

Ang Zenome Project: Whitepaper *blockchain-based genomic ecosystem*

Nikolay Kulemin

Sergey Popov

Alexey Gorbachev

Oktubre 6, 2017

Abstrak

Ang Industry 4.0 ay isang pamagat para sa kasalukuyang trend ng automation, scaling at data exchange sa manufacturing technologies. Kabilang dito ang artificial intelligence, virtual katotohanan, ang Internet ng mga bagay at pagtatasa ng Big Data. Ang mga genomics ay isang malinawkinatawan ng industriya 4.0 na nangangailangan ng paglutas ng maraming mga kagyat na problema, tulad ng imbakan at pagtatasa ng Big Data sa pagpapanatiling pampublikong access para sa mga mananaliksik at privacy para sa mga tao.

Sa kasalukuyan ay may problema sa hindi pagkakapantay-pantay sa industriya ng genome. Ibig sabihin na ang pangunahing bahagi ng personal na data ng genomic ay puro sa mga sentro ng data ng mga korporasyong genomic, gobyerno, pang-aham at medikal na institusyon at pharmaceutical companies. Bukod dito, may isang isyu ng legal na limitasyon ng pag-access sa personal na genomic data, pati na rin ang kawalan ng posibilidad para sa genomic data pamamahala at pagbabahagi. Ang pagbabagong-anyo ng genomic na data na ito ay nagbabawas ang pag-unlad sa isang bilang ng mga pang-aham at medikal na mga patlang.

Ang pag-unlad ng cryptocurrencies at mga teknolohiya na batay sa blockchain ay humahantong sa makabuluhang pagbabagong-anyo ng maraming mga ekonomiyang pang-ekonomiya. Aplikasyon ng blockchain na diskarte ay magiging isang lifeline na nagbibigay-daan upang i-upgrade ang pag-unlad ng personal na genomics. Gagawin ng bawat tao ang may-ari ng kanyang genetic data.

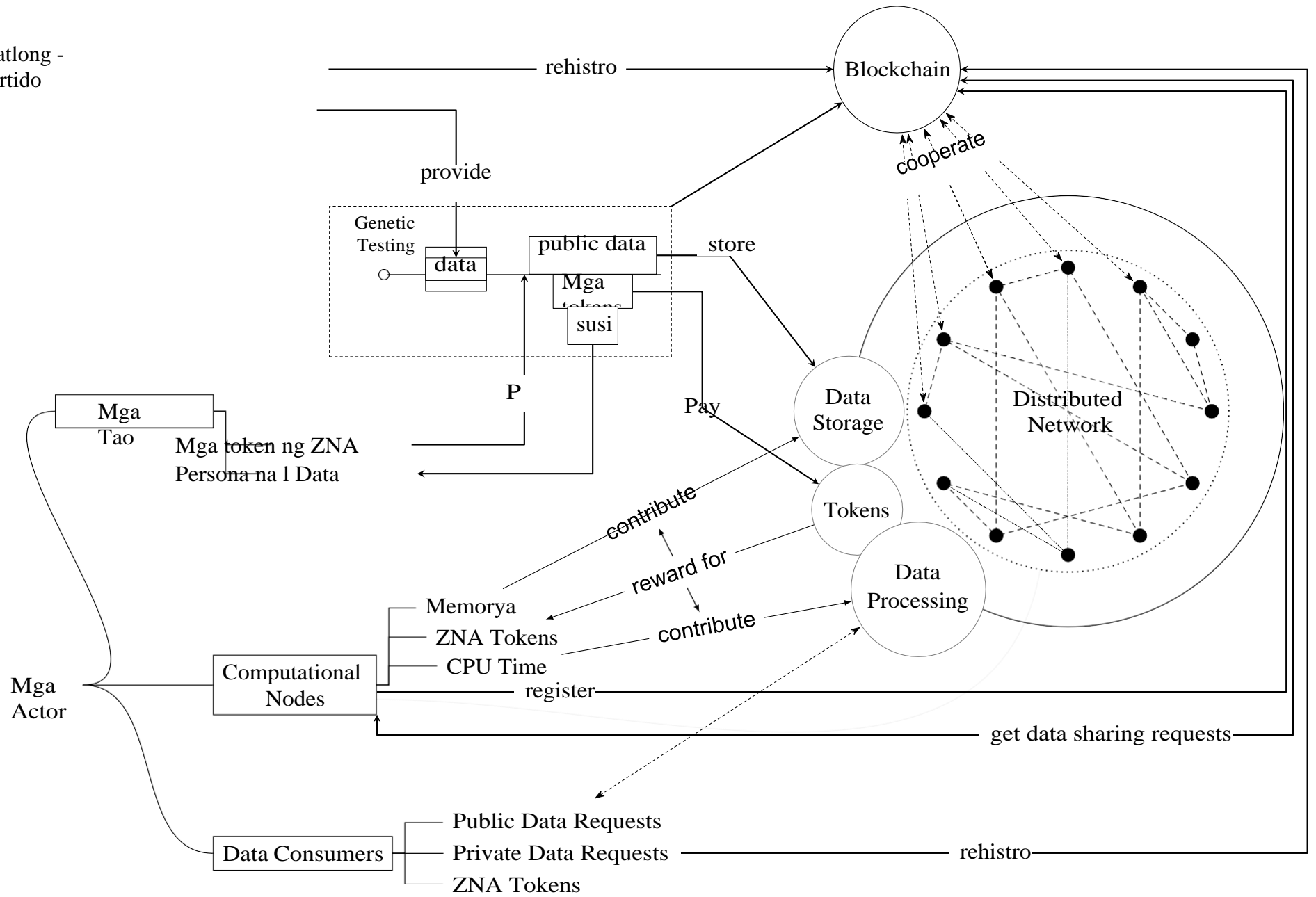
Ang proyektong Zenome ay isang desentralisado na database na hinimok ng blockchain ng genomic impormasyon. Sinusuportahan ng platform na ito ang posibilidad na pamahalaan ang iyong genomic data habang pinapanatili ang privacy at kakayahang gumawa ng kita mula sa pagbebenta ng access sa iba't ibang bahagi ng genome. Itatatag ang pantay na kondisyon para sa droga pag-unlad at para sa pag-unlad ng pang-aham at medikal na teknolohiya.

Ang Zenome ay isang bagong pang-ekonomiyang kapaligiran batay sa genomic data at blockchain teknolohiya. Ang pagpapatupad ng aming haka-haka modelo ay malulutas ang mga sumusunod na kahirapan:

- Paglikha ng isang imprastruktura para sa pagtatago ng Big genomic data gamit ang ibinahagi database
- Buksan ang access sa milyun-milyong genome ng tao sa buong mundo na may proteksyon sa pagkapribado
- Posibilidad para sa bawat tao na sumali sa siyentipiko at klinikal na pananaliksik at upang makinabang mula dito
- Pinasisigla ang pagpapahusay ng mga agham ng genom sa pag-unlad ng mga bansa at de-monopolisasyon ng genomic data sa mga binuo bansa

Ikatlong -
Partido

1



Ang Koponan

Alexey Gorbachev

Tagapagtatag

*Molecular biologist at Blockchain Enthusiast
Ph.D. ng molecular biology at biochemistry*

Unibersidad: Moscow State University, Kagawaran ng Molecular Biology.

Si Alexey ay may isang mahalagang pang-agham na background, kadalubhasaan sa negosyo, at proyekto pamamahala. Sa Zenome Alexey ay responsable para sa pangkalahatang pangitain at pag-unlad ng negosyo..

Email: alex@zenome.io

LinkedIn: <https://www.linkedin.com/in/alexey-gorbachev-24b5305b/>

Nikolay Kulemin

Tagapagtatag

Ph.D. in bioinformatics

Unibersidad: Moscow Institute of Physics and Technology,
Kagawaran ng Molecular at Biological Physics.

Espesyalista sa bioinformatics at matematika na biology, ay may karanasan sa akademiko pananaliksik at pagpapatupad ng mga pagpapaunlad sa industriya ng genome. Si Nikolay ay isang tagapagtatag ng isang kumpanya na bumubuo ng mga bagong algorithm para sa genomic analysis.

Email: nick@zenome.io

LinkedIn: <https://www.linkedin.com/in/nikolay-kulemin-50ab47a9/>

Vladimir Naumov

Genomic data scientist

Data scientist sa human genomics

Unibersidad: Pirogov Russian National Research Medical University

Siyentipiko sa GERO, 8 taon sa industriya ng genomics. Gumagana sa paglikha ng mga pipelines at pagbuo ng mga bagong paraan sa pagsusuri at visualisasyon ng genomic data.

Email: vov@zenome.io

LinkedIn: <https://www.linkedin.com/in/vladimir-naumov-8285a25b/>

Sergey Popov

Developer

Blockchain, P2P, distributed systems, purong matematika

Unibersidad: Moscow Institute of Physics at Technology,
Kagawaran ng General and Applied Physics

May karanasan sa panteorya ng pisika, teoretikal na impormasyon, purong matematika, pag-unlad ng mga sistema ng ibinahagi at pag-unlad ng smart-contract.

Email: sp@zenome.io

Dmitry Kwon

Tagapagturo

Ph.D. sa molecular biology, business development manager

Unibersidad: Moscow State University, kagawarn ng molecular biology

Espesyalista sa molekular genetika. Dmitry ay may isang makabuluhang pang-agham na background, kadalubhasaan ng teknolohiya ng genetic analysis, genomic at diagnostic market, ay may matagumpay na karanasan sa negosyo sa biotechnology at Dx top intl companies

.Email: dk@zenome.io

LinkedIn: <https://www.linkedin.com/in/dmitry-kwon-2763b119/>

MGA NILALAMAN

ANG KOPONAN

GENOME TOKENOMICS

KABANATA 1- GENOMICS

BACKGROUND 7

Genome (pahina 7). Ang International Human Genome Project (pahina 8). Pag-unlad ng pagtatasa ng genome (pahina 8). Epekto sa iba pang mga agham (p.9). Mga database ng genetic (pahina 10).

GENOMIC MARKET OVERVIEW 10

Genomic market growth dinamika (pahina 10). Mga lugar ng produkto (pahina 12).

MGA HAMON PARA SA GENOMICS..... 12

Mababang availability ng pagtatasa ng genome (pahina 13). Pagkompromiso sa privacy bilang isang presyo para sa paglahok (pahina 14). Multicentral na pang-agham at klinikal na pag-aaral (pahina 14). Mga materyal sa pag-iimbak ng Biobanks mula sa iba't ibang indibidwal (p. 14). Pagsasaayos ng analytical software para sa laganap na aplikasyon (pahina 15). Genomic interpretasyon ng data (pahina 15). Genomic interpretasyon ng data (pahina 16). Secure pagtatago ng napakalaking data (pahina 17). Patuloy ang database na-update na mga questionnaire (pahina 18). Ang malapit na hinaharap ng genomics (pahina 20).

MGA TALAKAYANG ISYU NG PERSONALANG GENOMICS.....20

Privacy (pahina 21). Pampubliko (p. 21). Ang karapatang magkaroon ng data ng genomic? (p. 22). Ang karapatang ma-access ang genomic impormasyon (pahina 23).

KABANATA 2 - KONSEPTO

ANG PROYEKTONG ZENOME24

Pilosopikong Pananaw (pahina 24). Ang Zenome Platform Ecosystem (pahina 25). Ang System Architecture's Pangkalahatang-ideya (pahina 25). Mga serbisyo ng genetic na serbisyo (p. 27).

MGA GAWAIN SA PLATAPORMANG ZENOME 28

Mula sa Mga Likas na Resource Nodes (p. 28). Tao / Gumagamit (pahina 29). Data Consumer (pahina 30). Serbisyo Tagapaglaan (p. 30).

GENOMIC DATA31

Mga uri ng data ng omics (genomic) (p. 31). Preprocessing ng genetic data (p. 31). Ang Problema sa Peking Data (pahina 32). Ang problema ng pagkakaakilanlan ng user batay sa genomic data (pahina 32). Imbakang ng genomic data (pahina 33).

PERSONAL NA PROFILES.....	33
<i>Pagtutukoy ng palatanungan (pahina 34). Mga query sa system. (p. 34). Secure transfer ng personal</i>	
SISTEMANG RATING	36
GAMITIN ANG KASO	36
<i>Indibidwal na gumagamit (pahina 36). Kalusugan (p. 37). Kumpanya (pahina 38). Ang pang-agham na komunidad (p. 38).</i>	

KABANATA 3 – TEKNIKAL NA BAHAGI

IPINAMAHAGING BAGAY	39
----------------------------------	-----------

Isang konsepto ng ibinahagi subsystem (pahina 39). Panloob na proseso (pahina 40).

PANGUNAHING IPINAMAHAGING SUBSYSTEMS NG PLATFORM	40
MABABANG ANTAS NG INTEROPERATION LAYER.....	41

Ibinahagi ang P2P network (pahina 41). Ibinahagi ang mga katangian ng pagpapatupad ng network (pahina 41). Palitan ng mensahe sa isang ipinamamahagi na kapaligiran (pahina 42)

BLOCKCHAIN.....	42
------------------------	-----------

Pangunahing impormasyon (pahina 42). Ang isang subsystem para sa pakikipagtulungan sa Blockchain (p. 43). Panloob na mga token (p. 43). Konsepto ng account (pahina 44).

IPINMAHAGING DATA STORAGE NETWORK.....	44
---	-----------

Ang mga prinsipyo ng ibinahagi sa operasyon ng network (pahina 44). Pagkatiwalaan ng pagiging maaasahan ng imbakan ng data (pahina 45). Imbakan ng data pagkatiwalaan sa privacy (pahina 45). Mga partikular na tampok ng pag-iimbak ng impormasyon sa genomic (pahina 45).

USER’S PERSONAL DATA.....	46
----------------------------------	-----------

Questionnaire specification (p. 46).

Panimula: Genome Tokenomics

Sa mga sumusunod na isang pangkalahatang paglalarawan ay ipapakita ng Zenome platform. Ang pangangailangan para sa mga token at benepisyo para sa inaasahang mamumuhunan ay tinalakay.

Para sa pinaka-bahagi, ang genomic na impormasyon ay naka-imbak sa mga database, tinustusan ng mga pamahalaan o malalaking korporasyon. Isa-isa, naglalaman ang bawat database hindi sapat na data upang gawin ang quantum leap patungo sa isang panahon ng genomics at katumpakan gamot. Kasabay nito, ang bawat database ay naglalaman ng napakaraming impormasyon na imposible para sa isang solong kumpanya na iproseso ang lahat ng iyon.

Lumilitaw na ang pagpapalitan ng impormasyon sa genetiko ay napakahalaga. Ang prospektadong genetic market ay dapat tiyakin ang proteksyon mula sa posi bleng maling paggamit at lalo na ang genetic diskriminasyon. Napakahalaga na mapanatili transparency at pantay na pag-access sa market na ito.

Ang pangkalahatang pagpapalitan ng impormasyon sa genetiko ay dapat matugunan ang mga sumusunod na isyu:

- Ang fragmentation ng genetic data.
- Ang limitadong pag-access ng mga siyentipiko, medika at kumpanya sa genetic data.
- Mababang affordability ng genetic testing.
- Ang kakulangan ng pagkapribado ng mga sumang-ayon upang buksan ang pagbabahagi ng pag-access ng kanilang genomic data.
- Hindi sapat na mapagkukunan ng computing.

Nilalayan ni Zenome ang paglikha ng personal na genomics infrastructure, na magbibigay-daanbang mga kalahok sa:

- Mag-upload ng impormasyon sa genetiko at kontrolin ito
- Ligtas na mag-imbak ng sariling genetic na impormasyon.
- Gumawa ng tubo sa pamamagitan ng pagbebenta ng access sa genetic data o bahagi nito.
- Sa ilalim ng genetic testing kapalit ng karapatang gamitin ang genetic information.

- Kumuha ng mga indibidwal na rekomendasyon sa pagkain o batay sa personal na pagsasanay sa genetic makeup.
- Gumamit ng iba pang mga serbisyo ng genetic.

Ang mga pangunahing customer ng genetic na impormasyon ay mga kumpanya na interesado sa genetic pag-target tulad ng Google, Facebook, Unilever at mga kumpanya ng parmasyutiko.

Sa loob ng Zenome Platform iba't ibang uri ng impormasyon, katulad ng genomic, personal at pinansiyal na data, ay magkakaugnay na magkakaugnay. Ang partikular na katangian ng bawat uri tinutukoy ang paraan upang mag-imbak ng impormasyon ng ganitong uri. Financial data, na kung saan kasama ang mga talaan ng mga transaksyon, ay naka-imbak sa blockchain. Anonymized genomic ang data ay naka-imbak sa ipinamamahagi na network. Ang personal na data ng kalahok ay pinananatili ang kanilang sariling mga computer lamang. Ang pag gamot ng data ng iba't ibang uri ay diffently nagbibigay ng privacy pati na rin ang scalability ng system.

Dahil ang lahat ng transaksyon ng data, kabilang ang pagbili at pagbebenta ng data, ay pinamamahalaan ng smart-kontrata, na sumasalamin sa desentra lisadong katangian ng platform, mga pakikipag-ugnayan maaari lamang isa ma ang mga balanse na nakaimbak sa mga blockchain. Gamit ang alinman sa mga pre-existing na token para sa layuning ito ay magresulta sa hindi makatwirang pag-asa sa pagtatasa ng panlabas na token o barya. Kaya, ang isang hiwalay na utility na token ay dapat na ibibigay sa kapangyarihan pang-ekonomiyang pakikipag-ugnayan sa platform. Ito, sa partikular, ay nangangahulugan na hindi mo magagawa bumili ng isang genetic na data sa «normal» ng pera, kailangan mong kumuha ng mga token muna.

Zenome DNA (ZNA) ay isang utility token sa Zenome platform. Ang tasa ng Ang ZNA ay nakatali sa tagumpay ng platform.

Sa whitepaper na ito tatalakayin namin sa karagdagang detalye ang pinaka-pinindot na mga problema sa ang lugar ng genomics, pati na rin ang solusyon Zenome platform ay nagsasangkot.

Kabanata 1

Genomics

Background

Sa seksyong ito, ang mga kahulugan ng mga salitang "genome" at "genomics" ay ibinigay. Ang kasaysayan ng unang pagsisikap ng genomic sequencing at ang paglitaw ng high-throughput sequencing technology (NGS) ay isinasaalang-alang. Ang mga pangunahing tagagawa ng reagents at kagamitan na ginamit upang makuha ang genomic data ay inilarawan. Mga isyu tungkol sa pag-iipon itinuturing na genomic data at binabawasan ang gastos ng pagsusuri. Isang pagsusuri ng kasalukuyang database ng genomic.

Genome

The genome Ang genome ay ang kumpletong hanay ng mga tagubilin ng genetic na natagpuan sa isang cell [1].

Ang genome ay naglalaman ng biological na impormasyon na kinakailangan para sa pag-unlad at paggana ng isang organismo. Ang genome ng tao ay binubuo ng linear double-helical Ang mga molekula ng DNA ay isinaayos sa 22 pares ng chromosomes at dalawang sex chromosomes - X at Y. Ang lahat ng impormasyon na nakapaloob sa isang genome ay naka-encode gamit ang quaternary code sa pamamagitan ng isang pagkakasunud-sunod ng 4 nucleotides na itinalaga A, T, C, at G. Ang term "Upang basahin ang isang genome" ay nangangahulugang "upang matukoy ang isang sequence ng nucleotide sa pamamagitan ng isang sequencing proseso "[2].

Ang indibidwal na pagkakasunud-sunod ng isang genome ay tumutukoy sa iba't ibang organism mga tampok, kabilang ang hitsura, pagkamaramdamin sa ilang mga sakit, kakayahan sa atletiko, metabolismo, mga kagustuhan sa nutrisyon, pagkakatugma sa mga kasosyo sa sekswal (ang kakayahan upang maisip ang mga bata), at marami pang iba.

Ang International Human Genome Project

Ang International Human Genome Project¹ ay inilunsad sa ilalim ng pangangasiwa ng NIH (National Institutes of Health) noong 1990 upang matukoy ang kumpleto pagkakasunud-sunod ng haploid na human genome. Ang unang pinuno ng proyekto ay isa sa ang mga tagahanap ng istruktura ng DNA, ang nagwagi ng Nobel na premyo na si James Watson. Ang isang sunod na pagkakasunud-sunod ng genome ng tao ay natapos sa kalagitnaan ng 2000 at na-publish sa simula ng 2001 sa journal Nature.

Ang halaga ng internasyonal na proyekto na nakumpleto na may pampublikong pondo ay humigit-kumulang sa \$ 3 bilyon. Noong 1998, isang pribadong kumpanya, Celera Genomics, ang sumali sa lahi sa pagkakasunud-sunod ang genome ng tao. Ang pinuno ng pribadong proyekto, na binuo sa parallel sa institusyon ng pamahalaan, ay bantog na siyentipiko at negosyante Si Craig Venter, na namamahala sa \$ 300 milyon sa mga pribadong pamumuhunan para sa Celera proyekto. Sa pamamagitan ng paggamit ng bagong shotgun sequencing diskarte at mas produktibo computational methods, ang sequence ng Craig Venter's genome ay na-publish halos sabay-sabay sa data na ginawa ng internasyonal na kasunduan sa 2001 [3] sa journal Science. Ang "buong" genome ng tao ay na-publish noong 2007, at ang ilang mga tao genomic rehiyon na mahirap sa pagkakasunud-sunod ay mananatiling hindi kilala.

Pag-unlad ng pag-aaral ng genome

Malawak na pamumuhunan, isang malaking bilang ng mga natitirang kalahok mula sa pang-agham na komunidad, at kumpetisyon sa mga pribado at pampublikong organisasyon Nagbigay ng malaking tulong para sa pagbuo ng mga teknolohiya ng pag-aaral ng genome. Bilang isang resulta, modernong sequencing teknolohiya tulad ng NGS (susunod na henerasyon sequencing) ² ay lumitaw na kasama ang isang bagong sangay ng agham na tinatawag bioinformatics, isang batang larangan ng pananaliksik sa intersection ng matematika, IT, at biology, na bumubuo ng mga diskarte at mga algorithm para sa pagtatasa ng malalaking biological datasets sa produktibo at computationally epektibong paraan.

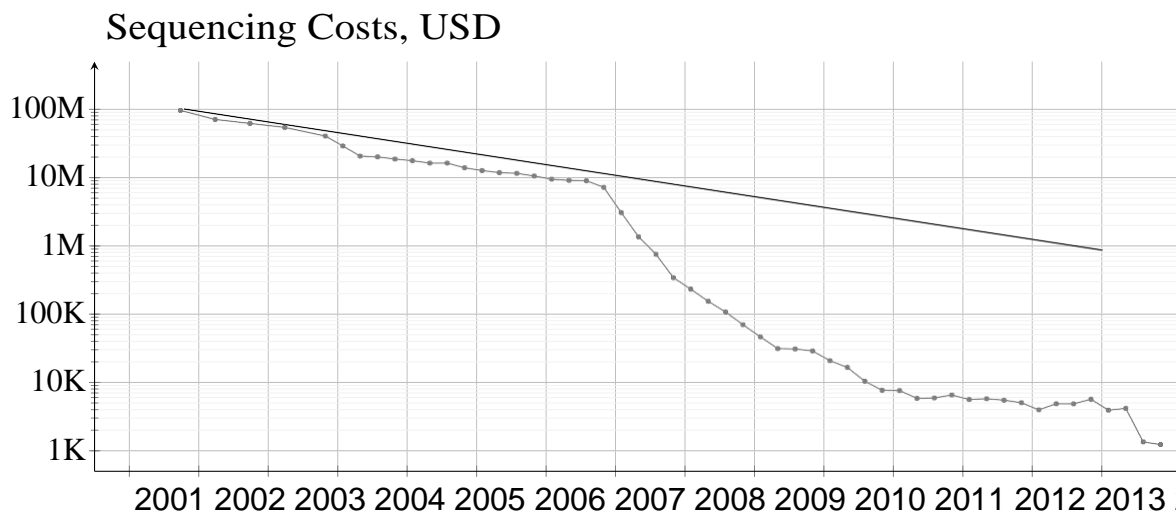
Ang paglitaw ng pangalawang at ikatlong henerasyon ng mga teknolohiya ng sequencing (NGS) ay may na humantong sa isang malakas na pagbabawas ng halaga ng pagtatasa ng genome. Habang kahit sa 2009 ang gastos ng isang buong pag-aaral ng isang genome ay tungkol sa 100,000 USD, sa kasalukuyan ang average na presyo para sa ang parehong pagsusuri ay bumaba sa humigit-kumulang na mas mababa sa 1,000 USD (tingnan ang Larawan 1 at Talahanayan 3).

¹https://en.wikipedia.org/wiki/Human_Genome_Project

²https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-you-will-learn/what-next-generation-dna-https://en.wikipedia.org/wiki/Moore's_law

Talahanayan 1: Mga developer ng kagamitan at mga reagent supplier para sa genomic sequencing. Kinuha ang kapitalisasyon ng merkado mula sa Yahoo Finance

Kumpanya	Mga Produkto	Capital-	Bansa
Illumina	Hardware, reagents, consumables, software	28.06 B	USA
Thermo Fisher Scientific	Hardware, reagents, consumables, software (isang bahagi ng negosyo)	68.98 B	USA
Oxford Nanopore Technologies	Hardware, reagents, consumables	534.41 M	UK
Pacific BioScience	Hardware, reagents, consumables	436,93 M	USA
Roche	Hardware, reagents, consumables, software (isang bahagi ng negosyo)	213.44 B	Switzerland
Agilent Technologies	Hardware, reagents, consumables, software (iasng bahagi ng negosyo)	19.32 B	USA



Anyo: Moore's Law at pagbabawas sa gastos ng genomic analysis. Ang isang makabuluhang drop sa mga presyo noong 2008 ay dahil sa pagdating ng mga susunod na henerasyon ng mga teknolohiya ng pagkakasunud-sunod (NGS) [4].

Epekto sa iba pang mga agham

Ang pag-unlad ng genomics (ang larangan ng agham na nag-aaral ng iba't ibang mga genome) ay may na humantong sa pagbabago ng maraming pang-agham na mga patlang, mula sa biology at antropolohiya sa medisina at kahit na agham panlipunan. Ang isang bilang ng mga nangungunang komersyal na kumpanya tulad ng Google, Apple, IBM, Amazon, at Alibaba ay nagtakda ng layunin ng paggamit genomics upang ayusin ang kanilang mga produkto at serbisyo ayon sa mga profile ng genomic ng kanilang mga customer. Ang mga naturang pagsasaayos ay magpapahintulot sa mga kumpanyang ito na maayos na tune user

relasyon at upang mahulaan ang mga pangangailangan at potensyal na aktibidad ng kanilang mga customer 3

Genetic databases

Ang pagbawas sa gastos ng pagkakasunod-sunod ay humantong sa isang pagpaparami ng pagtaas sa magagamit na data ng genomic. Halimbawa, ang kumpletong genome ng tao sa tinatawag na "raw data" na format ay maaaring kumatawan sa 50 GB hanggang 2 TB ng data (depende sa sequencing lalim kinakailangan). Upang mag-imbak tulad ng isang malaking halaga ng data ng genomic, espesyal na genomic database ay nilikha na naglalaman ng iba't ibang uri ng data, tulad ng raw data na nakuha mula sa genomic sequencers ("bumabasa" o "pagbabasa"), mga pagkakasunod-sunod ng mga gene at mga protina, mga hanay ng mga coding na rehiyon ng isang genome na tinatawag na exomes, at kahit na ang mga pagkakasunod-sunod ng buong genome (scaffolds); ang ilan sa mga database na ito ay naglalaman ng may kaugnayan sa clinically impormasyon pati na rin ang mga relasyon sa pagitan ng genetic traits at sakit. Karamihan ng mga database na ito ay pinamamahalaang sa gitna at tinustusan ng mga gobyerno o malaki mga korporasyon. Ang mga siyentipiko sa buong mundo ay kasangkot sa pagdaragdag ng bagong data sa ang mga database na ito, na nagpapagana ng mabilis na pag-update at pag-synchronise. Sa Talahanayan 2, ang ilan ang mga kilalang database na inilarawan.

Ang karamihan ng mga naturang database ay naka-host sa mga binuo bansa at naka-centrally pinamamahalaan at kontrolado ng mga pamahalaan. Access sa ilan sa mga database na ito ay limitado, kahit na para sa pang-angham na komunidad, o ay limitado sa pamamagitan ng komersyal mga subscription. Kahit na ang mga tagapagtatag ng mga database ng genomic ay nagsasabing sila ay ligtas at hindi nakikilalang nag-imbak ng genomic data, sa katotohanan, ang data na nakaimbak ay lamang pseudonymous, tulad ng sa ilang mga kaso, ang mga indibidwal ay nakilala batay sa kanilang genomic information [5].

Pangkalahatang ideya ng pangkalahatang merkado

Sa seksyon na ito, isang maikling pagsusuri ng mga teknolohiya ng genomic na teknolohiya ay ibinigay. Mga halimbawa ng mga pinakasikat na mga produkto ng genomic at mga kumpanya ang pagbibigay ng iba't ibang mga serbisyo ng genomic ay inilarawan.

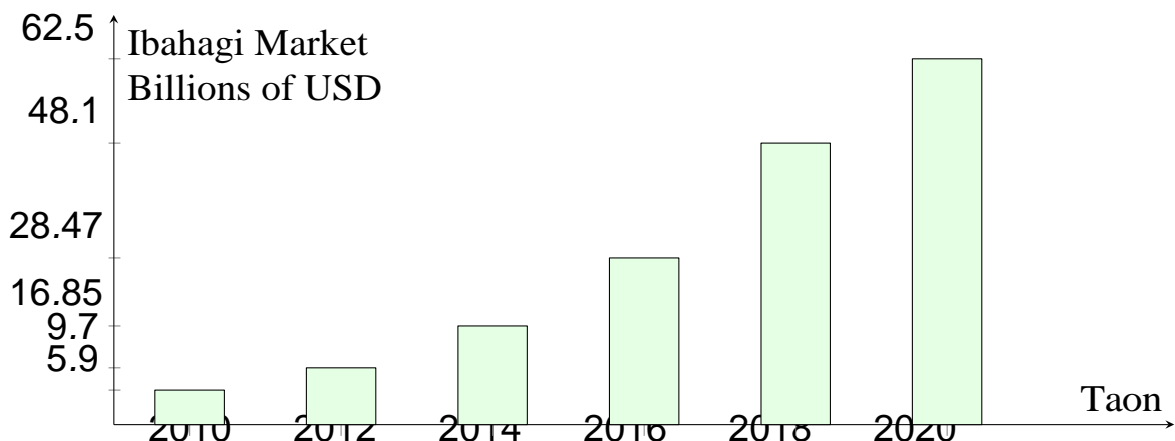
Genomic market growth dynamics

Ang merkado para sa mga teknolohiya ng genomic ay mabilis na lumalaki at lubos na maaasahan. Sa kasalukuyan, ang kabuuang dami ng merkado ay humigit-kumulang na \$ 25 bilyon na may halos sampung beses paglago, mula \$ 5.9 bilyon sa 2010 hanggang \$ 60 bilyon sa 2020 (hinulaang).

³<https://www.smeal.psu.edu/fcfe/documents/innovations-in-medical-genomics-pdf>

Talahanayan 2: Genomic databases

GenBank	http://exac.broadinstitute.org
May-ari: NCBI-NIH, USA Produkto: Genome sequences database Imbak: Higit sa 199,341,377 ibat ibang genome sequences	
ExaC	www.ncbi.nlm.nih.gov/genbank
May-ari: Broad Institute of MIT and Harvard, USA, ODC Open Database License (ODbL) Produkto: Exome Aggregation Consortium Imbak: 60,706 human exome samples/sequences	
UniprotKB	www.ebi.ac.uk/uniprot/
May-ari: EMBL-EBI, SIB, PIR, UK, Switzerland, USA Produkto: Open Knowledge Base. Manual expert curation. Proteins and genes sequences. Imbak: More than 555,100 manually reviewed and annotated record	
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/
May-ari: NCBI-NIH, USA Produkto: malayang magagamit na archive para sa interpretasyon ng klinikal na kahalagahan ng genomic variant para sa naiulat na kondisyon Naka-imbak: > 158 000 na isinumite ang mga interpretasyon, na kumakatawan sa > 125 000 na mga variant	
HGMD	http://www.hgmd.cf.ac.uk/ac/index.php
May-ari: QiaGen Produkto: Commercial database Naka-imbak: 208,368 mga rekord ng mutation ng tao na may mga anotasyon	
SNPedia	https://www.snpedia.com/index.php/SNPedia
May-ari: Open database Produkto: SNPedia is a wiki investigating human genetics Naka-imbak: 107,073 SNP at naka-link na mga tala	
1000 Genomes Project	www.1000genomes.org
May-ari: EMBL-EBI, Wellcome Trust layunin: hanapin ang karamihan sa mga variant ng genetiko na may mga frequency na hindi bababa sa 1% Nkaimbak: Higit sa 2,504 Mga sample / sequence ng Genome	
100000 Genomes Project	http://www.genomicsengland.co.uk/
May-ari: NHS, Government of UK Produkto: Ang database ng pamahalaan ng UK na naglalaman ng isang pagkakasunud-sunod ng 100,000 mga genome Nkaimbak: 32,642 Buong mga sequence ng genome	



Anyo 2: Genomic market dynamics paglago para sa panahon 2010-2024

Mga lugar ng produkto

Bilang isang pangunahing bahagi ng Industry 4.0, ang genomics ay may malawak na hanay ng potensyal mga application sa halos lahat ng mga larangan pang-ekonomiya. Pangunahing mga lugar ng produkto ng kasalukuyang genomics merkado:

- **NIPT (Non-Invasive Prenatal Testing)**

Paglalarawan:	Batay sa DNA ng pangsanggol na nagmula sa sanggol dugo. Maaaring makilala ng pagsusulit ang pangsanggol na aneuploidy [6] at kasarian ng isang sanggol sa unang bahagi ng anim na linggo sa isang pagbubuntis.
----------------------	--

Ibahagi ang Market:	4 bilyon USD
----------------------------	--------------

Mga Namumuno sa Market:	Illumina, Natera, Ariosa, Sequenom.
--------------------------------	-------------------------------------

- **PGS (Preimplantation genetic screening)**

Paglalarawan:	Batay sa DNA-microarray o NGS-sequencing genetic pag-profile ng mga embryo bago itanim sa IVF (in pamamaraan ng pagpapabunga (vitro fertilization))
----------------------	---

Ibahagi ang Market:	336.4M USD
----------------------------	------------

Mga Pinuno ng Market:	Illumina, Agilent Technologies
------------------------------	--------------------------------

- **DTC (Direct-to-consumer) genetic testing, SNP-genotyping**

Description:	Genetic test based on DNA-microarray SNP-genotyping for getting some following recommendations: ancestry, nutrition needs, carrier status, optimal exercise.
---------------------	--

Market Share:	2 Billions USD
----------------------	----------------

Market Leaders:	AncestryDNA, 23andMe, DNAfit, deCode genetics.
------------------------	--

- **Diagnostics company (Including oncogenomics)**

Paglalarawan:	Genetic test batay sa DNA-microarray SNP-genotyping for pagkuha ng ilang mga sumusunod na rekomendasyon: ninuno, nutrisyon mga pangangailangan, katayuan ng carrier, pinakamainam na ehersisyo.
----------------------	---

Ibahagi ang Market:	2 Bilyong USD
----------------------------	---------------

Mga Pinuno ng Market:	AncestryDNA, 23andMe, DNAfit, genetika deCode.
------------------------------	--

Mga Hamon para sa mga Genomics

Sa seksyon na ito, ang mga pangunahing problema sa mga kontemporaryong genomika ay isinasaalang-alang. Dapat malutas ang mga problemang ito upang ipatupad ang konsepto ng Genomics 2.0: ang nasa lahat ng dako na pagpapalawak at pagpapatupad ng personal genomic na teknolohiya..

Mababang availability ng pagtatasa ng genome

Ang mga gastos para sa iba't ibang uri ng pagtatasa ng genome ay nakalista sa Talahanayan 3 at, sa pangkalahatan, nagsisimula mula sa \$ 100. Ang sukat ng presyo na ito ay napakababa sa presyo ng genome reading 5 o 10 years ago [7].

Gayunpaman, ang presyo ng sequencing at bioinformatic interpretation ay nananatiling mataas sapat na ang pagbibigay ng pangkalahatang publiko na may access sa genomic analysis ay mahirap, lalo na sa pagbuo ng mga bansa.

Talaan 3: Uri ng Genome analysis

DNA microarray

Paglalarawan: pagtatasa ng 1-5 milyong pre-napiling SNPs
Saklaw ng presyo: \$ 100-500
impormasyon 0.033% ng buong genome
nagbibigay ng serbisyo 23andMe, AncestryDNA, DNAfit.

Genes panels

Saklaw ng presyo: \$ 100-2000
halaga ng impormasyon 0.001-1% ng buong genome
nagbibigay ng serbisyo Pathway Genomics, CeGaT.

Exome sequencing

Paglalarawan: Sequencing para sa coding bahagi ng isang genome (exome)
Saklaw ng presyo: \$ 250-3000
halaga ng impormasyon 2% ng buong genome
nagbibigay ng serbisyo BGI, CeGaT.

Whole Genome Sequencing

Saklaw ng presyo: \$ 600-10000
halaga ng impormasyon 80-98% ng buong genome
nagbibigay ng serbisyo BGI, FullGenomes, Human Longevity

Tandaan: sa katunayan, ang isang bahagi ng isang genome ay hindi susundan. Ang laki nito ay depende sa depth ng sequencing at ang mga detalye ng proseso ng paghahanda ng sample. Kaya ang terminong "Buong Genome Sequencing" talaga ay nangangahulugan ng pagkuha ng pagkakasunud-sunod ng bahagyang higit sa 80% ng isang genome.

Pagkuha ng isang malaking halaga ng impormasyon ng genomic (mga pagkakasunud-sunod ng mga genome na magkakasama na may phenotypic na katangian) para sa mga residente ng mga umuunlad na bansa ay labis mahalaga mula sa pananaw ng pagkuha ng pinakamalawak na posibleng pagkakaiba-iba ng genomic na impormasyon ay, gayunpaman, ay makabuluhang pasiglahin ang pag-unlad ng genomic malaking data market. Kahit na sa karamihan ng mga bansa na binuo, mas mababa kaysa sa 2% ng populasyon ay may undergone anumang genomic analysis (microarray, exome sequencing, o sequence ng buong-genome).

Pag-kompromiso sa privacy bilang isang presyo para sa paglahok sa bukas na biomedical investigations.

Ang pinapayuhang pahintulot para sa pagkolekta at pagproseso ng personal na data ay isang mahalagang isyu sa bawat biomedical research study. Anumang proyekto na kinasasangkutan ng pag-aaral ng genome ng tao ay nagsisimula sa koleksyon ng mga lagda mula sa mga indibidwal na nagpapatunay sa katotohanan na naiintindihan nila ang mga kahihinatnan at sumasangayon sa mga tuntunin ng pag-aaral ng pananaliksik. Ang anyo ng naturang pahintulot ay nag-iiba depende sa proyekto at maaaring kasama ang pagbibigay pahintulot na gamitin ang data sa mga proyekto sa hinaharap, ang mga kahihinatnan nito mahuhulaan.

Sa Personal na Proyektong Genome, pinangangasiwaan ng Harvard Medical School, kusang sumang-ayon ang mga kalahok na ang kanilang data at mga halimbawa ng kanilang genetic na material Maaaring gamitin nang maraming beses at maaaring magamit sa ibang mga laboratoryo. Ang mga kalahok ng proyektong ito ay partikular na alam na ang kanilang pagkakakilanlan ay maaaring di-kilala at ang kanilang pribadong data ay maaaring maging publiko. Nilalayon ng proyektong ito tiyakin na ang genomic data mula sa maraming mga tao hangga't maaari ay lantaran magagamit upang pasiglahin ang bagong pananaliksik at pagpapaunlad sa industriya ng genomika. Ang mga may-akda ng proyektong ito ay naniniwala na kung hindi kami nagbibigay ng bukas na access sa genomic na data at pagpapalitan ng impormasyon, kami ay nasa panganib na magtapos sa libu-libong mga nakahiwalay, pribado na nakaimbak ng mga koleksyon ng mga genomic data (mula sa mga pharmaceutical company, genomic corporations, at mga siyentipikong sentro), ngunit ang bawat isa sa mga magkahiwalay na database ay hindi maglalaman ng sapat na data upang paganahin ang mga natuklasan ng tagumpay.

Pagsasagawa ng pang-internasyonal na multicentral na siyentipiko at klinikal na pag-aaral.

upang magsagawa ng mga siyentipiko at klinikal na pag-aaral sa mga malalaking sampol mula sa iba't ibang populasyon mga grupo. Sa kasalukuyan, ang pagkolekta ng mga sample ng genomic mula sa mga indibidwal na may iba't ibang mahirap ang etniko na pinagmulan, dahil kailangang gumawa ng mga espesyal na proyekto, mga ekspedisyong isinasagawa, at pahintulot na nakuha mula sa mga lokal na regulator.

Ngayon, isang startup lamang ang sinusubukang matugunan ang problemang ito [https:// www.dnasimple.org/](https://www.dnasimple.org/) para sa isang maliit na bayad at sa pangako ng pagkawala ng lagda.

Paglikha ng biobanks na nagtatabi ng mga materyales mula sa iba't ibang indibidwal

Ang Biobanks ay nagsisilbi bilang mediators sa pagitan ng mga donor ng mga biological na materyales (dugo sample, buto sa utak ng buto, atbp.) at mga mananaliksik sa pamamagitan ng pagproseso ng mga materyales Nakuha at iniimbak ang mga ito para magamit sa hinaharap. Sa pangkalahatan, ang biobanks ay isang mahalagang tool para sa

ang pag-unlad ng isinapersonal na (katumpakan) na gamot at pag-unlad ng droga. Isa sa ang pinakamahalagang pag-andar ng biobanks ay ang koleksyon ng materyal na donor para sa paggamit sa hinaharap, kabilang ang dugo, utak ng buto, at kahit mga selula ng mikrobyo.

Sa kasalukuyan, ang biobanks ay aktibo na binuo sa maraming mga bansa. Gayunpaman, may posibilidad na ang pinaka-mahalagang mga halimbawa ay nabibilang sa karamihan sa mga financially secure biobanks, halimbawa, biobanks na kabilang sa malaki mga parmasyutiko na kumpanya, na hahantong sa hindi patas na pag-access sa materyal biobank para sa iba't ibang mga kategorya ng mga mananaliksik. Samakatuwid, may pangangailangan na lumikha ng biobanks sa maraming mga bansa at lungsod hangga't maaari hanggang sa limitasyon ng mga indibidwal na bioreservoirs para sa bawat indibidwal.

Pagproseso: limitado ang rate ng pag-aaral ng genome sa pamamagitan nito bioinformatic processing. Pagsasaayos ng analytical software para sa laganap na application.

Sa kasalukuyan, ang bilis ng pagkuha ng genomic data gamit ang mga teknolohiya ng pag-aaral ng genome (sequencing) ay mataas at outstrips ang bilis ng pagproseso ng mga data na ito. Kung isaalang-alang naming anumang malakihang pang-agham na pag-aaral ng isang malaking bilang ng mga genome, ang pang-eksperimentong mga hakbang ng pagkuha ng genomic na impormasyon account para sa hindi hihigit sa 20% ng pag-aaral tagal, habang ang pagpoproseso ng data ay binubuo ng isang mas malaking bahagi ng proyekto. Dito, ang pagpoproseso ng data ay nangangahulugang ang mga hakbang na kinuha mula sa pagkuha ng raw sequencing data upang maunawaan ang mga resulta at maghanap ng iba't ibang mga asosasyon.

Ang isa pang problema sa modernong software ay ang genomic analysis software ay na ginawa ng mga siyentipiko at para sa mga siyentipiko at sa gayo'y nangangailangan ng mga pagsasaayos na malawakan naaangkop sa mga doktor at para sa pangkalahatang mga mamimili. Bilang genomic analyzers ang laki ng USB stick ay umiiral na, ang paggamit ng personal na sequencer sa parehong ang paraan bilang isang personal na computer ay maaaring isaalang-alang hindi science fiction ngunit, sa halip, makatotohanang sa malapit na hinaharap.

Genomic data interpretation: matematiko modelo ng

Ang iba't ibang mga modelo at mga algorithm ay ginagamit upang mapagkunan ang mga panganib ng pagkakaroon ng mga sakit batay sa genetic data. Ang mga pangunahing uri ng mga modelong ito ay batay sa uri ng mana ay isinasaalang-alang: monogenic, polygenic at multifactorial. Pagtatasa sa mga panganib ng pagkakaroon ng multifactorial o komplikadong sakit ay nangangailangan ng accounting para sa impluwensiya at pagkagambala ng maraming mga gene pati na rin ang mga environmental factor. Para sa isang mas detalyadong paglalarawan ng mga magagamit na pamamaraan para sa pagtatasa ng mga panganib ng komplikadong mga sakit, tingnan ang [8, 9].

Upang bumuo ng isang bagong modelo para sa pagtatasa ng panganib na magkaroon ng sakit, ito ay kinakailangan upang kasangkot ang isang malaking bilang ng mga siyentipiko. Kailangan ding mag-imbetiga maraming mga publikasyon tungkol sa sakit na sinusuri, upang makilala ang uri nito mana, upang matukoy ang mga polymorphisms at mutations na nag-aambag sapag-unlad ng sakit, at upang bumuo ng "genomic algebra," ibig sabihin, isang hanay ng mga panuntunan para sa pagpapahalaga sa panganib. Kapag ang isang modelo ay itinatag at napatunayan sa Ang silico, klinikal na pag-aaral ay kailangang isagawa upang masuri ang pagiging magamit nito. Ang pamamaraan na ito ay kasalukuyang ang pinaka tumpak, ngunit ito ay mahal sa mga tuntunin ng dami ng oras at kinakailangan ang pagsisikap.

Genomic data interpretation: machine learning application.

Ang paggamit ng mga algorithm sa pag-aaral ng machine upang masuri ang mga panganib ng mga multifactorial disease ay malawakan na sinisiyasat, ngunit sa ngayon, dahil sa kawalan ng sapat bilang ng mga sample ng pagsasanay, umiiral na mga modelo ng matematika na binuo ng biological lumalabas ang mga siyentipiko ng mga diskarte sa pag-aaral ng machine.

Gayunpaman, ang pag-aaral ng machine ay ginagamit upang mahulaan ang ilang mga kumplikadong mga katangian ng katawan ng tao. Ang isang halimbawa ay ang hitsura ng hula sa gawain ni Craig Venter at mga kasamahan. Ang esensya ng kanilang trabaho ay kasangkot pag-aralan ang mga genome at humigit-kumulang 30,000 mga puntos ng facial data mula sa ilang libong boluntaryo. Batay sa data na nakuha, halimbawa ng pagsasanay para sa makina Ang pag-aaral ng mga algorithm ay binuo at mga dependency sa pagitan ng mga genomic na katangian at tinutukoy ang indibidwal na hitsura. Dahil sa gawaing ito, mayroon ang mga makina natutunan upang tumpak na ibalik ang anyo ng isang tao batay sa kanyang genomic data [10, 11].

Ang mga resulta ng proyektong ito ay nagbibigay-daan sa hula ng paglitaw ng isang criminal o ng isang hindi pa isinilang na bata sa mga unang yugto ng pagbubuntis. Sa pamamagitan ng pagkuha ng dugo sample mula sa isang buntis at pagkuha ng fetal DNA mula sa dugo, ang hitsura ng isang hindi pa isinilang na bata sa kanyang ika-18 na kaarawan ay maaaring tumpak hinulaang.

Upang ipatupad ang proyektong ito, hinikayat ni Craig Venter ang isa sa pinakamahasay na pag-aaral ng machine mga espesyalista mula sa Google, Franz Och, isang siyentipikong bituin sa computer na kilala bilang pinuno arkitekto ng Google Translate [12].

Sa kasalukuyan, ang pag-aaral ng machine ay hindi gaanong ginagamit para sa mga sakit, na napakalaking at tama ang nakabalangkas na mga halimbawa ay kinakailangan para sa pagsasanay. Ang paglikha ng isang komprehensibong database ng mga genome ng tao, pati na rin ang pagkakaroon ng detalyadong Ang mga questionnaire na nagpapakita ng mga kalagayan sa kalusugan ng mga indibidwal, ay maaaring magsulong ng pag-unlad ng computational training sa genomics at magresulta sa mataas na predictive katumpakan sa pagtukoy sa panganib ng pag-unlad ng sakit. Kasabay nito, ang mga data na ito ay maging pampubliko at magagamit sa lahat ng mga gumagamit ng system, hindi kasama ang posibilidad ng kanilang monopolisasyon. Ang availability na ito ay lubhang mahalaga, tulad ng konsentrasyon ng ang malalaking halaga ng data sa mga database ng mga korporasyon ay magresulta sa mga monopolyo sa

larangan ng pag-aaral ng genomic machine.

Ligtas na pag-iimbak ng napakalaking data

Napakahalaga ng personal na data ng seguridad; sinusubukan nating lahat na pigilan ang pagnanakaw ng data ng credit card, impormasyon sa seguro, mga numero ng account ng pagbabangko, at medical impormasyon ng anumang uri. Ang pagnanakaw ng impormasyon sa genomic ay maaaring tila hindi mahalaga para sa maraming mga tao ngayon. Gayunpaman, maaari itong humantong sa napakaseryosong mga bunga nito mahirap hulaan, halimbawa, ang posibilidad na ang isang bahagi ng isang indibidwal. Ang genomic sequence ay maaaring synthesized at nakatanim sa site ng isang krimen o isang gawa ng terorista.

Ang mga kasalukuyang solusyon para sa problemang ito ay may kasangkot na naka-encrypt na imbakan sa isang sentral na server, tulad ng [13], <https://www.pathway.com/>, <https://www.23andme.com/>, o <http://www.humanlongevity.com/>. Ang uri ng "closed" na imbakan ng data ay relatibong ligtas, ngunit ipinagbabawal nito ang posibilidad ng pagbabahagi ng data at pagbibigay ng access sa ang mga siyentipiko mula sa buong mundo, na isang mahalagang kondisyon para sa ang pag-unlad ng modernong genomic science.

Ang isa pang problema na nauugnay sa pag-iimbak ay tungkol sa malaking sukat ng isang genome mismo, at ang pagpaparami ng paglago ng halaga ng data ng genomic na magagamit, tulad ng higit pa at mas maraming indibidwal ang nasasailalim sa genome sequencing. Sa isang pag-aaral [14], noong 2025, ang kabuuang dami ng naka-imbak na genomic na data (na ibinigay sa laki ng isang kumpletong genome ay 100 Gb) ay hinuhulaan na maabot ang 40 exabytes kada taon at ang hinulaang genomic ay hinuhulaan upang maging isa sa mga pinakamalaking consumer ng imbakan at pagproseso ng impormasyon mga kapasidad.

Talaan 4: Apat na mga domain ng Big Data noong 2025.

Astronomy

Pagkuha ng Imbakan	25 zetta-bytes / year 1 EB / taon
Pagtatasa	Sa pagbabawas ng data sa kinaroroonan Real-time na pagproseso; Napakalaking volume;
Pamamahagi	Dedicated na mga linya mula sa antena sa server (600 TB / s)

Twitter

Pagkakaloob ng Imbakan	0.5-15 bilyong tweet / taon 1-17 PB / taon
Pagsusuri ng Pamamahagi	Pagpapatungkol sa paksa at sentimyento; Pagsusuri ng Metadata Maliit na yunit ng pamamahagi

YouTube

Pagkuha ng Imbakan	500-900 milyong oras / taon 1-2 EB / taon
Pagsusuri ng Pamamahagi	Mga limitadong kinakailangan Major bahagi ng bandwidth ng modernong gumagamit (10 MB / s)

Genomics

Pagkuha	1 zetta-bases/year
Pag-imbak	2-40 EB/year
Pagsususri	Heterogeneous data and analysis; Variant calling, ~ 2 trillion central processing unit (CPU) hours; All-pairs genome alignments, ~ 10,000 trillion CPU hours.
Pamamahagi	Many small (10 MB/s) and fewer massive (10 TB/s) data movement

Ang pangangailangan para sa pangkaraniwang database na patuloy na na-update questionnaires.

Kasabay nito, ang kakulangan ng isang pampublikong database na ipinatupad batay sa Ang konsepto ng ibinahagi na storage at paggamit ng open-source software ay maaaring humantong sa kabuuang dominasyon ng lugar na ito ng merkado ng mga kumpanya tulad ng Google o Amazon dahil sa kanilang makapangyarihang server⁴. Kung ang mga korporasyon o mga pharmaceutical company maging monopolista sa larangan ng genomic na impormasyon, susundin namin ang mabagal at mahal na pag-unlad ng gamot na may kasalukuyang mga diskarte sa paggamot na nagpapatuloy sa halip na isang hinaharap kung saan ang isang sakit ay maiiwasan kahit bago nito pag-unlad o sa pinakamaagang yugto ng pagpapakita nito.

⁴<https://cloud.google.com/genomics/>

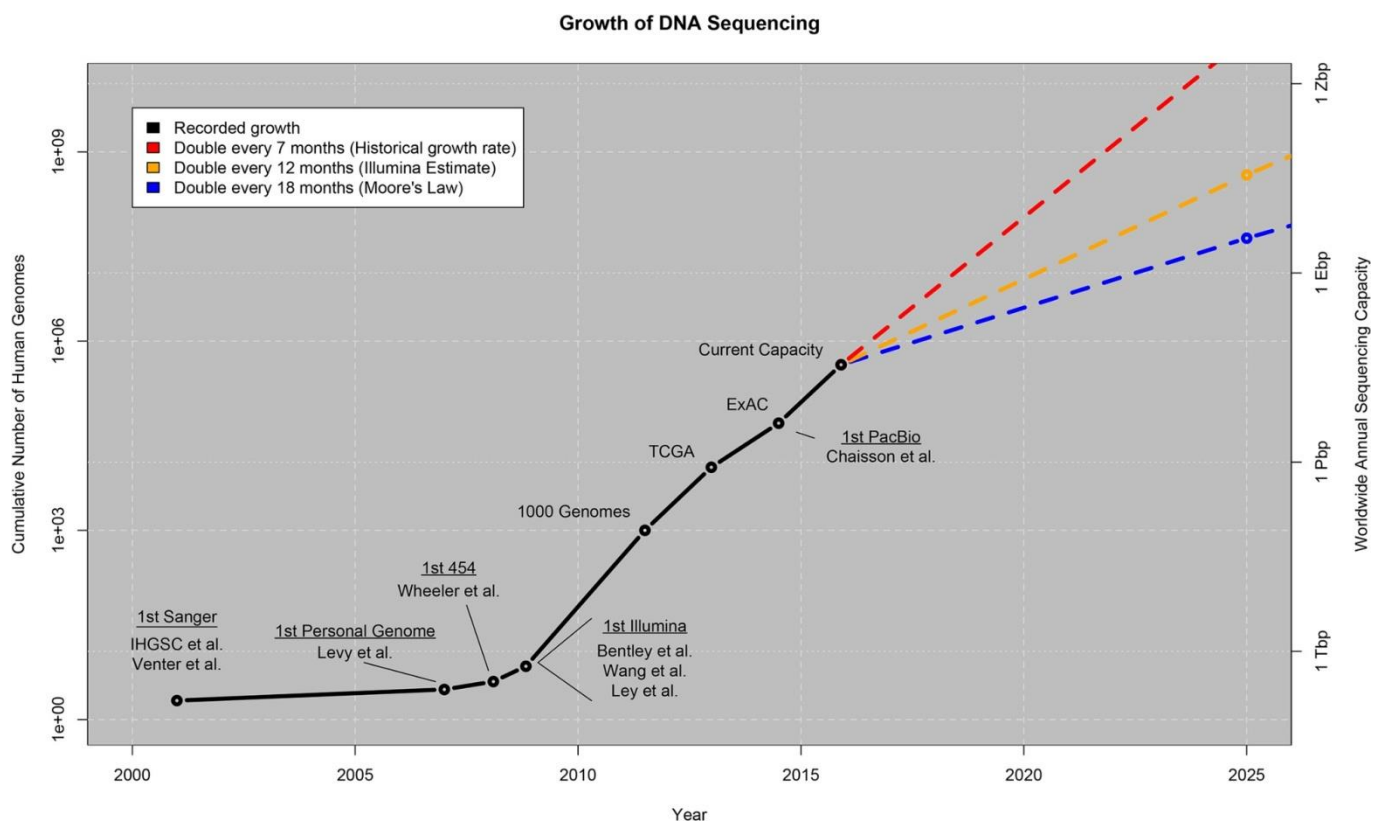


Figure 3: Paglago ng DNA sequencing data. Ang balangkas na ito ay nagpapakita ng paglago ng DNasequencing pareho sa kabuuang bilang ng mga human genome sequenced (kaliwang axis) at ang buong taon na kapasidad ng pagkakasunod-sunod (ang tamang axis: Tera-basepairs (Tbp), Mapa-basepairs (Pbp), Exa-basepairs (Ebp), at Zetta-basepairs (Zbp)). Ang mga halaga sa pamamagitan ng 2015 ay batay sa makasaysayang mga tala ng publikasyon na naka-highlight milestones sa sequencing (unang Sanger sequencing sa pamamagitan ng unang PacBio pantao genome na nai-publish) pati na rin ang tatlong mga kapuri-puri proyekto gamit ang malakihang sequencing: ang 1000 Genomes Project, na pinagsama-samang daan-daang genome ng tao sa pamamagitan ng 2012; Ang Kanser Genome Atlas (TCGA), na pinagsama-samang ilang libong tumor / normal na mga parating na genome; at ang Exome Aggregation Consortium (ExAC), na pinagsama-samang mahigit sa 60,000 exomes ng tao. Marami sa mga genome na nakakasunod sa petsa ay kumakatawan sa buong exomes sa halip na buong genomes, ngunit inaasahan namin ang ratio na ito upang maging mas pinaliit papunta sa buong genome sa hinaharap. Ang mga halaga na lampas 2015 ay kumakatawan sa aming mga pag-uulat sa ilalim ng tatlong posibleng mga curve ng paglago, tulad ng inilarawan sa ang pangunahing teksto. Kinuha mula sa [14].

Ang malapit na hinaharap ng genomics

Bilang isang konklusyon sa kabanatang ito, ito ay nagkakahalaga ng naglalarawan ng ilang mga hypothetical ngunit technically feasible perspectives at panganib na maaaring nahaharap sa genomic industriya at lipunan sa pangkalahatan:

- Pagbawas sa gastos ng pag-aaral ng genome at ang miniaturization ng genomic sumusunod na mga aparato pababa sa antas ng mga plug-in ng cell phone;
- Ang paputok na paglago ng data ng genomic at ang imbakan nito at ang paglitaw ng "Genomic hackers" at mga isyu sa privacy (pagprotekta sa data);
- Ang pamamahagi ng kumot ng genomic medicine at telehealth;
- Mga pagbabago sa industriya ng pagkain na may kinalaman sa pagpapatupad ng personalized nutrisyon batay sa mga genome;
- Ang pag-unlad ng isinapersonal na therapy ng gamot;
- Dating batay sa genomic compatibility;
- Pagkakakilanlang personalidad gamit ang genomic na impormasyon, kabilang ang kakayahang gumawa ng mga pagbabayad at makakuha ng mga serbisyo;
- Ang pagtaas sa average na lifespan sa lahat ng mga bansa, ang extension ng aktibo kahabaan ng buhay, at isang matibay na bias patungo sa mga matatandang populasyon, late pregnancies, at nagpapababa ng mga birthrate;
- Ang pagpapaunlad ng mga teknolohiya ng pag-edit ng genome;
- Disenyo ng hitsura para sa mga bata sa hinaharap at iba pang mga pagpapaunlad na mahirap isipin. Halimbawa, ang teknikal na posibilidad ng pagpili ng malusog mga bata at pagtukoy ng kanilang hinaharap na hitsura sa entablado at sa pamamagitan ng pagkuha ng fetal DNA mula sa dugo ng isang buntis na babae ay umiiral na [10].

Mga etikal na Isyu ng mga personal na genomics

Sa seksyon na ito, ang mga etikal na problema na nauugnay sa pag-unlad at Ang lahat ng pamamahagi ng mga genomic na teknolohiya ay itinuturing, tulad ng privacy, mga pampublikong database, bukas na pag-access para sa mga mananaliksik, posibleng maling paggamit at pagbabanta sa personal na kalayaan dahil sa pagpapalawak ng mga genomics.

Sa kasalukuyang puting papel, nagbibigay lamang kami ng maikling pagsusuri sa mga pangunahing problema at pananaw ng industriya ng genomic. Para sa isang mas detalyadong talakayan tungkol sa mga problema sa privacy at data ng seguridad, tingnan ang <https://www.smeal.psu.edu/fcfe/> mga dokumento / innovations-in-medical-genomics-pdf.

Privacy

Ang personal na impormasyon ng genomic ay napaka sensitibo para sa maraming tao. Gayunpaman, marami ang mga tao ay hindi lubos na nauunawaan na, batay sa kanilang genomic na impormasyon, ito ay posible upang matukoy ang kanilang buhay, kagalingan upang gumawa ng emosyonal na mga desisyon (manipulability ng paggawa ng desisyon), posibilidad na magkaroon ng iba't ibang kaisipan sakit at panganib ng biglaang kamatayan dahil sa, halimbawa, puso arrhythmia.

Ang ganitong impormasyon ay maaaring maging kapaki-pakinabang para sa recruitment ng trabaho, halalan pakikilahok, at pangangalaga ng seguro sa medikal. Mayroon ding posibilidad na ang isang masamang artista na nakakaalam ng pagkakasunud-sunod ng isang genome ay maaaring umalis ng mga fragment ng DNA katulad ng genome na iyon, halimbawa, ang lokasyon ng isang pagkilos ng terorista upang i-frame o ilegal na akusahan ang isang tao. Maaaring tanggihan ang isang medikal na paggamot (o kinakailangan upang magbayad ng isang mas mataas na bayad) o barred mula sa pagkuha ng isang ninanais na trabaho.

Ang mga korporasyon at pamahalaan ay maaaring sadyang impluwensiyahan ang mga desisyon ng isa at pagbili gamit ang kanilang kaalaman sa "mga kahinaan" sa isang genomic na impormasyon. Samakatuwid, ang pangangalaga ng pagkakasunud-sunod ng data ng genomic ay kinakailangan upang protektahan ang pantay mga karapatan ng iba't ibang kategorya ng mga tao.

Kasabay nito, ang ilang mga pag-aaral ay ginanap na nagpapahintulot sa pagkilala ng mga indibidwal na pagkakakilanlan batay sa kanilang hindi nakikilalang mga genome [5, 15].

Bukod dito, ang ilang mga kumpanya (<http://www.humanlongevity.com/media/>) ay nagtataglay machine-based na algorithm sa pag-aaral na maaaring tumpak na muling buuin ang hitsura ng isang indibidwal na gumagamit lamang ng kanyang genomic data [16].

Publicity

Ang ganitong paraan ay magbibigay-daan sa pagpapaunlad sa pamamagitan ng pagpigil ng gamot kung saan ang pag-aaral ng malaking halaga ng data ay nagbibigay-daan sa hula ng sakit pag-unlad (bago ang paglitaw nito), pagpapagana ng mga aksyon na nagpapataas ng habang-buhay at mapabuti ang kalidad ng buhay [17] pati na rin ang pagkakakilanlan ng mga donor sa buong mundo matugunan ang iba't ibang mga medikal na pangangailangan. Sa kasamaang palad, wala pang wastong mga solusyon ang kasalukuyang umiiral para sa pampublikong paggamit ng genomic na impormasyon habang pinapanatili ang indibidwal na privacy. Gayunpaman, ang ilang mga startup na nagtatrabaho sa larangan na ito gamit ang blockchain technology ay dapat ay nabanggit.

EncrypGen, isang startup na kamakailan-lamang na isinasagawa ng isang ICO, inilarawan ang umiiral na mga problema at ang relasyon sa pagitan ng privacy at publikidad gamit ang blockchains [18]. Gayunpaman, ang puting papel na nauugnay sa proyektong ito ay walang isang paglalarawan ng isang teknikal na pagpapatupad na lutasin ang problema ng privacy at availability.

Ang isa pang proyekto sa larangan na ito ay ang startup ng DNAbits [19], na ang tagapagtatag, Dror Samuel Brama, ay may patentadong pangkalahatang diskarte sa pag-iimbak at paglipat ng data gamit ang mga blockchain technology [20]. Gayunpaman, ang kumpanya na ito ay hindi technically ipinatupad ang konsepto nito sa loob ng nakaraang tatlong taon.

Ang karapatang magkaroon ng data ng genomic?

Sa kasalukuyan, walang lehislatibong kahulugan ng karapatang magkaroon ng sariling genetiko impormasyon. Sa ilang mga binuo bansa, kabilang ang USA, Germany, at Austria, ang mga mamamayan ay walang karapatan na ma-access at ariin ang kanilang genetic data sa konteksto ng interpretasyon nito [21]. Isang ahente, na kinakatawan ng isang manggagamot o kailangan ng medikal na sentro na may karapatang magbigay ng naturang impormasyon. Ang landas na ito ay ginagamit ng mga kumpanya Pathway Genomics sa USA at CeGaT sa Alemanya (<http://www.cegat.de/en/>).

Upang magsagawa ng pagtatasa ng genetiko, ang payo ng isang manggagamot na maaaring isang tagapagkaloob ng genetic testing ay kinakailangan, at tanging ang doktor na ito ay may karapatan na bigyang-kahulugan ang impormasyon na ibinigay ng genetic analysis.

Sa USA, mayroong mga service provider sa larangan ng "genetics for fun," tulad ng mga kompanya ng 23andMe at Ancestry.com, na nagbebenta ng mga pagsubok sa genetiko nang direkta sa dulo customer, ngunit ang mga kumpanyang ito ay maaari lamang magbigay ng impormasyon tungkol sa etnikong pinagmulan at ilang mga katangian na may kaugnayan sa kalusugan (halimbawa, mga katangian sa sports) at kakulangan ang pahintulot na magbigay ng pinaka mahalagang impormasyon sa medisina. Ang mga paghihigpit na ito, na ipinapataw ng mga regulator tulad ng FDA, huwag hadlangan ang kakayahan ng 23andMe magbenta ng access sa genetic data sa mga malalaking kumpanya ng pharmaceutical. Ang ilang mga naturang deal ay kilala: isang deal ay ginawa sa Genentech (isang subdibisyon ng parmasyutiko higante Roche) para sa \$ 60 milyon [22] para sa isang pag-aaral ng Parkinson's disease, at isang deal ay ginawa na may isa pang malalaking kompanya ng parmasyutiko, Pfizer, para sa isang pag-aaral ng nagpapaalab sakit sa bituka (hal., Crohn's disease) [23]. Ang ilang mga ulat din claim na 23andMe nagkaroon ng negosasyon sa Novartis sa pag-aaral ng Alzheimer's disease [24].

Kaya, kasalukuyan naming binibigyan ang mga malalaking kumpanya ng karapatang pamahalaan ang aming genomic impormasyon, iimbak ito, at upang kumita mula dito. Mga korporasyon, sa likod ng belo ng magandang intensyon, monopolize ang malaking data ng genomic, at hindi namin mahuhulaan kung paano ito. Ang monopolyo ay makakaimpluwensya sa mga presyo ng gamot sa hinaharap at nakapagpapagaling na pagtuklas

Ang karapatang ma-access ang genomic na impormasyon

Isa pang etikal na isyu ang dapat talakayin; Sa itaas, nabanggit namin na ang mga paglabag sa privacy at ang pag-access sa genomic na impormasyon ay maaaring gamitin ilegal, halimbawa, sa pamamagitan ng isang tagapag-empleyo. Ang isa ay maaaring fired o tinanggihan ng isang promosyon ng trabaho batay sa genetic impormasyon. Para sa mga may trabaho na may kaugnayan sa kaligtasan ng mga tao at mga sistema, tulad ng mga drayber ng trak o bus, mga piloto, mga operator ng istasyon ng atomikong kapangyarihan, o mga tao na may iba pang katulad na trabaho, ang kalagayan sa kalusugan ay mahalaga, at genomic maaaring maiwasan ng impormasyon ang isang aksidente o kahit isang kalamidad. Sa ilang mga propesyon, ang isang potensyal na panganib ay maaaring magbanta sa manggagawa sa halip na mga tagabantay, tulad ng isang karbon o brilyante na may mga problema sa baga. Para sa mga kasong ito, isang diskusyon na kinasasangkutan ang mga propesyonal at eksperto pati na rin ang pangkalahatang publiko ay kinakailangan upang bumuo ng legal mga pamantayan na kumokontrol sa paggamit ng genomic na impormasyon ng mga tagapag-empleyo.

Kabanata 2

Konsepto

Proyektong Zenome

Pilosopikong Pananaw

Ang pampublikong kamalayan ng genomic medicine ay nananatiling medyo mababa sa binuo at kahit na mas masahol pa sa pagbuo ng mga bansa. Ito ay nangangahulugan na ang mga tao, sapangkalahatan, mayroon maliit na pag-unawa sa mga posibleng benepisyo ng genomics pati na rin ang mga posibleng panganib na nauugnay dito. Sa maraming bansa nagresulta ito sa pagbuo ng labis-kumplikadong institutional na mga pamamaraan upang protektahan ang genetic information mula sa posibleng maling paggamit na, sa kabilang banda, humahadlang sa pag-unlad ng siyensiya. Ang Zenome Platform ay magtataas ng kamalayan sa genomic na gamot, kaya ang mga gumagamit maaaring gumawa ng mga nakakamalay na desisyon tungkol sa kanilang data. Upang matiyak na, ang Zenome Ang plataporma ay batay sa sumusunod na mga pangunahing prinsipyo:

Indibidwal na pagmamay-ari ng personal na impormasyon ng genomic Ang bawat kalahok May lahat ng mga karapatan para sa personal na data ng genomic.

Kalayaan ng pagpili Ang bawat kalahok ay nagpapasya kung paano ang indibidwal na impormasyon ng genetiko ay dapat gamitin. Ang isa ay maaaring magpasya kung o hindi na lumahok pang-agham / klinikal na pananaliksik.

Ang karapatang ibahagi Ang kalahok ay maaaring magbigay ng access sa genetic na impormasyon sa isang third party sa isang paraan na naghihigpit sa pagkopya ng data.

Privacy Ang pag-encrypt ng pribadong data ay imposible upang ma-access ang indibidwal na genetic impormasyon nang walang tahasang pahintulot ng gumagamit.

Ipinamamahagi ng imbakan ng data Ang ibinahagi na database architecture ay nagbibigay ng mataas kakayahang makamit at kasalanan sa pamamagitan ng pagtitiklop at pag-scale kakayahan

Ipinamamahagi ng data processing Pinoproseso ang data sa maraming nodes sa network sa parehong oras. Ang sinumang gumagamit ay maaaring maging isang node sa pamamagitan ng pagbibigay ng disk space at CPU oras sa network.

Scalability Ang platform architecture ay nagbibigay-daan sa mahusay na scalability at kakayahang umangkop ng ang sistema.

The Zenome Platform Ecosystem

Sa user na platform ng Zenome¹ ay nakikibahagi sa maraming uri ng iba't ibang mga pakikipag-ugnayan sa buong sistema. Ang mga pakikipag-ugnayan na ito ay nagaganap sa iba't ibang antas ng sistema, huwag makagambala sa isa't isa at maglakip ng iba't ibang mga pattern ng mga pakikipag-ugnayan. Kaya, dapat silang katawanin bilang mga natatanging entidad na may iba't ibang tungkulin.

Ang mga sumusunod na tungkulin ay magagamit sa platform ng Zenome:

(Pagkalkula / Pag-iimbak) Node na nagbibigay ng imbakan at kapangyarihan ng CPU para sa isang gantimpala.

Tao na nag-upload ng indibidwal na genetic na data sa platform at posibleng gumagamit ng mga serbisyong genetiko.

Analyst na interesado sa pag-aaral ng genetic na impormasyon sa platform. Mayo kumakatawan: isang siyentipikong datos, isang organisasyong siyentipiko at iba pa.

Tagapaglaan ng Serbisyo na nagpapatupad ng serbisyo ng genetic ng user-space (posibleng binayaran) sa platform. Talaga, ito ay isang organisasyon na gumagamit ng genetic data bilang isang bahagi ng negosyo nito.

Ang bawat gumagamit ay nakikibahagi sa isang bilang ng pakikipag-ugnayan ng iba't ibang uri, pagkuha sa iba't ibang mga tungkulin. Ang ilan sa kanila, tulad ng Service Provider at Analyst, nangangailangan ng mga espesyal na kaalaman, ngunit ang Node at Tao ay hindi.

Ang bawat tungkulin ay tatalakayin sa mga detalye sa ibang pagkakataon.

Pangkalahatang-ideya ng System Architecture

Ang Zenome platform ay isang ipinamamahagi na application na binubuo ng 3 pangunahing mga layer.

¹User in a broader sense, i.e. a person or software that runs on behalf of that person.

Network (at Access ng Data) Layer: nagbibigay ng isang antas ng abstraction na nagpapaloob sa pakikipag-ugnayan sa network at nagbibigay ng interface sa isang ipinamamahagi sa kapaligiran sa itaas na layer.

	Blockchain	DHT Kademlia
Data Storage Cost	High	Low
Data Immutability	True	False
Pagganap	Low	High
Deterministic Result	True	False

Ang layer na ito ay binubuo ng dalawang mga sistema ng ipinamamahagi ng isang ganap na naiibang likas na katangian:

Ipinamamahagi Ledger (batay sa blockchain) na nagtatala ng mga transaksyon sa pagitan ang mga kalahok sa isang napapatunayan at permanenteng paraan. Upang ma-access ang blockchain node Ang software ay nagpapatakbo ng naka-embed na Ethereum client.

Ibinahagi ang Hash Table Network (batay sa Kademlia protocol) na pinagsasama pisikal na node sa isang network ng overlay at nagbibigay-daan sa pagpapasa ng mensahe sa pagitan ng mga node at ibinahagi ang imbakan ng data.

GAWAIN | (Kinakalkula / Pag-iimbak) Node ay nagpapatakbo sa Network layer.

Ang Middleware ay naglalaman ng pamamahala ng account, pare-parehong interface sa seguridad mga tampok ng pinagbabatayan layer at API na may mataas na antas para sa software na tumatakbo sa Antas ng aplikasyon.

GAWAIN | **Service Provider** Ang Platform API ay nagbibigay-daan sa mga third-party na magsimula ng mga serbisyong genetiko sa platform. Application Layer Zenome application ng isang advanced na interface ng end-user na nagta-translate ng mga pagkilos ng user sa Middleware sa napaka-pare-parehong paraan. Ang interface ay napapalawak ng disenyo upang ang mga serbisyo ng genetic ay maaaring tumakbo natively sa softwarestack.

Application Layer Zenome application ng isang advanced na interface ng end-user na nagta-translate ng mga pagkilos ng user sa Middleware sa napaka-pare-parehong paraan. Ang interface ay napapalawak ng disenyo upang ang mga serbisyo ng genetic ay maaaring tumakbo natively sa softwarestack.

Mga serbisyo ng genetic na serbisyo

Ang merkado ng mga serbisyong genetiko ay kasalukuyang bumubuo sa mga sumusunod na lugar:

1. Pananaliksik at pag-aampon ng teknolohiya sa merkado.
2. Pagbibigay ng mga serbisyong diagnostic ng genomic.
3. Ang sertipikasyon ng gobyerno ng mga teknolohiyang genetiko.
4. Pagbuo ng legal na balangkas. Sa partikular, ang mga hakbang sa pambatasan sa pangalagaan ang impormasyon ng genetiko;

TANDAAN

The Ang istraktura ng merkado ay masyadong kumplikado. Ang ilang mga manlalaro ay, sa katunayan, pagbuo sa maraming direksyon upang mahanap ang kanilang lugar sa merkado sa mukha ng mabilis na pagtaas ng mga pangangailangan ng mamimili.

Ang mga pang-agham na korporasyon ay nagtatrabaho sa pagtuklas at pag-aampon ng bago mga teknolohiya sa merkado. Kabilang dito ang:

- Mga korporasyong parmasyutiko, mga biotechnological at diagnostic na kumpanya, tulad ng Pfizer at Myriad
- Mga kumpanya na bumuo at nagbebenta ng lahat ng kinakailangang kemikal na pandagdag reagents (tulad ng Life Technologies).

Ang IT-bioinformatic na kumpanya ay nakikibahagi sa pag-imbento at pagpapaunlad ng pamamaraan ng computational data processing. Ang mga manlalaro sa sektor na ito ay pa rinstruggling upang harapin ang mga uri at volume ng data na nakuha.

Ang mga siyentipiko at medikal na mga sentro ay naglalaro ng isang nangungunang papel sa probisyon at pagpapaunlad ng mga serbisyo ng genetic diagnostics.

Ang mga komersyal na laboratoryo ay nagbibigay ng mabilis, mahusay at kadalasan murang mga serbisyo ng genetic diagnostic. Sila ay may malaking pananalapi at kakayahan sa mapagkukunan.

Ang mga kompanya ng diagnostic ng genetic direct-to-Consumer ay nagdaragdag sa populasyon interes sa genetic diagnostics. Sa kasalukuyan ang segment na ito ay napakaliit, ngunit sa hinaharap, maaari itong lumaki sa isa sa nangungunang bahagi ng merkado at maaaring maging pinagtibay sa clinical practice.

Talahanayan 5: Paghahambing sa Mga Katulad na Produkto sa Market

	We	GeneCoin	Encrypgen	23andMe	Pathway Genomics	Snpedia (Promethease)	Human longevity
Decentralized	✓	✓	✓	-	-	-	-
Suitable for non-human organisms	✓	✓	✓	-	-	-	-
Customer is the owner of his data	✓	✓	✓	-	-	✓	-
Possibility to load your own data	✓	✓	✓	-	-	✓	-
Opened nonprivate data	✓	✓	-	-	-	-	✓
Performs its own data analysis	✓	✓	-	✓	✓	✓	✓
Provides a report for customers	✓	-	-	✓	✓	✓	-
Uses AI and Machine Learning	✓	-	-	-	-	-	✓
Sharing without transmitting huge data	✓	-	✓	-	-	-	-
Earn using your data	✓	-	-	-	-	-	-
Opened for scientists	✓	-	-	-	-	✓	-
Is a platform for other tools	✓	-	✓	-	-	-	-

Mga Tungkulin sa Platform ng Zenome

Mula sa Mga Likas na Resource Node

Ang sistema mula sa isang pananaw ng isang computational node

Node - isang kalahok na nagbibigay ng mga mapagkukunan ng kanyang computer (imbakan at CPU time) para sa layunin ng ibinahagi imbakan at pagproseso ng genetic data para sa gantimpala sa mga token ng ZNA.

Upang maging isang computational node, ang user ay nagpapatakbo ng Zenome software sa kanyang sarili computer at aktibo ang papel na Node gamit ang graphical user interface. Ang software dapat patuloy na tumakbo sa background.

TANDAAN | Ang isang command line version ng Zenome software ay ibinigay din na maaaring partikular na ginamit upang simulan ang computational Node.

Ang patakaran ng paglalaan ng mapagkukunan ng sistema at pamamahala ng gawain ay maaaring maging flexibly isinaayos ng may-ari ng node:

- pinakamalaki na laki ng imbakan na pinapahintulutang gamitin ang software
- isang patakaran ng paggamit ng computational resource:

Fixed CPU / GPU na paggamit ng bilang ng mga core at pinakamababang paglo-load (in porsyento) bawat bawat core.

Ang mga mapagkukunan ng paggamit ng CPU / GPU ay inilalaan sa software tulad ng isang paraan na hindi sila makagambala sa mga application ng user.

- Pumili ng computational na proseso (sa pamamagitan ng id) na magkakaroon ng pinakamataas na pagpapatupad prayoridad.
- Pansamantalang itigil ang node: hilingin ang network na ilipat ang data hindi magagamit sa ibang lugar sa iba pang mga node. Maghintay hanggang makumpleto ang paglipat ng data at idiskonekta.
- Patakbuhan nang ganap ang node: hilingin ang network na ilipat ang lahat ng data sa ibang mga node, maghintay hanggang makumpleto ang paglipat ng data at tanggalin ang data.

Tao / Gumagamit

Tao - isang indibidwal na handang magbigay ng kanyang genetic na impormasyon sa Zenome system upang gumawa ng tubo mula sa pagbebenta ng kanyang personal na data o sa gumamit ng mga serbisyong nakabatay sa genetika na magagamit sa platform.

Ang isang gumagamit ay nag-i-install ng Zenome software upang gumana sa personal na genetic na impormasyon.

Gamit ang graphical user interface, ang user ay may kakayahang:

- Lumikha ng personal na account
- Pamahalaan ang mga token: tumanggap, ilipat, gamitin, magbayad para sa imbakan ng data, gastusin sa bayad mga serbisyong genetiko.
- Mag-upload ng genetic data (ang format ng file ay kadalasang nakita ng awtomatikong).
- Pamahalaan ang personal na data: magbigay ng personal na data, punan ang isang palatanungan, tingnan ang isang listahan ng mga pinakatanyag na mga questionnaire.
- Makipagtulungan sa mga naka-target na alok gamit ang ligtas na mga suhestiyon subsystem.
- Gumamit ng mga serbisyo ng genetic at i-configure ang antas ng pagiging pribado para sa bawat serbisyo. isa-isa. Kapag nag-load ng data ng genetika, maaaring mapili ang antas ng privacy sa:

Buong Privacy Sa kasong ito ang data ay naka-imbak sa naka-encrypt na form, at ang buong presyo ay sisingilin para sa pagtatago ng naturang data.

Standard na Privacy Ang data ng genetic ay nakaimbak bilang mga fragment na ginagawang imposible kilalanin ang user. Ang bawat fragment ay nakaimbak sa system nang pantao. Impormasyon tungkol sa pagtutugma ng mga fragment sa mga user ID ay pribado. Sa kasong ito, ang imbakan ay mas mura dahil sa subsidyo.

Ang Data ng Pag-access ng Publiko ay naka-imbak nang pantao. Ang imbakan ay libre dahil sa mga subsidyo.

Pansin! Kahit na walang mga teknikal na paghihigpit laban sa pagtaas ng antas ng privacy, ang data sa hinaharap lamang ay apektado. Ang data na sabay na ginawa ng publiko hindi kailanman maging pribado muli. Isaalang-alang ito kapag nag-upload ka ng data sa unang pagkakataon.

TANDAN

Pansin! Kahit na walang mga teknikal na paghihigpit laban sa pagtaas ng antas ng privacy, ang data sa hinaharap lamang ay apektado. Ang data na sabay na ginawa ng publiko hindi kailanman maging pribado muli. Isaalang-alang ito kapag nag-upload ka ng data sa unang pagkakataon.

Data Consumer

Data Consumer — Data Consumer - isang siyentipiko, komersyal na kumpanya, siyentipikong organisasyon o ibang kalahok sa platform, na interesado sa pagtatasa ng genetic data gamit ang platform mga kapasidad. Ang mamimili ng datos ay maaaring gumawa ng mga kahilingan sa mga gumagamit at magtakda ng refundna babayaran sa mga tumutugon na mga gumagamit

Tandaan: mayroong ilang mga paghihigpit para sa mga query na idinisenyo upang pigilan sistema ng gumagamit de-anonymization at iba pang mga maling paggamit ng platform. Ang mga paghihigpit na ito nakasalalay din sa pagrango ng consumer sa system.

TANDAAN

Service Provider

Tagapaglaan ng Serbisyo - isang organisasyon na gumagamit ng genetic data sa negosyo nito at sa gayon nagpapatupad ng isang serbisyo ng gumagamit sa platform. Maaaring piliin ng gumagamit ang kanyang

sarili kung aling data ang nais niyang ibahagi sa serbisyo provider.

TANDAAN

Maabisuhan ang user kung magagamit ang hiniling na data upang makilala siya

TANDAAN

Halimbawa, hindi makita ng service provider ang higit sa 70% ng listahan ng mutasyon malinaw na form gamit ang direktang query o upang makuha ang impormasyon na tumutugma raw data (tulad ng fastq file) sa questionnaire ng gumagamit

Genomic data

Mga uri ng mga omics (genomic) na data

Sa loob ng balangkas ng konsepto, ang 3 uri ng omics (genomic) na data ay maaaring kilalang depende sa pagbabayad at ang halaga ng impormasyon:

- Buksan ang data, na hindi mahalaga para sa mga may-ari nito, ngunit mahalaga para sa siyentipiko.

Halimbawa: Genome of *Helicobacter pylori* bacteria strain.

- Buksan ang data, na mahalaga kapwa para sa may-ari nito at para sa kasunduan.

Halimbawa: Mga genome ng karamihan ng mga kalahok sa network.

- Pinaghihigpit na data na nakaimbak lamang sa loob ng network.

Para sa mga pinaghihigpitang proyekto ng iba't ibang mga komersyal at pampublikong institusyon.

Preprocessing ng genetic data

Ang paghawak ng NGS genomic data (at karamihan sa iba pang mga dataset omics) kadalasan binubuo ng dalawang malayang hakbang:

1. Preliminary processing ng "raw" data.

2. Dedikadong pagtatasa ng genetic sequences para sa pagbuo ng personal mga rekomendasyon o sa loob ng balangkas ng isang pananaliksik.

Isang reference genome - ay isang digital DNA dataset na binuo bilang isang

TANDAAN

Isang reference genome - ay isang digital DNA dataset na binuo bilang isang kinatawan halimbawa ng hanay ng mga genes 'ng species.

Ang preprocessing ng NGS genomic data ay kinabibilangan ng mga sumusunod na hakbang:

1. Alignment ng NGS bumabasa sa reference genome.

2. Paghahanap para sa mutations at iba pang mga pagkakaiba mula sa reference genome, at pag-save ng kanilang listahan sa gVCF na format.

TANDAAN | *Tandaan: Ang protocol ay pareho para sa data ng mga di-pantaong organismo.*

Siyempre, sa kasong ito ang isang naaangkop na reference genome ay kinakailangan. Kung ang personal na dataset ay kumakatawan sa resulta ng microarray genotyping technology (23andMe uri ng file), maaari rin itong mai-upload sa platform bilang gVCF simula noong katulad ng mga format ng data ng mga uri ng file na ito.

Ang Pekeng Problema sa Data

Kung ang genetic data ng iba pang (di-pantao) na organismo ay na-upload sa halip ng tamang genome (aksidente o kusa), ito ay makikita sa panahon ng preprocessing ng raw data at ang user ay aabisuhan. Kung sinasadya ng user nag-upload ng mga pekeng (pekeng) genetic data sa system, maaari itong makita gamit maraming kilalang pamamaraan ng pag-verify bago i-save ang mga ito sa imbakan.

TANDAAN | The Ang pang-ekonomiyang pag-uudyok upang pigilin ang pag-upload ng pekeng data ay ang pagbabayad para sa imbakan ng datos ay dapat gawin sa buong taon.year.

Ang problema ng pagkakakilanlan ng user batay sa genomic data

Buksan ang access sa genetic na impormasyon itinaas ang problema ng pagtukoy ng mga gumagamit sa pamamagitan ng kanilang genomic at iba pang data. Kung ang isang gumagamit ay nagpasiya na hindi lubusang ibunyag ang kanyang genetiko impormasyon, angkop na mga hakbang ay dapat gawin sa bawat hakbang ng pagproseso at pagtatago ng data. Upang malutas ang problema ng pagkakakilanlan ng user, ang pakikipag-ugnayan ay dapat na dinisenyo sa isang paraan na sa bawat yugto walang node maaaring matukoy ang pagmamay-ari ng genetic na materyal sa pamamagitan ng partikular na indibidwal, o kahit na ang lungsod kung saan nakatira ang indibidwal na ito.

TANDAAN | Ang mga pagkakaiba sa pagitan ng mga residente ng isang lungsod ay bumubuo ng halos 0.01%

Sa bawat yugto, ang layunin sa itaas ay nakamit sa iba't ibang paraan:

1. Sa preprocessing phase - sa pamamagitan ng paghahati ng source file sa mga bahagi upang ang average coverage ay mas mababa kaysa sa confidence threshold (6 na kopya).
2. Sa yugto ng imbakan - ang imbakan ng account na pira-piraso ng haba.

Imbakan ng genomic data

Ang genomic data ay matatagpuan sa isang ipinamamahagi na network batay sa DHT Kademia protocol. Ang mga kalahok na nagbibigay ng mga mapagkukunan para sa pagpapatakbo ng network na ito (tingnan ang mga detalye tungkol sa papel ng Node) tumanggap ng mga pagbabayad para dito sa mga token ng ZNA. Sa upang makatanggap ng pagbabayad, kailangan nila upang patunayan sa network na sila talaga pagtatago ng mga datos na ito. Ang pamamaraan ng pagsusuri na ito ay batay sa paggamit ng blockchain. Ginagamit ang enkripsiyon kapag kinakailangan.

NOTE | *tandaan: Ang pagsasama sa Storj at FileCoin node ay ipapatupad din..*

Tulad ng nabanggit na, ang data ay maaaring raw at naproseso.

Uri ng data	Raw	Processed
Features		
Format	fastq / bam	(lists of mutations) gvcf / vcf + bed / 23me(txt)
Size, Gb	50	2
Halaga	To improve the technology of sequencing and processing (for the equipment development market)	To conduct research, as well as to make a report.
Storage conditions		
Number of copies	At least 3 in the independent nodes	At least 5 in the independent nodes

Ang genomic data ay nakaimbak na nahahati sa mga fragment sa isang paraan na ang haba ng ang isang fragment ay hindi pinapayagan na unambiguously kilalanin na ito fragment nabibilang sa isang partikular na indibidwal.

TANDAAN: Ang impormasyon tungkol sa kung aling mga fragment ng genome ay bumubuo sa genome ng gumagamit pribado rin at maaaring makuha lamang sa pahintulot ng gumagamit.

Personal na profiles

Ang pagpuno ng mga personal na mga questionnaires ay nagpapataas ng makabuluhang paggamit ng genomic data. Ang mga gumagamit ay punan ang questionnaire gamit ang graphical interface.

If Kung ang ilang mga palatanungan ay naging popular, ang application ay nagsasabi sa gumagamit na punan ito. Ang bawat analyst ay maaaring lumikha ng kanyang sariling palatanungan at ilagay ito sa platform.

Pagtutukoy ng questionnaire

Ang isang bilang ng mga questionnaires ay maaaring malaki; samakatuwid, ito ay kinakailangan upang ipakilala ang konsepto ng pagtutukoy ng questionnaire..

Pagtutukoy ng questionnaire - ay isang kumpletong paglalarawan ng lahat ng larangan ng ang palatanungan at ang pinahihintulutang halaga ng mga larangan

Sa pormal, ang pagtutukoy ng palatanungan ay naglalaman ng sanggunian sa may-akda, isang paglalarawan at isang nakaayos na hanay ng mga talaan, na ang bawat isa ay tumutugma sa isang larangan ng palatanungan.

Ang mga patlang ay maaaring may ilang mga uri:

Ang numerong field Ang halaga ng patlang ay isang integer.

Maramihang pagpili Ang halaga ng patlang ay isang bilang ng isang sagot.

The Ang mga sagot sa mga uri ng mga katanungan ay ilalagay sa bukas na pag-access (walang anumang sanggunian sa user), dahil hindi sila nagbabanta sa privacy.

String field String field Ang halaga ng patlang ay isang string. Ito ay isang pribadong larangan, dahil ito

Pribadong block Pinapayagan ang paggawa ng anumang hanay ng mga patlang pribado, hindi alintana ng kanilang mga tunay na uri.

Queries to the system.

Ang istatistika ng data tungkol sa ilang mga katayuan ng genomic ay bukas sa publiko kung ang hindi nagpasya ang may-ari na i-encrypt ang mga ito. Bilang karagdagan, ang impormasyon tungkol sa Available din ang mga sagot sa mga questionnaire. Samakatuwid, alam ng lahat, halimbawa, ang bilang ng mga user ng network na may edad na 25 taong gulang o may mutasyon rs6025 (coagulation factor V).

Ang arkitektura ng system ay ginagawang imposible na kunin ang buong database:

Sa panahon ng paglikha ng mga nakakaugnay na mga katanungan, ang customer ay hindi makakuha ng access sa raw na data.

- Kasama sa pangunahing bayad ang isang limitadong bilang ng mga kahilingan bawat araw. Ang bayad para sa Ang karagdagang mga query na kasama sa quota ay lumalaki exponentially sa isang araw.
- Kung ang resulta ng nauugnay na query ay naglalaman ng mas mababa sa 100 mga gumagamit, ang result ay hindi ipagkakaloob sa customer.

Ang genome ay binubuo ng, ay pinahihintulatang ililipat. Walang analyst ang makakaalam kung anong mga uri ng data ang naka-encrypt.

Secure transfer ng personal na data

Ang proseso ng paglilipat ng personal na data sa pagitan ng mga kalahok sa sistema ay dapat ariin ang mga sumusunod na katangian:

1. Ang buong data na ipinapadala ay dapat na magagamit lamang sa mga mamimili at sa nagbebenta.
2. Ang pagpapadala ng mga token ay dapat lamang maganap kung ang data ay nagging Matagumpay na ipinadala.
3. Ang isang pagtatangka na ibenta ang hindi tamang data ay dapat makilala at mai-block.
4. Ang isang pagtatangka na sadyang maling akusahan ang nagbebenta sa pagbebenta ng hindi tama Dapat na maipahayag ang data.
5. Hindi dapat mapagkakatiwalaan ang paghahatid ng data sa ilang third party.

NOTE

The blockchain technology will be used for secure data transfer. However, it should be considered that storing (and transferring) large amounts of information to the blockchain is resource consuming. Therefore:

6. Pinapayagan na ilipat lamang ang isang maliit na halaga ng data sa pamamagitan ng blockchain. Ang natitirang data ay maaaring ipadala sa pamamagitan ng isang simpleng naka-encrypt channel ng komunikasyon.

Rating system

Ang platform ay lilikha ng rating para sa:

- Hiwalay para sa bawat genetic fragment o bloke ng personal na data
- Organisasyon
- Mga service provider
- Mga tagapagkaloob ng data (tulad ng mga laboratoryo ng DNA sequencing)

TANDAAN

There lang indibidwal na rating ng user, dahil ito ay aktwal na kumakatawan sa kabuuan ng kanyang mga genetic fragment rating, na hindi magagamit sa publiko. Kung aAng genetic data ay na-upload gamit ang account ng Samahan, pagkatapos ay ang unang rating ay awtomatikong tumaas ng rating ng samahan.

Ang mga kadahilanan na nakakaapekto sa rating ng mga fragment ng genetiko
Ang kumpirmasyon mula sa laboratoryo ay nagdaragdag ng rating ng na-upload na data sa proporsyon sa rating ng lab kung saan sila nakuha.

Ang pagkumpirma ay maaaring kumakatawan sa isang digital na lagda ng laboratoryo, o data upload ng laboratoryo mismo sa kahilingan ng kliyente.

Suriin para sa kadahilanan Nagbibigay-daan upang suriin ang impormasyon ng genetic ng gumagamit gamit ang paunang natukoy istatistika ng mga modelo ng mga frequency ng polymorphism at pag-uugnay ng genetiko.

Ang modyul na ito ay nasa ilalim ng pag-unlad.

Paglahok sa pananaliksik Ang pagtaas ng rating sa bilang ng matagumpay pananaliksik na kinasasangkutan ng fragment. Kung ang resulta ng impormasyon

Use cases

Indibidwal na gumagamit

Para sa mga indibidwal na gumagamit ay may isang pagkakataon upang makuha ang kanilang genomic na impormasyon at upang buksan ito sa isang pinagkukunan ng kita. Ang kumbinasyon ng genome at nito Ang pakikipag-ugnayan sa kapaligiran ay isang mahalagang mapagkukunan ng impormasyon. Ang aming platform ay magbibigay-daan sa gumagamit na ligtas na pamahalaan ang mapagkukunan na ito.

Nagbibigay ang platform ng kakayahang ligtas na mag-imbak at magbahagi ng impormasyon sa genetiko, na nagpapahintulot sa mga user na makatanggap ng iba't ibang mga serbisyong genetiko. Narito ang ilang halimbawa:

- mga ulat at rekomendasyon sa nutrisyon, mga panganib ng sakit, pagpapaganda, diyeta, fitness
- maghanap ng mga kamag-anak at paglilinaw ng mga ninuno
- mga dating serbisyo
- indibidwal na pagpili ng damit, sapatos, pag-setup ng klima ng tahanan, paglalakbay destinasyon at mga lugar ng paninirahan
- iba't ibang mga pagkakaiba-iba ng mga ulat ng genetic para sa isang grupo ng mga indibidwal, para sa halimbawa ng mga sports team o mga grupo ng nagtatrabaho
- halos lahat ng aspeto ng buhay ng tao ay naiimpluwensyahan ng genetika, kaya't tingnan natin kung ano ang mga bagong kumpanya ay maaaring makabuo ng paggamit ng aming platform

Bukod sa pluralidad ng mga serbisyo, ang isang gumagamit ay binibigyan ng isang pagkakataon upang kumita mula sa ang kanyang genetic uniqueness sa pamamagitan ng pagbibigay ng data ng questionnaire sa mga kumpanya para sa pangangailangan sa pagsasaliksik. Kaya, ang indibidwal na data kasama ang impormasyon ng genetiko ay nagging isang analogue sa mga kalakal o mga mapagkukunan ng mineral

Kalusugan

Ang modernong pangangalagang pangkalusugan at personalized na gamot ay hindi maaaring isipin nang hindi ginagamit ng mga teknolohiya ng genomic. Ang plataporma ay magpapahintulot sa mga pasyente na ligtas na magbahagi ng genetiko
 Ang impormasyon na ginagamit sa klinika na may mga medikal na tauhan:

- ang indibidwal na dosis at hindi pagpaparaan sa mga gamot (halimbawa, indibidwal dosing ng anticoagulant warfarin batay sa mga genetic na katangian)
- personal na katanggap na mga saklaw ng mga biochemical body parameter (halimbawa, PSA marker)
- genetic predisposition sa iba't ibang mga sakit (halimbawa, isang mataas na panganib ng macular pagkabulok at ang pangangailangan para sa karagdagang pananaliksik at pag-iwas)
- Pag-transplant at organ donation. Maaaring ligtas na ibabahagi ng mga user ang impormasyon tungkol sa uri ng kanilang mga HLA antigens na tumutukoy sa compatibility sa pagitan ng mga indibidwal sa panahon ng paglipat. Kaya, ito ay magiging Posible upang lumikha ng isang secure na database ng mga donor at mga boluntaryo upang i-save ang mga buhay sa pamamagitan ng paglipat

Kumpanya

Mayroong dalawang uri ng Mga Kumpanya, una na nagbibigay ng mga user na may mga serbisyo batay sa genomic data, habang ang ikalawa ay interesado sa pagkuha ng genetic data mula sa mga gumagamit upang magsagawa ng kanilang sariling pananaliksik.

Ang unang uri ay inilarawan sa zone ng mga gumagamit ng paggamit. Ang ikalawang uri ay maaaring inilarawan bilang pagbili ng mga gumagamit ng genomic data upang magsagawa ng kanilang sariling pananaliksik at upang mapabuti ang mga katangian ng consumer ng mga produkto, pag-target ng genetic ng mga produkto at advertising, ang ilang mga halimbawa na kung saan ay:

- Halimbawa, ang isang parmasyutikong kumpanya ay nagbabalak na magpalabas ng isang bagong gamut na gumaganap laban sa mutant na protina ng kanser. Maaaring mahanap ng kumpanya ang sistema ang mga gumagamit, na survived sa sakit, upang bayaran ang mga ito para sa genetic data at upang makuha ang dalas ng mutasyon sa isang gene encoding ang protina na naka-target sa pamamagitan ng aktibong sangkap
- Ang mga plano ng kumpanya ng mamimili upang magpasok ng isang bagong merkado at kailangang subukan kung paano ang mga gumagamit malasahan ang lasa ng produkto. Ito ay kilala na ang ilang mga sanhi ng pampalasa pagkagalit sa mga carrier ng isang partikular na genetic variant ng gustatory receptor gene. Ang kumpanya ay nagpapadala sa isang network ng isang alok upang pag-aralan ang mga carrier ng genetic na variant na ito at binibili ang isa pang pampalasa, o hinahanap ang dalas ng genetic variant na ito sa iba't ibang mga merkado at kaya mga target ang produkto para sa iba't ibang mga merkado.

Ang pang-agham na komunidad

Para sa pang-agham na komunidad ang sistema ay nagbubukas ng posibilidad ng pag-iimbak, pagbabahagi at gumaganap ng pananaliksik na may iba't ibang genomic data. Dahil ang platform ay hindi limitado sa pagtatrabaho sa mga genome ng tao, maaari itong magamit para sa ligtas na imbakan at ang pagproseso ng genomic data, halimbawa, na may kaugnayan sa agrikultura (halaman, hayop, microorganisms).

Sa pangkalahatan, ang pagkakaroon ng mga ekosistema ay humahantong sa pagpapayaman ng siyentipiko komunidad sa pamamagitan ng pag-access sa pangkalahatang data ng populasyon, kahit na walang sanggunian ang mga indibidwal na questionnaires. Bilang karagdagan, may pahintulot ng mga gumagamit, maaari nilang ay naging bahagi rin ng siyentipikong pananaliksik.

Nagbibigay din ang platform ng ipinamamahagi kapangyarihan ng computational, access kung saan ay payagan ang pagproseso ng malalaking volume ng genetic data (katulad ng AWS na inangkop sa trabaho na may genetic data).

KABANATA 3

Bahagi ng Teknikal

Ibinahagi na mga bagay

subsystem

Isang konsepto ng ibinahagi subsystem

(Distributed) Subsystem - isang koleksyon ng ilang bahagi ng system

(platform) mga elemento at mga proseso na maaaring kinakatawan mula sa object-oriented pananaw bilang isang entity na may isang natatanging pagkakakilanlan at exhibits isang mahusay na tinukoy

panlabas na pag-uugali.

Upang magbigay ng isang detalyadong katangian ng isang subsystem, ang mga sumusunod na aspeto ay dapat inilarawan:

Upang magbigay ng isang detalyadong katangian ng isang subsystem, ang mga sumusunod na aspeto ay dapat inilarawan:

1. Istruktura: ang mga elemento at mga proseso na bumubuo sa sistema.

2. Panlabas na Pag-uugali: pakikipag-ugnayan ng isang subsystem bilang buo sa iba

mga kalahok. Sa partikular:

- Mga interface: isang koleksyon ng mga posibleng query sa isang subsystem bilang isang buo.

- Mga Pagkilos: mga pagkilos na kinuha ng isang subsystem tungkol sa iba pang sistema

mga kalahok.

3. Panloob na Estado: ang pinagsamang panloob na estado ng isang subsystem.

TANDAAN

Ang prefix na "quasi-" ay nangangahulugang ang tunay na pakikipag-ugnayan ay nagaganap sa subsystem Ang mga elemento, na kung saan, ay magkakaroon ng kumplikadong pakikipag-ugnayan sa bawat isa bilang isang bahagi ng tunay na pakikipag-ugnayan na ito, upang ang lahat ng mga bagay na ito ay maituturing na pakikipag-ugnayan

Sa ibaba ay mag-drop kami ng mga pagkakaiba sa pagitan ng isang subsystem at ang representasyon nito bilang

Ang mga pakikipag-ugnayan sa isang subsystem ay maaaring kinakatawan bilang isang pagkakasunod-sunod ng medyo maliit

bilang ng mga pangunahing pagpapatakbo:

- Subsystem interface, iyon ay, ang mga aksyon ng iba pang mga kalahok tungkol sa subsystem.
- Ang mga aksyon tungkol sa iba pang mga kalahok sa platform.
- Mga panloob na proseso na nagbabago sa panloob na estado ng isang sistema.

Panloob na proseso

Panloob na proseso sa isang malawak na kahulugan - kumakatawan sa isang koleksyon ng lahat ng panloob mga proseso ng bawat elemento ng subsystem at mga pakikipag-ugnayan sa pagitan ng mga elementong ito.

Panloob na proseso (sa isang makitid na kahulugan) - kumakatawan sa mga proseso sa loob ng subsystem bilang kabuuan na nagbabago sa panloob na estado nito. Buong paglalarawan ng panloob Ang mga proseso ay naglalaman ng lahat ng pag-uugali ng subsystem maliban sa mga isyu ng partikular nito pagpapatupad.

Main na ipinamamahagi subsystems ng platform

1. • Pangunahing Sistema Layer (Kritikal Infrastructure)
2. • Antas ng Imbakan at Pagproseso ng Data
3. • Mga Interaksyon sa Mataas na Antas
4. Ang mga sumusunod na subsystems ay kasama sa loob ng platform:

Basic System Layer

Mababang Antas sa Interoperation Basic subsystem ng palitan ng mensahe sa pagitan mga node ng network. Pinapayagan din nito ang paglikha ng mga talahanayan ng hash na ipinamamahagi

Mga pahintulot Ang imprastruktura ng mga account at access sa pamamahala pribadong impormasyon.

2. Data Storage at Processing Level

Mga antas Mga antas ng Abstraction para ma-access sa ibinahagi file system. Pagproseso ng Infrastructure para sa isang ipinamamahagi na mga kalkulasyon.

3. Mga Interaksyon sa Mataas na Antas

Mga secure na query Isang subsystem na nagbibigay-daan sa pagbuo ng mga alok para sa pagbili genomic data na maaaring ipakita sa naaangkop na mga gumagamit lamang.

Buksan ang mga operasyon ng data Nagbibigay ng mga tool para sa operasyon nang bukas (hindi pribadong) data.

Nagbibigay ng Platform ng Mga Panlabas na Serbisyo ang API para sa pagkonekta sa panlabas na sentro pinamamahalaang mga serbisyo sa platform. Binubuo ang ligtas na paglipat ng data gamit ang maginoo web protocol.

Low-Level Interoperation Layer

Distributed P2P network

Ibinahagi ang P2P network Ang batayan ng platform ay P2P network batay sa Kademia protocol. Ang network ay nagsasangkot sa mga computer ng mga gumagamit na naka-install Zenome software.

Node - ay isang node ng ipinamamahagi P2P network na kumakatawan sa isang computer ng gumagamit na may Zenome software na naka-install dito.

NOTE Dahil dito, isang network ng overlay ay itinayo sa pagitan ng mga devicenakikilahok sa network. Ito ay kumakatawan sa isang virtual na network kung saan sa bawat isa Ang kalahok na isang 'NodeId' ay itinalaga, na walang kaugnayan sa totoong IP-address ng aparato. Ang bawat node ay nag-iimbak ng listahan ng mga "proximal" node, kung saan ang distansya sa pagitan ng mga node ay kinakalkula batay sa node " NodeId 'at hindi nauugnay sa karaniwang distansya.

Ang mga node ay nag-iimbak ng data sa pamamagitan ng paggamit ng mga spreadsheet na hash na ibinahagi.

Ibinahagi ang mga katangian ng pagpapatupad ng network

Ginamit ang isang nabagong protocol specification. Ang pangunahing pagkakaiba ay ang mga sumusunod:

Ang mga node ay maaaring makipagpalitan ng mga di-makatwirang mga mensahe sa pagitan ng bawat isa. Ang isang node ay magagawang

upang maglipat ng mensahe sa ibang node na alam lamang ang 'NodeId' nito.

- Maraming mga talahanayan ng hash ang maaaring umiiral sa loob ng network. Natukoy ang hash table sa pamamagitan ng string key.
- Iba't ibang mga patakaran ng pag-iimbak at pagtatanggal ng halaga ay maaaring itakda para sa iba't ibang mga talahanayan.
- Ang data na inilipat mula sa isang node sa isa pa ay naka-encrypt (tingnan sa ibaba).

Mensahe palitan sa isang ipinamamahagi na kapaligiran

Kaya, ang mga kalahok ng peer-to-peer network ay maaaring makipagpalitan ng mga sumusunod

:PING	Patunayan na ang isang node ay buhay pa rin
STORE(T,K,V)	I-imbak ang halaga na 'V' sa pamamagitan ng susi 'K' sa talahanayan 'T' sa isang node na natatanggap ang mensahe
FIND_NODE(N)	Ang isang node na tumatanggap ng mensahe ay magpapadala ng data tungkol sa mga node na pinakamalapit sa node na 'N' sa mga node na alam nito.
FIND_VALUE(T,K)	Kung ang pares '(K, V)' ay naka-imbak sa isang tumatanggap na node, ipadala ang halaga na 'V', iba pa ipadala ang data tungkol sa mga kilalang node na kung saan ay "mas malapit
SEND(M,N,D?)	Ipinapadala ang mensaheng 'M' na maaari ring maglaman ng data na 'D' sa node.'N'. Ang 'FIND_NODE' ay ginagamit upang mahanap ang node.

NOTE | Tandaan: Ang antas ng pakikipag-ugnayan na ito ay antas ng transportasyon

Blockchain

ang pangunahing impormasyon

Ang platform ay gumagamit ng Ethereum blockchain na kumakatawan sa isang solong desentralisado virtual machine (EVM). Ang nais na sistema ng lohika ay maaaring ipatupad gamit ang smart na mga kontrata

QUOTE | Ang isang kontrata ay isang koleksyon ng code (mga function nito) at data (estado nito) na namamalagisa isang partikular na address sa Ethereum blockchain.

- Panimula sa Smart Contracts (Solidity manual)

Kaya, ang smart na mga kontrata ay maaaring mag-imbak ng data. Para sa data sa blockchain Ang abstract type system na inilarawan sa itaas ay may bisa pa rin (ngunit may ilang mga paghihigpit).

Ang isang subsystem para sa pakikipagtulungan sa Blockchain

NOTE | Pansin: ang mga detalye ng pagpapatupad ay maaaring magkakaiba depende sa platform

Upang magkaloob ng access sa Blockchain, kasama ang node software ang buong pagpapatupad ng Ethereum Node na nagbibigay ng access sa pamamagitan ng 'JSON-RPC 2.0'.

Ang mga lihim na key ay naka-imbak sa naka-encrypt na imbakan. Ang isang user ay sinenyasan upang itakda ang password sa unang paglunsad ng application.

Kahit na ito ay technically magagawa upang gamitin ang isang umiiral na account, ito ay inirerekomenda

NOTE | Kahit na ito ay technically magagawa upang gamitin ang isang umiiral na account, ito ay inirerekomenda

Pinapayagan ng application interface ang:

1. Create new account
2. Import an existing account
3. Set up the backup of private storage

NOTE | Pansin: Inirerekomenda na i-set up ang cloud backup dahil ito ay magpapahintulot sa panatilihin ang isang access sa account kahit na sa kaso kapag ang computer ay pisikal na hindi magagamit.

NOTE | *Pinapayagan ng software ang pagkuha ng access sa Ethereum node command line. Ang function na ito ay pangunahing inilaan para sa pag-debug. Hindi inirerekomenda na gumamit ng isang utos linya kung hindi mo maintindihan ang mga layunin nito.*

Panloob na mga token

Ang mga ekonomikong pakikipag-ugnayan sa loob ng sistema ay ibinibigay gamit ang panloob na mga token ng ZNA. Kinakatawan nila ang mga wastong Ethereum token at maaaring mabibili at mabenta sa exchangebahay.

Konsepto ng account

Ang account sa platform ay nagpapahintulot sa user na makipag-ugnay sa system sa pamamagitan ng paglalaro ng ilang kasabay ng mga tungkulin. Ang bawat papel sa account ay tumutugma sa hiwalay na smart na kontrata sa ang blockchain..

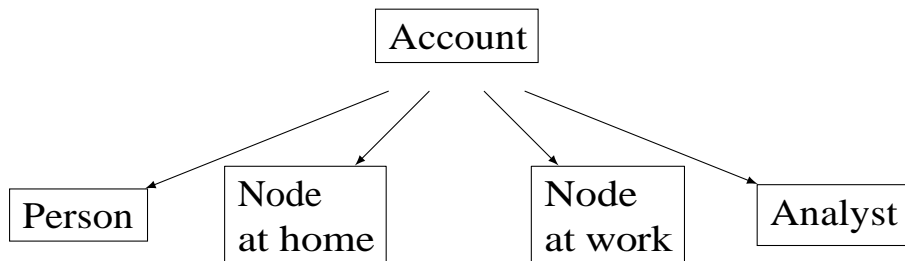


Figure 4: Ang account ng isang gumagamit na nagbigay ng kanyang genome (Tao role) sa platform, na gumagana bilang isang bioinformatician at sumusuporta sa dalawang computational node na matatagpuan sa bahay at sa trabaho.

Ang tungkulin - ay isang katangian ng kalahok ng sistema tungkol sa partikular na uri ng mga pakikipag-ugnayan kung saan siya ay tumatagal ng bahagi. Ang bawat kalahok ay maaaring maglaro ng ilang mga tungkulin sabay-sabay..

NOTE | Ang isang hiwalay na subsystem ay responsable para sa pamamahagi ng tungkulin.

Distributed data storage network

The principles of distributed network operation

Ibinahagi ang network ng imbakan ng data

Ang mga prinsipyo ng ibinahagi sa operasyon ng network Ang ibinahagi na network batay sa DHT Kademlia protocol ay ginagamit para sa pagtatago data ng mga arbitrary na uri. Ang mga node ng network na ito ay kinakatawan ng nakarehistro mga tagapagkaloob ng mapagkukunan, na kumita ng gantimpala para sa pagtatago ng data at pagsasagawa ng kailangan ng mga pag-compute. Ang kalahok na nag-upload ng data ay nagbabayad ng gantimpala para sa pag-iimbak ng data, ngunit, sa isang bilang ng mga mahahalagang mahalagang kaso ang mga singil na ito ay bahagyang o ganap na subsidized ng system. Halimbawa, mayroong isang mekanismo ng pagtustos ng pag-iimbak ng mahalagang data sa siyensiya. Ang yunit ng imbakan ay kinakatawan ng isang di-makatwirang block ng data at natatanging key sa pamamagitan ng kung saan maaari itong ma-access. The storage unit is represented by an arbitrary data block and unique key by which it can be accessed.

NOTE | Ang laki ng isang bloke ng data ay limitado upang maiwasan ang mga maling paggamit.

Suporta sa mga panlabas na serbisyo. Isang posibilidad ng pagsasama sa Storj at FileCoin node, pati na rin sa mga centrally managed storages, ay ipapatupad din.

Pagkatiwalaan ng imbakan ng data

Upang tiyakin ang pagiging maaasahan ng imbakan ng data, dahil sa ang katunayan na ang mga node ay maaaring malayang kumonekta sa at idiskonekta mula sa network, ang data ay naka-imbak nang nakapag-iisa maraming node nang sabay-sabay. Ang uri ng data na nakaimbak ay tumutukoy sa bilang ng node kung saan ang mga impormasyon ay nakakopya.

Talahanayan 6: Mga huwaran na parameter ng genetic data.

	Raw	Processed
Format	fastq/bam	gvcf/vcf + bed/23me(txt)
Size	50Gb	2Gb
Nodes	≥ 3	≥ 5

Kapag ang bilang ng mga node ay nagbabago, ang data na nakaimbak ay ipagkakalat muli upang masiyahan

ang kinakailangan para sa napakaliit na bilang ng mga node.

Dahil ang lahat ng nais na maging isang node sa ipinamamahagi na network, ito ay hindi ligtas na paniwalaan na ang node ay talagang nag-iimbak ng file kapag inaangkin ito. Kaya, ang tseke ng data retrievability ay pana-panahon na gumanap ayon sa nararapat protocol mula sa Security Layer. Ang data sa mga resulta ng pag-check ay ipinakilala sa ang blockchain at maaaring maging dahilan para sa pagpapalabas ng gantimpala o pagbabago ng node rating.

Data storage privacy assurance

Ang pagkapribado ng imbakan ng data sa naturang isang ibinahaging network ay nagiging posible dahil sa paggamit ng walang simetrya cryptography. Ang aktwal na paraan ng pag-encrypt ay tinutukoy ng kaukulang protocol ng seguridad.

NOTE | Dapat pansinin na ang mekanismo ng subsidizing ay hindi maaaring magamit para sa pagtatago

Specific features of storing genomic information

Ang pag-aari ng ilang fragment ng genome sa partikular na indibidwal ay maaaring hindi malinaw na nakilala kung ang fragment ay sa halip malaki. Ito ang dahilan kung bakit ang mga genome

sa ipinamamahagi na network ay nakaimbak na sa halip ay maikling mga fragment.

Ang pagkasira ng genome ay laging ginagawa batay sa reference genome. Para sa ang bawat reference genome (halimbawa, isa pang bersyon ng human reference genome

o isang genome ng organismo ng iba't ibang uri ng hayop) ang isang fragmentation ay pinili lamang minsan, at ang bawat fragment ay itinalaga na may identifier na natatangi para sa reference na ito genome.

User's personal data

Record - ay isang minimal na yunit ng palitan ng data. Hindi posible na ilipat (halimbawa, magbenta para sa isang gantimpala) isang bahagi ng impormasyon mula sa isang talaan. Ito ay posible na lumikha ng isang bagong rekord na naglalaman lamang ng isang bahagi ng data, ngunit sa tulad ng isang kaso na ito bagong record ay makabuluhang mas kawili-wiling para sa mga potensyal na mga mamimili dahil sa mababang rating nito. Tinutukoy ng scheme ng data kung aling impormasyon at kung saan dapat mag-record ang format

naglalaman ng. Paggamit ng scheme ay isang nababaluktot na kasangkapan para sa pag-iisa ng format ng palitan ng data

sa pagitan ng mga kalahok sa network. Ang tagatukoy ng scheme ng data ay maaaring anumang string na

nagbibigay-daan sa indibidwal na mahanap ang paglalarawan nito sa network, halimbawa, URL o smart address ng kontrata na may paglalarawan. Dapat itong maunawaan na pangunahing

Ang layunin ng pamamaraan ng data ay upang mapag-isa ang pagbuo ng kung ano ang kasama sa

ang data sa pagitan ng mga mamimili at ang mga nagbebenta ng mga data na ito.

Pagtutukoy ng Questionnaire

Ang pagpuno ng personal na mga questionnaires ay nagpapataas sa halaga ng data ng genomic sa kalahatan.

Ang isang malaking bilang ng mga questionnaires ay maaaring umiiral, sa gayon ito ay kinakailangan upang ipakilala ang

konsepto ng pagtutukoy ng questionnaire.

Ang pagtutukoy ng questionnaire - ay kumakatawan sa isang kumpletong paglalarawan ng lahat ng larangan ng

ang palatanungan at ang pinahihintulutang halaga ng mga larangang ito.

Sa pormal, isang detalye ng palatanungan ay naglalaman ng reference ng may-akda, isang paglalarawan

at nag-order ng listahan ng mga talaan na bawat isa ay tumutugma sa isang larangan ng questionnaire. Ang numerong field Ang sagot ng isang user ay dapat nasa loob ng pinahihintulutang halaga

saklaw ng [a, b]. Ang sagot ay naka-code ng unsigned integer, kung saan nagsisimula ang count

mula sa kaliwang hangganan ng saklaw:

Ang sagot sa ganitong uri ng tanong ay magagamit ng publiko dahil hindi ito

Pagtutukoy ng Questionnaire

Ang pagpuno ng personal na mga questionnaires ay nagpapataas sa halaga ng data ng genomic sa kalahatan. Ang isang malaking bilang ng mga questionnaires ay maaaring

umiiral, sa gayon ito ay kinakailangan upang ipakilala ang konsepto ng pagtutukoy ng questionnaire.

Ang pagtutukoy ng questionnaire - ay kumakatawan sa isang kumpletong paglalarawan ng lahat ng larangan ng ang palatanungan at ang pinahihintulatang halaga ng mga larangang ito.

NOTE

Ang sagot sa ganitong uri ng tanong ay magagamit ng publiko dahil hindi ito nagbabanta sa privacy.

Ang isang sagot sa ganitong uri ng tanong ay isang pribadong impormasyon.

Multiple choice Ang sagot ay naka-code ng unsigned integer, kung saan nagsisimula ang countmula sa kaliwang hangganan ng saklaw:

Ang sagot sa ganitong uri ng tanong ay magagamit ng publiko dahil hindi ito nagbabanta sa privacy. Ang isang sagot sa ganitong uri ng tanong ay isang pribadong impormasyon.

NOTE

Pansin: ang sagot sa ganitong uri ng tanong ay magagamit ng publiko dahil ito ay ginagawa hindi nagbabanta sa privacy

String answer Ang sagot ng gumagamit ay naka-imbak bilang isang string.

NOTE

Ang isang sagot sa ganitong uri ng tanong ay isang pribadong impormasyon. *Ang istrakturang ito ay ganap na naka-encrypt kung kinakailangan. Ito ay naka-imbak ng hindi bababa sa gumagamit computer. Ang isang user ay maaaring mag-upload ng naka-encrypt na backup sa blockchain kung gusto niya sa.*

Ang kahilingan ng personal na data mula sa mga gumagamit ay may ilang natatanging katangian:

- Hindi lahat ng mga gumagamit ay nagtutupad ng pamantayan ng partikular na pananaliksik. Upang suriin kung ang data ay angkop, ang isang access sa personal na impormasyon ay kinakailangan.
- Ang personal na impormasyon ay hindi dapat ilipat sa labas ng computer ng gumagamit alinman sa tahasan (direkta) o implicitly ("warrant canary").
- Sa isang kaso kung ang user ay nakakatugon sa pamantayan, makakatanggap siya ng isang alok ibahagi ang kanyang personal na data para sa isang gantimpala

CONCL. 1

Dahil hindi pinahihintulutan ang paglipat ng impormasyon, ang pagsuri ay dapat gumanap sa nakahiwalay na kapaligiran. Ang code na maipapatupad sa kapaligiran na ito magkaroon ng ganap na access sa pribadong impormasyon ngunit hindi maaaring makipag-ugnayan sa iba pang mga bahagi ng sistema.

CONCL. 2

Ang resulta ng pagpapatupad ng function na ihiwalay ay hindi pinahihintulutang ipadala bilang isang sagutin ang potensyal na customer dahil nagbabanta din ito sa privacy.

CONCL. 3 | Makatuwirang pumili ng pinaghihigpitang hanay ng personal na data, na kinakailangan upang makagawa ng isang konklusyon sa isang hiwalay na hakbang ng pag-check. Makikita ng user kung ano

CONCL. 4 | Ang desisyon ay batay sa screen ng alok ng palitan. Malinaw na inililipat ang data na nakalista sa screen ng alok ng palitan. Tanging ang data na hiniling na gumawa ng isang tseke ay maaaring mailipat

Kahilingan sa personal na data:

- Sa unang hakbang isang access ay hiniling sa personal na data ng user.

Ang mga pagkakakilanlan ng data na kung saan ay nakalista sa kahilingan na ito ay magagamit sa pag-check function.

- Sa pangalawang hakbang, ang code ng checking function ay papatayin sa loob ng isang nakahiwalay na kapaligiran, at ang resulta ng pagpapatupad nito ay kumakatawan sa nag-aalay na alok tungkol sa pagpapalit ng data o nag-aalok ng pagkansela. Sa huli kaso walang data ay inilipat kahit saan. Sa dating kaso ang isang user ay aabisuhan kapag ang alok ng palitan ay binuo, pagkatapos ay sinuri niya ang alok, ang listahan ng data na ginamit sa paglagay ng tsek at ang listahan ng data na ililipat kung ang user ay sumang-ayon sa alok. Kung tinatanggihan ng user ang alok, walang data na inililipat. Kung tinatanggap ng user ang alok, tanging ang data na tahasang ipinakita sa gumagamit ay inilipat.

Bibliography

- [1] NHGRI. *Talking Glossary of Genetic Terms. Word «Genome»*.
URL: <https://www.genome.gov/glossary/index.cfm?id=90>.
- [2] Venter J.C., Smith H.O., Adams M.D. “The Sequence of the Human Genome”. In: *Clinical Chemistry* 61.9 (2015), pp. 1207–1208.
URL: <http://clinchem.aaccjnls.org/content/61/9/1207.long>.
- [3] Adams M.D. Venter J.C. Smith H.O. “The Sequence of the Human Genome”.
In: *Science* 291.5507 (2001), pp. 1304–1351. ISSN: 0036-8075.
DOI: 10.1126/science.1058040.
URL: <http://science.sciencemag.org/content/291/5507/1304>.
- [4] Wetterstrand KA. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*.
URL: www.genome.gov/sequencingcostsdata.
- [5] Melissa Gymrek et al. “Identifying Personal Genomes by Surname Inference”.
In: *Science* 339.6117 (2013), pp. 321–324. ISSN: 0036-8075.
DOI: 10.1126/science.1229566.
URL: <http://science.sciencemag.org/content/339/6117/321>.
- [6] H. Christina Fan et al. “Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood”. In: *Proceedings of the National Academy of Sciences* 105.42 (2008), pp. 16266–16271.
DOI: 10.1073/pnas.0808319105.
URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2562413/>.
- [7] Sboner, Andrea and Mu, Xinmeng Jasmine and Greenbaum, Dov and Auerbach, Raymond K. and Gerstein, Mark B. “The real cost of sequencing: higher than you think!” In: *Genome Biology* 12.8 (Aug. 2011), p. 125. ISSN: 1474-760X.
DOI: 10.1186/gb-2011-12-8-125.
URL: <https://doi.org/10.1186/gb-2011-12-8-125>.
- [8] Rachel R. J. Kalf et al. “Variations in predicted risks in personal genome testing for common complex diseases”. In: *Genet Med* 16.1 (Jan. 2014), pp. 85–91. ISSN: 1098-3600.
DOI: 10.1038/gim.2013.80.
URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3883880/>.

- [9] Karen Norrgard. “Calculation of Complex Disease Risk”. In: *Nature* (2008). URL: <https://www.nature.com/scitable/topicpage/calculation-of-complex-disease-risk-756>.
- [10] Kathryn Nave. “How Craig Venter is fighting ageing with genome sequencing”. In: *WIRED UK* (2016). URL: <http://www.wired.co.uk/article/craig-venter-human-longevity-genome-diseases-ageing>.
- [11] Amalio Telenti et al. “Deep sequencing of 10,000 human genomes”. In: *Proceedings of the National Academy of Sciences* 113.42 (2016), pp. 11901–11906. DOI: 10.1073/pnas.1613365113. eprint: <http://www.pnas.org/content/113/42/11901.full.pdf>. URL: <http://www.pnas.org/content/113/42/11901.abstract>.
- [12] Luke Timmerman. “Google Translate Star Leaves Venter’s Human Longevity For Illumina-Backed Grail”. In: *Forbes* (2016). URL: <https://www.forbes.com/sites/luketimmerman/2016/09/27/google-translate-star-leaves-venters-human-longevity-for-illumina-backed-grail>.
- [13] João Sá Sousa et al. “Efficient and secure outsourcing of genomic data storage”. In: *BMC Medical Genomics* 10.2 (July 2017), p. 46. ISSN: 1755-8794. DOI: 10.1186/s12920-017-0275-0. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5547444/>.
- [14] Zachary D. Stephens et al. “Big Data: Astronomical or Genomical?” In: *PLOS Biology* 13.7 (July 2015), pp. 1–11. DOI: 10.1371/journal.pbio.1002195. URL: <https://doi.org/10.1371/journal.pbio.1002195>.
- [15] Erika Check Hayden. “Privacy protections: The genome hacker”. In: *Nature* (2013). URL: <http://www.nature.com/news/privacy-protections-the-genome-hacker-1.12940>.
- [16] Laure-Anne Pessina. “Reconstructing a face from DNA: an EPFL alumnus takes the stage at the 2016 TED Conference”. In: *School of Engineering (Federal Institute of Technology Lausanne)* (2016). URL: <http://sti.epfl.ch/page-129921-en.html>.
- [17] Melanie Swan. “Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen”. In: *Journal of Personalized Medicine* 2.3 (2012), pp. 93–118. ISSN: 2075-4426. DOI: 10.3390/jpm2030093. URL: <http://www.mdpi.com/2075-4426/2/3/93>.

- [18] Patrick Lin. “Blockchain: The Missing Link Between Genomics and Privacy?” In: *Forbes* (2017).
URL: <https://www.forbes.com/sites/patricklin/2017/05/08/blockchain-the-missing-link-between-genomics-and-privacy>.
- [19] Justin Zimmerman. “DNA Block Chain Project Boosts Research, Preserves Patient Anonymity”. In: *CoinDesk* (2014).
URL: <https://www.coindesk.com/israels-dna-bits-moves-beyond-currency-with-genes-blockchain/>.
- [20] D.S. Brama. “Method, System and Program Product for Transferring Genetic and Health Data”. US Patent App. 14/218,865. July 2015.
URL: <https://www.google.com/patents/US20150205929>.
- [21] Melanie Swan. *Blockchain: Blueprint for a New Economy*.
URL: <https://www.goodreads.com/book/show/24714901-blockchain>.
- [22] Matthew Herper. “Surprise! With \$60 Million Genentech Deal, 23andMe Has A Business Plan”. In: *Forbes* (2015).
URL: <https://www.forbes.com/sites/matthewherper/2015/01/06/surprise-with-60-million-genentech-deal-23andme-has-a-business-plan>.
- [23] “23andMe, Pfizer to Launch Inflammatory Bowel Disease Genetics Study”. In: *GenomeWeb* (2014).
URL: <https://www.genomeweb.com/clinical-genomics/23andme-pfizer-launch-inflammatory-bowel-disease-genetics-study>.
- [24] “Genetic Wild West: 23andMe Raw Data Contains 75 Alzheimer’s Mutations”. In: *Alzforum* (2017).
URL: <http://www.alzforum.org/news/community-news/genetic-wild-west-23andme-raw-data-contains-75-alzheimers-mutations>.