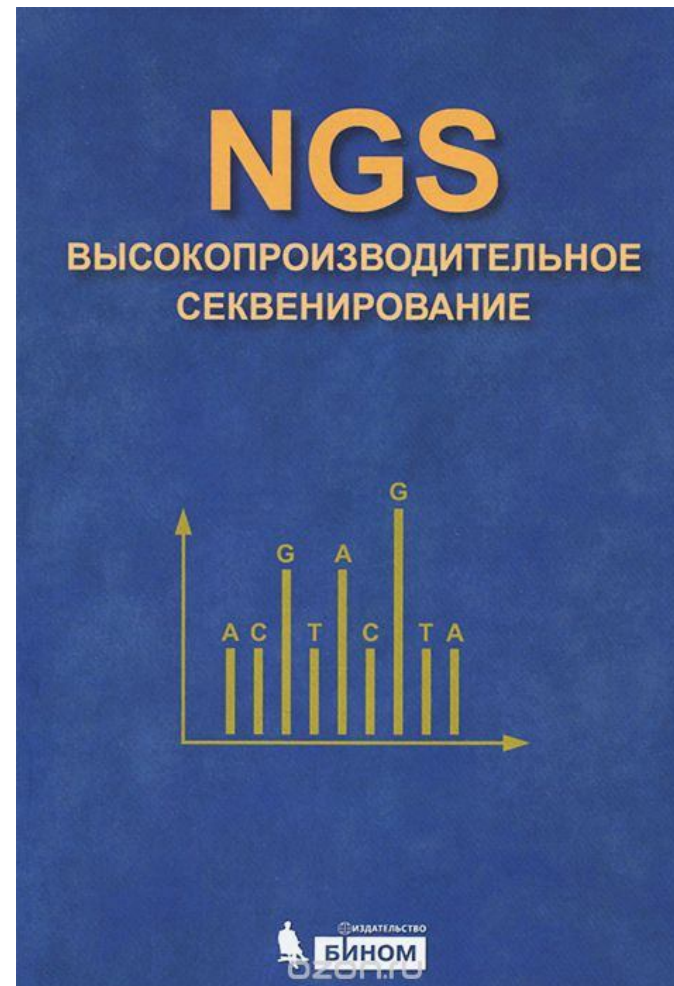
2001: Опубликована первая версия генома человека

2003: Одновременная публикация статей в Nature и Science, ознаменовавшая завершение проекта секвенирования генома человека

2004: Понимание того, что нужны принципиально новые методы секвенирования ДНК – Next Generation Sequencing -NGS



Секвенирование ДНК
Чемерис А.В., Ахунув Э.Д.,
Вахитов В.А.
М., Наука, 1999. 429 с.



NGS: высокопроизводительное
секвенирование
Ребриков Д.В., Коростин Д.О.,
Шубина Е.С., Ильинский В. В.
Год издания: 2014
Страниц: 232
М., БИНОМ. Лаборатория знаний



SOLiD



HeliScope



PacBio RS/RS II



MinION

Вес – 78 г

Мощность – 1 Вт

Размеры – 105x23x33 мм

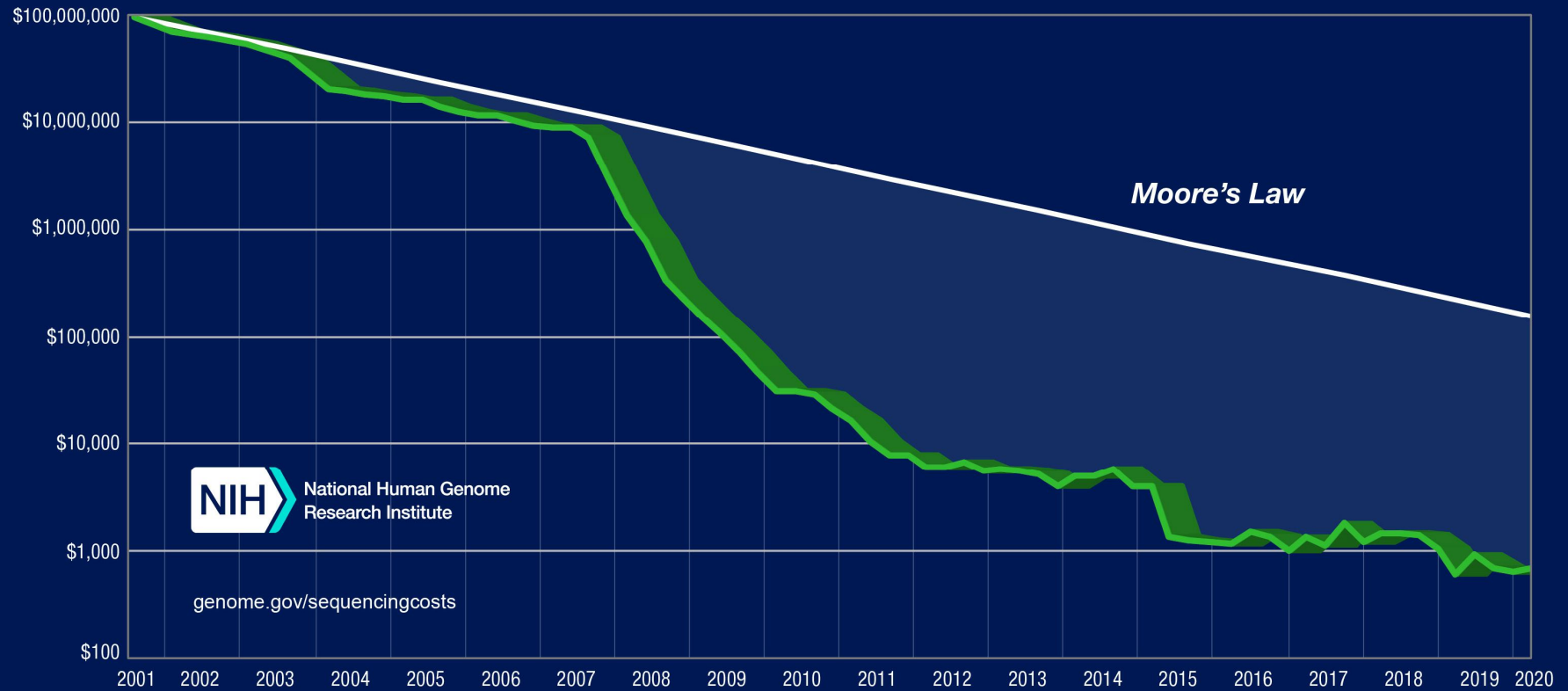
Количество нанопор – до 512

Интерфейс – USB 3.0 (до 5 Gbps)

Производительность – 10 - 20 Gbp за 48 часов

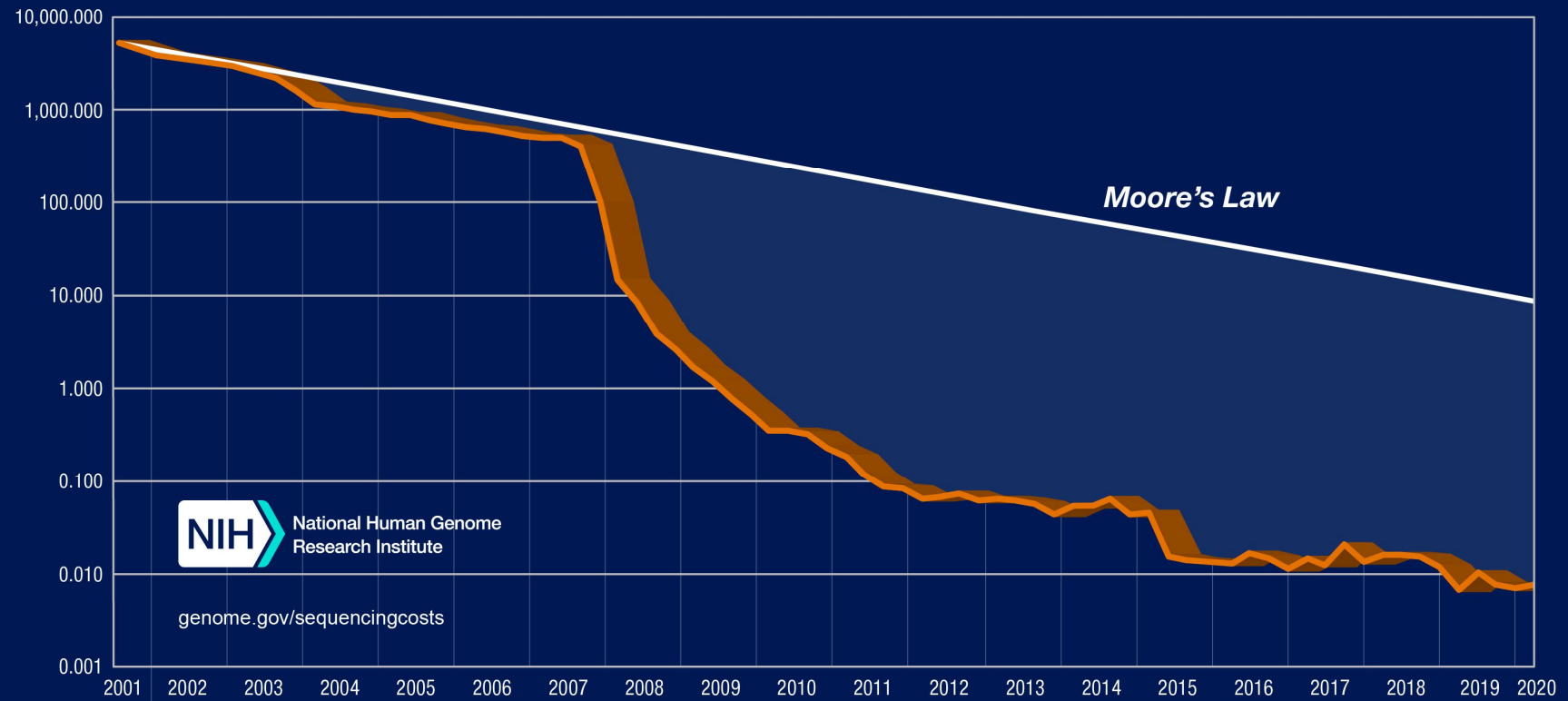


Cost per Human Genome

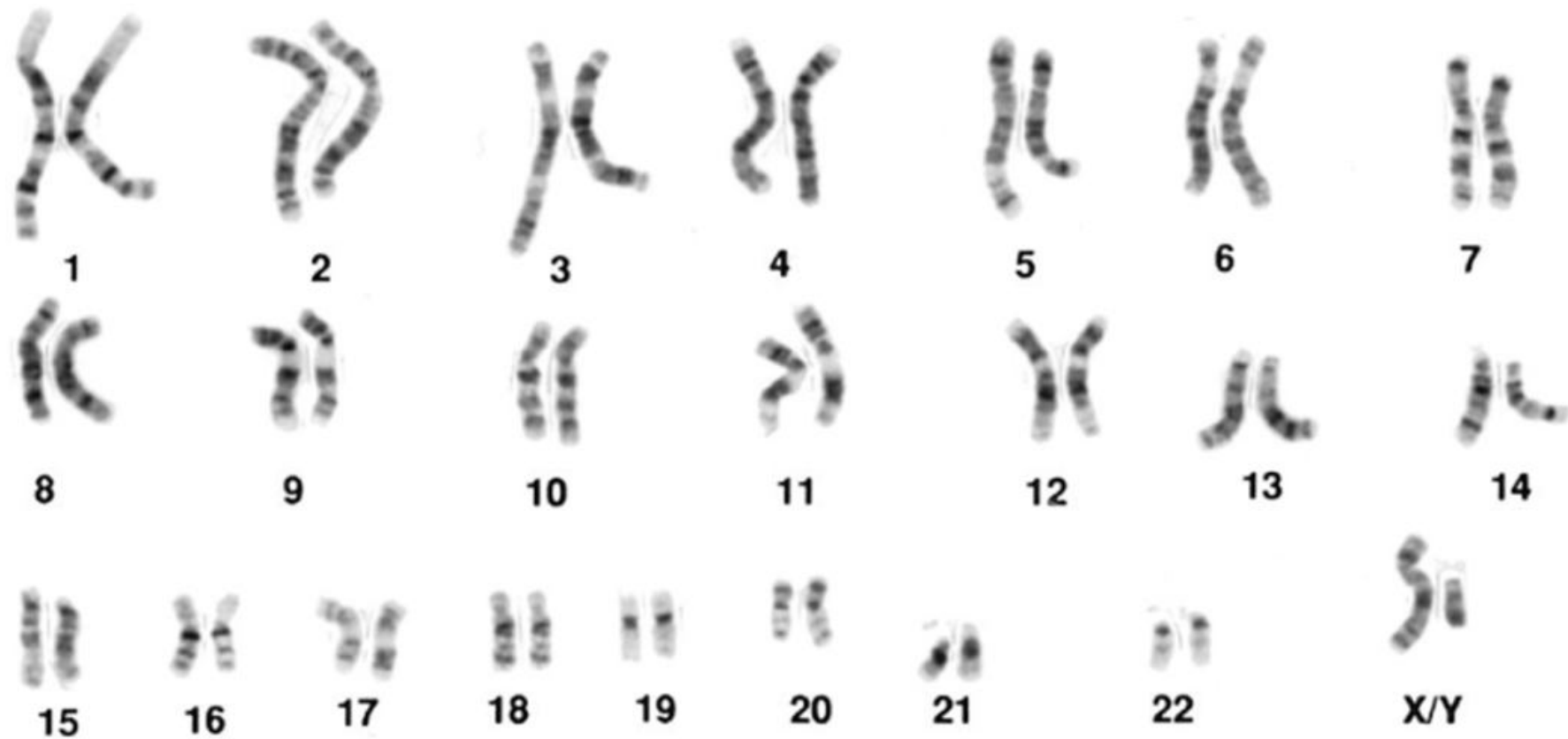


<https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>

Cost per Raw Megabase of DNA Sequence



<https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>



Размеры хромосом человека:

1 - 245 млн. п.н., 2 - 242 млн. п.н., 3 - 200 млн. п.н., 4 - 191 млн. п.н., 5 - 180 млн. п.н., 6 - 171 млн. п.н., 7 - 158 млн. п.н., 8 - 146 млн. п.н., 9 - 138 млн. п.н., 10 - 135 млн. п.н., 11 - 134 млн. п.н., 12 - 132 млн. п.н., 13 - 114 млн. п.н., 14 - 106 млн. п.н., 15 - 100 млн. п.н., 16 - 88 млн. п.н., 17 - 80 млн. п.н., 18 - 76 млн. п.н., 19 - 64 млн. п.н., 20 - 62 млн. п.н., 21 - 46 млн. п.н., 22 - 49 млн. п.н., хромосома X - 154 млн. п.н., хромосома Y - 57 млн. п.н.

**Итого – квазигапloidный ядерный геном человека ~ 3 млрд.п.н.
митохондриальная ДНК ~ 16500 п.н.**

Основные типы полиморфизма ДНК

Single-Nucleotide Polymorphism (SNP)

Однонуклеотидный полиморфизм (снипы)

5'-GGTTATCTGGCAGATAGA[A/C/G/T]GGTGACAACAGTCCCA-3'

Insertion/Deletion Polymorphism (Indels)

Инсерционно-делеционный полиморфизм (инделлы)

5'-XXXXXXXXXXXXXXXXXXXXN/-XXXXXXXXXXXXXXXXXXXXXXXX-3'

5'-XXXXXXXXXXXXXXXXXXXXA/AAAAAXXXXXXXXXXXXXXXXXXXX-3'

5'-XXXXXXXXXXXXXXXXXXXXNNNNNN/NNNXXXXXXXXXXXXXXXXXXXX-3'

Полиморфизм ДНК по числу повторов

5'-ttatctgGTTC GTTC GTTC GTTC GTTCgtcccatctg-3'

5'-ttatctgGTTC GTTC GTTC GTTC GTTC GTTC GTTCgtcccatctg-3'

Volume 586 Issue 7831, 29 October 2020



Perspective

Strategic vision for improving human health at The Forefront of Genomics

<https://doi.org/10.1038/s41586-020-2817-4>

Received: 30 June 2020

Accepted: 4 September 2020

Published online: 28 October 2020

 Check for updates

Eric D. Green¹✉, Chris Gunter¹, Leslie G. Biesecker¹, Valentina Di Francesco¹, Carla L. Easter¹, Elise A. Feingold¹, Adam L. Felsenfeld¹, David J. Kaufman¹, Elaine A. Ostrander¹, William J. Pavan¹, Adam M. Phillippy¹, Anastasia L. Wise¹, Jyoti Gupta Dayal¹, Britny J. Kish¹, Allison Mandich¹, Christopher R. Wellington¹, Kris A. Wetterstrand¹, Sarah A. Bates¹, Darryl Leja¹, Susan Vasquez¹, William A. Gahl¹, Bettie J. Graham¹, Daniel L. Kastner¹, Paul Liu¹, Laura Lyman Rodriguez¹, Benjamin D. Solomon¹, Vence L. Bonham¹, Lawrence C. Brody¹, Carolyn M. Hutter¹ & Teri A. Manolio¹

BOLD PREDICTIONS FOR HUMAN GENOMICS BY 2030

СМЕЛЫЕ ПРОГНОЗЫ ДЛЯ ГЕНОМИКИ ЧЕЛОВЕКА К 2030 ГОДУ

- 1. Генерация и анализ полной последовательности генома человека будет обычным делом для любой исследовательской лаборатории, став таким же простым, как проведение очистки ДНК.**
- 2. Биологическая функция(и) каждого человеческого гена будет известна; для некодирующих элементов в геноме человека такое знание будет скорее правилом, чем исключением.**
- 3. Общие особенности эпигенетического ландшафта и транскрипционных продуктов будут регулярно включаться в прогностические модели влияния генотипа на фенотип.**
- 4. Исследования в области геномики человека выйдут за рамки популяционных дескрипторов, основанных на исторических социальных конструктах, таких как раса.**
- 5. Исследования, включающие анализ последовательностей генома и связанной с ними фенотипической информации для миллионов людей-участников, будут регулярно демонстрироваться на школьных научных фестивалях.**

СМЕЛЫЕ ПРОГНОЗЫ ДЛЯ ГЕНОМИКИ ЧЕЛОВЕКА К 2030 ГОДУ *(продолжение)*

6. Регулярное использование геномной информации будет переходить от чего-то чрезвычайного к обычной практике в клиниках, делая геномное тестирование таким же рутинным, как и полный анализ крови.

7. Клиническая значимость всех встречающихся геномных вариантов будет легко предсказуема, что сделает диагностическое обозначение "вариант неопределенной значимости (VUS)" устаревшим.

8. Полная последовательность генома человека вместе с информативными аннотациями при желании будет надежно и легко доступна на его смартфоне.

9. Люди с наследственно разнообразным происхождением получат справедливую выгоду от достижений в области геномики человека.

10. Прорывные открытия приведут к передовым методам лечения, включающим геномные модификации для десятков генетических заболеваний.



The Nobel Prize in Chemistry 2020



© Nobel Media. Ill. Niklas Elmehed.

**Emmanuelle
Charpentier**

Prize share: 1/2



© Nobel Media. Ill. Niklas Elmehed.

Jennifer A. Doudna

Prize share: 1/2

The Nobel Prize in Chemistry 2020 was awarded jointly to Emmanuelle Charpentier and Jennifer A. Doudna "for the development of a method for genome editing."

Albany Medical Center Prize in Medicine and Biomedical Research - 2017

<http://www.amc.edu/Academic/AlbanyPrize/>

Albany Medical Center Prize Awarded to Distinguished Scientists

Five researchers, Emanuelle Charpentier, Ph.D., Jennifer A. Doudna, Ph.D., Luciano Marraffini, Ph.D., Francisco Juan Martínez Mojica, Ph.D., and Feng Zhang, Ph.D., were presented with this year's award.

Learn More



Лауреатами премии стали Эммануэль Шарпентье (Emmanuelle Charpentier) из Института инфекционной биологии Макса Планка (Max Planck Institute for Infection Biology) в Германии, Дженифер Дудна (Jennifer Doudna) из Калифорнийского университета в Беркли (University of California, Berkeley), Лучиано Маррафини (Luciano Marraffini) из Рокфеллеровского университета (The Rockefeller University) в Нью-Йорке, Франциско Моджика (Francisco J.M. Mojica) из Университета Аликанте (University of Alicante) в Испании и Фэн Чжан (Feng Zhang) из Гарвардского университета (Harvard University) и Massachusetts Institute of Technology.

CRISPR

Clustered Regularly Interspaced Short Palindromic Repeats

Расположенные регулярными кластерами короткие палиндромные повторы

DR - Direct Repeats

TREP - Tandem REPEATS

DVR - Direct Variable Repeats

SRSRs - Short Regularly Spaced Repeats

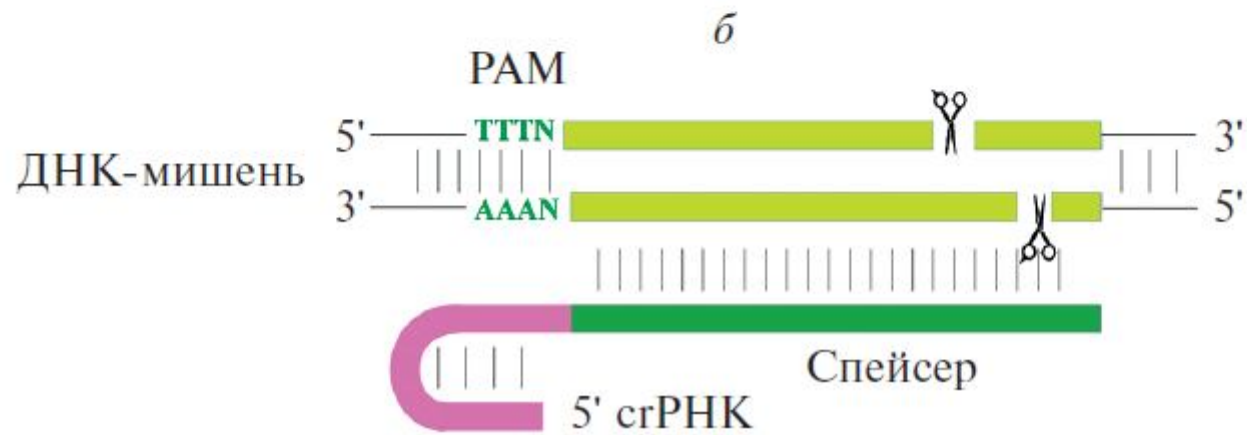
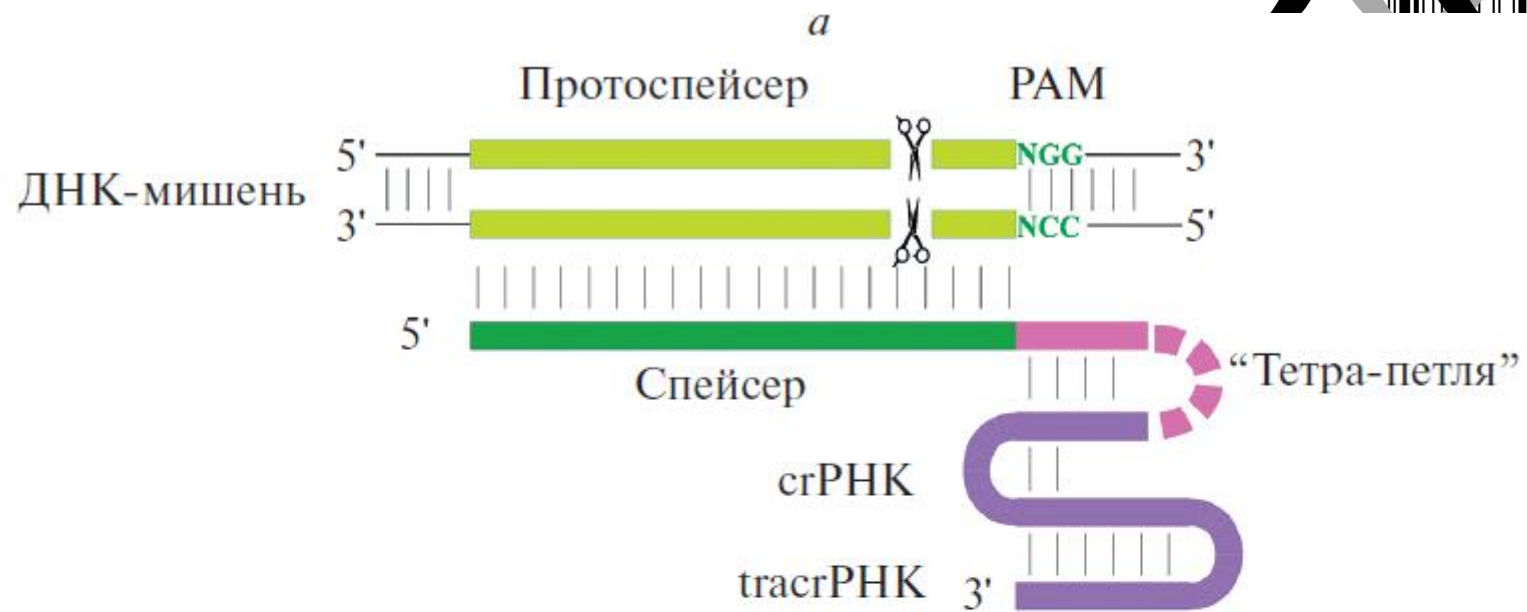
LCTR - Large Clusters of Tandem Repeats

SPIDR - Spacer Interspersed Direct Repeats

LTRR - Long Tandemly Repeated Repetitive sequences



**Упрощенная схема организации функционального CRISPR-локуса
(масштаб не соблюден)**



Этапы CRISPR/Cas-технологии геномного редактирования

Выбор задач и мишеней, под которыми понимается что предстоит сделать экспериментатору - нарушить ли работу какого-то конкретного гена либо изменить свойства кодируемого им белка, или внедрить некий новый ген. При этом существует несколько вариантов геномного редактирования на основе системы CRISPR/Cas: **а) нокаутные мутации (knock-out – KO)**, когда нарушается какой-либо ген, в том числе негативно влияющий на некий признак, важный для человека, сопровождаемые инсерциями или делециями ряда нуклеотидов в участке редактирования за счет негомологичного объединения репарируемых концов фрагментов ДНК в месте возникающего под действием Cas-нуклеазы двухцепочечного разрыва; **б) геномное редактирование, в ходе которого за счет гомологичной рекомбинации в определенные участки генома по местам двухцепочечных разрывов (получившее название нокин-вариант (knock-in – KI))** происходят встраивания иных нуклеотидных последовательностей в виде олигонуклеотидов или более протяженных фрагментов ДНК, как нарушающих работу целевого гена (это тоже разновидность нокаутирования того или иного гена), так и приводящих к появлению новых желательных признаков; **в) геномное редактирование без образования двухцепочечных разрывов ДНК и соответственно без делеций или инсерций в виде произведения однонуклеотидных замен (base editing – BE)** за счет направленного дезаминирования конкретных азотистых оснований с помощью «сшитых» с каталитически неактивными нуклеазами соответствующих дезаминаз (и некоторых других ферментов), обеспечивающее появление новых признаков, либо изменение имевшихся, включая нокаутирование генов вследствие образования терминирующих кодонов; **г) так называемое прайм-редактирование (prime editing – PE)**, когда за счет образующихся поочередно ников и действия сшитой с нуклеазой обратной транскриптазы происходит замена определенного участка ДНК. Помимо редактирования геномов или отдельных азотистых оснований с помощью CRISPR/Cas-технологии можно осуществлять также регуляцию работы отдельных генов в виде **CRISPRa (activation)** и **CRISPRi (interference)** вариантов, мишенью для которых служат промоторные области.

Дизайн гидовых РНК (гидРНК или РНК-гидов) под выбранные задачи и мишени, проводящийся с помощью компьютерных программ

Создание генно-инженерных конструкций, несущих необходимые для редактирования компоненты в виде подходящих Cas нуклеаз и гидРНК, а также донорной ДНК или образование из Cas нуклеаз и гидРНК соответствующих рибонуклеопротеидных частиц

Доставка этих CRISPR/Cas-компонентов в ядро клетки.

Проверка правильности редактирования генома с помощью тех или иных методов, включая полногеномное секвенирование.

УДК 577.21

**ДИЗАЙН РНК-ГИДОВ ДЛЯ CRISPR/Cas
РЕДАКТИРОВАНИЯ ГЕНОМОВ РАСТЕНИЙ**

© 2020 г. Г. А. Герашенков^{a, *}, Н. А. Рожнова^a, Б. Р. Кулуев^a, О. Ю. Кирьянова^{a, b},
Г. Р. Гумерова^a, А. В. Князев^a, З. Р. Вершинина^a, Е. В. Михайлова^a, Д. А. Чемерис^a,
Р. Т. Матниязов^a, Ан. Х. Баймиев^a, И. М. Губайдуллин^c, Ал. Х. Баймиев^a, А. В. Чемерис^a

Программный продукт
(по алфавиту)

Benchling
Breaking Cas
CCTop
CGAT
CHOPCHOP
CLD
CRISPOR
CRISPRdirect
CRISPR-GE^d
CRISPR GRNA Design Tool
CRISPR MultiTargeter
CRISPR-P
CRISPR-Local
CRISPR-PLANT
CRISPR Primer Designer
CRISPR RGEN Tools
CRISTA
CT-Finder
CRISPR-DT
CRISPR-RT
DESKGEN
Dharmacon CRISPR Design Tool
E-CRISP
GT-Scan
PhytoCRISP-Ex
Synthego CRISPR Design Tool
WheatCrispr

Программы дизайна sgRNA,
работающие on-line (по алфавиту):

CasBLASTR (<http://www.casblastr.org>); CasOT (<http://casot.cbi.pku.edu.cn>); COSMID (<https://crispr.bme.gatech.edu>); CREATE Designer (<http://www.thebioverse.org>); CRISPCut (<http://web.iitd.ac.in/crispcut/webserver/index.html>); CRISPETA (<http://crispeta.org.eu>); CRISPR4P (<http://bahlerweb.cs.ucl.ac.uk/cgi-bin/crispr4p/webapp.py>); CRISPR-Cas9 guide RNA design checker (https://eu.idtdna.com/site/order/designtool/index/crispr_sequence); CRISPR-DO (<http://cistrome.org/crispr/>); CRISPR Efficiency Predictor (<http://www.flyrnai.org/evaluateCrispr/>); CRISPR-ERA (<http://crispr-era.stanford.edu>); CRISPR-FOCUS (<http://cistrome.org/crispr-focus/>); CrispRGold (<http://crisprgold.mdc-berlin.de>); CRISPR Mapper (<http://crdd.osdd.net/servers/crisprge/mapper.php>); CRISPR.ML (<https://crispr.ml>); CRISPRoff (<https://rth.dk/resources/crispr/crisproff/>); CRISPR Optimal Target Finder (<http://targetfinder.flycrispr.neuro.brown.edu>); CRISPR-PN (<http://www.crispr-pn.net>); CRISPRScan (<http://www.crisprscan.org>); CRISPR sgRNA Design Tool (<http://www.genscript.com/gRNA-design-tool.html>); CRISPR-SKIP (<http://song.igb.illinois.edu/crispr-skip>); CRISPy CHO (<http://staff.biosustain.dtu.dk/laeb/crispy>); CRISPyS (<http://multicrispr.tau.ac.il>); CRISPy-web (<https://crispy.secondarymetabolites.org>); CROP-IT (<http://www.adlilab.org/CROP-IT/cas9tool.html>); DeepCRISPR (<http://www.deepcrispr.net>); EuPaGDT (<http://grna.ctegd.uga.edu>); FORECasT (<https://partslab.sanger.ac.uk/FORECasT>); GB CRISPR Tools (<https://gbcloining.upv.es/tools/crisprs>); ge-CRISPR (<http://bioinfo.imtech.res.in/manojk/gecrispr/index.php>); GPP Web Portal (<https://portals.broadinstitute.org/gpp/public>); Green Listed (<http://greenlisted.cmm.ki.se>); grID (<http://crispr.technology>); Guide RNA Generator (<http://penchovsky.atwebpages.com/applications.php?page=48>); GUIDES (<http://guides.sanjanalab.org/#/>); GuideScan (<http://guidescan.com>); inDelphi (<https://www.crisprindelphi.design/about>); Mojo Hand (<http://www.talendesign.org>); Off-Spotter (<https://cm.jefferson.edu/Off-Spotter/>); PAVOOC (<https://pavooc.me>); sgRNA Designer (CRISPRko) (<https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design>); Stupar Lab's CRISPR Design (<http://stuparcrispr.cfans.umn.edu/CRISPR/>); ZiFiT (<http://zifit.partners.org/ZiFiT>)

Программы дизайна sgRNA,
работающие автономно (по
алфавиту):

Azimuth (<https://github.com/MicrosoftResearch/Azimuth>); CasFinder (<http://arep.med.harvard.edu/CasFinder>); CASPER (<https://github.com/trinhlab/casper>); Crisflash (<https://github.com/crisflash/crisflash>); CRISPR-Analyser (<https://github.com/htgt/CRISPR-Analyser>); CRISPRer (<http://jstacs.de/index.php/CRISPRer>); CRISPRO (<https://gitlab.com/bauerlab/crispro>); CRISPR-offfinder (<https://sourceforge.net/projects/crispr-offfinder-v1-2/>); CRISSPRpred (<https://github.com/khaled-rahman/CRISPRpred>); CRISPRseek (<http://bioconductor.org/packages/release/bioc/html/CRISPRseek.html>); DeepCas9 (<https://github.com/lje00006/DeepCas9>); FlashFry (<https://github.com/aaronmck/FlashFry>); GRIBCG (<https://sourceforge.net/projects/gribcg/>); MENTHU (<http://genesculpt.org/menthu>); pgRNAFinder (<https://github.com/xiexiaowei/pgRNAFinder>); predictSGRNA (<http://www.ams.sunysb.edu/~pfkuan/software.html#predictsgm>); sgRNACas9 (<https://sourceforge.net/projects/sgrnacas9>); WU-CRISPR (<http://crispr.wustl.edu>).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
1	+/+	+/-			+/-		+/-	+/+			+	+/-	+/+	+	+	+	+			+/+		+/+				+	
2	+/-	+/-		+			+/-					+/-	+/-		+	+		+	+	+/+	+	+/-	+/-	+	+		
3	+/+	+/-				+	+/-		+	+		+/+	+/-			+	+			+/+		+/-				+	
4	+/+	+/-		+				+/-				+/-	-/+		+	+	+		+	+/+			+/+			+	
5	+/+	+/-			+/-	+	+				+	+/+	+/+		+	+			+	+/-					+		
6	+/+	+/-			+/+		+		+		+	-/+			+		+			+/+	+	+/-					
7	+/+	+/-					+/+	+/+	+		+	+/+		+	+	+				+/+						+	
8	+/+	+/-		+	+/+			+/-	+		+	+/-	+/+		+				+	+/+		+/-	+/-				
9	+/+	+		+			+		+			+/+			+	+				+/+	+				+		
10	+/-				+/-		+/-		+		+	+/-	+/-	+	+	+	+	+		-/+	+			+			
11	+/-	+/-	+/-		+/+		+/+		+	+		+/-								+/+		+/-				+	
12	+/-			+	+/-		+/-	+/-				+/+	+/-	+		+				+/+			+/-				
13	+/-						+/-				+	+/+	+/+			+				+/+		+/-					
14	+/-											+/+	+/+			+			+	+/+	+	+/-	+	+/-			
15	+/-												+/+		+	+				+/+						+	
16	+/-					+			+			+/-			+	+	+			+/+						+	
17	+/-							+/-				+/-			+	+				+/+					+	+	+
18	+/-							+/-	+			+/+			+				+	+/+						+	+
19	+/+	+/+	-/+				-/+					+/+	+/-		+	+				+/+						+	
20	+/-											-/+	+/+	+	+	+				+/+	+					+	+

1) CRISPR-P 2.0 \ CRISPR Local; 2) CRISPR-GE; 3) CRISPR RGEN Tools; 4) CRISPOR; 5) Benchling; 6) DESKGEN; 7) CCTop; 8) CHOPCHOP; 9) Breaking Cas; 10) E-CRISP; 11) CT-Finder \ CRISPR-DT \ CRISPR-RT; 12) CRISPR MultiTargeter; 13) CGAT; 14) CRISPRdirect; 15) CRISPR-PLANT; 16) GT-Scan; 17) CRISTA; 18) PhytoCRISP-Ex. 19) AsCRISPR; 20) SNP-CRISPR

A – wtSpCas9 nuclease/orthologues and Cas9 mutants; **B** – Cpf1 (Cas12a) / Cas12b; **C** – C2c2 (Cas13a) / CasX (cas12e); **D** – Custom PAM; **E** – nickases / FokI-Cas9; **F** – nuclease-deaminase; **G** – (proto)spacer length / seed length; **H** – 5'-end of gRNA/in vivo transcription promoter *in vivo*; **I** – mismatches;

J – indels in spacers and protospacers; **K** – GC-content in protospacers; **L** – input of DNA through the paste / as a file; **M** – input of individual genes using an identifier (gene name, Accession Number) / input of DNA using the genome coordinates; **N** – multiple sequences; **O** – ranking of sgRNA; **P** – off-target sites; **Q** – microhomology; **R** – flanking regions; **S** – restriction sites; **T** – both DNA strands / edited region (exon, intron, intergenic spacer); **U** – presence of the TTT(T) sequence; **V** – GC-content in protospacers / secondary gRNA structure (constant and variable parts); **W** – oligonucleotides and primers for cloning / PCR detection; **X** – sgRNA validation; **Y** – demo version (manual); **Z** – off-line.



ДНК-криминалистика



Биомика, 2018, Том 10, № 1, 85-140



БИОМИКА/BIOMICS

<http://biomics.ru>



ЭВОЛЮЦИЯ ПОДХОДОВ К ДНК-ИДЕНТИФИКАЦИИ ЛИЧНОСТИ

Чемерис Д.А.¹, Сагитов А.М.², Аминев Ф.Г.³, Луценко В.И.³, Гарафутдинов Р.Р.¹, Сахабутдинова А.Р.¹,
Василов Р.Г.⁴, Алексеев Я.И.⁵, Сломинский П.А.⁶, Хуснутдинова Э.К.¹, Чемерис А.В.¹

Биомика, 2019, Том 11, № 3, 282-314



БИОМИКА/BIOMICS

ISSN 2221-6197 <http://biomicsj.ru>



ДНК-КРИМИНАЛИСТИКА – ЗАРОЖДЕНИЕ, СОВРЕМЕННОСТЬ И ПЕРСПЕКТИВЫ

Анисимов В.А.¹, Гарафутдинов Р.Р.², Сагитов А.М.³, Сахабутдинова А.Р.²,
Хуснутдинова Э.К.², Аминев Ф.Г.¹, Чемерис А.В.²



БИОМИКС/БИОМИКА

ISSN 2221-6197 <http://biomicsj.ru>

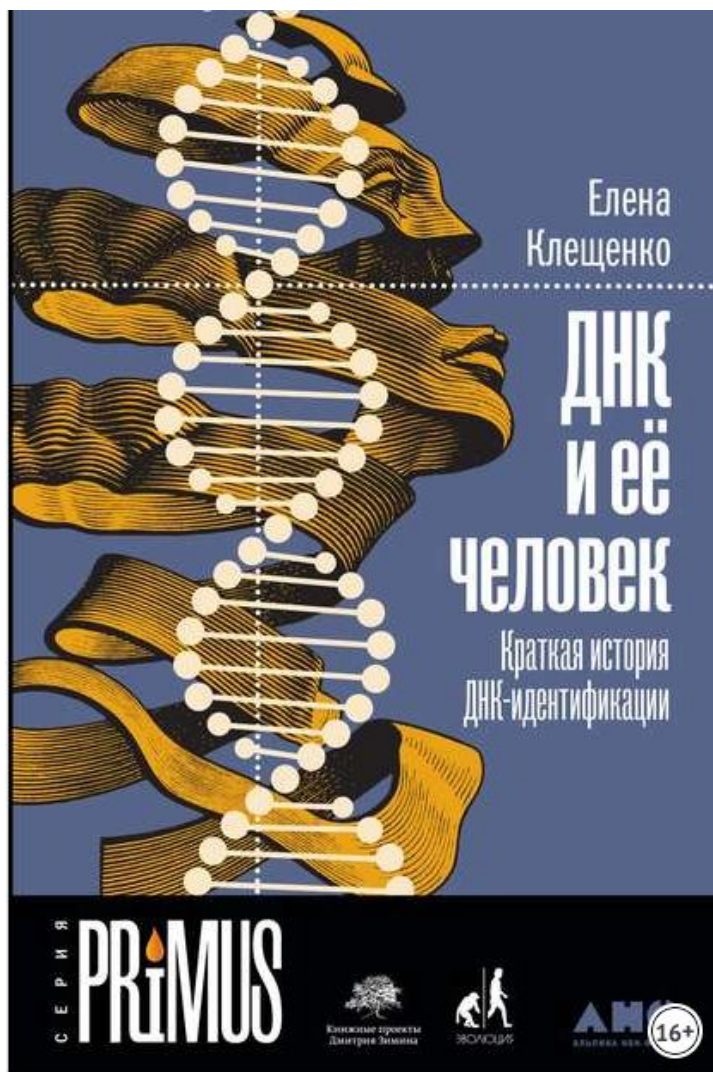


МИКРОДИПЛОТИПЫ КАК НОВЫЕ МАРКЕРЫ ДЛЯ ДНК-ИДЕНТИФИКАЦИИ ЛИЧНОСТИ

¹Чемерис Д.А., ²Гарафутдинов Р.Р., ³Сагитов А.М., ⁴Сагитова М.А., ⁵Михайленко К.И.,
⁶Зубов В.В., ⁷Василов Р.Г., ⁸Сломинский П.А., ⁹Анисимов В.А., ²Хуснутдинова Э.К.,
^{10,11}Алексеев Я.И., ¹¹Курочкин В.Е., ¹¹Лавров Г.С., ¹²Воробьев А.А., ⁹Аминов Ф.Г., ²Чемерис А.В.



ДНК- криминалистика



Стр. 247

«...надо признать, что включение того или иного объема ДНК-данных в биометрию паспортов всего мира – скорее всего, лишь вопрос времени. Слишком уж удобен этот метод идентификации».



ELSEVIER

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

Technical Note

A new digital approach to SNP encoding for DNA identification

Ravil R. Garafutdinov^{a,*}, Assol R. Sakhabutdinova^a, Petr A. Slominsky^b, Farit G. Aminev^c, Alexey V. Chemeris^a

Digitization	Binary encoding	1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	0000
	Hexadecimal encoding	8	4	2	1	C	A	9	6	5	3	0
Graphics	One-dimensional (linear) format											
	Two-dimensional format											

Linear



Цифровизация данных по однонуклеотидному полиморфизму ДНК и проект записи ДНК-идентификаторов личности в виде геномного штрих-кода

TTTGTGGGGTTATCTGGCAGATAGA[A/C/G/T]GGTGACAACAGTCCCATCTGGTGAA

ACGT	ACGT	ACGT	ACGT	ACGT	ACGT	ACGT	ACGT	ACGT	ACGT	ACGT
A A/A-	C C/C-	G G/G-	T T/T-	A C	A G	A T	C G	C T	G T	- -
A	C	G	T	AC	AG	AT	CG	CT	GT	
1	1	1	1	2	2	2	2	2	2	0
1000	0100	0010	0001	1100	1010	1001	0110	0101	0011	0000
8	4	2	1	C	A	9	6	5	3	0

Стандартная оцифровка нуклеотидов

A – 00
C – 01
G – 10
T – 11

8 бит

По нашему принципу оцифровки для одного снипа достаточно всего 4 бит (binary digit) информации

C AC GT C AG G T G C A CG G C AC GT CT T A T T AT AC AT T AT AC CG AC T AC T T CG AC GT
AC CG G GT CT CG T GT AC G CG A AC

01001100001101001010001000010010010001100011001001001100001100100100110000110101000110000001000110011100100100011
00111000110110000011100000100010110110000111100011000100011010101100001001111000010011010001100

Наиболее оптимальной представляется невидимая магнитная форма записи из «0» и «1»



линейный формат

индивид А



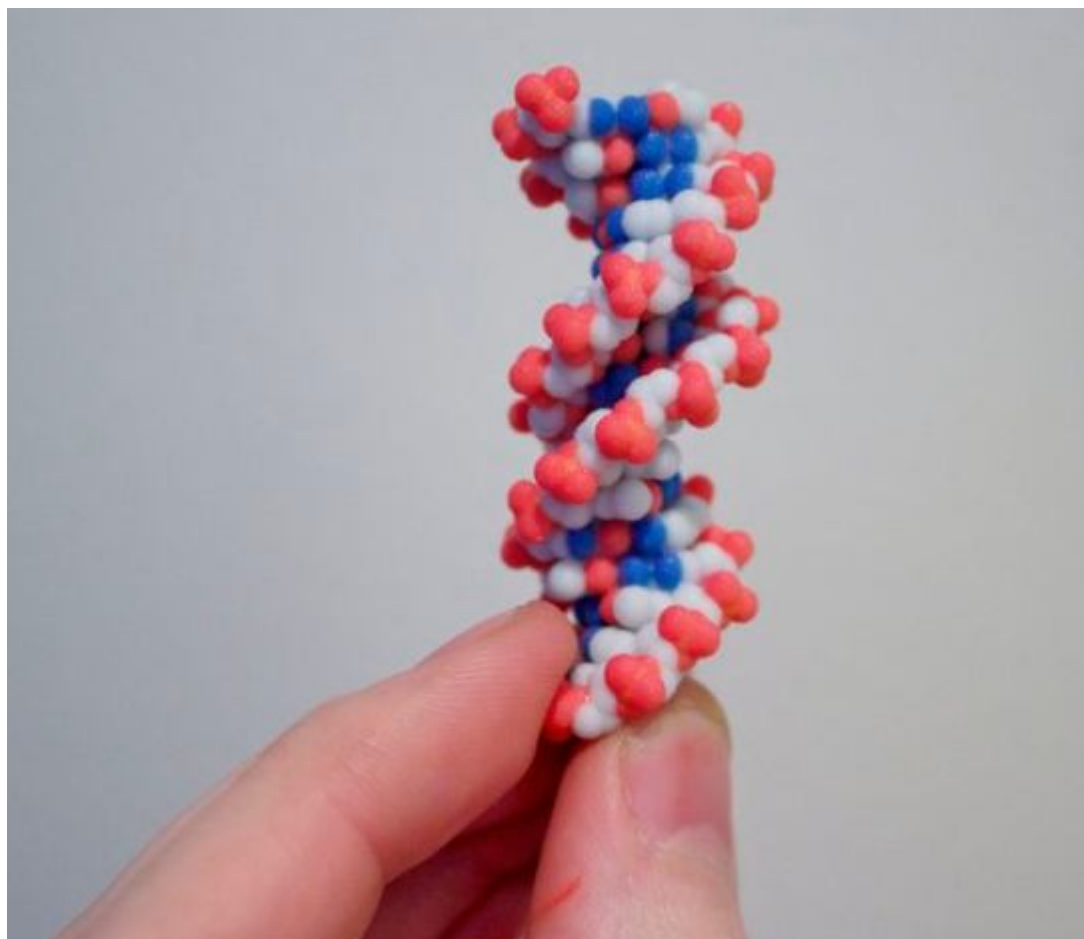
индивид В



двумерный формат

Объем файла хранения в базе данных для одного человека не более 1 килобайта!

Красным цветом показаны отличия между индивидами



Благодарю за внимание