

Projekt Zenome: Whitepaper

*genski ekosustav zasnovan na
blockchain tehnologiji*

Nikolay Kulemin

Sergey Popov

Alexey Gorbachev

Listopad 6, 2017

Sažetak

Industrija 4.0 je naziv za trenutačni trend automatizacije, mjerenja i razmjene podataka u proizvodnoj industriji. Uključuje umjetnu inteligenciju, virtualnu stvarnost, Internet stvari i analiziranje Big Data podataka. Genomika je živopisan predstavnik industrije 4.0 koji zahtjeva hitno rješavanje mnogih problema kao što su pohrana i analiza Big Data podataka koji istovremeno znanstvenicima pruža javni pristup podacima, a privatnost korisnicima.

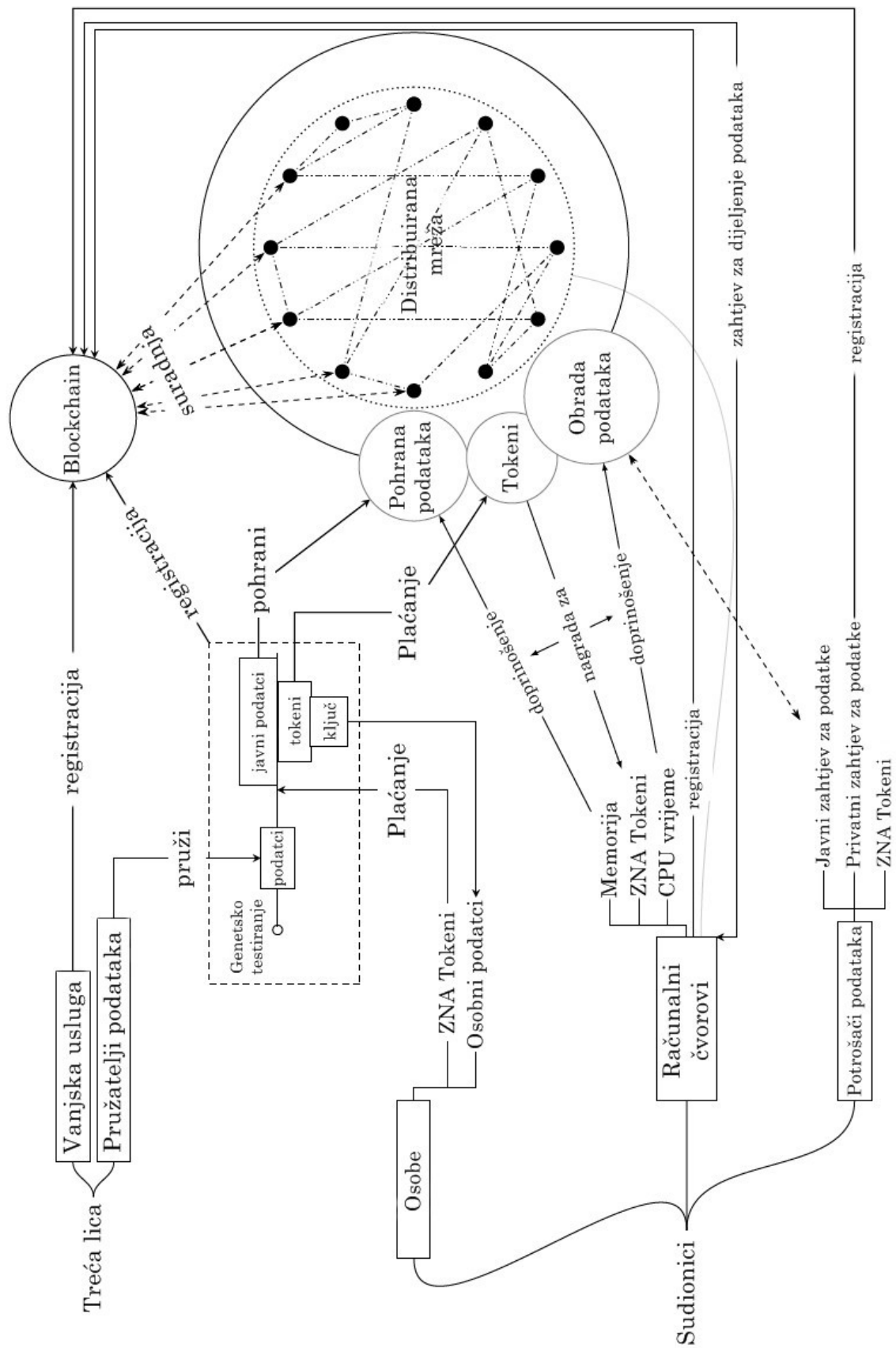
Trenutačno svjedočimo nejednakosti u genomskoj industriji. To znači da je veliki dio osobnih genomskih podataka koncentriran u podatkovnim centrima genomskih korporacija, vlada, znanstvenih i medicinskih ustanova i farmaceutskih tvrtki. Povrh toga, postoji problem pravnih ograničenja pristupa osobnim genomskim podacima kao i nedostatak mogućnosti upravljanja i dijeljenja genomskih podataka. Ovakva vrsta monopola nad genomskim podacima usporava razvoj na više znanstvenih i medicinskih područja.

Razvoj kripto valuta i blockchain tehnologije doveli su do značajne transformacije u mnogim ekonomskim područjima. Primjena blockchain načina pristupa problematici biti će osnovica napretka na polju osobnih genomskih podataka. Omogućiti će svakoj osobi da bude vlasnik svojih genetskih podataka.

Projekt Zenome je decentralizirana baza genomskih podataka zasnovana na blockchain tehnologiji. Ova platforma istovremeno omogućava upravljanje vašim genomskim podacima, zadržavanje privatnosti i mogućnost zarade prodajom pristupa različitim dijelovima genoma. Osigurati će jednake uvijete za razvoj lijekova i razvoj znanstvenih i medicinskih tehnologija.

Zenome je novi ekonomski okoliš zasnovan na genomskim podacima i blockchain tehnologiji. Implementacija našeg konceptualnog modela će riješiti sljedeće poteškoće:

- Stvaranje infrastrukture za pohranu Big genomskih podataka korištenjem distribuirane baze podataka
- Slobodan pristup milionima ljudskih genoma širom planete istovremeno osiguravajući zaštitu privatnosti
- Omogućiti će svakoj osobi da sudjeluje u znanstvenim i kliničkim istraživanjima i da zaradi na tome
- Potaknuti razvoj genomskih znanosti u zemljama u razvoju i smanjiti monopol koje razvijene zemlje drže nad genomskim podacima



Tim

Alexey Gorbachev

Osnivač

Molekularni biolog i blockchain entuzijast

Dr. molekularne biologije i biokemije

Sveučilište: Moscow State University, Odjel za molekularnu biologiju

Alex posjeduje zavidnu znanstvenu pozadinu, stručnost u poduzetništvu i upravljanju projektima. U Zenome Projektu Alex je odgovoran za opću viziju i razvoj projekta.

Email: alex@zenome.io

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

Nikolay Kulemin

Osnivač

Dr. bioinformatike

Sveučilište: Moscow Institute of Physics and Technology, Odjel za molekularnu i biologijsku fiziku

Specijalist u bioinformatičkoj i matematičkoj biologiji, posjeduje iskustvo u akademskim istraživanjima i primijenjenom razvoju u genomskoj industriji. Nikolay je osnivač tvrtke koja se bavi razvojem novih algoritama za genomsku analizu.

Email: nick@zenome.io

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

Vladimir Naumov

Znanstvenik genomskih podataka

Znanstvenik za podatke u ljudskom genomu

Sveučilište: Pirogov Russian National Research Medical University

Stručnjak za bioinformatiku i analizu podataka. Bivši glavni tajnik za sigurnost u iBinomu, bioinformatički znanstvenik u GERO, osmogodišnje iskustvo u genomskoj industriji. Radi na okupljanju farmaceutskih ispitnih grupa i razvoju novih načina analize i vizualizacije genomskih podataka.

Email: vov@zenome.io

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

Sergey Popov

Razvijač

Blockchain, P2P, distribuirani sustavi, čista matematika

Sveučilište: Moscow Institute of Physics and Technology, Odjel za opću i primijenjenu fiziku

Iskustvo u teoretskoj fizici, teoretskoj informatici, matematici, razvoju distribuiranih sustava i pametnih ugovora.

Email: sp@zenome.io

Dmitry Kwon

Savjetnik

Doktor molekularne biologije, savjetnik za razvoj poduzetništva

Sveučilište: Moscow State University, Odjel molekularne biologije

Stručnjak za molekularnu genetiku. Dmitry ima izuzetnu znanstvenu pozadinu, ekspertize iz tehnologije za genetsku analizu, genomska i dijagnostička tržišta, uspješno poslovno iskustvo u biotehnologiji.

Email: dk@zenome.io

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

SADRŽAJ

TIM

GENOMSKA TOKENOMIJA

POGLAVLJE 1 - GENOMIKA

1.1. POZADINA.....9

Genom (str. 9) The International Human Genome Project (str. 10) Razvoj genomske analize (str.10) Utjecaj na ostale znanosti (str. 12) Genetske baze podataka (str. 12)

1.2. PREGLED GENOMSKOG TRŽIŠTA.....13

Dinamika rasta genomskog tržišta (str.13) Područje primjene proizvoda (str. 14)

1.3. IZAZOVI ZA GENOMIKU.....15

Niska dostupnost genomskih analiza (str. 15) Kompromitiranje privatnosti kao cijena sudjelovanja u otvorenim biomedicinskim istraživanjima (str.17) Provođenje međunarodnih multicentralnih znanstvenih i kliničnih studija (str. 18) Osnivanje biobanaka za pohranu materijala različitih pojedinaca (str. 18) Obrada: brzina genomskih analiza je ograničena njihovim bioinformatičkom obradom. Podešavanje analitičkog softvera za masovnu uporabu (str. 20) Tumačenje genomskih podataka: matematički modeli rizika razvoja bolesti (str. 20) Tumačenje genomskih podataka: aplikacije za učenje uređaja (str. 21) Sigurno pohranjivanje vrlo velikih količina podataka (str. 22) Potreba za zajedničkom bazom podataka sa stalno ažurirajućim upitnicima (str. 23) Genomika u bliskoj budućnosti (str. 24)

1.4. ETIČKI PROBLEMI U OSOBNOJ GENOMICI.....24

Privatnost (str. 24) Publicitet (str. 25) Pravo na vlastite genomske podatke (str. 26) Pravo pristupa genomskim informacijama (str.27)

POGLAVLJE 2 - KONCEPT

2.1. PROJECT ZENOME.....28

Filozofski pogled (str. 28) Ekosustav platforme Zenome (str. 29) Pregled arhitekture sustava (str. 30) Tržište genetskih usluga (str.31)

2.2. ULOGE U ZENOME PLATFORMI.....33

Iz perspektive resursnog čvora (str. 33) Osoba/Korisnik (str. 34) Korisnik/Konzument podataka (str. 35) Pružatelj usluga (str. 35)

2.3. GENOMSKI PODATCI.....35

Vrste genomskih podataka (str. 35) Predobrada genetskih podataka (str. 36) Problem lažnih podataka (str. 37) Problem otkrivanja identiteta na osnovu genomskih podataka (str. 37) Pohrana genomskih podataka (str. 38)

2.4. OSOBNI PROFILI.....39

Specifikacija upitnika (str. 39) Upiti/ispitivanje sustava (str. 40) Siguran prijenos osobnih podataka (str. 40)

2.5. SUSTAV VREDNOVANJA (REPUTACIJE).....41

2.6. OBLICI/NAČINI UPORABE.....42

Individualni korisnik (str. 42) Zdravlje (str.42) Tvrtka (str. 43) Znanstvena zajednica (str. 43)

POGLAVLJE 3 – TEHNIČKI DIO

3.1. DISTRIBUIRANI OBJEKTI.....45

Koncept distribuiranog podsustava (str. 45) Unutarnji proces (str. 46)

3.2. GLAVNI DISTRIBUIRANI PODSUSTAVI PLATFORME.....46

3.3. INTERPRETACIJSKI NIVO NISKE RAZINE.....47

Distribuirana P2P mreža (str. 47) Značajke primjene distribuirane mreže (str. 48) Razmjena poruka u distribuiranom okruženju (str. 48)

3.4. BLOCKCHAIN.....49

Osnovne informacije (str. 49) Podsustav za rad sa blockchainom (str. 49) Tokeni sustava (str. 50) Koncept računa (str. 50)

3.5. MREŽA DISTRIBUIRANE POHRANE PODATAKA.....51

Načela rada distribuirane mreže (str. 51) Osiguranje pouzdanosti pohrane podataka (str.52) Osiguravanje privatnosti pohranjenih podataka (str. 52) Specifične značajke pohranjivanja genomskih informacija (str. 53)

3.6. OSOBNI PODATCI KORISNIKA.....53

Specifikacije upitnika (str. 53)

Uvod: Genomska Tokenomija

U nastavku će biti predstavljen opći opis Zenome platforme. Raspraviti će se potreba za tokenima i dobitima za potencijalnog ulagača.

U većini slučajeva, genomske podatke su pohranjeni u bazama podataka koje financiraju vlade ili velike tvrtke. Samostalno, ni jedna baza podataka ne posjeduje dovoljnu količinu podataka da ostvari kvantni skok prema eri genomike i precizne medicine. U isto vrijeme, svaka baza podataka posjeduje toliku količinu podataka da je nemoguće za jednu tvrtku da ih sama obradi.

Čini se da je razmjena genetičkih informacija od presudne važnosti.

Potencijalno genomsko tržište mora osigurati zaštitu od moguće zlouporabe, a pogotovo od genetske diskriminacije. Osobito je važno zadržati transparentnost i jednakopravan pristup ovom tržištu za sve.

Globalna razmjena genetičkih informacija mora riješiti sljedeće probleme:

- Fragmentaciju genetičkih podataka
- Ograničen pristup podacima za znanstvenike, medicinare i tvrtke
- Nisku dostupnost genetičkih ispitivanja
- Nedovoljne informatičke resurse

Zenome namjerava izgraditi vlastitu genomsku infrastrukturu koja će učesnicima omogućiti:

- Učitavanje genetičkih informacija i kontrolu nad njima
- Sigurno pohranjivanje osobnih genetičkih informacija
- Zarađivanje prodajom pristupa genetičkom materijalu ili dijelu istoga
- Podvrgavanju genetičkim ispitivanjima u zamjenu za korištenje genetičkih

podataka

- Razvoj personaliziranih prehrambenih planova, program treninga na osnovu genetske slike
- Korištenje genetičkih usluga

Glavni kupci genetskih informacija su tvrtke zainteresirane u razvoj genetsko ciljanih proizvoda Google, Facebook, Unilever i farmaceutske tvrtke.

Unutar Zenome platforme različite vrste podataka se neraskidivo isprepleću. Različite vrste podataka se pohranjuju na različite načine, ovisno o njihovim osobinama. Financijski podatci koji sadržavaju zapise o transakcijama su pohranjeni na blockchainu. Anonimizirani genetički podatci su pohranjeni u distribuiranoj mreži. Osobni podatci učesnika (korisnika) su pohranjeni samo na njihovim računalima. Upravljanje podacima različitih vrsta na različite načine će jamčiti privatnost i proširivost sustava.

Obzirom da su svi prenosi podataka, uključujući kupovinu i prodaju, upravljani pametnim ugovorima, u interakcije mogu biti uključena samo stanja pohranjena na blockchainu. Upotreba bilo kakvih trenutno već postojećih tokena u ovu svrhu bi rezultirala bespotrebnom ovisnosti o vrijednosti tog vanjskog tokena. Iz tog razloga će u opticaj biti pušten poseban token koji će pokretati ekonomiju ove platforme.

Zenome DNA (ZNA) je token platforme Zenome. Vrijednost ZNA tokena je ovisna direktno o uspjehu same platforme.

U ovom projektu (whitepaperu) ćemo do u detalje raspraviti najveće probleme u području genomike, kao i rješenja koja pruža platforma Zenome.

Poglavlje 1

Genomika

1.1. Pozadina

U ovom odjeljku ćemo dati definicije termina “genom“ i “genomika“. Razmotriti će se i povijest prvog pokušaja genomskog sekvenciranja i pojava tehnologije sekvenciranja nove generacije (NGS). Biti će opisani i glavni proizvođači reagensa i opreme koja se koristi za dobivanje genomskih podataka. Razmotriti će se i pitanja vezana za gomilanje genomskih podataka i smanjenje cijene analiziranja istih. Biti će dan osvrt i na trenutnu genomsku bazu podataka.

Genom

Genom je skup svih genetičkih instrukcija unutar stanice [1].

Genom sadrži biološke informacije nužne za razvoj i funkcioniranje organizma. Ljudski genom se sastoji od linearnih duplih spiralnih lanaca DNA molekula organiziranih u 22 para kromosoma plus dva spolna kromosoma X i Y. Sve informacije sadržane u genomu su kodirani kvartarnim kodom kroz nizove sačinjene od 4 nukleotida koji se označavaju sa A, T, C i G. Izraz “čitati genom“ znači “odrediti redoslijed nukleotida koristeći proces sekvenciranja“ [2].

Individualna sekvenca genoma definira čitav niz organskih osobina, uključujući

izgled, podložnost različitim bolestima, fizičke predispozicije, metabolizam, preferenciju određene hrane, kompatibilnost sa seksualnim partnerima (sposobnost začeća djeteta) i mnoge druge.

The International Human Genome Project

The International Human Genome Project¹ je pokrenut pod nadzorom NIH (National Institute of Health) 1990 sa zadaćom otkrivanja kompletne sekvence genoma haploidnog čovjeka. Prvi vođa projekta je bio jedan od otkrivača sastava DNA, dobitnik Nobelove nagrade, James Watson.

Skica sekvence ljudskog genoma je kompletirana sredinom 2000 i objavljena početkom 2001 u časopisu Nature. Troškovi ovoga međunarodnog projekta, financiranog javnim sredstvima, su bili oko 3 milijarde USD. 1998, privatna tvrtka, Celera Genomics se je pridružila utrci u otkrivanju ljudskog genoma. Vođa ovog privatnog projekta, koji se razvijao paralelno sa vladinim institucijama, bio je poznati znanstvenik i poduzetnik Craig Venter koji je uspio skupiti 300 miliona USD u privatnim ulaganjima u Celeras Project. Koristeći novi pristup sekvenciranju i računalnim metodama, sekvenca Craig Venterovog genoma je objavljena skoro istodobno sa podacima predstavljenim od strane međunarodnog konzorcija 2001[3] u časopisu Science. "Cijeli" ljudski genom je objavljen 2007, dok su neki dijelovi ljudskog genoma koji su teški za sekvencionirati ostali nepoznati.

Razvoj genomske analize

Opsežna ulaganja, veliki broj izvanrednih sudionika iz znanstvene zajednice, i natjecanje između privatnih i javnih organizacija su osigurali značajan poticaj za razvoj tehnologije analize genoma. Kao rezultat pojavila se je moderna tehnologija sekvenciranja nove generacije (NGS)² zajedno sa novom granom znanosti nazvanom bioinformatika. Bioinformatika je svojevrsno sjecište

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

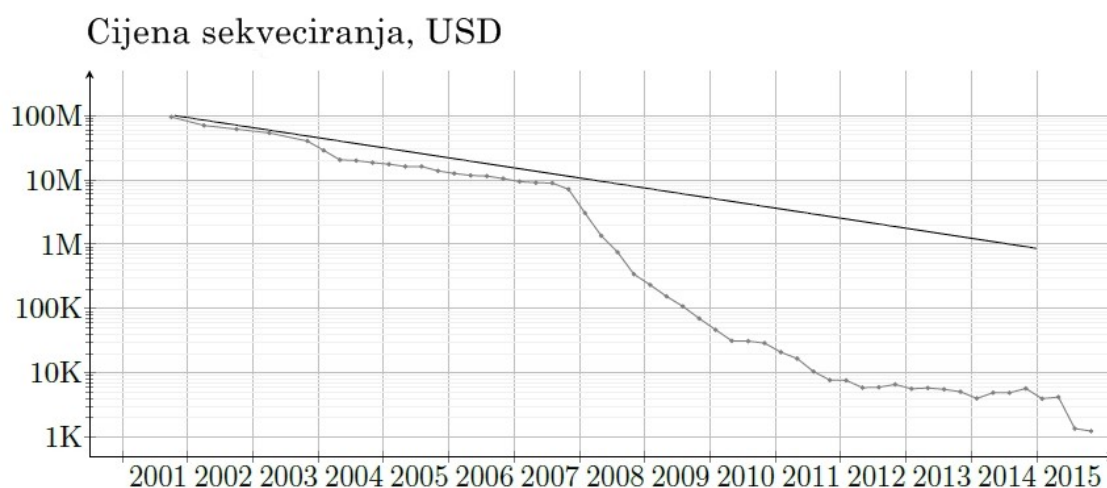
2 <https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-you-will-learn/what-next-generation-dna->
https://en.wikipedia.org/wiki/Moore's_law

matematike, IT i biologije. Bioinformatika razvija tehnike i algoritme za analiziranje ogromnih bioloških setova podataka na produktivan i računalno učinkovit način.

Pojavljivanje druge i treće generacije tehnologija za sekvenciranje (NGS) dovelo je do ogromnog pada troškova u genomskoj analizi. 2009 je trošak za punu analizu genoma iznosio 100 000 USD dok je danas taj trošak sveden na manje od 1000 USD (Slika 1 i Tabela 3).

Tabela 1: Razvijači opreme i opskrbljivači reagensa za genomsko sekvencioniranje. (Tržišna vrijednost preuzeta sa Yahoo Finance)

Tvrtka	Proizvod	Kapitalizacija	Država
Illumina	Hardver, reagensi, potrošni materijal, softver	28.06 B	USA
Thermo Fisher Scientific	Hardver, reagensi, potrošni materijal, softver (dio palete proizvoda)	68.98 B	USA
Oxford Nanopore Technologies	Hardver, reagensi, potrošni materijal	534.41 M	UK
Pacific BioScience	Hardver, reagensi, potrošni materijal	436.93 M	USA
Roche	Hardver, reagensi, potrošni materijal, softver (dio palete proizvoda)	213.44 B	Švicarska
Agilent Technologies	Hardver, reagensi, potrošni materijal, softver (dio palete proizvoda)	19.32 B	USA



Slika 1: Moorov zakon i smanjenje cijene genomske analize. Značajan pad cijene 2008 godine je rezultat pojave nove generacije tehnologije za sekvencioniranje (NGS) [4]

Utjecaj na ostale znanosti

Razvoj genomike (grane znanosti koja proučava različite genome) je doveo do transformacije na mnogim drugim znanstvenim poljima, od biologije i antropologije do medicine pa čak i društvenih znanosti. Veliki broj jakih svjetskih tvrtki, kao što su Google, Apple, IBM, Amazon i Alibaba, su postavili sebi zadatak upotrebe genomike za prilagodbu svojih proizvoda i usluga specifičnom genomskom profilu svojih kupaca. Takva prilagodba će omogućiti tim tvrtkama prisniji odnos sa kupcima i predviđanje njihovih potreba³.

Genetske baze podataka

Smanjenje cijene sekvenciranja je dovelo do eksponencijalnog povećanja u raspoloživosti genomskih podataka. Na primjer, kompletni ljudski genom u tzv. "raw podatkovnom formatu" može predstavljati od 50 GB do 2 TB podataka (ovisno o zahtijevanoj dubini sekvenciranja). Za potrebe pohrane tako velikih količina podataka razvijena je posebna genomski baza podataka koja sadrži različite vrste podataka kao što su, raw podatci preuzeti iz sekvenciranja, sekvence genoma i proteina, exoma, pa čak i sekvence cijelih genoma. Neke od tih baza podataka sadrže klinički relevantne podatke kao i podatke o vezi između genetskih osobina i bolesti. Većinom baza podataka, trenutno, upravljaju velike tvrtke ili vlade. Znanstvenici širom svijeta su uključeni u dodavanje novih vrsta podataka u ove baze što će omogućiti brzo ažuriranje i sinhronizaciju. U Tablici 2, prikazane su neke takve baze podataka.

Većina takvih baza podataka su smještene u razvijenim zemljama te su upravljane i kontrolirane od strane države. Pristup nekim od tih baza podataka je zabranjen, čak i za znanstvenu zajednicu ili je ograničen pretplatom. Iako osnivači baza podataka tvrde da su podatci pohranjeni sigurno i anonimno, u stvarnosti su ti podatci pohranjeni samo pod pseudonimima i, u nekim slučajevima, se pojedinac može bez problema identificirati na osnovu tih podataka [5].

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

1.2.Pregled genomskog tržišta

U ovom odjeljku biti će predstavljena kratka analiza tržišta genomske tehnologije. Biti će predstavljeni primjeri najpopularnijih genomskih proizvoda i tvrtki koje nude različite genomske usluge.

Dinamika rasta genomskog tržišta

Tržište genomskih tehnologija ubrzano raste i vrlo je obećavajuće. Ukupni tržišni volumen iznosi, trenutno, oko 25 milijardi USD, sa skoro deseterostrukim iznosom rasta sa 5.9 milijardi USD u 2010 na 60 milijardi USD u 2020 (predviđanje).

Tabela 2: Genomske baze podataka

GenBank	http://exac.broadinstitute.org
Vlasnik:	NCBI-NIH, USA
Proizvod:	Baza podataka genomskih sekvenci
Pohranjeno:	Više od 199 341 377 različitih genomskih sekvenci
ExaC	www.ncbi.nlm.nih.gov/genbank
Vlasnik:	Vlasnik: Broad Institute of MIT and Harvard, USA, ODC Open Database License
Proizvod:	Proizvod: Exome Aggregation Consortium
Pohranjeno:	Pohranjeno: 60 706 nizova ljudskih exoma
UniportKB	www.ebi.ac.uk/uniprot/
Vlasnik:	Vlasnik: EMBL-EBI, SIB, PIR, UK, Švicarska, USA
Proizvod:	Proizvod: Otvorena baza znanja, Sekvence bjelančevina i gena
Pohranjeno:	Pohranjeno: Više od 555 100 ručno pregledanih i zabilježenih zapisa
ClinVar	https://www.ncbi.nlm.nih.gov/clinvar/
Vlasnik:	Vlasnik: : NCBI-NIH, USA
Proizvod:	Proizvod: Besplatna arhiva za tumačenje kliničkog značaja genomskih varijacija
Pohranjeno:	Pohranjeno: >158 000 provedenih tumačenja što predstavlja > 125 000 varijacija
HGMD	http://www.hgmd.cf.ac.uk/ac/index.php
Vlasnik:	Vlasnik: : QiaGen
Proizvod:	Proizvod: Komercijalna baza podataka
Pohranjeno:	Pohranjeno: 208 368 zapisa o ljudskim mutacijama

SNPedia

<https://www.snpedia.com/index.php/SNPedia>

Vlasnik: Vlasnik: : Otvorena (slobodna) baza podataka

Proizvod: Proizvod: SNPedia je wiki projekt za istraživanje ljudske genetike

Pohranjeno: Pohranjeno: 107 073 SNP i povezanih zapisa

1000 Genomes Project

www.1000genomes.org

Vlasnik: Vlasnik: : EMBL-EBI, Wellcome Trust

Proizvod: Cilj: Pronalazak što više genetskih varijanti sa frekvencijom od najmanje 1%

Pohranjeno: Pohranjeno: Više od 2504 uzorka/sekvenci genoma

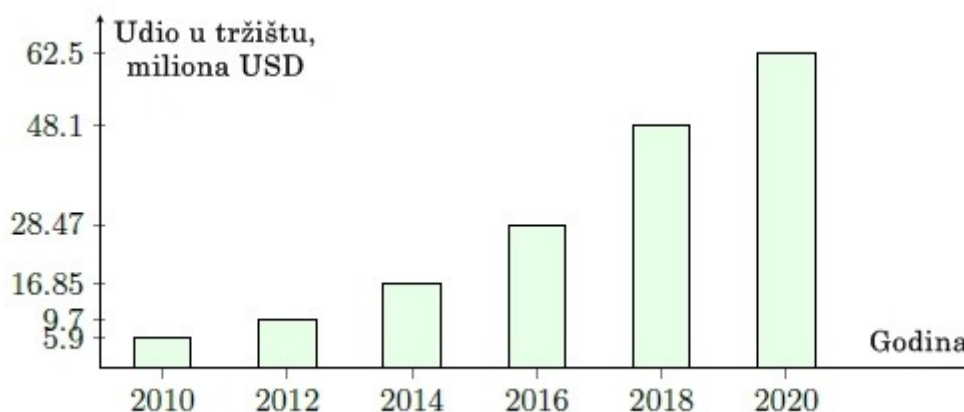
100000 Genomes Project

<http://www.genomicsengland.co.uk/>

Vlasnik: Vlasnik: : NHS, Vlada Velike Britanije

Proizvod: Proizvod: Baza podataka vlade Velike Britanije sa 100000 genomskih sekvenci

Pohranjeno: Pohranjeno: 32 642 kompletne genomske sekvence



Slika 2: Dinamika rasta genomskog tržišta za period 2010-2024

Područje primjene proizvoda

Ključna komponenta Industrije 4.0, genomika ima široki spektar potencijalne primjene u skoro svim područjima ekonomije. Glavna područja primjene proizvoda na trenutnom genomskom tržištu:

- NIPT (Neinvazivno prenatalno ispitivanje)

Opis:	Zasnovano na fetusnom DNA koji kola u majčinoj krvi. Ispitivanje može identificirati fetusnu aneuploidiju[6] i spol fetusa već od 6. tjedna trudnoće.
Tržišni udio:	4 milijarde USD
Tržišni lideri:	Illumina, Natera, Ariosa, Sequenom.

- PGS (Predimplantivno genetsko provjeravanje)

Opis:	Zasnovano na DNA mikropostrojima ili HGS profiliranju embrija genetskom sekvencom prije primjene IVF zahvata (in verito oplodnje)
Tržišni udio:	336.4 milona USD
Tržišni lideri:	Illumina, Agilent Technologies.

- DTC (Direktno potrošaču) genetsko ispitivanje, SPN-genotipizacija

Opis:	Genetsko ispitivanje zasnovano na DNA mikropostrojima SPN-genotipizacije za određivanje: podrijetla, prehrambenih smjernica, nositelja nasljednih bolesti, optimalnog plana tjelovježbe
Tržišni udio:	2 milijarde USD
Tržišni lideri:	AncestryDNA, 23andMe, DNAfit, deCode genetics.

- Dijagnostičke tvrtke (uključujući onkogenomiku)

Opis:	Različite vrste dijagnostičkih postupaka, uključujući i genetičke. Sekvencioniranje genskih panela, exoma, kompletnih genoma, likvidne biopsije.
Tržišni udio:	16 milijarde USD
Tržišni lideri:	Pathway Genomics, Human Longevity, Inc, Laboratory Corporation of America, Quest Diagnostics.

1.3. Izazovi za genomiku

U ovom odjeljku ćemo razmatrati glavne probleme suvremene genomike. Ti problemi se moraju riješiti da bi se mogao uvesti koncept Genomike 2.0, sveprisutnog širenja i primjene osobnih genomskih tehnologija

Niska dostupnost genomskih analiza

Cijene različitih vrsta genomskih analiza su dane u Tablici 3 i, općenito gledano, koštaju od 100 USD pa na više. Cijene su krajnje niske u usporedbi sa istim tim cijenama prije 5 do 10 godina[7].

Međutim, unatoč tome, cijene sekvenciranja i bioinformatičkog tumačenja su još

uvijek previsoke za većinu populacije, naročito u zemljama u razvoju.

Tablica 3: Vrste genomske analize

DNA mikropostroji

Opis:	Analiza 1-5 miliona prethodno odabranih SNP-a
Cjenovni rang:	\$ 100-500
Količina informacije:	0.033% cijelog genoma
Pružatelji usluge:	23andMe, AncestryDNA, DNAfit.

Genski paneli

Cjenovni rang:	\$ 100-2000
Količina informacije:	0.001-1 % cijelog genoma
Pružatelji usluge:	Pathway Genomics, CeGaT.

Sekvencioniranje exoma

Opis:	Sekvencioniranje kodnog dijela genoma (exoma)
Cjenovni rang:	\$ 250-3000
Količina informacije:	2 % cijelog genoma
Pružatelji usluge:	BGI, CeGaT.

Sekvenciranje kompletnog genoma

Cjenovni rang:	\$ 600-10000
Količina informacije:	80-98 % cijelog genoma
Pružatelji usluge:	BGI, FullGenomes, Human Longevity.

Bilješka: u stvarnosti dio genoma neće biti sekvenciran. Veličina ovisi o dubini sekvenciranja i detaljima procesa pripreme uzorka. Dakle, termin "Sekvenciranje cijelog genoma" u biti znači dobivanje sekvence od malo više od 80% genoma.

Dobivanje velike količine genomske informacije za žitelje zemalja u razvoju je izuzetno važno sa stajališta dobivanja što veće raznolikosti genomske informacije koje, će za uzvrat, značajno stimulirati razvoj geneomskog Big Data tržišta. Čak se je i u razvijenim zemljama samo 2% stanovništva podvrgnulo bilo kakvom obliku genomske analize (mikropostroji, sekvenciranje exoma ili sekvenciranje cijelog genoma).

Kompromitiranje privatnosti kao cijena sudjelovanja u otvorenim biomedicinskim istraživanjima

Informacijski obrazac za pristanak na pribavljanje i obradu osobnih podataka je osnovni problem u svakoj biomedicinskoj istraživačkoj studiji. Bilo koji projekt, koji uključuje studiju ljudskog genoma, započinje sa prikupljanjem potpisa individua kojim potvrđuju da su svjesni posljedica i da su suglasni sa uvjetima

istraživačke studije. Forma takvog obrazca za pristanak ovisi o projektu i može uključivati davanje dozvole na korištenje podataka i u budućnosti. Posljedice ovakvog pristupa mogu biti vrlo nepredvidive.

U Personal Genome Project (Projektu osobnog genoma), kojeg nadgleda Harvard Medical School (Medicinska škola sa Harvarda) sudionici dobrovoljno daju pristanak da se njihov genetički materijal može koristiti više puta i da bude dostupan i drugim laboratorijima.

Sudionici ovog projekta su posebno informirani da se može dogoditi da se njihov identitet otkrije, te da podatci postanu dostupni u javnosti. Ovaj projekt nastoji osigurati genomske podatke, od što je moguće većeg broja ljudi, koji će biti slobodno dostupni u nadi stimuliranja novih istraživanja i razvoja u genomskoj industriji. Pokretači ovog projekta vjeruju da riskiramo da završimo sa tisućama izoliranih, privatno pohranjenih zbirki genomskih podataka (do farmaceutskih tvrtki, genomskih korporacija i znanstvenih centara) ako ne osiguramo slobodan pristup genomskim podacima i razmijeni informacija. Ni jedna od izoliranih zbirki podataka neće imati dovoljnu količinu podataka da sama potakne nova revolucionarna otkrića.

Provođenje međunarodnih multicentralnih znanstvenih i kliničkih studija

Za provođenje novih istraživanja i razvoja na polju genomike nužno je provesti znanstvene i kliničke studije na velikom uzorku prikupljenom od različitih populacijskih grupa. Trenutno je prikupljanje uzoraka od pojedinaca sa različitim etničkim pozadinama komplicirano jer se moraju pokrenuti posebni projekti, organizirati ekspedicije te dobiti dozvole od lokalnih regulacijskih tijela.

Trenutno postoji samo jedan startup koji nastoji riješiti ovaj problem <https://www.dnasample.org/> u zamjenu za malu naknadu i obećanje anonimnosti.

Osnivanje biobanaka za pohranu materijala različitih pojedinaca

Biobanke služe kao posrednici između darivatelja biološkog materijala (uzorci krvi, koštane srži...) i znanstvenika na način da obrađuju i pohranjuju dobiveni

materijal za buduću upotrebu. Općenito gledano, biobanke su ključni alat za napredak osobne (precizne) medicine i razvoj lijekova. Jedna od najvažnijih funkcija biobanke je prikupljanje donorskog materijala, krvi, koštane srži pa čak i stanica zametaka za buduću upotrebu.

Trenutno se biobanke aktivno razvijaju u mnogo zemalja. Međutim postoji vjerojatnost da će najvrjedniji uzorci pripadati financijski najjačim biobankama (npr. biobankama u vlasništvu velikih farmaceutskih tvrtki) što će dovesti do nejednakog pristupa materijalu za različite kategorije istraživača. Iz tog razloga je potrebno osnivanje biobanaka u što je moguće većem broju zemalja i gradova.

Obrada: brzina genomskih analiza je ograničena njihovim bioinformatičkom obradom. Podešavanje analitičkog softvera za masovnu uporabu

Trenutno je brzina prikupljanja podataka korištenjem tehnologije genomske analize visoka i nadmašuje brzinu obrade podataka. Ako promotrimo bilo koju veliku znanstvenu studiju, na velikom broju genoma, vidjeti ćemo da dobavljanje genomske informacije ne oduzima više od 20% vremena trajanja studije. Sve ostalo vrijeme otpada na obradu podataka.

U ovom slučaju obrada podataka podrazumijeva skup koraka od pribavljanja sirovih sekvenciranih podataka pa sve do tumačenja rezultata i traženja zajedničkih poveznica.

Drugi problem sa modernim softwareom je činjenica da su software za genomske analize napisali znanstvenici za znanstvenike. Takvom softwareu je potrebna prilagodba da bude široko prihvaćen među liječnicima i općom populacijom. Kako genomski analizator veličine USB sticka već postoji, uporebu vlastitog sekvencera na isti način kao i osobnog računala ne trebamo gledati kao znanstvenu fantastiku nego nešto sasvim normalno u bliskoj budućnosti.

Tumačenje genomskih podataka: matematički modeli rizika razvoja bolesti

Za određivanje rizika razvoja bolesti na osnovu genetičkih podataka koriste se različiti modeli i algoritmi. Osnovni tipovi ovih modela se temelje na vrsti

nasljednosti koje se promatra: monogenska, poligenska i multifaktorijalna. Pri procjeni rizika razvoja multifaktorijalne ili složene bolesti potrebno je voditi računa o utjecaju i miješanju mnogo gena kao i na faktore okoline. Za detaljnije opise dostupnih metoda za procjenu rizika od kompleksnih bolesti pogledajte [8, 9].

Za razvoj ovog modela, za procjenjivanje rizika razvoja bolesti, nužno je uključiti veliki broj znanstvenika. Također je nužno istražiti što više već objavljenih publikacija, vezanih za bolest koju se misli analizirati, da bi se identificirao tip nasljednosti, polimorfizme i mutacije koji utječu na razvoj bolesti te razviti “genomsku algebru” tj. skup pravila za utvrđivanje rizika. Kada se model uspostavi i potvrdi trebaju se provesti klinička ispitivanja da se procijeni njegova primjenjivost. Trenutno je ovaj pristup najtočniji, ali je skup gledano sa strane izdvojenog vremena i truda.

Tumačenje genomskih podataka: aplikacije za učenje uređaja

Uporaba algoritama za učenje uređaja za procjenu rizika multifaktorijalnih bolesti se opsežno istražuje, ali, za sada, zbog malog broja trenažnih uzoraka nisu u rangu postojećih matematičkih modela koje su razvili biolozi.

Međutim, učeni uređaji se već koriste kod predviđanja određenih složenih osobina ljudskog tijela. Primjer je pojava predviđanja u radu Craiga Ventera i njegovih kolega. Srž njihovog rada koji obuhvaća analiziranje genoma je oko 30 000 facijalnih točaka skupljenih od tisuća dobrovoljaca. Na osnovu prikupljenih podataka razvijeni su trenažni algoritmi za učenje uređaja te određena ovisnost genomskih osobina i individualnog izgleda. Na osnovu ovoga rada, uređaji su naučili točno rekonstruirati izgled osobe na osnovu njegovih ili njenih genomskih podataka[10, 11].

Rezultati ovog projekta omogućuju predviđanje kriminalnog ponašanja ili fizičkog izgleda nerođenog djeteta tijekom ranog stadija trudnoće. Uzimanjem krvi trudnice i izoliranjem fetalne DNA, iz te krvi, može se točno utvrditi izgled nerođenog djeteta na njegov ili njen 18 rođendan.

Da bi ostvario ovaj projekt Craig Venter je vrbovao jednog od najboljih stručnjaka za učenje uređaja iz Googlea, Franza Ocha, kompjuterskog stručnjaka koji je bio glavni arhitekt Google Translatea [12].

Učenje uređaja trenutno nije u širokoj uporabi u istraživanju bolesti jer mu je potreban vrlo velik i pravilno složen uzorak za učenje. Stvaranje opsežne baze podataka ljudskog genoma, kao i mogućnost detaljnog upitnika koji će odraziti zdravstveno stanje pojedinca, može potaknuti razvoj računalnog treninga u genomici koji će rezultirati visokom točnošću u predviđanju rizika razvoja bolesti. Istovremeno, ovi podatci bi bili javni, dostupni za slobodnu uporabu svim korisnicima sustava, isključujući na taj način mogućnost monopola nad njima. Ovakva vrsta dostupnosti je krajnje važna jer bi koncentriranje velike količine podataka u bazama korporacija rezultiralo monopolom na polju genomike i učenja uređaja.

Sigurno pohranjivanje vrlo velikih količina podataka

Sigurnost osobnih podataka je vrlo važna, svi mi pazimo da naša kreditna kartica ne bude ukradena, štitimo podatke, brojeve bankovnih računa i medicinske informacije bilo kojeg oblika. Krađa genomskih podataka se trenutno većini ljudi čini nevažnom. Međutim, može dovesti do niza vrlo ozbiljnih posljedica koje su teške za predvidjeti, npr. mogućnost da se sintetizira dio nečije genomske sekvence te se podmetne na mjesto zločina ili terorističkog napada.

Trenutna rješenja ovoga problema uključuju enkripciju podataka na centralnom serveru, kao što su [13], <https://www.pathway.com/>, <https://www.23andme.com/> ili <http://www.humanlogevity.com/>. Ovakav način “zatvorene” pohrane podataka je relativno siguran, ali sprječava mogućnost dijeljenja podataka i pristupa podacima znanstvenicima iz cijelog svijeta, što je od izuzetne važnosti za razvoj moderne genomike.

Još jedan problem vezan za pohranjivanje podataka je veličina samoga genoma i eksponencijalni rast količine genomskih podataka (sve više i više ljudi se podvrgava sekvenciranju). Prema jednom istraživanju [14] predviđa se da će do 2025, ukupni volumen pohranjenih podataka (ako uzmemo u obzir da je veličina cijelog genoma oko 100 Gb) dostići 40 eksabajta godišnje, a pohrana genomskih podataka će postati najveći potrošač prostora za pohranu i kapaciteta za obradu podataka.

Tablica 4: Četiti Big Data domene 2025. godine

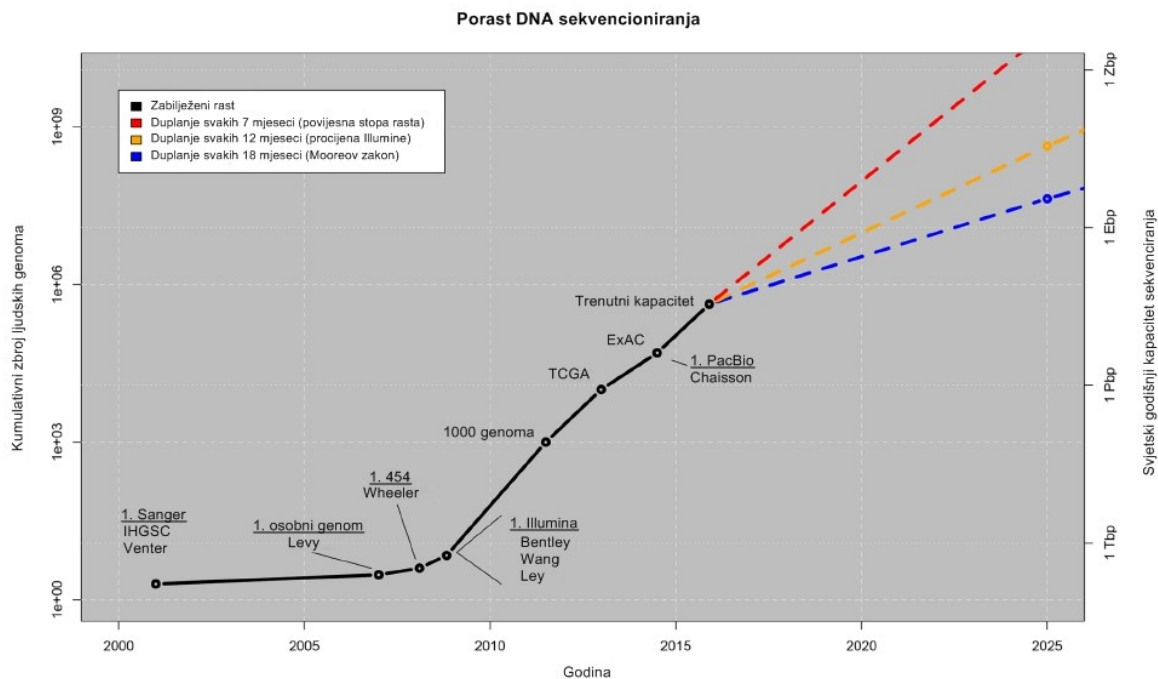
Astronomija	
Akvizicija:	25 zetabajta/godina
Pohrana:	1 EB/godina
Analiziranje:	Smanjenje podataka na licu mjesta, Obrada u stvarnom vremenu, Ogromni volumeni
Prijenos:	Stalna veza od antene do servera
Twitter	
Akvizicija:	0.5-15 milijardi tweetova godišnje
Pohrana:	1-17 PB/godina
Analiziranje:	Analiza naslova i podataka, Analiza metapodataka
Prijenos:	Male distribucijske jedinice
YouTube	
Akvizicija:	500-900 miliona sati godišnje
Pohrana:	1-2 EB/godina
Analiziranje:	Ograničeni zahtjevi
Prijenos:	Glavni udio u propusnosti modernih veza (10 MB/s)
Genomika	
Akvizicija:	1 zetabajt godišnje
Pohrana:	2-40 EB/godina
Analiziranje:	Heterogena analiza podataka, Povezivanje varijanti, oko 2 tiliuna procesorskih (CPU) sati rada, Poravnanje svih genomskih parova, oko 10 000 triliuna CPU radnih sati
Prijenos:	Mnogo malih (10 MB/s) i nekoliko većih (10 TB/s) prijenosa podataka

Potreba za zajedničkom bazom podataka sa stalno ažurirajućim upitnicima

Nedostatak javnih baza podataka zasnovanih na konceptu distribuirane pohrane i istovremeno niska stopa uporabe softvera otvorenog karaktera (open-source) mogu dovesti do apsolutne prevlasti velikih tvrtki, kao što su Google ili Amazon, na ovom području (prvenstveno radi snažnih servera ovih tvrtki⁴). Ukoliko korporacije ili farmaceutske tvrtke postanu monopolisti na polju genomskih informacija, biti ćemo svjedoci sporog i skupog razvoja medicine koja će ustrajati na trenutnom pristupu liječenju umjesto da se okrene budućnosti u kojoj će se bolesti moći spriječiti prije samog njihovog pojavljivanja ili u najranijim fazama

⁴ <https://cloud.google.com/genomics/>

njihove manifestacije.



Slika 3: Porast podataka DNA sekvencioniranja. Ovaj graf pokazuje porast DNA sekvencioniranja: ukupni broj sekvenciranih ljudskih genoma (lijeva os) i svjetski godišnji kapacitet sekvencija (desna os: parovi Tera reda veličine (Tbp), parovi Peta reda veličine (Ptb), parovi Eksa reda veličine (Ebp) i parovi Zeta reda veličine (Zbp)). Vrijednosti kroz 2015. godinu su zasnovani na povijesnim podacima iz publikacija sa naznačenim prekretnicama (od prve Sanger sekvence pa do prve objeve PacBio ljudskog genoma) kao i tri primjerna projekta koji su koristili sekvencioniranje velikih razmjera: Projekt 1000 genoma (1000 Genomes Project) koji je skupio stotine ljudskih genoma do 2012; Atlas genoma raka (The Cancer Genome Atlas - TCGA) koji je skupio nekoliko tisuća tumor/normalnih genomskih parova; i Konzorcij za skupljanje eksoma (Exome Aggregation Consortium - ExAC) koji je skupiopreko 60 000 ljudskih eksoma. Većina, do danas, sekvenciranih genoma su kompletni eksomi, a ne kompletni genomi, ali očekuje se da će se omjer u budućnosti prevagnuti na stranu kompletnih genoma. Vrijednosti poslije 2015. predstavljaju naše tri moguće projekcije rasta, kako je i opisano u glavnom tekstu. Preuzeto iz [14].

Genomika u bliskoj budućnosti

Kao zaključak ovoga poglavlja, vrijedno je opisati neke hipotetske, ali tehnički izvodljive perspektive i opasnosti sa kojima se genomska industrija i društvo

općenito mogu suočiti:

- Smanjenje cijene genomske analize i minijaturizacija uređaja za genomsko sekvenciranje na red veličine mobilnih telefona.
- Eksplozivan rast genomskih podataka i njihova pohrana, pojava “genomskih hakera“ i problemi vezani za privatnost (zaštita podataka)
- Promjene u prehrambenoj industriji uključujući implementacije personaliziranih planova prehrane temeljenih na genomima
- Razvoj personaliziranih terapija lijekovima
- Upoznavanje zasnovano na genetskoj kompatibilnosti
- Osobna identifikacija pomoću genomskih informacija, uključujući mogućnost obavljanja uplata i plaćanja te korištenje raznih usluga
- Produljenje prosječnog životnog vijeka u svim državama, produljenje aktivne dugovječnosti i jaka tendencija prema staračkoj populaciji, kasne trudnoće i sve manji broj poroda
- Razvoj tehnologija za uređivanje genoma
- Izgled i oblikovanje nerođene djece i ostali trendovi koje je teško i zamisliti. Na primjer, tehnički izvedive mogućnosti odabira zdravog djeteta i određivanje njegovog budućeg izgleda bilo u fazi embrija bilo iz izoliranja fetalne DNA iz krvi trudnica već postoje [10].

1.4. Etički problemi u osobnoj genomici

U ovom odjeljku će se razmotriti etički problemi vezani za razvoj i sveprisutnu pojavu genomskih tehnologija kao što su privatnost, javne baze podataka, slobodan pristup podacima za znanstvenike, moguće zlouporabe i rizike vezene za osobne slobode.

U trenutnom idejnom projektu (whitepaperu) dajemo samo kratak pregled glavnih problema i perspektiva genomske industrije. Za detaljnije rasrave o privatnosti i sigurnosti podataka pogledajte.

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

Privatnost

Osobne genomske informacije su vrlo osjetljivo pitanje za mnogo ljudi. Međutim, većina ljudi ne razumije u potpunosti da se, na osnovu njihovih genomskih informacija, može odrediti njihov životni vijek, sklonost donošenju emocionalnih odluka (manipuliranje donošenjem odluka), vjerojatnost razvoja različitih mentalnih bolesti i rizik od iznenadne smrti zbog, recimo, srčanog infarkta.

Takve informacije mogu biti nepogodne za prijave na posao, sudjelovanje na izborima i određivanje cijene zdravstvenog osiguranja. Postoji, također, mogućnost podmetanja uzgojenog genomskog materijala nevinog pojedinca na mjesto zločina ili terorističkog napada. Osobi se na osnovu tih informacija može i uskratiti liječenje (ili naplatiti viša naknada) ili zabraniti obavljanje željenog posla.

Korporacije i vlade mogu namjerno utjecati na odluke i kupovine pojedinca koristeći znanje o "slabostima" u pojedinačevim genomskim informacijama. Prema tome, zaštita genomskih podataka je nužna da bi se zaštitila jednaka prava različitim kategorijama ljudi.

U isto vrijeme su provedene neke studije koje su omogućile otkrivanje identiteta pojedinaca na osnovu njihovog anonimnog genoma [5, 15].

Čak što više, neke tvrtke (<http://www.humanlongevity.com/media>) posjeduju algoritme temeljene na učenju uređaja koji mogu točno rekonstruirati izgled pojedinca samo na osnovu njegovih ili njenih genetskih podataka [16].

Publicitet

Takav pristup će omogućiti razvoj preventivne medicine koja će analizom velike količine podataka predvidjeti razvoj bolesti (prije nego se razviju) omogućujući na taj način poduzimanje mjera koje će produljiti životni vijek i povećati kvalitetu života [17] kao i omogućiti svjetsku identifikaciju donora za različite medicinske potrebe. Na žalost, trenutno još uvijek ne postoji kvalitetno rješenje koje bi omogućilo javnu uporabu genomskih podataka, a da ne naruši privatnost vlasnika podataka. Međutim, vrijedno je spomenuti neke startupe, bazirane na tehnologiji blockchaina, koji rade na pronalaženju rješenja za taj problem.

EncrypGen, startup koji je nedavno proveo ICO, je opisao probleme i veze između

privatnosti i javne uporabe blockchaina [18]. Međutim, u idejnom projektu, koji je objavljen, nedostaje opis tehničke implementacije rješenja koje bi istovremeno riješilo probleme privatnosti i dostupnosti.

Još jedan projekt na ovom polju je DNAbits startup[19] čiji je osnivač, Dror Samuel Brama, patentirao okvirni pristup pohrani podataka i njihovom prometu koristeći blockchain tehnologiju [20]. Međutim, u zadnje tri godine, ova tvrtka još uvijek nije tehnološki uspjela implementirati ovu tehnologiju.

Pravo na vlastite genomske podatke

Trenutno ne postoji pravno regulirana definicija prava na posjedovanje vlastitog genetičkog materijala. U nekim razvijenim državama, uključujući USA, Njemačku i Austriju, građani nemaju pravo pristupa i posjedovanja njihovih genetičkih podataka u kontekstu tumačenja i obrade istih[21]. Potrebno je angažirati posrednika, liječnika ili medicinsku ustanovu, sa pravom pružanja takvih informacija. Ovu činjenicu koriste tvrtke Pathway Genomics u USA i CeGaT u Njemačkoj (<http://www.cegat.de/en/>).

Da bi se podvrgli genetskoj analizi, potreban je savjet liječnika koji ujedno može biti davatelj usluge genetskog ispitivanja. Ujedno, samo taj liječnik ima pravo tumačenja podataka dobivenih tom genetskom analizom.

U USA postoje davatelji usluga na polju “genetike za razonodu“, kao što su 23andMe i Ancestry.com, koji prodaju genetske testove izravno potrošaču, ali su ovlaštene da daju samo informacije o etničkom podrijetlu i određenim zdravstveno vezanim osobinama (na primjer, sportske osobine), ali ne i medicinski najrelevantnije informacije. Ova ograničenja, nametnuta od strane regulatorskih tijela kao što je FDA, ne sprečavaju tvrtke kao što je 23andMe da prodaju pristup svojim pohranjenim genetskim podacima velikim farmaceutskim tvrtkama. Poznati su neki takvi primjeri: posao sklopljen sa Genentech (odjel farmaceutskog giganta Rochea) vrijedan 60 miliona USD [22] za ispitivanje Parkinsonove bolesti. Posao je sklopljen i sa drugom velikom farmaceutskom tvrtkom Pfizer za studiju Chronove bolesti [23]. Neki izvještaji također tvrde da je tvrtka 23andMe vodila pregovore i sa Novartisom za studiju Alzheimerove bolesti[24].

Prema tome, možemo vidjeti da se trenutno pogoduje velikim tvrtkama kojima

dajemo za pravo da raspolažu našim genetskim podacima, da ih pohranjuju i zarađuju na njima. Korporacije, iza vela dobrih namjera, monopoliziraju genomske podatke. Još ne znamo kako će ta monopolizacija utjecati na budućnost cijena lijekova i medicinskih otkrića.

Pravo pristupa genomskim informacijama

Potrebno je raspraviti još jedno etičko pitanje koje se nameće. Ranije smo već naveli da se kršenje privatnosti i pristup genomskim informacijama može iskoristiti u ilegalne svrhe, recimo od strane poslodavca. Može se dogoditi da poslodavac na osnovu genomskih podataka otpusti ili uskrati napredovanje na poslu. Za sve koji obavljaju poslove koji imaju veze sa sigurnošću ljudi i sustava, kao što su vozači kamiona i autobusa, piloti, uposlenici nuklearnih elektrana, ili ljudima sa sličnim zanimanjima zdravstveno stanje je od ključne važnosti i genomske informacije bi mogle spriječiti nesreće ili čak i katastrofe. U nekim profesijama, potencijalne opasnosti bi mogle ugroziti radnika prije nego promatrača kao što su rudari, u rudnicima uglja i li dijamantina, sa plućnim problemima. Za takve slučajeve, nužno je provesti raspravu, u kojoj će sudjelovati profesionalci i stručnjaci kao i opća populacija, a koja će rezultirati donošenjem pravnih standarda za reguliranje upotrebe genomskih podataka zaposlenika od strane poslodavca.

Poglavlje 2

Koncept

2.1. Project Zenome

Filozofski pogled

Javna osviještenost o genomskoj medicini je još uvijek jako niska u razvijenim zemljama i još lošija u zemljama u razvoju. To znači da ljudi, općenito, ne razumiju moguće dobrobiti genomike kao ni opasnosti koje su vezane za nju. U mnogim zemljama je to dovelo do prekomplikiranih institucionalnih procedura koje služe za zaštitu genetskih informacija od mogućih zlouporaba. U jednu ruku te procedure štite podatke i privatnost, a u drugu usporavaju znanstveni napredak.

Platforma Zenome će nastojati podignuti osviještenost javnosti o genomskoj medicini da bi korisnici mogli donijeti svjesne odluke vezane za njihove podatke. Da bi stvarno uspio u tom naumu Projekt Zenome se zasniva na sljedećim osnovnim principima:

Individualno vlasništvo nad osobnim genomskim podacima Svaki sudionik ima pravo na svoje osobne podatke

Sloboda izbora Svaki sudionik sam odlučuje kako će se njegovi osobni podatci koristiti. Svatko će moći sam odlučiti želi li sudjelovati u znanstvenim / kliničkim istraživanjima.

Pravo dijeljenja Učesnik može dopustiti pristup osobnim genetičkim informacijama trećoj strani na način da spriječi mogućnost kopiranja podataka

Privatnost Enkripcija osobnih podataka onemogućava pristup individualnim podacima bez izričitog odobrenja korisnika.

Distribuirana pohrana podataka Arhitektura distribuiranih baza podataka pruža visoku razinu dostupnosti i tolerancije greške kroz umnožavanje i mogućnost proširivosti.

Distribuirana obrada podataka Podatci se obrađuju u mnogo mrežnih čvorova u isto vrijeme. Bilo koji korisnik, koji mreži osigura prostor na tvrdom disku i CPU snagu, može postati čvor.

Proširivost Arhitektura platforme omogućava odličnu proširivost i fleksibilnost sustava

Ekosustav platforme Zenome

U Zenme platformi korisnik⁵ je uključen u različite vrste interakcija kroz sustav. Te se interakcije odvijaju na različitim razinama sustava, ne miješaju se međusobno i koriste različite obrazce interakcije. S obzirom na to, trebaju biti predstavljene kao različita tijela koja imaju različite uloge.

Sljedeće uloge su moguće u Zenome platformi:

(Računanje/Pohrana) Čvor koji pruža kapacitet pohrane i procesorsko (CPU) vrijeme u zamjenu za nagradu.

Osoba predavanje individualnih genetičkih informacija platformi i mogućnost korištenja genetskih usluga.

Analitičar je zainteresiran za analiziranje genetskih informacija na platformi. To može biti znanstvenik za obradu podataka, znanstvena organizacija itd.

Davatelj usluge koji implementira genetsku uslugu (naplativu po želji) na platformi. U osnovi, ovo je organizacija koja koristi genetske podatke kao

⁵ Korisnik u širem smislu, tj. osoba ili software koji radi u ime osobe

dio posla.

UKRATKO

Svaki korisnik je uključen u više interakcija različitih vrsta, preuzimajući na sebe različite uloge. Neke od njih, kao recimo Pružatelj usluga i Analitičar, zahtijevaju posebno znanje, za razliku od Čvora i Osobe koji ne moraju imati neko posebno znanje.

Svaka od ovih uloga će biti detaljno opisana u nastavku.

Pregled arhitekture sustava

Zenome platforma je distribuirana aplikacija koja se sastoji od tri glavna nivoa.

Mrežni nivo (i nivo pristupa podacima): osigurava nivo sažimanja koja zatvara interakciju mreže i osigurava sučelje distribuiranog okruženja gornjim nivoima platforme.

	Blockchain	DHT Kademlia
Cijena pohrane podataka	Visoka	Niska
Nepromijenjivost podataka	DA	NE
Performanse	Niske	Visoke
Deterministički rezultat	DA	NE

Ovaj nivo se sastoji od dva distribuirana sustava potpuno različite prirode:

Distribuirana lista (bilanca) koja bilježi transakcije između sudionika na dokaziv i stalan/nepromjenjiv način. Za pristup blockchainu, čvorov software koristi ugrađenog Ethereumovog klijenta.

Distribuirana Hash tablica mreže (zasnovana na Kademlia protokolu) raspoređuje fizičke čvorove u isprepletenu mrežu koja omogućava prijenos poruka između čvorova i distribuiranu pohranu podataka.

ULOGA

Čvor za računanje/pohranu radi na Mrežnom nivou

Middleware upravlja računima, posjeduje sučelje za rješavanje sigurnosnih značajki temeljnog nivoa i API visoke razine za software koji se izvodi na

Aplikacijskom nivou.

ULOGA

Pružatelj usluge radi na Middlewareu. Takozvana External Services Platform API omogućava trećim licima da započnu genetske usluge na platformi

Aplikacijski nivo Zemone aplikacija posjeduje napredno sučelje za krajnjeg korisnika koji prosljeđuje i prevodi korisnikove radnje Middlewareu na konzistentan način. Sučelje je po dizajnirano da bude proširivo na način koji će omogućiti gentičkim uslugama da slobodno koriste svoj software.

Tržište genetskih usluga

Tržište gnetских usluga se trenutno razvija u sljedećim područjima:

1. Istraživanje i prihvaćanje tehnologije na tržištu.
2. Pružanje genomskih dijagnostičkih usluga.
3. Državno certificiranje genetskih tehnologija.
4. Razvoj pravnih okvira, konkretno legislativnih mjera vezanih za sigurnost genetskih informacija.

BILJEŠKA

Struktura tržišta je poprilično složena. Pojedini sudionici se, u biti, razvijaju u više smjerova u namjeri da pronađu svoje mjesto na tržištu brzo rastućih potreba potrošača.

Sljedeći sudionici su zastupljeni na tržištu genetskih usluga:

Znanstvene korporacije rade na razvoju i prihvaćanju novih tehnologija na tržištu. To uključuje:

- Farmaceutske korporacije, biotehnoške i dijagnostičke tvrtke, kao što su Pfizer i Myraid
- Tvrtke koje razvijaju i prodaju potrebne kemijske nadomjesne reagense (kao što je Life Technologies).

IT-bioinformatičke tvrtke su uključene u pronalaženje i razvoj metoda računalne obrade podataka. Sudionici ovoga sektora se još uvijek bore da

izađu na kraj sa tipovima i količinom prikupljenih podataka.

Znanstveni i medicinski centri igraju vodeću ulogu u pružanju i razvoju genetičkih dijagnostičkih usluga.

Tržišni laboratoriji pružaju brze, učinkovite i obično relativno jeftine dijagnostičke usluge. Posjeduju ogromne financijske i resursne rezerve.

Genetičko-dijagnostičke tvrtke usmjerene direktno na potrošača povećavaju zainteresiranost javnosti za genetičku dijagnostiku. Trenutno je ovaj segment još uvijek mali, ali u budućnosti može prerasti u jedno od vodećih dijelova tržišta te je moguće da bude prihvaćen i u kliničkoj praksi.

Tabela 5: Usporedba sa sličnim proizvodima na tržištu

	Mi (Zenome)	GeneCoin	EncrypGen	23andMe	Pathway Genomics	Snpedias (rethase)	Human longevity
Decentraliziran	+	-	-	-	-	-	-
Pogodan za neljudske organizme	+	-	-	-	-	-	-
Korisnik je vlasnik svojih podataka	+	-	-	-	-	+	-
Mogućnost postavljanja svojih podataka	+	-	+	-	-	+	-
Slobodni neprivatizirani podatci	+	-	-	-	-	-	+
Obavlja vlastitu analizu podataka	+	+	-	+	+	+	+
Osigurava izvješće korisniku	+	-	-	+	+	+	-
Koristi AI i učenje uređaja	+	-	-	-	-	-	+
Dijeljenje bez prijenosa velikih količina podataka	+	-	+	-	-	-	-
Mogućnost zarade na svojim podacima	+	-	-	-	-	-	-
Otvoren za znanstvenike	+	-	-	-	-	+	-
Platforma i za ostale alate	+	-	+	-	-	-	-

2.2. Uloge u Zenome platformi

Iz perspektive resursnog čvora

Sustav iz perspektive računalnog čvora

Node (Čvor) – je sudionik sustava koji pruža resurse svoga računala (prostor za

pohranu podataka i procesorsko vrijeme) u svrhu distribuirane pohrane i obrade genetskih podataka u zamjenu za nagradu u ZNA tokenima.

Da bi se postalo računalnim čvorom, korisnik mora izvoditi Zenome software na svome računalu i aktivirati ulogu Čvora u grafičkom korisničkom sučelju. Software se mora besprekidno izvoditi u pozadini.

BILJEŠKA | Dostupna je i verzija Zenome softwarea sa komandnim promptom što zna biti jako korisno, naročito kod aktiviranja računalnog Čvora

Način dodjele i iznos resursa koji se dodjeljuju, kao i upravljanje zadacima je vrlo fleksibilno i može biti konfigurirano od strane vlasnika čvora:

- najviši kapacitet pohrane koji je softwareu dozvoljen da koristi
- način uporabe računalnih resursa:

Nepromjenjiva (fiksna) uporaba CPU/GPU resursa, broj jezgri i maksimalno učitavanje (u postotcima) po svakoj jezgri

Promjenjiva (dinamička) uporaba CPU/GPU je kontrolirana Zenome softwareom na način da ne remeti rad ostalih korisnikovih aplikacija.

- Odabir računalnog procesa (po ID) koji će imati najviši prioritet izvođenja
- Privremeno gašenje čvora: poslati zahtjev mreži da se podatci, koji nisu dostupni nigdje drugo, prenesu ostalim čvorovima. Čekati dok se ne završi prijenos podataka te se onda odspojiti.
- Gašenje čvora u potpunosti: poslati zahtjev mreži da izvrši prijenos svih podataka ostalim čvorovima. Čekati da se prijenos završi, a zatim pobrisati podatke.

Osoba/Korisnik

Osoba – je individua voljna podijeliti svoje genetske informacije Zenome sustavu u namjeri da zaradi na prodaji tih informacija ili u namjeri da koristi neki od dostupnih genetskih servisa platforme.

Korisnik instalira Zenome software za rad sa svojim osobnim genetskim informacijama.

Koristeći grafičko sučelje, korisnik može:

- Otvoriti osobni korisnički račun/profil
- Upravljati tokenima: primiti, prebacivati, koristiti, platiti prostor za pohranu, potrošiti na neki od genetskih servisa koji se naplaćuju
- Učitati genetske podatke (format dokumenta se automatski prepoznaje)
- Upravljati osobnim podacima: dostaviti osobne podatke, popunjavati upitnike, izlistati najpopularnije upitnike
- Raditi sa ciljanim ponudama koristeći zaštitni podsustav
- Koristiti genetske servise i postavljati nivo privatnosti za svaki servis ponaosob

Pri učitavanju genetskih podataka, nivo privatnosti se može izabrati između:

Full Privacy U ovom slučaju podatci su pohranjeni u kriptiranom obliku, a za pohranu takvih podataka se naplaćuje puna cijena.

Standard Privacy Genetski podatci su pohranjeni u fragmentima što onemogućuje identificiranje korisnika. Svaki fragment je otvoreno pohranjen u sustavu. Informacija koja se odnosi na povezivanje fragmenata sa korisnikovim ID je privatna. U ovom slučaju pohrana je jeftinija zbog subvencija.

Public Access Podatci su otvoreno pohranjeni. Pohrana je besplatna zbog subvencija.

BILJEŠKA

Pozor! Iako nema tehničkih ograničenja na povećanje nivoa privatnosti, ove promjene se mogu primijeniti samo na buduće podatke. Podatci koji su jednom bili proglašeni javnima ne mogu ponovno biti proglašeni privatnima. Dobro razmislite o ovome prije nego učitajte podatke po prvi put.

Kupac/Konzument podataka

Konzument podataka (Data Consumer) – je znanstvenik, poslovna tvrtka, znanstveni organizator ili neki drugi korisnik platforme koji je zainteresiran za analizu podataka koristeći kapacitete platforme. Konzument podataka može

korisnicima poslati zahtjev za pristup njihovim genetskim podacima i za uzvrat im platiti naknadu za korištenje.

BILJEŠKA | Postoje neka ograničenja vezana za obavijesti i upite, koja su stvorena da spriječe otkrivanje pravoga identiteta korisnika sustava i ostale zlouporabe. Ova ograničenja ovise i o kozumentovu ranku u sustavu.

Pružatelj usluga

Pružatelj usluga (Serice Provider) – organizacija koja koristi genetske podatke u poslovanju i na taj način implementira korisničke usluge na platformi.

BILJEŠKA | *Korisnik ne može sam birati podatke koje želi ili ne želi podijeliti sa pružateljima usluga. Korisnik će biti obaviješten u slučaju da, podatci koje je podijelio, mogu biti iskorišteni da ga se identificira.*

BILJEŠKA | Na primjer, pružatelji usluga ne mogu vidjeti više od 70% mutacijske liste u čistoj formi koristeći izravan upit ili dobiti informacije koji se poklapaju sa raw podacima (kao što su fastq datoteke) sa korisnikovim upitnikom.

2.3. Genomski podatci

Vrste genomskih podataka

Unutar okvira koncepta, postoje 3 vrste genomskih podataka koji se međusobno razlikuju po vrsti plaćanja i informacijskoj vrijednosti:

- Slobodni podatci koji nisu važni svojim vlasnicima, ali su važni znanstvenicima

Primjer: Genom *Helicobacter pylori* bakterijskog soja

- Slobodni podatci koji su vrijedni i njihovim vlasnicima i konzorciju.

Primjer: Genomi većine mrežnih sudionika.

- Ograničeni podatci koji su pohranjeni unutar mreže.

Za projekte sa ograničenim pristupom raznih komercijalnih i javnih institucija.

Predobrada genetskih podataka

Rukovanje NGS genomskim podacima (i većinom drugih genomskih skupova podataka) se obično sastoji od dva nezavisna koraka:

1. **Preliminarna obrada “raw“ podataka.**
2. **Detaljna analiza genetskih sekvenci** sa ciljem razvoja osobnih preporuka ili analiza unutar okvira istraživanja.

BILJEŠKA | **Referentni genom** – je digitalni DNA skup podataka složenim na način da predstavljaju reprezentativni primjer skupa gena određene vrste.

Predobrada NGS genomskih podataka uključuje sljedeće korake:

1. Poravnanje NGS očitavanja sa referentnim genomom
2. Traženje mutacija i ostalih razlika u odnosu na referentni genom i pohrana liste dobivenih očitavanja u gVCF formatu.

BILJEŠKA | *Protokol je isti i za obradu podataka neljudskih organizama. U tom slučaju, naravno, mora se koristiti odgovarajući referentni genom.*

Ukoliko skup osobnih podataka predstavlja rezultat mikropostrojne genotipne tehnologije (23andMe tip dokumenta) opet se može učitati na platformu kao gVCF dokument jer su formati podataka ovih dokumenata slični.

Problem lažnih podataka

Ukoliko se učitaju (slučajno ili namjerno) podatci ostalih (neljudskih) organizama umjesto ispravnog genoma, ovo će biti otkriveno prilikom predprocesiranja raw podataka i korisnik će biti obaviješten. Namjerno učitanje iskrivljenih (lažnih) genetskih podataka u sustav se može, prije pohranjivanja podataka u pohranu, otkriti koristeći mnoge dobro poznate metode za provjeru.

BILJEŠKA | S ekonomske strane će se korisnike navoditi da se suzdrže učitavanja lažnih podataka na način da će se usluga pohrane podataka plaćati unaprijed za cijelu godinu.

Problem otkrivanja identiteta na osnovu genomskih podataka

Slobodan pristup genetskim informacijama otvara problem identifikacije korisnika na osnovu njihovih genetskih podataka. Ukoliko korisnik odluči da ne želi potpuno ograničiti pristup svojim genetskim podacima moraju se, pri svakoj etapi obrade podataka, poduzeti odgovarajuće mjere zaštite. Da bi se riješio problem korisničke identifikacije, interakcija se mora uspostaviti na način, koji neće dozvoliti ni jednom čvoru, pri bilo kojoj etapi obrade, da utvrdi vlasništvo nad genetskim materijalom, pa čak ni grad u kojem vlasnik podataka živi.

BILJEŠKA | Razlike između stanovnika jednog grada čine približno 0.01% sekvence.

Pri svakoj fazi, gore navedeni cilj se ostvaruje na različite načine:

1. U predobardnoj fazi – dijeljenjem izvornog dokumenta na dijelove na način da prosječna pokrivenost bude niža od praga pouzdanosti (6 kopija)
2. U fazi pohrane – fragmentiranje računa za pohranu po duljini

Pohrana genomskih podataka

Genomski podatci su pohranjeni u distribuiranoj mreži zasnovanoj na DHT Kademlia protokolu. Učesnici koji pružaju resurse za rad mreže (pogledajte detalje vezane za Čvorove) su plaćeni ZNA tokenima. Da bi ostvarili pravo da budu plaćeni moraju dokazati mreži da su stvarno pohranili podatke. Procedura provjere je zasnovana na korištenju blockchaina. Ukoliko je potrebno koristi se i enkripcija.

Biti će provedena i integracija Storj i FileCoin čvorova.

Kako je već navedeno, podatci mogu biti raw i obrađeni.

Vrsta podatka	Raw	Obrađeni
Osobine		
Format	Fastq/bam	(lista mutacija) Gvcf/vcf+bed/ 23me(txt)
Veličina, Gb	50	2
Vrijednost	Unaprjeđenje tehnologije sekvenciranja i obrade (za tržište opreme)	Provođenje istraživanja i izrada izvješća
Uvjeti pohrane		
Broj kopija	Najmanje 3 u međusobno neovisnim čvorovima	Najmanje 5 u međusobno neovisnim čvorovima

Genomski podatci su podijeljeni u fragmente i pohranjeni su na način da duljina fragmenta ne dozvoljava točnu identifikaciju vlasnik podataka.

Informacije koje govore čijeg su genoma fragmenti sastavni dio su također privatne prirode i mogu se dobiti samo uz dozvolu vlasnika.

2.4. Osobni profili

Ispunjavanje osobnog upitnika značajno povećava mogućnost primjene genomskih podataka. Korisnik popunjava upitnik koristeći grafičko sučelje.

U slučaju da neki upitnik postane popularan, aplikacija će potaknuti korisnika da ga ispuni. Svaki analizator može sastaviti svoj upitnik i postaviti ga na platformu.

Specifikacija upitnika

Broj upitnika može biti ogroman. Iz toga razloga je nužno uvesti koncept specifikacije upitnika.

Specifikacija upitnika – je kompletan opis svih polja upitnika i dozvoljenih vrijednosti tih polja.

BILJEŠKA | Formalno gledajući, specifikacija upitnika sadrži poveznicu na autora, opis i uređeni skup naredbi, od kojih svaka odgovara jednom polju upitnika.

Postoji nekoliko tipova polja:

Numeričko polje (Numeric Field) Vrijednost polja je cijeli broj

Višeznačni izbor (Multiple choice) Vrijednost polja je broj ili odgovor (tekst).

BILJEŠKA | Odgovorima na ovu vrstu pitanja će biti dozvoljen slobodan pristup (bez poveznica na korisnika koji je dao odgovore na njih) jer ne predstavljaju nikakvu opasnost za otkrivanje identiteta korisnika.

Polje poveznice (String Field) Vrijednost koje polje prima je poveznica. Ovo je privatno polje jer može potencijalno ugroziti privatnost korisnika na osnovu sadržaja koje je u njemu.

Privatni blok (Private Block) Dozvoljava proglašenje bilo kojega polja privatnim bez obzira o kojoj se vrsti polja radi.

Upiti/ispitivanje sustava

Statistički podatci koji se odnose na određene genomske statuse će biti slobodni (otvoreni) za javnost ako ih korisnik nije odlučio kriptirati. Informacije koje se odnose na odgovore upitnika će također biti slobodne prirode. Dakle, svatko će, na primjer, znati broj korisnika mreže koji su stari 25 godina i imaju mutaciju rs6025 (koagulacijski faktor V).

Arhitektura sustava će učiniti nemogućim preuzimanje cijele baze podataka:

- Tijekom stvaranja povezanih upita korisniku će biti onemogućen pristup raw podacima
- Osnovna naknada će uključivati samo određeni dio zahtjeva dnevno. Naknada za dodatne upite će rast eksponencijalno kroz dan.
- Ukoliko rezultat povezanih upita bude sadržavao manje od 100 korisničkih odgovora, rezultat neće biti dostavljen kupcu.

U slučaju da korisnik potpuno kriptira svoje podatke sam odlučuje kome će, dozvoliti prijenos podataka, uključujući i fragmente od kojih se njegov/njen genom sastoji. Ni jedan analitičar neće biti u stanju saznati koje su vrste podataka kriptirane.

Siguran prijenos osobnih podataka

Proces prijenosa osobnih podataka između sudionika sustava bi trebao posjedovati sljedeće osobine:

1. Mogućnost prijenosa kompletnog skupa podataka treba biti dozvoljena samo kupcu i prodavaču.
2. Prijenos tokena bi se trebao izvršiti samo u slučaju uspješnog prijenosa podataka.
3. Pokušaj prodaje netočnih podataka bi se trebao moći identificirati i blokirati.
4. Pokušaj namjernog, lažnog, optuživanja prodavača za prodaju lažnih podataka treba biti objelodanjen.
5. Prijenos podataka se ne bi smio povjeriti trećem licu.

BILJEŠKA | Za siguran prijenos podataka će se koristiti blockchain tehnologija. Međutim mora se uzeti u obzir da pohrana i prijenos velikih količina informacija na blockchain zahtjeva veliku potrošnju resursa. Iz tog razloga:

6. Je dozvoljen prijenos malih količina podataka kroz blockchain. Ostatak podataka se može prenijeti kroz jednostavni kriptirani komunikacijski kanal.

2.5. Sustav vrednovanja (reputacije)

Platforma će omogućiti vrednovanje:

- Odvojeno za svaki genetski fragment ili blok osobnih podataka
- Organizacije
- Pružatelje usluga
- Pružatelje podataka (npr. laboratoriji za sekvenciranje DNA)

BILJEŠKA

Neće biti vrednovanja (ocjenjivanja) individualnih korisnika jer bi to predstavljalo zbroj njegovih/njenih ocjena fragmenata, što nije dostupno javnosti. U slučaju da se podatci učitaju koristeći račun/profil Organizacije tada će početa reputacija biti automatski uvećana za reputaciju organizacije.

Faktori koji utječu na ocjenu genetskih fragmenta

Potvrda laboratorija povećava reputaciju učitanih podataka proporcionalno ocjeni laboratorija u kojem su dobiveni. Potvrdu može predstavljati digitalni potpis laboratorija ili podatci učitani od strane laboratorija na zahtjev klijenta.

Provjera vjerodostojnosti Dozvoljava provjeru korisničkih genskih informacija koristeći predefimirani statistički model polimorfne frekvencije i genetskog povezivanja. Ovaj modul je trenutno u fazi razvoja.

Sudjelovanje u istraživanju Reputacija se povećava sa brojem uspješnih istraživanja koja su uključivala uporabu fragmenta. Ukoliko se rezultat ispitivanja informacija učini nevjerojatnim, reputacija opada.

2.6. Oblici/načini uporabe

Individualni korisnik

Za individualne korisnike postoji mogućnost da svoje genomske informacije pretvore u izvor zarade. Kombinacija genoma i njegova interakcija sa okolišem je vrijedan informacijski resurs. Naša platforma će omogućiti korisniku sigurno upravljanje ovim resursom.

Platforma pruža mogućnost sigurnog pohranjivanja i dijeljenja informacija dopuštajući korisniku pristup raznim genetskim uslugama. Evo par primjera:

- Izvješća i preporuke prehrane, rizik od bolesti, kozmetika, prehrana, fitness (vježbanje)
- Potraga za rodbinom i utvrđivanje podrijetla
- Servisi za upoznavanje
- Individualni odabir odjeće, obuće, postavka klimatizacije u kući, turistička odredišta i područje prebivanja
- Različite varijacije genetskih izvještaja za grupe ili individualce, za npr. sportski tim ili radnu grupu
- Skoro svaki aspekt ljudskog života je pod utjecajem genetike pa željno iščekujemo da vidimo koje će tvrtke izrasti koristeći našu platformu

Osim mnoštva usluga, korisniku je pružena mogućnost i da zaradi na svojoj genetskoj jedinstvenosti pružajući odgovore na upitnike tvrtkama u svrhu različitih istraživanja. S obzirom na to, možemo povući neku analogiju između osobni podataka, u kombinaciji sa genetskim informacijama, i robe ili mineralnih resursa.

Zdravlje

Suvremena zdravstvena njega i personalizirana medicina se ne mogu zamisliti bez genomske tehnologije. Platforma će omogućiti pacijentima da sigurno dijele svoje genetske informacije koje se koriste u klinikama s medicinskim osobljem:

- Individualno doziranje i netolerancija na lijekove (npr. individualno doziranje antikoagulanta varfarin na osnovi genetskih karakteristika)
- Osobni prihvatljivi rasponi biokemijskih parametara (npr. PSA markeri)
- Genetske sklonosti različitim bolestima (npr. visok rizik od makularne degeneracije i potreba za dodatnim istraživanjem i prevencijom)
- Transplantacija i doniranje organa. Korisnici mogu sigurno dijeliti informacije koje se tiču njihovih HLA antitijela koja određuju kompatibilnost između individua pri transplantaciji. S obzirom na sve navedeno, biti će moguće stvoriti sigurnu bazu podataka donora i dobrovoljaca koji će spašavati živote tijekom transplantacije.

Tvrtka

Postoje dvije vrste Tvrtki. Prva pruža korisnicima usluge zasnovane na njihovim genomskim podacima dok je druga zainteresirana za dobivanje podataka od korisnika za provođenje vlastitih istraživanja.

Prva vrsta je opisana u poglavlju o oblicima/načinima uporabe. Druga vrsta se može opisati kao kupac genomskih podataka za provođenje vlastitih ispitivanja, poboljšavanje potrošačkih osobina proizvoda, genetski ciljano razvijanje proizvoda i reklamiranje. Neki od primjera su:

- Na primjer, farmaceutska tvrtka namjerava izbaciti lijek koji djeluje protiv mutirane kancerogene bjelančevine. Tvrtka može među korisnicima sustava pronaći korisnike, koji su preživjeli bolest, platiti im za njihove podatke, proučiti podatke i naći frekvenciju mutacija unutar gena koji kodira bjelančevinu.
- Tvrtka planira ući na novo tržište i za to joj je potrebno provesti ispitivanje kako korisnici percipiraju okus proizvoda. Poznato je neki okusi mogu izazvati odbojnost kod nosilaca određene genske varijacije ukusnog perceptivnog gena. Tvrtka, preko mreže, pošalje ponudu u namjeri da prouči nosioce ove genetske varijacije i napravi novi okus ili pronade frekvenciju ove genetske varijacije za različita tržišta i s obzirom na rezultate cilja proizvodom druga tržišta.

Znanstvena zajednica

Za znanstvenu zajednicu sustav otvara mogućnosti pohranjivanja, dijeljenja i provođenja istraživanja sa različitim genomskih podacima. Platforma nije ograničena samo na rad s ljudskim genomima pa se može koristiti i za sigurnu pohranu i obradu genomskih podataka relevantnih, recimo, za poljoprivredu (biljke, životinje, mikroorganizme).

Općenito gledano, postojanje ovakvog ekosustava vodi, kroz pristup podacima opće populacije, do obogaćivanja znanstvene zajednice čak i bez upućivanja na individualne upitnike. Uz to, korisnik svojim pristankom može postati dio znanstvenih istraživanja.

Platforma također pruža distribuiranu računalnu snagu, koja će omogućiti obradu velikih količina genetskih podataka (slično kao ASW koji je prilagođen da

radi sa genetskim podatcima).

Poglavlje 3

Tehnički dio

3.1. Distribuirani objekti

Koncept distribuiranog podsustava

Distribuirani podsustav (Distributed Subsystem) – skup nekoliko elemenata sustava (platforme) i proces koji se može prikazati iz objektno orijentirane perspektive kao sustav koji ima jasno definiran identitet i ispoljava jasno definirano, izvana vidljivo, ponašanje.

Da bi detaljno predstavili osobine podsustava, trebaju se opisati sljedeći aspekti:

1. **Struktura:** elementi i procesi koji tvore sustav.
2. **Vanjsko ponašanje:** odnos podsustava kao cjeline sa ostalim sudionicima. Naročito:
 - **Sučelja:** skupovi mogućih zahtjeva prema podsustavu kao cjelini.
 - **Djelovanje:** djelovanje podsustava se tiče i ostalih sudionika sustava.
3. **Unutarnje uređenje:** skup unutarnjih stanja podsustava.

Podsustav se može predstaviti kao kvaziobjekt sa kojim ostali sudionici mogu uzajamno djelovati.

Prefiks “kvazi“ znači da se odvija stvarna interakcija sa elementima podsustava koji, za uzvrat, imaju složenu međusobnu interakciju kao dio stvarne interakcije pa se sve to kombinirano može smatrati interakcijom sa nekim skupnim objektom.

Ispod ćemo navesti razlike između podsustava i njegov prikaz u obliku kvaziobjekta.

Odnos sa podsustavom se može predstaviti kao niz relativno malog broja osnovnih informacija:

- Sučelje podsustava, koje sa gledišta podsustava, predstavlja djelovanje ostalih sudionika podsustava.
- Djelovanja ostalih sudionika platforme.
- Unutarnji procesi koji mijenjaju unutarnje stanje sustava.

Unutarnji proces

Unutarnji procesi (u širem smislu) – predstavljaju skup unutarnjih procesa svih elemenata podsustava i interakcija među njima.

Unutarnji procesi (u užem smislu) – predstavljaju procese unutar podsustava kao cjeline koji mijenjaju njegovo unutarnje stanje. Puni opis unutarnjih procesa se sastoji od ponašanja svih podsustava isključujući pitanja vezana za njegovu specifičnu implementaciju.

3.2. Glavni distribuirani podsustavi platforme

Platforma se sastoji od sljedećih organizacijskih nivoa:

- Osnovnog nivoa sustava (kritična infrastruktura)
- Nivo pohrane i obrade podataka
- Interakcije visoke razine

Sljedeći podsustavi su sadržani unutar platforme:

1. Osnovni nivo sustava

Interoperabilnost niske razine Osnovni podsustav razmjene poruka između mrežnih čvorova. Također omogućava izradu distribuiranih hash tablica.

Autorizacija Infrastruktura za upravljanje profilima i pristupom privatnim informacijama.

2. Nivo pohrane i obrade podataka

Pohrana Nivo apstrakcije za pristup distribuiranom podatkovnom sustavu.

Obrada Infrastruktura za distribuiranu obradu podataka.

3. Interakcije visoke razine

Sigurnosna ispitivanja Podsustav koji omogućava formiranje ponuda za kupovinu genomskih podataka čiji je prikaz ograničen samo za određene korisnike.

Rukovanje slobodnim podacima Osigurava alate za rukovanje slobodnim (ne privatnim) podacima

Vanjske usluge Platforma pruža API za vanjsko povezivanje centralno upravljivih sustava platforme. Osigurava siguran prijenos podataka koristeći konvencionalne internet protokole.

3.3. Interpretacijski nivo niske razine

Distribuirana P2P mreža

Osnova platforme je P2P mreža zasnovana na Kademlia protokolu. Mreža uključuje računala korisnika koji imaju instaliran Zenome software.

Čvor – je čvor distribuirane P2P mreže koji predstavlja korisničko računalo sa instaliranim Zenome softwareom na njemu.

S obzirom na navedeno, napravljena je preklapajuća mreža između uređaja koji sudjeluju u mreži. Preklapajuća mreža predstavlja virtualnu mrežu u kojoj je svakom sudioniku dodijeljen “NodeId” (ID Čvora), a koji nema veze sa stvarnom IP adresom uređaja. Svaki čvor pohranjuje spisak (listu) “bliskih” čvorova, gdje se udaljenost među čvorova računa na osnovu čvorovih IDa i nije povezana sa relativnom udaljenošću među čvorovima.

Čvorovi pohranjuju podatke koristeći distribuirane hash tablice.

Značajke primjene distribuirane mreže

Koristi se protokol izmijenjenih tehničkih karakteristika. Glavne razlike su:

- Čvorovi mogu međusobno izmjenjivati proizvoljne poruke. Jedan čvor ima mogućnost poslati poruku drugom samo na osnovu poznavanja ID (NodeId) tog čvora.
- Unutar mreže može postojati više hash tablica. Hash tablica se identificira na osnovu ključa-poveznice.
- Za različite tablice se mogu postaviti različita pravila pohrane i brisanja rezultata.
- Podatci koji se prenose između čvorova su šifrirani (donji tekst).

Razmjena poruka u distribuiranom okruženju

Učesnici peer-to-peer (P2P) mreža mogu razmjenjivati slijedeće poruke:

PING	Potvrdi da su svi čvorovi aktivni
STORE(T,K,V)	Pohrani vrijednost “V” pod ključem “K” u tablici “T” u čvoru koji prima poruku
FIND_NODE(N)	Čvor koji prima poruku će, između čvorova koje zna, poslati podatke koji se odnose na čvorove koji su najbliži čvoru “N”.
FIND_VALUE(T,K)	Ako je par “(K,V)” pohranjen u čvoru primatelju, pošalji vrijednost “V”, a u suprotnom pošalji podatke koji se odnose na poznate čvorove koji su “bliže” dokumentu.
SEND(M,N,D?)	Šalje poruku “M” koja sadrži podatak “D” čvoru “N”. “FIND_NODE” se koristi za lociranje čvorova.

Ovaj nivo interakcije je transparentni nivo platforme.

3.4. Blockchain

Osnovne informacije

Platforma koristi Ethereum blockchain koji u osnovi predstavlja zasebni decentralizirani stroj (EVM – Ethereum Virtual Machine). Željena logika sustava se može implementirati koristeći pametne ugovore (smart contracts).

A contract is a collection of code (its functions) and data (its state) that resides at a specific address on the Ethereum blockchain.

- Introduction to Smart Contracts (Solidity manual)

Ugovor je skup linija koda (njegovih funkcija) i podataka (njegovih stanja) koji se nalaze na određenoj adresi na Ethereum blockchainu.

- Uvod u pametne kontakte (Solidity priručnik)

S obzirom na navedeno, pametni ugovori mogu pohranjivati podatke. Za podatke u blockchainu vrijedi, ranije navedeni, sažeti tip sustava (međutim sa nekim ograničenjima).

Podsustav za rad sa blockchainom

Pozor: implementacijski detalji se mogu razlikovati ovisno o vrsti platforme koja se koristi. Donji opis se odnosi na PC platformu.

Da bi osigurao pristup Blockchainu, software čvora sadrži punu implementaciju Ethereum Nodea osiguravajući spomenuti pristup uz pomoć “JSON-RPC 2.0“.

Tajni ključ je pohranjen u šifriranoj pohrani. Pri prvoj uporabi aplikacije korisnik mora odabrati i postaviti zaporku.

BILJEŠKA | Iako je tehnički moguće koristiti postojeći račun, preporučljivo je napraviti novi.

Sučelje aplikacije dozvoljava:

1. Otvaranje novog računa
2. Učitavanje postojećeg računa
3. Postavljanje sigurnosne kopije privatne pohrane (podataka)

BILJEŠKA | Pozor: Preporučljivo je imati cloud sigurnosnu kopiju s obzirom da ćete na taj način zadržati podatke za pristup računu čak i u slučaju da vam je računalo fizički nepostupno.

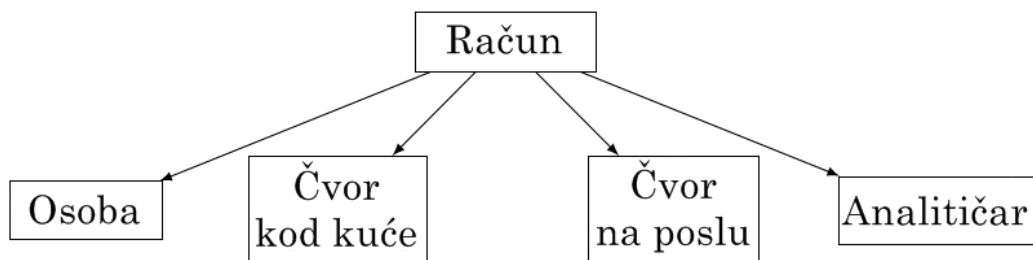
BILJEŠKA | *Software dozvoljava pristup Ethereumovu naredbenom retku. Ova funkcija je u prvom redu namijenjena za otkrivanje i otklanjanje pogrešaka. Nije preporučljivo koristiti naredbeni redak ukoliko ne razumijete njegovu svrhu.*

Tokeni sustava

Ekonomska interakcija unutar sustava se odvija korištenjem ZNA tokena. Oni predstavljaju validne Ethereum tokene i mogu se kupovati i prodavati na burzama/mjenjačnicama.

Koncept računa

Račun u platformi dozvoljava korisniku da komunicira sa sustavom istovremeno igrajući više uloga. Svaka uloga u računu odgovara različitom pametnom ugovoru u blockchainu.



Slika 4: Račun korisnika koji je podijelio svoj genom (uloga Osobe) na platformi, koji radi kao bioinformatičar i kao potpora računalnim čvorovima smještenim kod kuće ili na poslu.

Uloga – je osobina sudionika sustava koja se odnosi na točno određen tip interakcije u kojoj sudjeluje. Svaki sudionik može igrati više uloga istovremeno.

BILJEŠKA

Za dodjelu uloga je zadužen poseban podsustav.

3.5. Mreža distribuirane pohrane podataka

Načela rada distribuirane mreže

Distribuirana mreža temeljena na DHT Kademlia protokolu se koristi za pohranu podataka osjetljive prirode. Čvorove ove mreže predstavljaju registrirani pružatelji resursa koji su nagrađeni za pohranjivanje i obradu podataka. Sudionik koji je učitao svoje podatke plaća naknadu za pohranu tih podataka, ali u brojnim, naročito važnim, slučajevima ove naknade su djelomično ili potpuno subvencionirane od strane sustava. Npr., postoji mehanizam subvencioniranja pohrane znanstveno vrijednih podataka.

Jedinica pohrane je predstavljena osjetljivim podatkovnim blokom i jedinstvenim ključem pomoću kojega mu se može pristupiti.

BILJEŠKA

Veličina podatkovnog bloka je ograničena iz razloga sprečavanja zlouporabe.

Podrška vanjske usluge. Biti će implementirana i mogućnost integriranja

Storj i FileCoin čvorova, kao i nekih centralno upravljivih servisa za pohranu podataka.

Osiguranje pouzdanosti pohrane podataka

Da bi se osigurala pouzdanost pohrane podataka, u vidu činjenice da se čvorovi mogu slobodno spajati na i otpajati sa mreže, podatci su pohranjeni neovisno na nekoliko čvorova istovremeno. Vrsta podataka određuje broj čvorova na kojima se istovremeno kopira.

Tablica 6: Primjeri parametara genetskih podataka

	Raw	Obraden
Format	fastq/bam	gvcf/vcf + bed/23me(txt)
Veličina	50Gb	2Gb
Čvorovi	>3	>5

Kada se broj čvorova promijeni, pohranjeni podatci se ponovno distribuiraju da se zadovolji zahtjevu minimalnog broja čvorova.

Kako bilo tko, tko poželi, može biti čvor u distribuiranoj mreži, nije sigurno da svaki čvor uistinu i pohrani podatke za koje tvrdi. Iz tog razloga, periodično se vrši provjeravanje dostupnosti podataka prema odgovarajućem protokolu iz Sigurnosnog nivoa. Podatci provjere se šalju u blockchain i na osnovu rezultata se čvorove nagrađuje ili mijenja njihova reputacija.

Osiguravanje privatnosti pohranjenih podataka

Privatnost pohranjenih podataka je u takvoj distribuiranoj mreži moguća zbog primjene asimetrične kriptografije. Stvarna metoda enkripcije je određena odgovarajućim sigurnosnim protokolom.

BILJEŠKA | Treba naglasiti da se mehanizam subvencija ne može primijeniti na pohranu kriptiranih podataka.

Specifične značajke pohranjivanja genomskih informacija

Nedvojbeno povezati određene fragmente sa identitetom odgovarajuće osobe se može samo u slučaju ako je fragment jako velik. Iz ovoga razloga se genom u distribuiranoj mreži pohranjuje u kratkim fragmentima.

Fragmentiranje genoma se uvijek vrši na osnovu referentnog genoma. Fragmentiranje se za svaki referentni genom (npr drugu verziju ljudskog referentnog genoma ili genoma organizma druge vrste) bira samo jednom, a svakom fragmentu je dodijeljen identifikator koji je jedinstven za taj referentni genom.

3.6. Osobni podatci korisnika

Zapis – je najmanja jedinica na burzi razmijene podataka. Nije moguće prenijeti (npr., prodati za nagradu) samo dio informacije sadržane u zapisu. Moguće je napraviti novi zapis koji će sadržavati samo dio informacije, ali u tom slučaju će taj zapis biti značajno manje interesantan potencijalnim kupcima iz razloga njegove niske reputacije.

Schema podataka određuje koju informaciju i koji format bi zapis trebao sadržavati. Uporaba shema je fleksibilan alat za objedinjavanje formata podataka među sudionicima mreže. Identifikator sheme podataka može biti bilo koja poveznica koja omogućava pojedincu da pronađe svoj opis na mreži, npr., URL ili adresu pametnog ugovora sa opisom. Treba shvatiti da je primarna svrha podatkovne sheme objedinjavanje koncepta, što je sve uključeno u podatke, u interakciji kupaca i prodavača tih podataka.

Specifikacije upitnika

Ispunjavanje upitnika znatno podiže vrijednost genomskih podataka. Potencijalno može postojati veliki broj upitnika pa je nužno uvesti koncept specifikacije upitnika.

Specifikacija upitnika – predstavlja kompletan opis svih polja upitnika i dozvoljenih vrijednosti tih polja.

Formalno, specifikacija upitnika se sastoji od autorove reference, opisa i liste redoslijeda zapisa od kojih se svaki podudara sa točno određenim poljem

upitnika. **Numeričko polje A** korisnikov odgovor mora biti unutar dozvoljenih vrijednosti domene [a,b]. Odgovor je kodiran cijelim brojem bez predznaka, a brojenje počinje sa lijeve strane domene:

BILJEŠKA | Odgovor na takvu vrstu pitanja je javno dostupan s obzirom da ne ugrožava privatnost.

Višestruki izbori Korisnikovi odgovori su kodirani kao cijeli broj bez predznaka te tako uređeni predstavljaju serijski broj odgovora sa liste upita. U slučaju da je vrijednost jednaka nuli, korisnik je preferirao da ne odgovori na pitanje.

BILJEŠKA | Pozor: Odgovor na takvu vrstu pitanja je javno dostupan s obzirom da ne ugrožava privatnost.

Odgovor poveznica Korisnikov odgovor je pohranjen kao poveznica.

BILJEŠKA | Odgovor na takvu vrstu pitanja je privatna informacija.

Ispunjeni upitnik – je podatkovna struktura koja sadrži korisnikove odgovore na pitanja upitnika.

BILJEŠKA | Ova struktura je, ukoliko je to nužno, potpuno kriptirana. Pohranjena je na barem jednom korisničkom računalu. Korisnik može učitati kriptiranu sigurnosnu kopiju na blockchaih ako to želi.

Osobni podatci koji se zahtijevaju od korisnika imaju neke posebne osobine:

- Ne zadovoljavaju svi korisnici kriterije svakog istraživanja. Da bi se provela provjera da li su podatci zadovoljavajući, potreban je pristup osobnim informacijama.

- Osobne informacije ne smiju biti prenošene van korisnikova računala ni eksplicitno (direktno) ni implicitno (“warrant canary“ metoda).
- U slučaju da korisnik zadovoljava uvijete, primiti će ponudu da podijeli svoje podatke u zamjenu za nagradu.

ZAKLJ. 1	S obzirom da prijenos podataka nije dozvoljen, provjera bi trebala biti izvedena u izoliranom okruženju. Kod izvediv u ovom okruženju ima puni pristup privatnim informacijama, ali ne može komunicirati sa ostalim dijelovima sustava.
ZAKLJ. 2	Nije dozvoljeno slanje rezultata ove izolirane funkcijske izvedbe potencijalnom kupcu jer bi to moglo ugroziti privatnost korisnika.
ZAKLJ. 3	Razumno je odabrati ograničen skup osobnih podataka, koji su nužni za donošenje zaključka u zasebnom koraku. Korisnik će, na zaslonu ponude razmjene, moći vidjeti na osnovu kojih podataka se donosi odluka.
ZAKLJ. 4	Podatci koji su namijenjeni prijenosu su izričito navedeni na zaslonu ponude razmjene. Samo podatci za koje je tražena provjera mogu biti prenošeni implicitno, a lista tih podataka će također biti dostupna korisniku.

Zahtijevanje osobnih podataka:

- U prvom koraku zahtijeva se pristup osobnim korisničkim podacima. Samo će identifikatorima podataka koji su izlistani u zahtjevu biti dozvoljena funkcija provjere.
- U drugom koraku se, u izoliranom okolišu, izvodi program funkcije za provjeru, a rezultat predstavlja formuliranu ponudu koja se odnosi ili na razmjenu podataka ili na otkazivanje ponude. Ukoliko se radi o otkazivanju ponude podatci neće bit prenošeni.

Ukoliko se radi o razmijeni podataka korisnik je obaviješten da je formulirana ponuda za razmjenu podataka, korisnik provjerava ponudu,

provjerava listu podataka koja se je koristila tijekom provjere te listu podataka koji će biti prenošeni ukoliko se korisnik složi sa ponudom.

Ukoliko korisnik odbije ponudu podatci neće biti prenošeni.

Ukoliko korisnik prihvati ponudu, prenositi će se samo izričito navedeni podatci.

Bibliografija

- [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.aaccjnl.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>