

KIRJALLISUUSKATSAUS:  
GENOMILÄÄKETIETEEN KEHITYS 2010-LUVULLA

*Jussi Kosunen*

Syventävien opintojen opinnäytetyö

Lääketieteen koulutusohjelma

Itä-Suomen yliopisto

Terveystieteiden tiedekunta

Lääketieteen laitos / kliininen kemia

Elokuu 2019

ITÄ-SUOMEN YLIOPISTO, Terveystieteiden tiedekunta  
Lääketieteen laitos  
Lääketieteen koulutusohjelma  
KOSUNEN, JUSSI M.: Kirjallisuuskatsaus: Genomilääketieteen kehitys 2010-luvulla  
Opinnäytetyö, 57 sivua  
Opinnäytetyön ohjaajat: professori Kari Pulkki ja dosentti Esa Hämäläinen  
Elokuu 2019

---

Avainsanat: bioinformatiikka, DNA, geeni, Illumina, Ion Torrent, Nanopore, nukleotidien tehosekvensointi, lainsäädäntö, etiikka

Geeniteknologian mullistaneet, jo yli kymmenen vuoden ikään yltäneet tehosekvensointimenetelmät ovat luoneet uusia mahdollisuuksia lääketieteelliselle tutkimukselle ja sairauksien diagnosoinnille. Yksittäisistä geenitesteistä siirrytään yhä kattavampiin ja tarkempiin paneelitutkimuksiin, jonka lisäksi aiemmin tutkimuskäyttöön jääneet koko eksomin ja genomien sekvensoinnit alkavat osoittaa diagnostista hyötyä. Käyttöön vakiintuneet toisen sukupolven sekvensointialustat ovat saaneet uusia kilpailijoita kolmanneksi sukupolveksi nimettyjen reaaliaikaisten, puolijohteita ja nanohuokosia käyttävien sekvensointimenetelmien tultua markkinoille. Kaiken lisäksi teknologian kehityksen lieveilmiönä esiin on noussut suoraan kuluttajille geenitestejä markkinoivia kansainvälisiä yrityksiä, jotka ovat siirtäneet terveystiedon käsittelyn lääkärin vastaanotolta kotisohvalle.

Kokonaisten genomien tehosekvensointi tuottaa valtavia määriä dataa, jonka analysointi vaatii aikaa ja tehokkaita työkaluja. Geeniteknologian rinnalla kehittynyt bioinformatiikka onkin keskeisessä asemassa nykyaikaisen sekvensointidatan tulkitsemisessä. Tiedon paljouden vuoksi olemmekin siirtymässä genomiikan aikakaudelta ”big data” -aikaan, jossa geenitestauksen pullonkaulana ei olekaan enää sekvensointimenetelmien hitaus. Lääketieteellisesti hyödyllisen tiedon löytäminen miljoonista muuttujista vaatii uusia tietoteknisiä ratkaisuja.

Suomen lainsäädäntö on viime vuosina seurannut tiiviisti Euroopan unionin asettamaa mallia terveystietojen käytöstä. Vaikka nimenomaan geenitestejä koskevaa, yhtenäistä lainsäädäntöä ei ole, asettavat säädetyt lait kuitenkin toimintakehykset sekä potilastyölle että lääketieteelliselle tutkimukselle. Tästä huolimatta yhä laajenevat geenitestit altistavat sekä tutkijat että tutkittavat uusille eettisille ongelmille geenitiedon huolellisesta ja tarkoituksenmukaisesta käytöstä.

UNIVERSITY OF EASTERN FINLAND, Faculty of Health Sciences  
School of Medicine  
Medicine

KOSUNEN, JUSSI M.: Developments in genomic medicine in the 2010s: a review

Thesis, 57 pages

Tutors: Kari Pulkki, professor and Esa Hämäläinen, docent

August 2019

---

Keywords: bioinformatics, DNA, gene, Illumina, Ion Torrent, Nanopore, high-throughput sequencing, legislation, ethics

With over a decade in use, high-throughput sequencing methods have revolutionized gene technology and created new opportunities for medical research and disease diagnosis. Genetics are shifting from singular tests to more comprehensive and accurate gene panel assays, and in addition whole exome and genome sequencing have started to display clinical validity after being used mainly as tools for research. Well established second generation sequencing platforms have gained new competitors on the market from real time, semiconductor and nanopore sequencing methods dubbed as the third generation of sequencing. To add to all that, new internationally operating companies have started to market gene tests directly to consumers as a byproduct to the technological advances. This has caused a shift in the handling of health information from the clinic to the patient's own home.

High-throughput sequencing of whole genomes creates an enormous amount of data, which takes time and powerful tools to analyze. Bioinformatics, which developed alongside gene technology, is a key factor in interpreting contemporary sequencing data. Due to the wealth of overall sequencing data we are moving from the decade of genomics into the age of "big data", where the limiting factor in genetic testing is no longer insufficient sequencing power. Finding medically actionable information from millions of variables requires novel information technology solutions.

Finnish legislation has these past few years been closely following the European Union's model on the use of health information. Even with the lack of a comprehensive legislation addressing gene tests, the laws in effect set the frame of operation in both patient care and medical research. Regardless, the ever expanding gene tests expose both the researcher and the subject to new ethical problems concerning the careful and purposeful use of genetic information.

# SISÄLTÖ

## TERMISTÖ

1. JOHDANTO	6
2. UUDET SEKVENSOINTIMENETELMÄT	8
2.1. Yleistä genetiikasta ja geeniteknologiasta	8
2.2. Sekvensoinnin ensimmäinen sukupolvi	10
2.3. Sekvensoinnin toinen sukupolvi	12
2.3.1. 454-pyrosekvensointi	12
2.3.2. Illuminan CRT-sekvensointi	14
2.3.3. SOLiD-ligaasisekvensointi	15
2.4. Sekvensoinnin kolmas sukupolvi	16
2.4.1. HeliScope-sekvensointi	16
2.4.2. PacBio RS II – reaaliaikainen sekvensointi	17
2.4.3. Ion Torrent – puolijohdesekvensointi	18
2.4.4. Oxford Nanopore MinION – nanohuokossekvensointi	19
2.5. Uuden sukupolven sekvensoinnin käyttöaiheet	22
2.5.1. Kohdennettu sekvensointi	23
2.5.2. Eksomisekvensointi	23
2.5.3. Koko genomin sekvensointi	25
2.6. Kuluttajamarkkinoiden geenitestit	27
3. BIOINFORMATIIKAN KEHITYS	30
3.1. Bioinformatiikan historia lyhyesti	30
3.2. Bioinformatiikka uuden sukupolven sekvensoinnissa	33
4. GEENITEKNOLOGIAA KOSKEVA LAINSÄÄDÄNTÖ	35
4.1. Yleiskatsaus lainsäädännöstä	35
4.1.1. Lainsäädännön viimeaikaiset muutokset	35
4.1.2. Potilasta koskeva lainsäädäntö	37
4.1.3. Lääkäreitä ja tutkimusta koskeva lainsäädäntö	40
4.2. Laajojen geenitestien eettiset näkökannat	42
5. POHDINTA	45
LÄHTEET	49

## TERMISTÖ

APS	Adenosiini-5'-fosfosulfaatti, pyrosekvensoinnissa käytettävä energinen yhdiste
ATP	Adenosiinitrifosfaatti, tärkeä solujen energia-aineenvaihdunnan yhdiste
BP	Base pair, emäspari, jolla tarkoitetaan myös yhtä tutkittavaa sekvenssin emästä
cDNA	Komplementaarinen DNA, yksijuosteisesta RNA:sta käänteiskopioitu vastine-DNA
CE	Conformité Européenne, EEA:n standardien mukainen sertifiointi
CGH	Comparative genomic hybridization, näyte- ja kontrolli-DNA:ta vertaileva tutkimus
ChiP	Chromatine immunoprecipitation, DNA-proteiinikomplekseja hyödyntävä tutkimus
CRT	Cyclic reversible termination, Illumina-alustoilla käytettävä sekvensointimenetelmä
Cy3,Cy5	Syaniini-3/5-metioniini, synteettinen bioteknologiassa käytettävä väriaine
ddNTP	Dideoksinukleotidi, Sangerin menetelmässä käytettävä, muokattu emäs
<i>de novo</i>	Termistä <i>novus</i> (lat.), uutena tai aivan alusta tehtävä toimenpide
dNTP	Deoksinukleotidi, DNA-synteesiin käytettävä, vapaa emäs
DTC	Direct-to-Consumer, suoraan kuluttajalle markkinoitavat geenitestit
FISH	Fluorescence <i>in situ</i> hybridization, fluoresoivia DNA-koettimia käyttävä tutkimus
gDNA	Genominen DNA, tumansisäinen, koko perimän kopion sisältävä DNA
HTS	High-throughput sequencing, katso NGS.
Indel	Insertion-deletion, lyhyt emästen lisäys tai poistuma
<i>k</i> -mer	Laskennallisesti käsiteltävä emäsjärjestysjakso, jonka pituus on <i>k</i>
miRNA	Micro RNA, lyhyt proteiinia koodaamaton ja mRNA:n toimintaa säätelevä RNA-jakso
mRNA	Messenger RNA, DNA:n transkription tuottama lähetti-RNA
mtDNA	Mitokondrionaalinen DNA, rengasmaisen mitokondrionsisäinen DNA
NGS	Next-generation sequencing, uuden sukupolven sekvensointimenetelmä
PCR	Polymerase chain reaction, sykleittäin suoritettava tehokas DNA:n monistusreaktio
PPi	Pyrofosfaatti, ei-orgaaninen ATP-molekyylistä irtoava fosfaattiryhmä
SD	Segmental duplication, DNA-jakson peräkkäisesti esiintyvä jaksomonistuma
SMRT	Single molecule real-time, yhden DNA-molekyylin reaaliaikainen sekvensointi
SNP	Single nucleotide polymorphism, yhden emäksen muutos toiseksi, esim. A → C
UDI	Unique device identifier, laitteistokohtainen yksilöntunniste
WES	Whole exome sequencing, koko eksomin kattava geenitutkimus
WGS	Whole genome sequencing, koko perimän kattava, laaja geenitutkimus

## 1. JOHDANTO

2000-luku on ollut lääketieteellisen geenitutkimuksen kulta-aikaa. Uusien, nopeampien ja tehokkaampien DNA-sekvensointimenetelmien tulo markkinoille on mahdollistanut nopeamman ja halvemman analyysin yksittäisen henkilön tai potilaan perimästä. Alun alkaen pääosin tutkimuskäyttöön soveltunut genomisekvensointi on osoittautunut sopivaksi kliiniseen käyttöön. Eritoten harvinaisten tai muutoin vaikeasti todennettavien sairauksien diagnostinen sekvensointi on nykyisin mahdollista. Sekvensointiteknologian kehitys on kuitenkin ollut niin nopeaa, etteivät monetkaan viranomaistahot ja lainsäädäntö ole pysyneet muutoksessa mukana. Tätä ilmiötä käyttävät hyväksi kansainvälistä kauppaa käyvät yritykset, jotka tarjoavat tasoltaan vaihtelevia geenitestejä yhä halpevaan hintaan suoraan kuluttajille. Ilmiön ennustetaan lisäävän terveydenhuollon kuormitusta etenkin perinnöllisyysneuvonnan tarpeen lisääntyessä, sillä läheskään kaikki suoraan kuluttajille geenitestejä myyvät (Direct-to-Consumer, DTC) yritykset eivät tarjoa selkokielistä konsultaatioapua tai lausuntoja saaduista tuloksista. Testattavaksi tarjotaan mm. omaa suku- ja väestöhistoriaa, erilaisia perinnöllisiksi oletettuja soveltuvuuksia sekä terveyttä ennustavia geenitutkimuksia. Suomessa muutokseen on jo havahduttu, sillä vireillä oleva kansallinen genomitietohanke pyrkii arkipäiväistämään uuden perimätiedon käytön yksilön ja yhteisön hyväksi. Hankkeella on pitkä matka edessään ja moni ongelma kaipaa ratkaisua; yksi tärkeimmistä on terveystieteen ammattilaisten tietotaidon nykyaikaistaminen.

Vuosituhaten vaihteen jälkeen valmistui tähän mennessä maailman suurin biologinen yhteistyöhanke [1]. 1990-luvun alussa aloitettu Human Genome Project (HGP) sai päätöksensä vuonna 2003, ja projektin työryhmä julkaisi loppuraportin ihmisen eukromatiinin sekvensoinnista vuonna 2004 [2]. 13 vuotta kestänyt hanke maksoi hieman alle 3 miljardia Yhdysvaltain dollaria, ja sen työryhmä koostui kahdestakymmenestä yliopistosta sekä tutkimuskeskuksesta ympäri maailman [3]. Hankkeen tarkoituksena oli saada selville ihmisen eukromatiinin DNA-järjestys, ulkopuolelle jäi kromosomien sentromeereissä ja telomeereissä esiintyvä heterokromatiini. Tulokseksi saatua tietoa ihmisen genomista ja kromosomiston fyysisestä kartoituksesta luovutettiin vapaaseen käyttöön jo projektin alusta lähtien, ja hankkeen päätyttyä kansainvälisellä tiedeyhteisöllä oli ensimmäistä kertaa käytössään ihmisen referenssigenomi. Mutta mitä ihmisen perimästä saatiinkaan selville?

Raportissaan [2] International Human Genome Sequencing Consortium (IHGSC) kertoi löydöksistään seuraavaa: ihmisen eukromatiini, joka kattaa n. 90 % tuman sisäisestä DNA:sta, koostuu n. 2,88 miljardista emäsparista. Heterokromatiini mukaan laskettuna ihmisgenomin koko on n. 3,08

miljardia emäsparia. Laskentatavasta riippuen ihmisellä oletettiin olevan n. 20 000–25 000 geeniä, mutta myöhemmissä tutkimuksissa luku on laskenut n. 19 000 proteiinisynteesiin osallistuvaan geeniin [4]. Lisäksi ihmisgenomi sisältää muiden kädellisten ohessa suuren määrän segmenttimonistumia (segmental duplication, SD) eli peräkkäin toistuvia lähes identtisiä DNA-jaksoja. Segmenttien rakenne on altis yksilön fenotyypin muokkaaville geenimutaatioille, joten myöhemmissä tutkimuksissa [5-6] on arvioitu segmenttimonistuma-alueiden mutaatioiden edistäneen ihmisapinoiden evoluutiota. Näiden löydösten ja projektin muun datan analysointi jatkuvat yhä.

Teknologiselta kannalta hankkeen etenemisen ominaispiirteinä voidaan mainita kiihtyvä teknologian kehitys, jonka ansiosta sekvensointikustannukset laskivat projektin aikana lähes sadasosaan siitä, mitä ne olivat 1990-luvulla [7]. Kiihdyttääkseen teknologian kehitystä National Human Genome Research Institute (NHGRI) loi haasteen kansainväliselle tiedeyhteisölle kehittää tapa sekvensoida ihmisgenomi tuhannen dollarin hintaan [7-8]. Haasteen kannustimena on myös toiminut vuodesta 2004 jaettavat apurahat, joilla tuetaan kehittyneiden sekvensointiteknologioiden tutkijoita.

Tällä hetkellä koko ihmisgenomin sekvensointi 1 %:n virhemarginaalilla maksaa alle 1 500 dollaria [9]. Alle kymmenessä vuodessa sekvensointikustannukset ovat romahtaneet lähes tuhannesosaan lähtötasosta. Tämän on mahdollistanut vuodesta 2008 alkaen käyttöön vakiintunut uudenlainen sekvensointiteknologia. Teknologian ansiosta yksilön koko eksomin, eli kaikkien proteiineja koodittavien geenien, tai koko perimän tutkimus on mahdollista [10]. Aiemmin enintään tutkimuskäyttöön soveltunut menetelmä saa yhä useampia kliinisiä käyttöaiheita sitä mukaa kun uusia geenilöydöksiä raportoidaan.

Tässä katsauksessa perehdytään systemaattisen haun avulla uuden sukupolven sekvensointiteknologian perusteisiin, em. teknologian mahdollistamiin geenitesteihin ja niiden tulkintaan sekä eettisiin ja juridisiin kysymyksiin teknologian käytöstä. Materiaali on kerätty Itä-Suomen yliopiston kirjastosta saatavilla olleista painetuista julkaisuista sekä yliopiston sähköisen kirjaston verkkotietokantojen hakupalvelun avulla.

## 2. UUDET SEKVENSOINTIMENETELMÄT

### 2.1. Yleistä genetiikasta ja geeniteknologiasta

Geenisekvensoinnilla tarkoitetaan eliöiden periytyviä ominaisuuksia ohjelmoivien rakenteiden, kaksijuosteisten DNA-säikeiden, emäsjärjestyksen selvittämistä. DNA koostuu neljästä nukleotidista eli kolmiosaisesta molekyylistä, joissa nk. runko-osaan eli pentoosisokeriin (deoksiriboosi) kiinnittyneenä on fosfaattiryhmä ja yksi neljästä erilaisesta emäsryhmästä. Emäsryhmät jaetaan puriineihin, joita ovat adeniini (**A**) ja guaniini (**G**), sekä pyrimidiineihin, joihin lukeutuvat sytosiini (**C**), tymiini (**T**) sekä RNA:ssa esiintyvä urasiili (**U**). RNA on DNA:ta lähes identtisesti muistuttava makromolekyyli, jossa runko-osana on riboosisokeri ja emäsryhmistä tymiinin sijaan esiintyy urasiili.

Sekä DNA että RNA muodostavat peräkkäisiä, helminauhamaisia nukleotidijuosteita fosforidiesterisidosten avulla. Sidos on kahden nukleotidirungon 3. hiiliatomin hydroksyyllisivuketjun (-OH) ja 5. hiiliatomin fosfaattisivuketjun (-PO<sub>4</sub>) välinen. Tämä antaa DNA- ja RNA-juosteille selkeän, epäsymmetrisen kulkusuunnan. Suuntaa merkitään tieteellisissä teksteissä 5' → 3' -suunnaksi, jota perimän lukuun, muokkaukseen ja korjaukseen osallistuvat entsyymit noudattavat. Soluissa DNA esiintyy luonnollisessa muodossaan kaksoiskierteenä eli heeliksiä, jossa kaksi DNA-juostetta on toisiinsa vastakkain yhdistyneenä emäsryhmien välisten vetysidosten avulla. Vastakkain asettuvat emäkset pariutuvat Watson-Crickin emäsparisäännön mukaisesti, jossa parin muodostavat A ja T, sekä C ja G [11]. RNA-molekyyllissä tymiinin roolin korvaa urasiili, jolloin pari muodostuu A:n ja U:n kesken. RNA esiintyy pääosin yksijuosteisena ja on laskostunut erilaisiksi silmukoiksi omien emästensä kesken, mikä mahdollistaa molekyylin lukuisat erilaiset roolit solun toiminnan säätelyssä [12].

Ihmisillä ja muilla aitotumaisilla perimä eli genomi, joka kattaa kaiken periytyvän DNA:n, on tumakotelon suojaamana ja tiiviisti pakkautuneena histoniproteiinien sitomaksi kromatiiniksi. Solunjakautumista edeltäen kromatiini pakkautuu erillisiksi kromosomeiksi, joita ihmisillä tyypillisesti on 23 paria. Kromosomeja 1–22 kutsutaan autosomeiksi ja sukupuolikromosomeja X ja Y allosomeiksi. Solujen energiantuotantoon erikoistunut soluelin, mitokondrio, sisältää oman bakteeriplasmidia muistuttavan, rengasmaisen DNA-juosteen (mtDNA). Mitokondrioiden DNA:n periytymisen on oletettu olevan yksinomaan maternaalista eli äidiltä peräisin oleva mtDNA periytyy lapselle. Poikkeuksia toki on raportoitu, mutta niiden esiintyminen on hyvin harvinaista. Tutkimustiedon lisään-



tyessä on havaittu, että alkiokehityksen alkuvaiheissa useampi biologinen prosessi pyrkii tuhoamaan paternaalista mtDNA:ta. [13]

Ihmisillä on arvioitu olevan 19 000 – 20 000 proteiinisynteesiin osallistuvaa geeniä. Yksittäinen geeni merkitsee tiettyä DNA-emäsjärjestysjaksoa, joka ohjelmoi yhtä tai useampaa transkriptiotuotetta. Aktiivisesti proteiinisynteesiin osallistuvat geenit sijaitsevat löyhemmin pakatussa eukromatiinissa, kun taas vähemmän tunnettu heterokromatiini on tiiviimmin pakkautunutta ja sen oletetaan osallistuvan enemmänkin perimän kolmiulotteisen rakenteen säätelyyn. Proteiinisynteesiä ohjaavassa geenissä esiintyy emäskolmikkoja, kodoneja, jotka vastaavat yksittäisiä polypeptidiin liitettäviä aminohappoja. Neljän eri emäksen käyttö kolmikkona mahdollistaa  $4^3$  eli 64 erilaista kodonia, jotka ohjelmoivat 20 eri aminohappoa. Ylimäärä kodonivaihtoehtoja mahdollistaa yksinkertaisen virheensietokyvyn, jolloin yksittäisten emästen vaihtuminen tai puuttuminen eivät aina lamauta proteiinisynteesiä. Kolme kodonia on tunnistettu nk. lopetuskoodeiksi, jotka pysäyttävät proteiinisynteesin. [14]

Vuonna 1958 lääketieteen Nobelin jakaneet George Beadle ja Edward Tatum kehittivät geenien toiminnasta ”yksi geeni – yksi entsyymi” -hypoteesin. Koeasetelmassaan he altistivat leipähome *Neurospora crassan* röntgensäteilylle, ja altistuksen saaneiden homeiden jälkeläiset osoittivat geenimutaatioiden aiheuttamia entsyymien tuotantovirheitä [15]. Sittemmin hypoteesi on tarkentunut muotoon ”yksi geeni – yksi polypeptidi.” Kaikki proteiinit eivät ole entsyymejä, ja useat proteiinit koostuvat polypeptidien muodostamista useista alayksiköistä. Hypoteesille on myös poikkeuksensa silloin kun geeni silmukoidaan vaihtoehtoiseen muotoon, se koodaa solun säätelyyn osallistuvia lyhyitä RNA-juosteita tai geenimutaatio saa aikaan poikkeavia transkriptiotuotteita. [16]

Perimän tutkimus on perinteisesti jaettu karyo- ja molekyylikaryotyypitykseen sekä uudempiin molekyylogeneettisiin tutkimuksiin. Karyotyypityksessä potilaasta otetaan jakautumiskykyisiä, tumallisia soluja, jotka käsitellään solusyklin pysäyttävillä kemikaaleilla ja värjätään esimerkiksi Giemsa-väriaineella. Tiivistyneet ja värjäytyneet DNA-säikeet, kromosomit, tutkitaan valomikroskopiolla. Menetelmän erotustarkkuus mahdollistaa enintään kromosomien määrän, koon ja värjäytyneiden raitojen lukumäärän tarkastelun. Molekyylikaryotyypitys on aikojen saatossa kehittynyt tehokkaaksi tavaksi tutkia kromosomien rakennepoikkeamia muutaman tuhannen emäsparin erotustarkkuudella. Rutiinikäyttöön ennättänyt, tiettyjä ennalta määriteltäviä emäsjärjestyksiä tunnistava FISH (fluorescence *in situ* hybridization) pystyy visualisoimaan laajan kirjon kromosomiston muutoksia käyttämällä fluoresoivia merkkiaineita. Käyttämällä useita erilaisia koettimia yhtä aikaa pystytään FISH-

menetelmällä tunnistamaan yksittäisten kromosomien rakennepoikkeamia 5–10 tuhannen emäksen erotustarkkuudella. Suomalaisen tutkijoiden kehittämä CGH (comparative genomic hybridization) eli vertaileva genomien hybridisaatiotutkimus kykenee selvittämään koko genomien laajuisia rakennetta ja lukumäärämuutoksia. FISH-menetelmästä poiketen pysyvät nukleotidikoettimet kiinnitettynä koealustalle, ja potilaan värileimattu DNA-näyte kilpailee koettimista kontrolli-DNA:n kanssa. Taulukoksi aseteltu aCGH (array-CGH) yhdistettynä yhden emäksen muutoksia (single nucleotide polymorphism, SNP) tunnistaviin lisäkoettiin onkin tarkin perinteisistä molekyylikaryotyypitutkimuksista. Nykyisillä molekyylogeneettisillä menetelmillä DNA-näytteiden koostumus saadaan selvitettyä yksittäisten emästen tarkkuudella. Näytteeksi riittää vain vähäinen määrä, ~50 – 1 000 nanogrammaa, puhdasta DNA:ta. [17, 18]

## 2.2. Sekvensoinnin ensimmäinen sukupolvi

Ensimmäisen sukupolven sekvensointimenetelmää kutsutaan Sangerin menetelmäksi englantilaisen kemistin, Frederick Sangerin (s. 1918 – k. 2013), mukaan. Kaksi kemian Nobel-palkintoa elämänsä aikana saanut Sanger kehitti kollegoineen tavan selvittää DNA-molekyylin emäsjärjestys entsyymaattisella ”chain termination” -tekniikalla 1977 [19]. Tämä oli seuraava kehitysaskel Sangerin aikaisemman ”Plus and Minus” -menetelmän [20] ja samanaikaisesti kehitetyn Maxam & Gilbertin kemiallisen sekvensoinnin [21] jatkoksi.

Sangerin menetelmässä käytetään dideoksinukleotideja (ddNTP), eli yksittäisiä emäsmolekyyliä, joista puuttuu DNA-molekyylien peräkkäisen ketjuttamisen eli polymeerasireaktion mahdollistava hydroksyyli-ryhmä deoksiriboosin 3'-hiilestä. Sekvensointia varten tutkittava yksijuosteinen DNA-näyte eli ”templaatti” laitetaan reaktioliuokseen, jossa on DNA-polymeraasientsyymiä, kaikkia neljää nukleotidia (dNTP) irrallisina sekä yhden nukleotidin dideoksimuotoa. Yksi reaktioon osallistuvasta nukleotidista on oltava jollain tavalla leimattu myöhempää havainnointia varten. Reaktiossa polymeerasientsyymi rakentaa DNA-näytteelle komplementtijuostetta liittäen siihen emäsparisäätön mukaisesti saatavilla olevia vapaita emäksiä, kunnes ketjun päähän liitetty dideoksinukleotidi pysäyttää reaktion. Tulokseksi saadaan useita eripituisia DNA-pätkiä, ”fragmentteja”, jotka kaikki päättyvät samaan ennalta valittuun dideoksinukleotidiin. Reaktio toistetaan kullekin neljälle emäkselle erikseen ja näytteet erotellaan rinnakkain geelielektroforeesilla. Eripituiset DNA-fragmentit kulkeutuvat geelissä sähkövirran avulla siten, että lyhimmat ja molekyylipainoltaan pienimmät fragmentit ovat nopeampia ja isommat hitaampia. Näin voidaan päätellä tutkitun DNA-näytteen emäs-

järjestys lukemalla geeliin piirtyviä fragmenttivyohtykykkeitä alhaalta ylöpäin. Menetelmää käytettiin ensimmäisessä DNA-genomin sekvensoinnissa vuonna 1977, jossa selvitettiin DNA-bakteriofagi  $\phi$ X174:n (*phi-X-174*) koko perimä [22].

1980-luvun puolivälissä julkisuuteen tuotu PCR-menetelmä mahdollisti kaksijuosteisen DNA:n sekvensoinnin Sangerin menetelmällä. PCR eli polymeerasiketjureaktio kykenee monistamaan yksi- tai kaksijuosteista DNA:ta eksponentiaalisesti käyttämällä korkeaa lämpötilaa sietävää DNA-polymeraasientsyymiä lämmitys- ja jäähtyissykleissä [23]. Tämä mahdollistaa yksi- ja kaksijuosteisen DNA:n sekvensoinnin hyvin pienistä näytemääristä. Tavallisessa PCR-reaktiossa käytetään kahta polymeerasientsyymiä aktivoivaa aluketta (primer), jotta monistettavan DNA:n molemmat juosteet voidaan hyödyntää yhtä aikaa. Sangerin menetelmässä käytetään vain yhtä aluketta, jolloin polymeerasireaktio tuottaa satunnaisesti dideoksinukleotideihin pysähtyneitä, eripituisia DNA-fragmentteja. Sangerin sekvensointia PCR-menetelmän avulla kutsutaan sykliseksi sekvensoinniksi [24].

Vuonna 1981 julkaistu ”shotgun”-sekvensointi (vapaasti käännettynä haja-ammuntasekvensointi) osoittautui korvaamattomaksi menetelmäksi sekvensoida suuria pituuksia DNA:sta, ja mahdollisti näin genomilaajuisten sekvensointihankkeiden aloituksen. Shotgun-sekvensoinnissa pitkät DNA-juosteet pilkotaan joko suunnitellun pituisiksi tai satunnaisiksi fragmenteiksi, käyttämällä esimerkiksi restriktioentsyymejä tai mekaanisia voimia. DNA-fragmentit sekvensoidaan satunnaisessa järjestyksessä ja saadut emäsjärjestykset liitetään peräkkäin niiden sisältämien päällekkäisyysjaksojen (contig, termistä *contiguous*) avulla [25]. Menetelmä on yhä olennainen osa uuden sukupolven sekvensointialustojen toimintaa.

Emästen fluoresenssimerkkauksen [26], herkemmin sekvensointiin sopivien polymeerasientsyymien muokkauksen [27] sekä kapillaarielektroforeesin [28] lisäksi Sangerin menetelmään ovat nykyaikaisten, pitkälle automatisoitujen ensimmäisen sukupolven sekvensointialustojen ydin. Kapillaarielektroforeesi nopeuttaa DNA-fragmenttien kokojärjestyksen selvittämistä ja mahdollistaa useamman DNA-näytteen yhtäaikaisen sekvensoinnin. Fluoresenssivärien käyttö mahdollistaa emäsjärjestyksen nopean tunnistamisen suoraan tietokoneen avulla, ja menetelmä onkin aktiivisessa käytössä useammassa uudemman sukupolven sekvensointialustassa.

Sangerin menetelmää käytetään yhä sekvensoinnin kultaisena standardina sen luotettavuuden ja tarkkuuden vuoksi. Uudemman sukupolven sekvensointitulosten herkkyyden ja tarkkuuden parantamiseksi useat laboratoriot käyttävät Sangerin menetelmää ylimääräisenä työvaiheena [29].

## 2.3. Sekvensoinnin toinen sukupolvi

Toisen sukupolven sekvensoinnilla (kirjallisuudessa myös uuden sukupolven sekvensointi, tehosekvensointi, ”next-generation sequencing” (NGS) ja ”high-throughput sequencing” (HTS)) tarkoitetaan automatisoituja, laajakirjoisia ja nopeita sekvensointimenetelmiä, jotka poikkeavat Sangerin menetelmästä. Sekvensointialustoja yhdistävät oheiset kolme yksinkertaistettua toimintaperiaatetta: ① DNA-näytekirjaston luominen polymeerasientsyymien avulla ja näytteiden kiinnitys kiinteälle levylle (flow cell), ② syklinen sekvensointireaktion tuottaminen nestevirtauksen avulla sekä ③ sekvensoinnin aiheuttamien molekulaaristen tapahtumien tallentaminen optisten laitteiden avulla. Ennen sekvensointia DNA-näytekirjastot pilkotaan shotgun-sekvensointimenetelmän tapaan satunnaisiksi fragmenteiksi ja rikastetaan PCR-menetelmällä. Sekvensoinnista saadut useat miljoonat emäsjärjestysluvut (read) kootaan ja kartoitetaan yhtenäisiksi kokonaisuuksiksi. Valtavan kokoinen sekvensointidata analysoidaan algoritmipohjaisten tietokoneohjelmien avulla ja löydökset tulkitaan ajantasaisen konsensustiedon mukaisesti. Kukin uuden sukupolven sekvensointialusta koostuu monimutkaisesta entsymologian, kemian, korkeatarkkuuksisen optiikan, tietokonelaitteistojen ja tietotekniikkaohjelmistojen saumattomasta vuorovaikutuksesta. [30]

Sekvensointiteknologian uuden sukupolven voidaan sanoa saaneen alkunsa vuonna 2005, kun kaupalliseen käyttöön julkaistiin pyrosekvensointiin perustuva suuren mittakaavan sekvensointialusta. Vuodesta 2007 lähtien Roche Applied Sciencen omistukseen siirtynyt 454 Life Sciences onnistui yhdistämään emulsio-PCR-menetelmän ja pikolitrakokoisissa reaktiokammioissa tapahtuvan massiivisen rinnakkaissekvensoinnin yhteen laitteistokokonaisuuteen. Sekvensointialusta toi tutkijoiden käyttöön 100-kertaisen tehonlisäyksen Sangerin menetelmään verrattuna ja kuusi kertaa halvemmilla käyttökustannuksilla. [30, 31]

### 2.3.1. 454-pyrosekvensointi

Pyrosekvensointi perustuu DNA-synteesin aikana vapautuvien sivutuotteiden ja useiden entsyymien ketjureaktion aiheuttaman bioluminesenssin optiseen havainnointiin. Sekvensoitavan DNA-fragmentin emäsjärjestys selvitetään reaaliajassa käyttämällä syklistä synteesireaktiota ja syöttämällä reaktioon yksi kokeiltava nukleotidi kerrallaan. Reaktion keskeisinä biomolekyyleinä käytetään neljää entsyymiä, DNA-polymeeraasia, ATP-sulfurylaasia, lusiferaasia ja apyraasia, sekä edellä maini-

tuille entsyymeille spesifejä substraatteja adenosini-5'-fosfosulfaattia (APS) ja lusiferiinia. DNA-polymeraasin liittäessä vapaan nukleotidin fosfodiesterisidoksella osaksi kasvavaa DNA-juostetta jää reaktiosta tähteeksi pyrofosfaattiryhmä (PPi). ATP-sulfurylaasientsyymi luo pyrofosfaatista ja APS:sta ATP-molekyylin vaihtamalla APS:n fosfosulfaatin trifosfaatiksi. Reaktioliuoksen lusiferaasi muuttaa lusiferiinin oksilusiferiiniksi ATP-molekyylin toimiessa substraattina, ja reaktiossa vapautuu valoa ATP-molekyylien lukumäärän mukaisesti. Valosignaali ja sen sijainti tallennetaan optisesti. Lopuksi apyraasientsyymi katalysoi reaktioon osallistumattomat nukleotidit ja substraatit monofosfaateiksi, mikä estää niitä osallistumasta seuraaviin reaktioihin. Reaktiosyklejä toistetaan jokaista neljää emästä koettaen, kunnes näytetemplaattien emäsjärjestys on selvitetty. [32]

454 Life Sciencen sekvensointialustan DNA-näytekirjasto valmistetaan genomisesta DNA:sta (gDNA), joka paineilmalla nebulisoimalla pilkkoutuu n. 300–1 000 emäksen (base pair, bp) pituisiksi yksijuosteisiksi DNA-fragmenteiksi. Sekvensointialukkeen ja biotiinimolekyylin sisältämät sovitinjaksot (adapter) lisätään DNA-fragmenttien päihin ja kukin fragmentti kiinnitetään *streptavidinilla* päällystettyihin, magneettisiin DNA:n kaappaushelmiin. Tavoitteena on kiinnittää yksi DNA-kirjaston fragmentti yhteen helmeen, ja koko näytekirjasto emulsoidaan PCR-vesi-öljy-seokseen. DNA-fragmentin kiinnittäneet helmet pysyvät toisistaan erillään emulsion vesipisaroissa ja PCR-reagenssit jakautuvat emulsion öljyfaasiin, mikä mahdollistaa kaikkien näytefragmenttien klonaalisen rikastamisen yhdellä kertaa [33]. PCR:n jälkeen näyterikastetut helmet kerätään talteen ja sijoitetaan flow cell -levyn pikolitrakokoisiin reaktiokennoihin. Reaktiokennot ovat kooltaan niin pieniä, että niihin mahtuu kerrallaan vain yksi helmi.

Pyrosekvensointimenetelmän vahvuutena pidetään sen muihin toisen sukupolven sekvensointialustoihin verrattuna pitkiä fragmenttien lukupituuksia (500–1 000 emästä). Sekvensointiajo kestää 10–23 tuntia kerrallaan, mikä on nopeahkoa verrattuna muihin toisen sukupolven alustoihin. Menetelmän heikkoutena ja tunnettuna virhelähteenä pidetään epävarmuutta tunnistaa saman emäksen peräkkäisjaksoja; useamman nukleotidin yhtäjaksoinen ketjuttaminen synteisireaktiossa aiheuttaa epätasaisesti skaalautuvia, voimakkaita valosignaaleja. Tämä altistaa insertio-deleetiovirheille (indel), joiden esiintyvyys on n. 0,5–1 %:n luokkaa. Lisäksi menetelmä jää suorituskyvyssään ja käyttökustannuksissaan muista menetelmistä jälkeen. [30, 34] Laittevalmistaja Roche lopetti pyrosekvensointialustansa tuen vuonna 2016, mutta mm. bioteknologiayritys QIAGEN jatkaa pyrosekvensointialustojen markkinointia [35].

### 2.3.2. Illuminan CRT-sekvensointi

Solexa, joka siirtyi Illuminan omistukseen 2007, esitteli vuonna 2006 tavan sekvensoida DNA käyttämällä palautuvasti polymeraasireaktion pysäyttäviä fluoresoivia nukleotideja ("reversible dye termination") ja näitä varten räätälöityä DNA-polymeraasientsyymiä [36]. Sekvensointitapaa voi kutsua myös "cyclic reversible termination" eli CRT-menetelmäksi. Menetelmän erikoispiirteinä ovat kemiallisesti poisleikattavien DNA:n 3'-hiilen esteiden ja fluoresoivien sivuketjujen käyttö jokaista koetettavaa emästä kohden sekä flow cell-levylle muodostettavien fragmenttiryppäiden valmistus siltaavalla PCR-menetelmällä. Alustalla pystytään sekvensoimaan ihmisen koko genomi tarkasti, tehokkaasti ja edullisesti. [34, 37]

Illuminan sekvensointimenetelmässä näytekirjasto luodaan pilkkomalla DNA- tai RNA-näyte hydrodynamisesti n. 125–600 emäksen pituisiksi fragmenteiksi. Fragmenttien päihin kiinnitetään erilliset sovitinjaksot, jotka toimivat polymeraasireaktion aloitusjaksoina, lyhyinä tunnistuskoodeina sekä vastakappaleina flow cell -levyllä oleville sovittimille. Alustalla, jolla näytteiden rikastus ja sekvensointi tapahtuvat, on tiheään istutettuna lyhyitä oligonukleotidisovittimia ja niiden vastakappaleita vierekkäin. Monistamista varten näytefragmentit hybridisoidaan levyllä oleviin sovittimiin ja syklinen PCR-reaktio rakentaa sovittimia pitkin näytekirjaston fragmenteista komplementteja. PCR:n denaturaatiovaiheessa alustalleen kiinnittynyt fragmentti taipuu siltamaiseksi ja hybridisoi-tuu vieressään olevaan sovittimen vastakappaleeseen. Seuraavassa vaiheessa polymeraasientsyymi rakentaa uuden komplementaarisen juosteen alkuperäisen fragmentin viereen. Ketjureaktion jatkuessa näytekirjasto monistuu eksponentiaalisesti omiksi yksijuosteisiksi fragmenttiryhmiin, joiden vieressä on vastaavasti suuri määrä komplementaarisia fragmentteja.

Sekvensointi tapahtuu askel kerrallaan DNA-polymeraasin liittäessä fluoresoivan ja 3'-päästään muokatun nukleotidin näytetemplaattiin. Flow cell -levystä otetaan laservalaisu kuva, jonka mukaan selvitetään jokaisen fragmenttiryhmän sen hetkisen emäksen väri ja sijainti. Tämän jälkeen nukleotidin fluoresoiva sivuketju ja 3'-pään este leikataan kemiallisesti pois, mikä mahdollistaa polymeraasireaktion jatkamisen. Irtonaiset fluoresenssivärit ja reaktioon osallistumattomat molekyylit huuhdellaan pois ja uusi sykli aloitetaan. Yhdessä reaktiosykliissä voidaan kuulustella kaikkia neljää nukleotidia yhtä aikaa, sillä kerrallaan vain yhtä nukleotidia kiinnittyy yhteen fragmenttiryhmään. Reaktiosykliä jatketaan, kunnes kaikkien fragmenttien emäsjärjestys on selvitetty. Tämän jälkeen sekvensointi suoritetaan käänteisenä käyttäen fragmenttiryppäisiin muodostuneita komplementtijuosteita.

Illuminan sekvensointialusta on tämänhetkinen markkinajohtaja tehonsa, tarkkuutensa ja edullisuutensa vuoksi. Yhdellä näyteajokerralla saadaan luettua 4–6 miljardin emäksen järjestys 300–600 emäksen lukupituuksilla. Sekvensointiajot kestävät laitteistosta ja käyttötarkoituksesta riippuen 21 tunnista 10 vuorokauteen, mikä juontuu sekvensointimenetelmän ominaisesta asteittaisesta etenemistavasta. Kirjallisuudessa alustan virhelähteiksi mainitaan substituutiovirheet, jotka johtuvat osittain emästen fluoresenssivärien sekoittamisesta toisiinsa ( $A \rightarrow C$  sekä  $G \rightarrow T$ ) ja osittain sekvensointireaktion siirtymävaiheiden (phasing) arvaamattomuudesta. Virheiden esiintyvyys on n. 0,1–0,2 %. [30, 34, 38]

### 2.3.3. SOLiD-ligaasisekvensointi

Vuonna 2007 Applied Biosystems julkaisi oman sekvensointialustansa, jonka toimintaperiaatteena on oligonukleotidien tunnistaminen ligaasientsyymien avulla. ”Sequencing by oligo ligation and detection” eli akronyymina SOLiD käyttää ainutlaatuista kahden emäksen yhtäaikaista tunnistusmenetelmää. Menetelmässä DNA-ligaasientsyymi koettelee kahdeksan nukleotidin pituisia, fluoresoivia oktameerikoettimia (probe) näytetemplaatin komplementiksi. Oktameeri koostuu kolmesta osasta: ① kahdesta tunnistukseen tarvittavasta normaalista nukleotidista, ② kemiallisesti pois leikattavasta fluoresoivasta päätyosasta ja ③ näiden molempien keskellä olevasta neutraalien nukleotidien runko-osasta. Jokainen oktameerikoetin on värikoodattu yhdellä neljästä fluoresoivasta väristä ja jokaisessa koettimessa on kaksi tunnistusnukleotidia, mikä johtaa 16 toisistaan poikkeavan koettimen käyttöön sekvensointikemiassa. Kahden nukleotidin yhtäaikainen tunnistus lisää myös sekvensoinnin lukusyvyyttä ja tarkkuutta ilman ylimääräisiä työvaiheita. [39]

SOLiD-menetelmässä näytekirjasto luodaan emulsio-PCR-tekniikalla pyrosekvensointialustojen tapaan. Ennen näytekirjaston rikastamista fragmenttien päihin liitetään ligaasientsyymien toiminnalle keskeisten primer-alukkeiden vastakappaleet. Näyterikastetut helmet sijoitetaan flow cell-levyn reaktorikammioihin, joihin ne kiinnittyvät näytefragmenttien 3'-pään sovittimien avulla. Sekvensointisyklin aluksi näytefragmenttiin hybridisoidaan ligaasientsyymien tarvitsema aluke. Jokaisessa syklistä kaikki 16 oktameerikoettinta kilpailevat ligaasientsyymiin kiinnittymisestä. Ligaasientsyymi katalysoi kahden DNA-juosteen välille fosfodiesterisidoksen ainoastaan, jos DNA-juosteet hybridisoituvat näytefragmenttiin virheettömästi. Onnistuneesti liggeratun oktameerin väri tunnistetaan op-

tisesti ja sen fluoresoiva loppuosa poistetaan kemiallisesti. Primer-alukkeesta kasvavaan juosteketjuun jää kahden tunnistetun emäksen lisäksi kolme tunnistamatonta kohtaa, eli sykleittäin tunnistetaan näytefragmentin emäkset 1 ja 2, 6 ja 7, 11 ja 12, 16 ja 17 ja niin edelleen, kunnes reaktiosarja on käyty loppuun. Seuraavassa syklien sarjassa käytetään eripituista primer-aluketta, jolloin ligaa-sientsyymi tunnistaa jo edellisessä syklissä selvitetyn yhden emäksen lisäksi sen vieressä olevan tuntemattoman emäksen. Menetelmää käyttämällä kaikki näytteen emäkset tunnistetaan vähintään kaksi kertaa sekvensointiajon aikana. [39, 40]

SOLiD-alustalle erityispiirteenä on lyhyiden 35–75 emäksen pituisten näytefragmenttien käyttö sekvensoinnissa. Lyhyitä juosteita uudelleensekvensoidaan kymmeniä kertoja, mikä tuottaa satoja miljoonia luettuja nukleotideja sekvensointiajon aikana. [40] Sekvensointiajon pituudeksi mainitaan 6–8 vuorokautta. Mittavan lukusyvyyden vuoksi SOLiD-alustan SNP-erotuskyky on erinomainen, mutta lyhyiden lukupituuksien ja erikoisen emästunnistusmetelmän vuoksi alusta on altis substituuti virheille. Virheiden esiintyvyys on n. 0,06–0,1 %:n luokkaa. [30, 34] Alustaa valmistanut Applied Biosystems siirtyi vuonna 2014 Thermo Fisherin omistukseen.

## **2.4. Sekvensoinnin kolmas sukupolvi**

Kirjallisuudessa kolmannen sukupolven sekvensoinnilla tarkoitetaan yksittäisen nukleinihappomolekyylin reaaliaikaista sekvensointia (single molecule real-time sequencing, SMRT) ja puolijohdesekvensointia. [41] SMRT-tekniikoita yhdistävä toimintaperiaate on näytekirjaston klonaalisen rikastamisen puuttuminen, mikä teoriassa eliminoi PCR:sta johtuvat virheet ja vähentää sekvensointiin kuluvaan aikaan sekä resursseja. Puolijohdesekvensoinnissa emäsjärjestys päätellään optisen havainnoinnin sijaan seuraamalla molekyyllitasolla tapahtuvia pieniä sähköjännitteen muutoksia. Sekvensointimenetelmiä esitetään yksinkertaistettuna kuvaajassa 1 [s. 21].

### **2.4.1. HeliScope-sekvensointi**

Ensimmäisen SMRT-sekvensointialustan julkaisi Helicos Biosciences vuonna 2008. HeliScope-menetelmässä restriktioentsyymillä pilkotut, monistamattomat DNA-fragmentit kiinnitetään sekvensointilevylle sovitinjuosteita käyttäen. Sovitinjuosteet koostuvat polyA-häntäosasta, jossa on fluore-



soiva merkkiaine, ja levyllä oleva sovittimen vastinkappale on kovalenttisesti levyyn kiinnitetty polyT-juoste. Ennen sekvensointireaktion aloittamista levyllä kiinnittyneiden näytefragmenttien sijainti tallennetaan optisesti polyA-häntäosien fluoresenssisignaalin mukaan. HeliScope-alustan sekvensointireaktio muistuttaa Illuminan ”reversible dye termination”-menetelmää, jossa DNA-polymeraasi liittyy yksitellen fluoresenssimerkityn ja 3’-päästään estetyn nukleotidin näytetemplaattiin. Illuminan menetelmästä poiketen kaikki koetettavat nukleotidit on leimattu samalla aallonpituudella fluoresoivalla Cy5-värillä. Sovitinjuosteiden merkkiaine Cy3 fluoresoi eri aallonpituudella, joten se ei kuvaudu sekvensoinnin aikana. Kuvantamisen jälkeen fluoresenssimerkki ja 3’-pään este poistetaan kemiallisesti ja levyn huuhtelun jälkeen seuraava nukleotidi tuodaan reaktioon. [42]

Menetelmällä tuotetaan lyhyillä lukujaksoilla (27–50 emästä) satoja miljoonia lukuja per sekvensointiajo. Näin laaja mittakaava tosin tarkoittaa, että kaikkia nukleotidien lisäyksiä ei havaita, ja varsinkin emästen toistojaksojen tunnistaminen on heikkoa. Tämä altistaa deletiovirheille, joiden esiintyvyys lähteestä riippuen on 7–15 %. [34, 42]

#### **2.4.2. PacBio RS II – reaaliaikainen sekvensointi**

Pacific Biosciences toi vuonna 2010 markkinoille ensimmäisen aidosti reaaliaikaisen sekvensointimenetelmän, joka perustuu yksittäisten DNA-polymeraasientsyymien toiminnan seuraamiseen. PacBio RS II-alustassa sekvensointia varten räätälöity DNA-polymeraasi on kiinnitetty ”zero-mode waveguide”-reaktiokammion pohjaan. Kyseinen rakenne on nanometrikoon aallonjohdin eli eräänlainen neulansilmäapertuuri, joka tarkkailee hyvin kapeaa aluetta polymeraasientsyymien kohdalla. DNA-näytteen emäsjärjestys selvitetään käyttämällä neljällä värillä fluoresoivia vapaita nukleotideja, joiden fluoresenssimerkki on kiinnitettyinä nukleotidin 5’-fosfaattiryhmään. Sekvensoinnin aikana DNA-näyte etenee polymeraasientsyymiä pitkin reaktiokammion pohjassa. Polymeraasientsyymi liittyy nukleotidin kerrallaan osaksi kasvavaa komplementtijuostetta, jolloin laservalossa ilmenevä emäksen ominaisväri tallennetaan videolle. Nukleotidien välille katalysoitua fosfodiesterisidos saa fluoresenssimerkin irtoamaan ja poistumaan aallonjohtimen tunnistusalueelta. Tämä mahdollistaa kaikkien neljällä värillä fluoresoivien nukleotidin yhtäaikaisen käytön reaktiokammiossa. Polymeraasientsyymien toimintaa ei hidasteta tai pysäytetä sekvensoinnin aikana eli kyseessä on aidosti reaaliaikainen sekvensointimenetelmä. [43]

Sekvensointialusta käyttää nk. SMRT-kennoja, joissa on useita tuhansia zeptolitrakokoisia ( $10^{-21}$  l) reaktiokammioita vierekkäin. DNA-näytekirjastoja ei rikasteta klonaalisesti vaan näytteet pilkotaan natiivimuodossaan ja muokataan päistään silmukoiduiksi SMRT-bell-templaateiksi [44]. Silmukoituja näytteitä voidaan uudelleensekvensoida useita kertoja peräkkäin, mikä lieventää alustalle ominaista indel-virhealttiutta (11–15 %). [30, 34, 41] PacBio RS II-alustan sekvensointiajot ovat lyhyitä, n. 4 tuntia kerrallaan, ja näytteiden kertalukupituudet ovat suuret (8 500–14 000 emästä). Yksi SMRT-kenno tuottaa ajon aikana n. 50 000–60 000 lukua, ja laitteisto voi käyttää yhtä aikaa 16 eri kennoa. Nopeudestaan ja pitkistä lukupituuksista huolimatta alustan käyttökustannukset ovat vielä suuret.

### 2.4.3. Ion Torrent – puolijohdesekvensointi

Life Technologies julkaisi vuonna 2010 ensimmäisen Ion Torrent-puolijohdesekvensointia hyödyntävän alustan, Personal Genome Machinen (PGM). Siinä missä muut aikaisemmat sekvensointialustat käyttävät optiikkaa sekvensointireaktion havainnointiin, tunnistaa PGM polymeerasireaktiossa vapautuvia vetyioneja eli protoneja ( $H^+$ ) [45]. Menetelmä muistuttaa sekä näytekirjaston valmistelultaan, että DNA-synteesin sivutuotteiden havainnoinniltaan Rochen 454-pyrosekvensointia. Näytekirjasto valmistetaan emulsio-PCR-menetelmällä [33] ja näyterikastetut helmet sijoitetaan CMOS-piirilevyllä oleviin miljooniin pieniin reaktorikennoihin. Jokaisen kennon pohjalla on ioniherkkä IS-FET (ion-sensitive field-effect transistor) -komponentti, joka muuttaa kennonsisäisen ionipitoisuuden vaihtelun sähköiseksi signaaliksi. Sekvensointi tapahtuu vaiheittain syöttämällä reaktioon yksi nukleotidi kerrallaan ja tallentamalla kennokohtainen signaalinmuutos.

Alusta ei vaadi ylimääräisiä, kalliita reagensseja eikä laitteiston hintaa nosta korkeatarkkuuksinen optiikka. Menetelmä on myös nopea, sillä laitteistosta riippuen sekvensointiajo kestää vain 2–8 tuntia kerrallaan. Sekvensointialustan lisäominaisuutena on sen skaalautuvuus; puolijohdepiirejä vaihtamalla voi valita laitteen käyttökohteen lyhyistä uudelleensekvensointiprojekteista koko genomien sekvensointiin. PGM tuottaa n. 400 emäksen fragmenttipituuksilla yhden miljardin verran lukuja per sekvensointiajo, kun taas 2012 julkaistu Ion Proton pystyy n. 10 miljardin lukuihin fragmenttipituuksien ollessa puolet lyhyempiä eli n. 200 emästä. Menetelmän tunnettuna virhelähteenä on pyrosekvensoinnistakin tuttu emästen peräkkäisjaksojen heikentynyt tunnistaminen. Tällöin polymeerasiin liittämien emästen lukumäärä ja kennokohtainen signaalin voimakkuus eivät täysin skaalaudu. Tämä johtaa n. 1 %:n indel-virheisiin. [30, 34, 41]

#### 2.4.4. Oxford Nanopore MinION – nanohuokossekvensointi

Nanohuokossekvensointi (nanopore sequencing) on viimeaikaisin kaupalliseen käyttöön ennättänyt sekvensointitekniologia. Menetelmä perustuu DNA- tai RNA-molekyylien kuljettamiseen ahtaan, nanometrikokoisen huokosrakenteen läpi elektroforeesin avulla. Nanohuokonen jakaa sekvensointi-reaktorikammion kahteen osaan, toimien näiden välisenä porttina ja mahdollistaen sähköisen potentiaalieron muodostumisen kammioiden välille. Huokosessa kulkeutuessaan DNA- tai RNA-juoste aiheuttaa emäs kerrallaan pieniä muutoksia kammioiden välisessä ionivirtauksessa, mikä havainnoidaan sähköisenä signaalina. Teoriassa tämä mahdollistaa pitkien nukleotidijuosteiden emäsjärjestyksen selvittämisen nopeasti, ilman kallista optista laitteistoa ja näytteiden rikastamista PCR:lla. Toimintaperiaatteeltaan tekniologia on karkeasti jaettuna kahteen luokkaan; biologisia transmembraaniproteiineja (poriineja) sekä puolijohdehuokosia (solid-state pore) käyttäviin alustoihin. Edellä mainituista toistaiseksi vain biologista alustaa nähdään kaupallisessa käytössä. [46]

Nanohuokosteknologian käyttö ennustettiin jo 1990-luvulla [47]. Tuolloin osoitettiin *Staphylococcus aureuksen*  $\alpha$ -hemolysiiniproteiinin ( $\alpha$ -HL) kyky kuljettaa yksittäisiä yksijuosteisia DNA-molekyyliä solukalvojen läpi, tosin ilman mahdollisuutta selvittää näytteiden emäsjärjestyksestä. Sekvensointikäyttöön nanohuokosmenetelmä alkoi soveltua vasta vuodesta 2012 lähtien, jolloin kehitettiin menetelmä hidastaa nukleotidijuosteiden kulkua huokosen läpi. Ilman hidastavaa tekijää DNA liikkuu nanohuokosessa noin yksi emäs mikrosekunnissa, mikä on liian nopeaa havainnoitavaksi. Käyttämällä  $\phi$ 29-bakteriofagin (*phi-29*) DNA-polymeraasia nopeudenrajoittimena hidastuu DNA:n kulku murto-osaan normaalista, ja täten emäsjärjestyksen selvittäminen on mahdollista. [48] Toinen laajalti tutkittu biologinen nanohuokonen on *Mycobacterium smegmatiksen* poriini A (MspA). Sen toiminta on samankaltainen  $\alpha$ -hemolysiiniin verrattuna, mutta proteiinin nanohuokonen on vieläkin ahtaampi, mikä teoriassa lisää emästen erottelutarkkuutta.

Vielä tuotekehitysasteella olevalla puolijohdehuokosteknologialla on useita haasteita ratkaistavana. Grafeenista, pii-, alumiini-, boori- ja molybdeeniyhdisteistä pystytään jo nyt valmistamaan ultraohuita nanohuokoskalvoja, jotka teoriassa sopivat biomolekyylien translokaation (i. siirtymän) tarkkailuun. Menetelmillä tosin esiintyy tiettyjä ominaisheikkouksia, kuten esimerkiksi DNA:n taipumusta takertua grafeenihuokoseen, piyhdistehuokosten rappeutumista elektrolyyttiliuoksessa sekä liian korkea signaali-kohinasuhde verrattuna biologisiin nanohuokosiin. DNA:n siirtymäaika puolijohdehuokosen läpi on myös vielä liian nopeaa luotettavan sekvensointituloksen saamiseksi. [46, 49]

2014 rajoitettuun ennakkokäyttöön julkaistu Oxford Nanopore Technologies:n MinION on taskukokoinen, biologista nanohuokosteknologiaa käyttävä sekvensointialusta. Tietokoneen USB-porttiin liitettävä laite hyödyntää kertakäyttöisiä sekvensointikennoja, joissa on 512 kappaletta neljän nanohuokosen virtauskanavaa. Valmistusteknisistä syistä kaikki kennon nanohuokokset eivät ole aktiivisia, vaan keskimäärin 60 % kaikista 2 048 nanohuokosesta soveltuu sekvensointiin. Alustan erityispiirteinä ovat sen huomattavan suuret lukupituudet (10 000–60 000 emästä) ja reaaliaikainen sekvensoinnin analysointi ja säätely. 48 tunnin sekvensointiajon teoreettinen maksimitulos on n. 10 miljardia emäslukua, mikä ei aivan riitä genomitason tutkimuksiin. [50] Suuremman skaalan sekvensointiin Oxford Nanopore Technologies tarjoaa ennakkokokeiluun GridION- ja PromethION-alustoja. [51, 52]

Alustan näytekirjasto luodaan pilkkomalla kaksijuosteinen DNA-näyte suuriksi fragmenteiksi, joiden molempiin päihin liitetään erityiset moottoriproteiinin sisältävät nukleotidisovittimet. Sekvensoinnin aloittava, Y:n muotoinen nk. johtava sovitin (leader adapter) sisältää entsyymaattisen moottoriproteiinin ja reaktorikammiossa sijaitsevan kiinnityssovittimen vastakappaleen. Toinen sovittimista on hiusneulasilmukaksi laskostettu nukleotidijuoste, joka sisältää erityisen ”hairpin”-proteiinin. Valmis näyteseos pipetoidaan elektrolyyttiliuoksella kyllästettyihin reaktorikammioihin, ja näytteiden Y-sovitinpäät hakeutuvat nanohuokosten äärelle sijoitettuihin vastakappaleisiin.

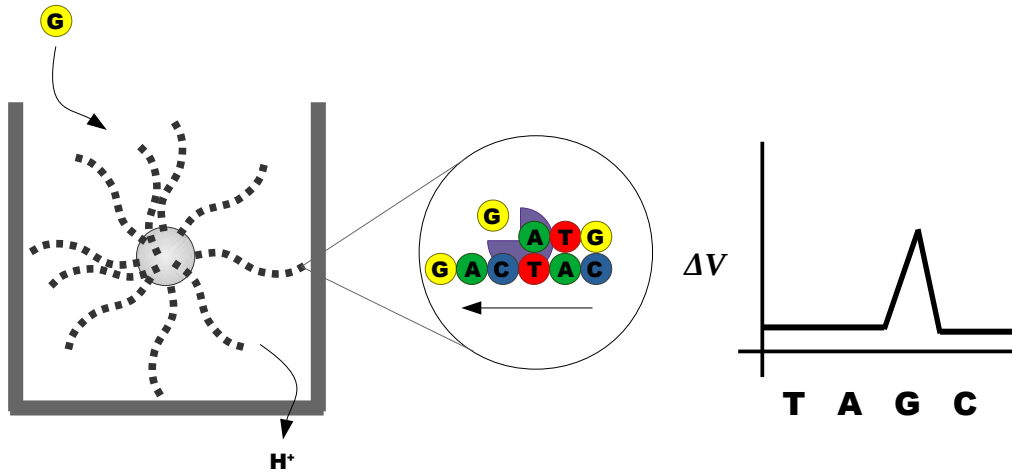
Sekvensointi aloitetaan kytkemällä sähkövirta kammioiden välille, jonka seurauksena Y-sovittimen yksijuosteinen 5'-pää ohjautuu nanohuokoseen. Sovittimen moottoriproteiini alkaa purkaa DNA-näytteen kaksoisjuostetta yksi emäs kerrallaan, jolloin näytetemplaatti kulkeutuu ajan mittaan nanohuokosen läpi. Näytteen ajautuessa hiusneularakenteeseen asti aukaisee toisen sovittimen proteiini silmukan, jonka jälkeen näytteen komplementtijuoste pääsee kulkeutumaan nanohuokosessa. Sekvensoinnin aikana saatava signaalivirta tallennetaan 3 000 Hz:n taajuudella ja näytteen emäsjärjestys selvitetään 3–6 nukleotidin pituisina ”k-mer”-jaksoina. Jaksot tunnistetaan käyttämällä tilastomalleja aiemmin tallennetuista, tunnetuista k-mer-signaaleista. Pelkästään templaattijuosteen sekvensointia kutsutaan kirjallisuudessa ”1D”-sekvensoinniksi, kun taas sekä templaatti- että komplementtijuosteiden peräkkäistä sekvensointia kutsutaan ”2D”-sekvensoinniksi. [53]

Sekvensointimenetelmän tuoreus näkyy sen varsin korkeana virhealttiutena, ja kirjallisuudessa alustan indel-virheet sekä virheelliset emäsmääritykset vaihtelevat (10–20 %). [46, 49, 53] Sekvensoin-

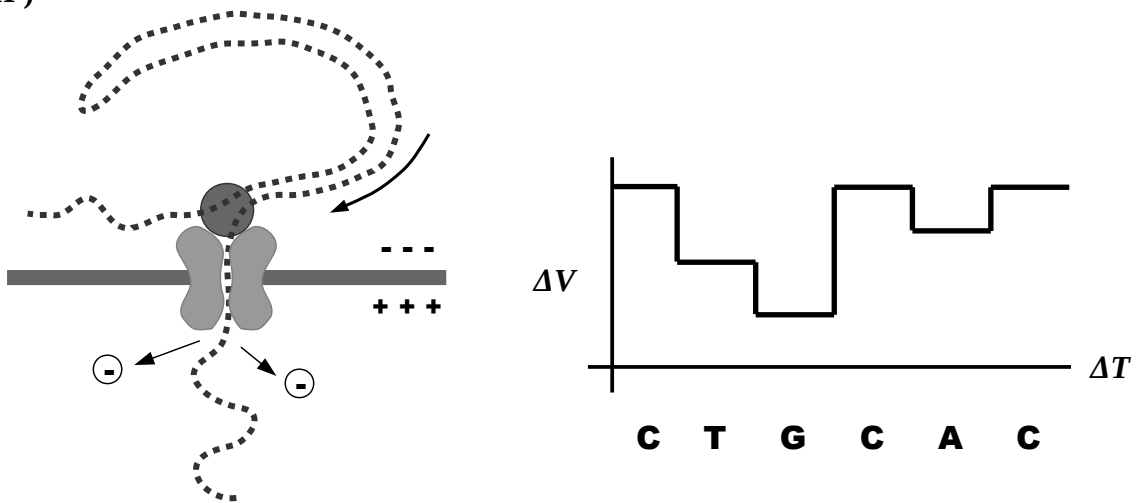
tikemian ja ohjelmistojen kehittymisen myötä virhetaajuudet ovat laskeneet (1D = 14,5 %, 2D = 7,5 %). [54]

### Kuvaaja 1. Kolmannen sukupolven sekvensointimenetelmiä yksinkertaistettuna

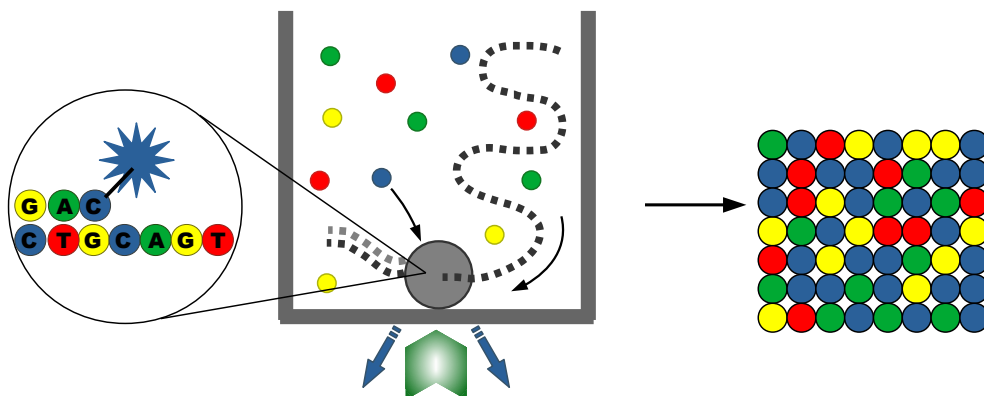
I)



II)



III)



- I )** Ion Torrent PGM-alustan toiminta yksinkertaistettuna. Vapaan emäksen lisäys näytehelmen sisältävään reaktorikammioon vapauttaa protoneja ( $H^+$ ) polymeerasireaktion seurauksena. Reaktorikammion ioniherkkä transistori muuttaa reaktioliuoksen pH-muutoksen sähkösignaaliksi. Positiivinen signaali tulkitaan syklin sen hetkiseksi emäkseksi, kuvan tapauksessa guaniiniksi.
- II)** Oxford Nanopore Technologies:n MinION-alustassa kaksijuosteinen DNA-näyte purkautuu asteittain moottoriproteiinin katalysoimana. DNA-juoste kulkeutuu nanohuokosessa ja osittain tukkii reaktorikammioista toiseen kulkeutuvan ionivuon. Jännitemuutokset tulkitaan emäsjärjestykseksi.
- III )** PacBio RS II-alustan toiminta perustuu tauottomaan polymeerasientsyymien seurantaan. Reaktorikammioon kohdistettu laser (kuvassa vihreä nuoli) saa aikaan sen hetkisen, entsyymissä kulkevan emäksen väriaineen mukaisen värisignaalin. Signaali videoidaan ja videon perusteella määritetään DNA-juosteen emäsjärjestys.

## 2.5. Uuden sukupolven sekvensoinnin käyttöaiheet

Modernit sekvensointialustat soveltuvat teoriassa, ja usein myös käytännössä kaikkeen mahdolliseen geenitutkimukseen joko yhdistettynä muihin menetelmiin tai pelkästään tehosekvensoinnin kautta. Bioinformatiikkatyökaluilla on nykyisin mahdollista visualisoida tutkittavan DNA:n rakennetta alkaen yksittäisistä emäsmuutoksista ja laajentuen kromosomien lukumäärä- ja rakenneanalyysiin. Genomisen DNA:n lisäksi sekvensoitavaksi soveltuvat myös muun muassa yksilön epigenomi ja transkriptomi. Epigenomin sekvensoinnissa selvitetään geeniekspressiota säätelevien DNA:n ja histonien metylaation astetta kohdesoluissa. Bisulfiitti- ja ChiP-sekvensointien (chromatin immunoprecipitation) avulla onkin löydetty useita syöpäsairauksien diagnosointiin, käyttäytymiseen ja hoitovasteeseen liittyviä onkogenejä ja geenikandidaatteja. Transkriptomi sen sijaan kuvaa kohdesolun näytteenottohetkellä tapahtunutta geenien ekspressiota sekvensoimalla kaikki solunsääinen RNA. Komplementaariseksi DNA:ksi (cDNA) muunnetut mRNA ja miRNA ovat olleet tutkimuksen kohteina varsinkin lääkeaineiden ja syöpähoitojen vasteita arvioidessa. Edellä mainitut tutkimuskohteet ovat vielä valtaosin kliinisen käytön ulkopuolella laajempien kokeiden ja tutkimusten puuttuessa. [55] On kuitenkin hyödyllistä erottaa toisistaan ensisijaisesti tutkimuskäyttöön soveltuvat ja diagnostiset menetelmät, joita molempia tässä osiossa tarkastellaan.

22.5.2019 tehdyn internethaun perusteella julkisessa terveydenhuollossa NGS-menetelmällä suoritettavia tutkimuksia tarjoavat suurien keskussairaaloiden yhteydessä toimivat laboratoriot, kuten esim. HUSLAB, Fimlab, ISLAB, TYKS Laboratorio sekä Nordlab. Hakuun sisällytettiin laboratoriorien ohjekirjoissa luetellut ”NGS”-hakusanalla löytyvät tutkimukset.

### 2.5.1. Kohdennettu sekvensointi

Uuden sukupolven sekvensointimenetelmät soveltuvat erityisesti kohdennettuun sekvensointiin sekä koko genomien sekvensointiin. Kohdennetussa sekvensoinnissa potilaan perimästä eristetään etukäteen valitut ehdokasgeenit tai geeniekspressioon vaikuttavat alueet, joita halutaan tutkia. Valitut DNA-alueet tai RNA-tyypit sekvensoidaan kaikki yhdellä kertaa, ja kun analysoitava alue on tarkoin mitoitettu, kasvaa sen lukupeitto ja sitä mukaa tulosten luotettavuus. Menetelmä on sekä ajallisesti että kustannuksellisesti tehokasta, kun sekvensointiin kuluvat resurssit riippuvat tutkittavien nukleotidijaksojen määrästä sekä alueen kaappaukseen räätälöidyn välineistön spesifisyydestä. Saadut tulokset ovat myös verrattain yksinkertaiset tulkita jopa vanhempia, käytössä vakiintuneita diagnoosityökaluja käyttäen. Menetelmän hyötyjen vuoksi kohdennettu sekvensointi onkin ensimmäinen kliiniseen käyttöön soveltuva NGS-menetelmä. Tarkoin mitoitettu ja joustamaton paneelitutkimus tosin asettaa menetelmälle rajoitteita. Tutkimus voi jäädä puutteelliseksi, mikäli tutkittavalla alueella on suuria rakennepoikkeamia, geeninsiirtymiä tai kopiolukuvaihteluita (copy number variation, CNV), joita ei ole voitu ennustaa. Lisäksi tutkimuksen tilaajalla tulisi olla jo valmiiksi rajattuna diagnoosivaihtoehdot, joita lähdetään selvittämään. Väärin mitoitettu tai epäselviä tuloksia tuottanut paneelitutkimus voi usein johtaa uusiin, laajempiin jatkotesteihin, jolloin kohdennetun sekvensoinnin tuoma aika- ja resurssihyöty menetetään. [56-57]

Kohdennettu sekvensointi soveltuu erityisesti vahvan genotyyppi-fenotyyppi-yhteyden omaavien geenivarianttien tunnistamiseksi. Näitä ovat mm. yhteen geeniin sidonnaiset (monogeeniset), muista sairauksista selvästi poikkeavat (heterogeeniset), Mendeliaanisesti periytyvät ja matalan alleelifrekvenssin perinnölliset sairaudet. Paneelitutkimukseen soveltuvat geenialueet valikoidaan jo aiemmissä tutkimuksissa tehtyjen löydösten perusteella, eli uusia ja tuntemattomia sairauksia ei menetelmällä ole tarkoituskaan diagnosoida. Tutkimustiedon lisääntyessä myös geenipaneelien kohteet lisääntyvät, mikä tosin lisää diagnoosivaihtoehtojen lisäksi tulosten tulkinnan epävarmuustekijöitä. [58] Nykyisin monet kansainvälisesti toimivat laboratoriot tarjoavat eri kokoisia geenipaneelitutkimuksia mm. terveys-, väestöhistoria- ja lifestyle-tarkoituksiin [kts. kappale 2.6].

### 2.5.2. Eksomisekvensointi

Koko eksomin sekvensoinnissa potilaan genomisesta DNA:sta eristetään proteiinisynteesiin keskeisesti liittyvät eksonialueet, eli se on käytännössä yksi kohdennetun sekvensoinnin sovelluksista. Ek-

somitutkimuksessa saadaan tietoa mahdollisista proteiinisynteesiin vaikuttavista geenivarianteista, ja koska eksonit kattavat vain n. 1-2 % koko genomista on se tutkimusmenetelmänä n. 90 % halvempi ja nopeampi vaihtoehto koko genomien sekvensoimiselle [56]. Kohdennetun sekvensoinnin vuoksi menetelmällä ei voi todeta eksoneja ympäröivien DNA-rakenteiden kuten intronien, promootorialueiden tai RNA:n silmukoimisesta vastaavien ”splicing”-alueiden rakennepoikkeamia. Lisäksi eksonien kaappaukseen käytettävän reagenssivälineistön valinta vaikuttaa tutkimustuloksiin, jolloin eksomiin voi jäädä aukkoja tai heikon lukusyvyyden omaavia toistolukujaksoja. Toistaiseksi käytettävissä ei ole reagenssipakettia, joka kattaisi täysin kaikki geenialueet riittävällä lukusyvyydellä. [57-58] Eksomisekvensointi voi myös paljastaa potilaan terveyteen vaikuttavia, vielä oireettomia sivulöydöksiä, geenivirheiden kantajuuksia tai geenivariantteja, joille ei vielä voi osoittaa terveysvaikutuksia. Tämän vuoksi eksomisekvensointiin voi liittyä myös eettisiä ongelmakohtia [kts. kappale 4.2].

Kliinisessä käytössä eksomisekvensointi soveltuu diagnosoimaan epävarmojen sairauksien tunnistamiseen, harvinaisten perinnöllisten sairauksien diagnostiikkaan sekä syöpäkudoksen geneettiseen profilointiin. [55,57] Yksittäisten geenivirheiden diagnosoimiseksi menetelmä on liian laaja ja työläs, mutta muutoin käyttökelpoinen niissä tapauksissa, joissa millään muulla perinteisellä menetelmällä ei diagnosia ole saatu selville. Esimerkiksi pediatriassa lasten ja nuorten lisäksi voidaan yleensä sekvensoida molempien vanhempien eksomit, jolloin nk. triotutkimuksessa selvitetään, onko lapsella vanhemmilta peritty geenivirhe tai *de novo* -mutaatio. Menetelmää voidaan käyttää mm. kehityshäiriöiden, -viivästymien ja epäselvien oireyhtymien geneettisen syyn selvittämiseksi. [59] Eksomisekvensoinnista saadun tiedon tulkitseminen ei ole mahdollista ilman kattavien geenivariaatitietokantojen hyödyntämistä. Näihin lukeutuvat myös populaatiokohtaiset normaalivariaatioiden kokoelmat, joita käytetään harmittomiksi oletettujen variaatioiden suodattamiseksi. Suuren datamäärän analysointi vaatii bioinformatiikkatyökalujen ja tehokkaiden tietokoneiden käyttöä [kts. luku 3].

Tutkimuskäytössä eksomisekvensointi on huomattavasti houkuttelevampi vaihtoehto koko genomien sekvensointiin verrattuna, ja se tuottaakin jatkuvasti uusia löydöksiä sekä Mendeliaanisesti periytyvien että monitekijäisten sairauksien geneettisestä taustasta. Uusien löydösten lisäksi monet jo aiemmin sairauksiin liitetyt geenit joudutaan ehkä uudelleenluokittelemaan sitä mukaa kun uusia geneettisiä yhteyksiä löydetään. [60]



### 2.5.3. Koko genomien sekvensointi

Koko genomien sekvensointi (whole genome sequencing, WGS) on tutkimuksena hypoteesiton; tiettyjen muutosten etsimisen sijaan voidaan vapaasti tutkia, mikä tutkittavan geneettisessä profiilissa on erilaista referenssiin verrattuna. Menetelmä on eksomisekvensointia herkempi genomien rakennemuutosten, kuten geenien translokaatioiden (siirtyminen paikasta toiseen), insertioiden, poistumien sekä geenienvälisten rakenteiden poikkeamien tunnistamiseksi, minkä lisäksi se kattaa myös eksomisekvensoinnillakin löydettävät variantit. [57] Nimestään huolimatta menetelmä ei kata aivan koko perimää, vaan sekvenssidata usein sisältää matalan lukupeiton aiheuttamia katkoksia ja sekvensointialustasta riippuvia ominaisvirheitä. [41] WGS-menetelmälle aiemmin ominaista toistojaksojen, kuten kopiolumuutosten ja lyhyiden nukleotiditoistumien, heikkoa tunnistamista on pystytty lievittämään lisäämällä näytefragmenttien pituutta ja muuttamalla sekvensoinnin analyysialgoritmeja. [61] Menetelmä on selvästi eniten aikaa ja resursseja vaativin tutkimusvaihtoehto, sillä tarpeeksi luotettavan tuloksen saamiseksi on tutkittavan perimän n. 3,2 miljardia emästä sekvensoitava yli 30-kertaisella lukusyvytydellä. Saatujen jopa satojen gigatavujen kokoisten tiedostojen analysointi vaatii eksomisekvensoinnin tapaan tehokkaita tietoteknisiä työkaluja. Genomista pystytään löytämään miljoonia yksilökohtaisia eroja referenssikvansseihin verrattuna [62], ja suurena haasteena onkin tutkittavan terveyteen vaikuttavien geenimuutosten löytäminen normaalivariaation seasta. Tulosten tulkintaa vaikeuttavat myös lukuisat monitekijäisten sairauksien ilmenemistä sekoittavat tekijät, joita luetellaan taulukossa 1 [s. 26]. Riskinä on myös löytää eksomisekvensoinnin tapaan oireettomia, satunnaisia mutaatioita tai korkean riskin periytyviä sairauksia, joista tutkittava ei ole ollut tietoinen.

Sekvensoinnin tuottaman valtavan datamäärän analyysiin käytettävän ajan ja eksomisekvensointiin verrattuna vaatimattoman lisähyödyn vuoksi ei WGS:n laajempi kliininen käyttö ole vielä realistista. Rajatuissa tapauksissa menetelmä voi kuitenkin paljastaa sellaisia geeniekspressioon vaikuttavia mutaatioita tai rakennepoikkeamia, joita eksomisekvensoinnilla ei voida löytää. Lääketieteellisessä tutkimuksessa koko genomien tarkka läpikäyminen voi mm. auttaa löytämään uusia kohteita lääkeshoidoille, tehostaa syöpähoitoja ja paljastaa uusia perinnöllisiä tekijöitä yleisten sairauksien taustalta. [55-57]

**Taulukko 1. Esimerkkejä geenitestien tulkintaa vaikeuttavista tekijöistä**

Fenokopio	Yksilöllä on sairaus, jonka genotyyppi on kuitenkin erilainen verrattuna muihin saman sairauden kantajiin
Vähentynyt penetranssi	Sairaudelle altistavan geenimuutoksen kantajat eivät kaikki sairastu
Taudille altistavien alleelien yleisyys	Sairauden kantajuus on populaatiossa yleistä, jolloin sairauteen liittyvien alleelien merkitystä on vaikea arvioida
Lokusheterogeenisuus	Sairauden oireiden taustalla vaikuttavat useat eri lokuksissa sijaitsevat geenit
Alleeliheterogeenisuus	Sairaudelle altistuminen riippuu saman geenin eri alleelien eroavaisuudesta
Pleiotropia	Sama geeni vaikuttaa useaan eri ominaisuuteen tai sairauden ilmenemiseen eri kudoksissa
Epistaasi	Sairauden ominaisuudet johtuvat kahden eri geenin alleelien yhteisvaikutuksesta
Sairautta säätelevät mutaatiot	Sairaudelle altistava mutaatio vaikuttaa vain osittain geenin toimintaan
Siteerattu julkaisusta: Perola M. <i>Terveiden ja sairauden genetiikka – monitekijäiset taudit ja ominaisuudet</i> . Kirjassa: Aittomäki K, Moilanen J, Perola M. <i>Lääketieteellinen genetiikka</i> , 1. painos. Helsinki: Kustannus oy Duodecim 2016. ISBN: 978-9516564671.	

## 2.6. Kuluttajamarkkinoiden geenitestit

Uusien geeniteknologioiden edelläkävijämaassa, Yhdysvalloissa, käynnistyi 2010-lukua lähestyttäessä useita henkilökohtaiseen testaukseen erikoistuneita laboratorioita. Nämä yritykset alkoivat markkinoida edullisia geenitestejä suoraan kuluttajille käyttämällä perinteisiä molekyylikaryotyyp-pimenetelmiä ja uuden sukupolven sekvensointia. Tuolloin sylkinäytteestä tarjottiin testattavaksi mm. sairauksien kantajuutta, sairastumisriskin arviointia, lääkeaineiden sopivuutta sekä joidenkin tartuntatautien vastustuskyvyn arviointia. Yritysten toimintamalliksi mainitaan testien markkinoiminen suoraan kuluttajalle, joka sitten pyytää terveystietä ammattilaista tilaamaan ja tulkitsemaan testin, tai sitten asiakas suorittaa koko prosessin itse. [63] Nopeasti kasvanut ala on aiheuttanut huolta asiantuntijoiden ja viranomaisten tahoilla varsinkin geenitiedon huolimattomasta käytöstä ja testien laadusta. Testausmenetelmien akkreditoinnin lisäksi asiantuntevan perinnöllisyysneuvonnan saata-vuus tai puute ovat olleet keskeisinä puheenaiheina kirjallisuudessa. Puutteellista tai jopa virheellistä terveystietoa saava asiakas voi ajautua epätoivon partaalle, varsinkin jos tiedon puutteessa tulok-sia tulkitaan väärin. DTC-yritysten puolestapuhujien mukaan kuluttajien lisääntyneet vaihtoehdot terveystiedon hankinnasta osoittavat positiivista kehitystä yksilöiden voimaannuttamisen ja autonomin kannalta. Kriitikot sen sijaan näkevät geenitestaukset ilman asianmukaista neuvontaa jopa hai-tallisena ilmiönä, varsinkin kun tuloksista huolestunut asiakas ajautuu vääjäämättä julkisen terveydenhuollon piiriin. [64-65]

Vuonna 2010 Yhdysvaltain elintarvike- ja lääkeainevirasto FDA (Food and Drug Administration) määräsi kaupalliset genomitestit luokiteltavaksi lääkinnällisten laitteiden joukkoon, mikä vaatii asianmukaiset käyttöoikeudet ennen markkinointia. Seuraavien neljän vuoden aikana lähes kaikki seitsemästätoista geenitestejä markkinoivasta yrityksestä sulki palvelunsa, osa poistuen kokonaan markkinoilta. Jäljelle jääneet yritykset supistivat testitarjontansa vanhemmuus- ja sukututki-muksiin tai kehittyivät edelleen suuremmiksi bioteknologiayrityksiksi. [66] Marraskuussa 2013 FDA jälleen kielsi internetin välityksellä geenitestejä myyvää 23andMe-bioteknologiayritystä markkinoimasta henkilökohtaisia genomipalveluita. Kiellon taustalla oli FDA:n kannalta puutteellinen näyttö yrityksen tarjoamien palveluiden diagnostisesta hyödystä ja jo aiemmin geenitesteille asetettu luokitus lääkinnällisistä laitteista. Huhtikuussa 2017 FDA kuitenkin hyväksyi 23andMe:n Genetic Health Risk (GHR)-testit kymmenen tunnetun sairauden diagnosoimiseksi kuluttajille [67]. 23andMe on tähän mennessä ainoa DTC-geenitestejä tarjoava yritys, jolla on FDA:n myöntämä diagnosti-  
stisten geenitestien markkinointilupa.

Nykyisin valtaosa DTC-yrityksistä toimivat kansainvälisesti. Vuonna 2016 julkaistussa tutkimuksessa löydettiin 246 DNA-testejä myyvää yritystä, jotka markkinoivat terveys- ja väestöhistoriatiesten lisäksi mm. ravitsemukseen, urheilukykyyn, lasten lahjakkuuteen ja vanhemmuuteen liittyvää testausta. [68] Monella edeltä mainituista testivaihtoehdoista ei ole riittävää tieteellistä näyttöä, ja yrityksiä on syytetty liiallisten lupauksen markkinoinnista. Väestöhistoriatiesten suosio kuitenkin on voimakkaassa nousussa. Vuonna 2017 arvioitiin kuluttajien teettäneen yli 12 miljoonaa DNA-testiä, joista suurin osa tehtiin USA:ssa. [69] Vuoden 2019 alussa luku oli jo yli kaksinkertaistunut yli 26 miljoonaan testiin. [70] Suurimpina vaikuttajina kuluttajamarkkinoilla toimivat mm. yhdysvaltalaiset AncestryDNA, 23andMe, Gene by Gene sekä israelilainen MyHeritage. Suomalainen diagnostisia geenitestejä tarjoava Blueprint Genetics toimii myös maailmanlaajuisesti, ja on akkreditoitu Suomen lisäksi Yhdysvalloissa CAP- ja CLIA-kriteerein. Muun muassa HUSLAB teettää osan geenipaneelitutkimuksistaan yrityksen kautta alihankintana. [71-72]

Euroopassa DTC-testausta koskevat lait ja säädökset mainitaan hajanaisiksi [73]. Katsauksessa kerrotaan, että Euroopassa geenitestaus on tähän mennessä aina yhdistetty terveydenhuoltoon ja siten ollut jäsenvaltion oman lainsäädännön alaista. Geenitestaus ulkoisena palveluna onkin uusi ilmiö, joka periaatteessa on jäsenvaltion oman biolääketiedettä, bioetiikkaa tai geenitestejä koskevan lainsäädännön alaista. Terveydenhuollon ulkopuolinen, mutta terveystietoa käsittelevä DNA-testaus on mm. Saksassa ja Ranskassa lailla kiellettyä. Perinnöllisyysneuvonnan pakollisuudesta geenitestiin yhteydessä on myös valtiokohtaisia eroja. Suomessa ja 15 muussa EU-jäsenvaltiossa geenitesteihin liittyvästä neuvonnasta säädetään Oviedo-yleissopimuksen mukaisesti, ja vaikka Suomen lakiin ei artiklan 12 mukaisesti ole tehty erillistä perinnöllisyysneuvontaa koskevaa säädöstä, ajavat muut kansainväliset sopimukset saman asian. Vuonna 2022 voimaan astuva IVD-asetus tulee oletettavasti muuttamaan geenitestiin tarjontaa EU:ssa [kts. kappale 4.1.1].

Suomessa jo vuonna 2011 julkaistussa kannanotossaan [74] valtakunnallinen sosiaali- ja terveysalan eettinen neuvottelukunta ETENE neuvoo kuluttajan geenitestejä harkitsevia pohtimaan tarkoin mihin on ryhtymässä. Jokaisen testejä tilaavan tulisi ymmärtää saatavan tiedon laadun, luotettavuuden ja tulkinta-avun vaihtelun mahdollisuus. Pelkän elämänhallinnan parantamisen vuoksi ei kaupallisiin geenitesteihin suositella ryhtyttävän, eikä myöskään lasten ja nuorten testaukseen. Suomen itenäisyyden juhlarahasto SITRAn Taloustutkimukselle vuonna 2013 teettämän kyselytutkimuksen [75] mukaan selvä enemmistö (61 %) vastanneista (n=2017) on kiinnostunut selvittämään omia perinnöllisiä riskitekijöitään. 45 % vastanneista oli myös sitä mieltä, että täysi-ikäisten tulisi saada päättää itse geenitestauksesta ja yli puolet (54 %) haluaisivat geenitietonsa hänen itsensä yksinomi-

seen omistukseen. Vuonna 2020 suunnitellusti käynnistyvä genomikeskus ja sitä tukeva lainsäädäntö [kts. kappale 4.1.1] tulevat myös hyvin todennäköisesti vaikuttamaan kuluttajamarkkinoiden geenitestaukseen, ja varsinkin yksilön oikeuteen saada asiantuntevaa neuvontaa genomitiedosta.

### 3. BIOINFORMATIIKAN KEHITYS

#### 3.1. Bioinformatiikan historia lyhyesti

Bioinformatiikka omana tieteenalanaan on varsin tuore ilmiö, vaikkakin sen juuret johtavat 1960-luvulle saakka. Siinä missä tietotekniikkaa on menestyksekkäästi käytetty matemaattisten tieteiden työkaluna, on biologi joutunut ensin jollain tavalla löytämään laskettavaksi soveltuvia muuttujia. Vasta molekyyliatasolla toimivat biologiset mallit ovat soveltuneet tietokoneiden laskettavaksi, ja tuolloinkin tutkija on usein joutunut luomaan omat työkalunsa.

Vuonna 1962 alun perin proteiinien aminohappojärjestyksiä laskemaan luotua COMPROTEIN-ohjelmaa [76] pidetään ensimmäisenä *de novo*-”assembler” eli kokoamisohjelmana. Reikäkorteille koottu ja FORTRAN-kieleen perustuva ohjelma kokoaa lyhyistä aminohapposekvensseistä suurempia, yhtenäisiä kokonaisuuksia vertailemalla sekvenssien päällekkäisyyksiä. Samaa periaatetta noudattavat lukuisat myöhemmin käyttöön otetut sekvenssiohjelmistot. Aminohapposekvenssien selvitys loi alustan uudelle tavalle selittää biologisia malleja. Lajien sukulaisuussuhteita vertaileva paleogenetiikka sai alkunsa proteiiniperheiden yhteneväisyystutkimuksista, ja sen tavoitteena on rakentaa lajien evoluutiolle sukupuu [77]. Työlääksi osoittautunut aminohapposekvenssien käsin vertailu johti ajan mittaan linjaus- ja yhtäläisyyslaskentaan räätälöityjen tietokonealgoritmien luomiseen. Vuonna 1970 julkaistuun Needleman-Wunsch -”aligner-” eli rinnastusalgoritmiin perustuva ja useita sekvenssejä yhtä aikaa vertaileva CLUSTAL on yhä käyttökelpoinen rinnastustyökalu [78], jonka ensimmäinen versio julkaistiin vuonna 1988. Ohjelma soveltuu proteiini- ja nukleotidisekvenssien vertailuun, ja sen uusimmat versiot toimivat myös internet-selaimilla.

Molekyylibiologia koki tieteenalana paradigmanmuutoksen 1960-luvun lopulla, kun DNA:n tarkka rooli geneettisen tiedon siirrossa ratkaistiin. Francis Crickin hypoteesin pohjalta, jossa DNA:sta siirtokopioitu RNA ohjelmoi proteiinien aminohappojärjestyksen, selvitettiin DNA-emäskolmikkojen eli kodonien geneettinen koodi [79]. Tiedon myötä pystyttiin siirtymään entistä tarkempaan ja yksityiskohtaisempaan sekvenssiojennimenetelmään eli DNA-sekvenssointiin. Allan Maxamin ja Walter Gilbertin ja sittemmin Frederick Sangerin kollegoineen kehittämät yksittäisiä nukleotideja manipuloivat sekvenssiojennimenetelmät loivat myös tarpeen tietotekniselle avulle. Sekvenssoinnista saatujen tuhansien nukleotidien paikkatiedon tallentaminen, rinnastus ja vertailu eivät olleet enää mahdollista käsityönä, mikä johti ensimmäisten sekvenssointityökalujen kehittämiseen. Sangerin sekvenssoin-

nista saatujen DNA-jaksojen analysointiin, manipulointiin ja kommentoimiseen luotu Stadenin paketti [80] on ensimmäinen DNA-sekvensointiin soveltuva, yhtenäinen tietokoneohjelmisto.

Sekä biokemian että tietotekniikan kehitys kokivat huomattavaa kasvua 1970-luvun päätyttyä, kun ensimmäiset mikrotietokoneet tulivat markkinoille. Vuonna 1984 julkaistu GCG oli ensimmäinen sekvenssianalyysiä varten kehitetty ohjelmistopaketti, joka oli kehitetty DEC VAX-11-keskustietokoneen ajettavaksi [81]. Lyhyessä ajassa julkaistiin myös ensimmäiset työpöytäkäyttöön soveltuvat sekvenssien kokoamis- ja analyysiohjelmit Apple II-, CP/M- ja IBM:n PC-alustoille [82-84].

Kansainvälistä tietojenvaihtoa helpottamaan kehitetyt European Molecular Biology Laboratoryn (EMBL) sekvenssikirjasto [85] ja GenBank® [86] olivat ensimmäisiä tietokonepohjaisia geeni- ja proteiinisekvenssitietokantoja. Alun perin siirrettävällä medially ja lähiverkoissa toimineet tietokoneet kannustivat tutkijoita jakamaan löydöksensä julkisesti jo ennen internet-aikaa. 1980-luvun loppuun mennessä nopeasti lisääntynyt sekvenssidata pakotti tutkijat kehittämään entistä tehokkaampia sekvenssien etsintä- ja vertailutyökaluja. Yhä käytössä olevat FASTA [87] ja BLAST [88] loivat pohjan modernille tietokantapohjaiselle bioinformatiikalle.

1990-lukua kutsutaan genomiikan aikakauden aluksi. Kyseisen vuosikymmenen aikana aloitetut kokonaisten genomien sekvensointiprojektit olivat mahdollisia huippuunsa kehitetyn, automatisoidun Sangerin sekvensoinnin ansiosta. Massiivisten projektien myötä kehitettiin myös lukuisia genomien kokoamiseen soveltuvia ohjelmistoja, mm. Phred-Phrap-Consed-ohjelmistopaketti [89], TIGR [90] ja EULER [91]. Tim Berners-Leen kollegoineen vuonna 1992 käynnistämän kansainvälisen WWW-palvelun [92] myötä sekvensointiprojektien tiedon jakaminen ja julkisten tietokantojen käyttö yksinkertaistuivat. Aiemmin työläiden, komentoriviltä ajettavien BBS-purkkien (bulletin board system, sähköinen ilmoitustaulu) sijaan käyttöön tulivat helppolukuiset, graafiset verkkosivustot. Verkkoresursseiksi julkaistiin muun muassa seuraavat edelleen käytössä olevat sivustot: GenBank (1992), EMBL (1993), NCBI (1994), Genomes (1995), PubMed (1997) ja Human Genome (1999) [93].

2000-luvulla ihmisgenomin julkaisun ja ensimmäisten uuden sukupolven sekvensointitekniikoiden käyttöönoton myötä alkoi myös eksponentiaalinen bioinformatiikkatyökalujen kehitys. Vuoteen 2017 mennessä arvioitiin olleen yli 23 000 bioinformatiikkaohjelmaa ja -tietokantaa, joista n. 82 % on suunniteltu vain yhden tehtävän tai alustan työkaluksi [94]. *Gauthier J ym. (2018)* [93] lisäksi toteavat, että nykyisin sekvensointityökaluja on niin paljon, että on vaikea valita itselleen sopivaa. Kaupallisten, alustakohtaisten ohjelmistojen lisäksi julkisesti on tarjolla avoimen lähdekoodin il-

maisia ohjelmistoja kuten EMBOSS [95]. Taulukossa 2 [s. 32] listataan sekä julkisesti saatavia että kaupallisia NGS-alustoille tarkoitettuja analyysiohjelmistoja.

<b>Taulukko 2. Julkisesti saatavia ja kaupallisia analyysityökaluja NGS-alustoille</b>	
<u>Julkisesti saatavia ohjelmistoja</u>	Osoitteessa
<b>Rinnastustyökaluja</b>	
BWA	<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>
Novoalign	<a href="http://www.novocraft.com">http://www.novocraft.com</a>
Bowtie 2	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
<b>Synnyntäisten varianttien tunnistimia</b>	
GATK (UnifiedGenotyper)	<a href="https://software.broadinstitute.org/gatk/download/">https://software.broadinstitute.org/gatk/download/</a>
SAMtools	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>
Varscan2	<a href="http://varscan.sourceforge.net/">http://varscan.sourceforge.net/</a>
SNVMix	<a href="https://shahlab.ca/projects/snvmix/">https://shahlab.ca/projects/snvmix/</a>
<b>Somaattisten (syöpä-) varianttien tunnistimia</b>	
SomaticSniper	<a href="http://gmt.genome.wustl.edu/packages/somatic-sniper/">http://gmt.genome.wustl.edu/packages/somatic-sniper/</a>
JoinSNVMix2	<a href="https://shahlab.ca/projects/JointSNVMix/">https://shahlab.ca/projects/JointSNVMix/</a>
Mutect	<a href="https://github.com/broadinstitute/mutect">https://github.com/broadinstitute/mutect</a>
<b>Kopiolukumutosten tunnistimia</b>	
ExomeCNV	<a href="http://sourceforge.net/projects/exome-cnv/">http://sourceforge.net/projects/exome-cnv/</a>
CONTRA	<a href="http://sourceforge.net/projects/contra-cnv/">http://sourceforge.net/projects/contra-cnv/</a>
CNVnator	<a href="http://sv.gersteinlab.org/cnvnator/">http://sv.gersteinlab.org/cnvnator/</a>
RDXplorer	<a href="http://sourceforge.net/projects/rdxplorer/">http://sourceforge.net/projects/rdxplorer/</a>
<b>Rakennemuutostyökaluja</b>	
Delly	<a href="https://github.com/dellytools/delly">https://github.com/dellytools/delly</a>
Breakdancer	<a href="http://breakdancer.sourceforge.net/">http://breakdancer.sourceforge.net/</a>
PINDEL	<a href="http://gmt.genome.wustl.edu/packages/pindel/">http://gmt.genome.wustl.edu/packages/pindel/</a>
CREST	<a href="http://www.stjude.com/research/site/lab/zhang">http://www.stjude.com/research/site/lab/zhang</a>
<b>Kaupallisia ohjelmistoja</b>	
Avadis NGS	CLCBIO Genomic Workbench
DNASStar Lasergene	Genomatix
Sequencher	Softgenetics Nextgene
Siteerattu julkaisusta: Oliver GR, Hart SN, Klee EW. <i>Bioinformatics for Clinical Next Generation Sequencing</i> . Clinical Chemistry 2015; 61: 124-135. Linkit päivitetty 9.5.2019.	



### 3.2. Bioinformatiikka uuden sukupolven sekvensoinnissa

Nykyisen bioinformatiikan pullonkaulana on sekvensointitutkimuksista saatava valtava tietomäärä ja sen suhteen rajalliset käsittelyresurssit. ”Big data” eli massadata on yleisnimitys jatkuvasti kasvavalle ja järjestelemättömälle informaatiolle, jonka analysointiin ei voi soveltaa perinteisiä menetelmiä. Modernien sekvensointialustojen tuottama, jopa teratavujen kokoinen tietomäärä sopii hyvin massadatan määritelmään. Arvioidaan, että vuoteen 2025 mennessä sekvensoinnista saatujen geenivarianttien tunnistamiseen tarvitaan n. kaksi triljoonaa prosessorituntia ja että seuraavan vuosikymmenen aikana sekvensoidaan jopa kaksi miljardia ihmisgenomia. Massadatan louhintaan on ehdotettu pilvipalveluiden, prosessoriparviin, näytönohjainprosessorien tai uudelleenohjelmoitavien mikroprosessorien käyttöönottoa [96].

Modernien sekvenssianalyysiohjelmistojen mainitaan toimivat kolmivaiheisesti [97-98]. Ensimmäisessä vaiheessa sekvensointialustan tuottama raakasignaali muunnetaan tunnistettavaksi nukleotidien sarjaksi. Uuden sukupolven sekvensointialustoilla esiintyy ominaisvirheitä, jotka aiheuttavat joko väärin tunnistettuja nukleotideja tai indel-virheitä. Tämän vuoksi DNA-fragmentit sekvensoidaan useaan kertaan, jotta saadun sekvenssin mittava lukusyvyys (30-100x) auttaisi tilastollisessa virheenkoroauksessa [61]. Ensimmäiseen vaiheeseen kuuluu myös tunnistettujen nukleotidien automaattinen laatupesteytys, jota käytetään väärin tunnistettujen emästen tai vastaavasti SNP:ien merkitsemiseksi.

Toisessa vaiheessa sekvensoinnista saadut lyhyet lukujaksot rinnastetaan referenssigenomin mukaisesti, ja sekvenssille suoritetaan algoritminen virheenkoroaus. Referenssigenomi on mosaiikkisesti rakennettu ja tämänhetkisesti virheettömin kuvaus kokonaisesta ihmisgenomista, johon sekvensoinnista saatuja lukujonoja verrataan. Viimeaikaisin referenssigenomi on 20.3.2019 päivitetty GRCh38, joka perustuu Human Genome Projectiin osallistuneiden ihmisten perimään [99]. Jos tutkittava DNA-alue on tuntematonta, voidaan rinnastuksen sijaan vaihtoehtoisesti käyttää *de novo*-koamisalgoritmeja. Toisen vaiheen aikana myös tunnistetaan potilaan geenivariantit, joita voivat aiheuttaa SNP:t, muutaman nukleotidin indel-muutokset tai laajemmat DNA:n rakennemuutokset kuten translokaatiot, inversiot tai kopiolumuutokset. Haluttujen varianttien tunnistamiseksi on suodatus- ja tunnistusalgoritmien valinta oltava huolellista.

Viimeisessä vaiheessa yhteen koottua genomiprofiilia tulkitaan siihen sopivien ”annotaatioiden” eli sekvenssikuvausten mukaisesti [97]. Sekvenssikuvaukset haetaan useasta variaatiotietokannasta ja

erityisesti proteiinisynteesiin vaikuttavat rakennepoikkeamat lasketaan algoritmisesti osana analyysiä. Suurista väestötutkimuksista koottujen populaatiofrekvenssien pohjalta voidaan potilaalta pois sulkea väestössä yleisesti esiintyviä, oletetusti harmittomia alleelimuutoksia. 1000 Genomes ([www.internationalgenome.org](http://www.internationalgenome.org)) [62] ja Exome Aggregation Consortium (<http://exac.broadinstitute.org/>) [100] ovat esimerkkejä nk. normaaleista väestötietokannoista. Näyttöön ja tapaustutkimuksiin perustuva Online Mendelian Inheritance in Man ([www.omim.org](http://www.omim.org)) [101] on yksi vanhimmista genetiikkaa koskevista tutkimustietokannoista, ja se sisältää yhteensä lähes 25 000 merkintää geeni-variaatioista ja geeni-fenotyypikuvauksista. 2016 julkaistu ClinVar ([www.ncbi.nlm.nih.gov/clinvar/](http://www.ncbi.nlm.nih.gov/clinvar/)) [102] on uusi, OMIM:in tavoin kliinisesti merkittäviä geeni-fenotyypikuvauksia sisältävä, näyttöön perustuva tietokanta. Edellä mainittuja tietokantoja käytettäessä on kuitenkin huomioitava, että jopa 25 % tietokannoissa esiintyvistä varianteista on havaittu väärinluokitelluiksi, minkä vuoksi sekvenssikuvausten manuaalinen läpikäyminen on yhä osa analyysiä [98].

Proteiinien rakenteisiin mahdollisesti vaikuttavat genomimuutokset lasketaan esimerkiksi Variant Effect Predictor (VEP, [www.ensembl.org/info/docs/tools/vep/index.html](http://www.ensembl.org/info/docs/tools/vep/index.html)) [103] -algoritmillä. Kyseiset ohjelmat pyrkivät tunnistamaan ja merkitsemään proteiinisynteesin ennen aikaiseen pysähtymiseen, aminohapposubstituutioon tai muuhun geeniekspression poikkeamaan johtavia transkription muutoksia. Ohjelmien virhelähteinä mainitaan niiden joustamaton ja osittain kokeellinen tapa tulkita mahdollisia muutoksia, ja esimerkiksi useita eri transkriptiotuotteita koodaavan geenin virheiden tulkinta voi antaa useita eri virhearvioita. Sekvenssianalyysin kolmannen vaiheen helpottamiseksi löytyy useita kaupallisia ja avoimen lähdekoodin ohjelmistoja, jotka auttavat geenivarianttien tulkinnassa, visualisoinnissa ja löydösten vahvistamisessa. Oikein käytettyinäkin annotaatiolähteet tuottavat valtavia listoja dataa, jonka suodattaminen ja tärkeysjärjestykseen laittaminen lankeavat laitteiston käyttäjälle [98]. Tulosten tulkinnasta ja raportoinnista vastaa yhä ihminen, eikä mikään vielä viittaa siihen, että tilanne tulisi vielä lähitulevaisuudessa muuttumaan.

## 4. GEENITEKNOLOGIAA KOSKEVA LAINSÄÄDÄNTÖ

### 4.1. Yleiskatsaus lainsäädännöstä

Suomessa perinnöllisyyslääketiedettä ja lääketieteellistä genetiikkaa koskeva lainsäädäntö on hajautettu useisiin lääketiedettä, viranomaistoimintaa ja yksilön oikeuksia koskeviin lakeihin. Yksin omaan geenitestejä koskevaa lainsäädäntöä ei ole, vaan Suomessa noudatetaan Euroopan unionin mallia yleisistä lääketieteellisten laitteiden ja tutkimusten säädöksistä. Suomessa Sosiaali- ja terveysministeriön (STM) lupa- ja valvontavirasto Valvira valvoo ihmiskehosta peräisin olevien näytteiden käyttöä lääketieteellisessä hoidossa ja tutkimuksessa. Valviran asema on kirjattu lakiin terveydenhuollon laitteista ja tarvikkeista (2010/629) säädöskokoelman 2017/936 myötä [104]. Samassa säädöskokoelmassa lakiin lisättiin maininnat EU:n uusista lääkinnällisten laitteiden asetuksista. Viimeaikaisimmat katsauksen aiheeseen liittyvät muutokset lainsäädännössä käsittävät viime vuosina voimaan tulleita EU:n direktiivejä ja asetuksia, sekä niiden pohjalta Suomen lakiin tehtäviä sovelluksia.

#### 4.1.1. Lainsäädännön viimeaikaiset muutokset

Euroopan unionin asetukset 2017/745 lääkinnällisistä laitteista (MD-asetus) [105] sekä 2017/746 *in vitro* –diagnostisista lääkinnällisistä laitteista (IVD-asetus) [106] korvaavat kolme aikaisempaa lääkinnällisten laitteiden EU-direktiiviä (90/385/EEC, 93/42/EEC ja 98/79/EC). Uudet asetukset määrittävät tarkat vaatimukset lääkinnällisistä laitteista EU:n jäsenmaiden välisille sisämarkkinoille. Keskeisinä tavoitteina ovat potilasturvallisuuden vahvistaminen sekä laitevalmistajien ja ammattiharjoittajien vastuun lisääminen kilpailukyvyyn kärsimättä. Samalla asetukset vaimentavat jäsenmaiden oman lainsäädännön vaikutusta yhteisiä päätöksiä vastaan sekä varmistavat, että sovitut määräykset tulevat voimaan kaikkialla samassa aikataulussa. Asetuksissa määrätään käyttöön useita laadun valvonnan työkaluja, mukaan lukien CE-merkinnät, yksilöllistetyt laitetunnisteet (UDI-tunnisteet), uudet sertifikaatit ja markkinoinnin jälkeisen valvonnan järjestäminen. Markkinoitavien tuotteiden on kuuluttava Euroopan yhteiseen lääkinnällisten laitteiden tietokantaan (EUDAMED). Lisäksi jäsenmaita veloitetaan nimeämään uudet ilmoitetut laitokset, joille oikeutetaan MD- ja IVD-asetusten määräämien vaatimusten arviointi.

IVD-asetuksessa määritetään aiempaa selkeämmin geenitestien asema diagnostisina tuotteina. Käytettävien testien tieteellinen, analyttinen ja kliininen kompetenssi tulee osoittaa entistä tarkemmin, ja ilmoitettujen laitosten roolia laaduntarkkailussa vahvistetaan. IVD-laitteiden käyttöriskiarviointi uudistetaan, ja käyttöön tulee neljä luokitusta: matalimman riskin A, B, C ja korkeimman riskin D. Laitevalmistajan tulee julkaista yhteenveto tuotteensa turvallisuudesta ja käyttöön soveltuvuudesta aina, kun tuo markkinoille C- tai D-luokan valmisteen. Ainoastaan A-luokan laitteiden markkinointiin ei tarvita ilmoitetun laitoksen ennakkoarviointia.

Lääkinnällisille laitteille annetaan kolmen vuoden ja IVD-laitteille viiden vuoden siirtymäajat, joiden aikana vanhempien direktiivien mukaisia laitteita saa vielä markkinoida. Asetuksien mukainen aikajana on seuraava:

2019	Ensimmäiset ilmoitetut laitokset nimetään.
5/2020	MD-asetus astuu voimaan.
5/2022	IVD-asetus astuu voimaan.
5/2024	Vanhojen direktiivien mukaiset MDD- ja IVDD-sertifikaatit vanhenevat.
5/2025	Ainoastaan uusien asetusten mukaiset laitteet saavat olla markkinoilla.

Suomessa laki terveydenhuollon laitteista ja tarvikkeista (2010/629) [107] muistuttaa jo nykyisessä muodossaan pitkälti edellä mainittuja EU-asetuksia. Lain pykälässä 24 säädetään yleiset vaatimukset lääkinnällisten laitteiden ammattimaisesta käytöstä ja käyttäjistä. Pykälät 9, 18, 53 (a) ja 54 määräävät CE-merkinnän käytöstä, valvontaviranomaisen asemasta sekä eurooppalaisen laitteistotietokannan käytöstä. Näin ollen kynnys asetusten tulevasta soveltamisesta on jo valmiiksi matala.

Sosiaali- ja terveysministeriön vuonna 2015 käynnistämä genomistrategia [108] tähtää genomitiedon lisääntyneeseen ja turvalliseen käyttöön osana normaalia terveydenhuoltoa. Tavoitteeseen liittyy myös yksilökohtainen terveyshyödyn tavoittelu omaa ja väestön genomitietoa hyödyntäen. Hankkeeseen liittyvät uusi genomilaki (STM071:00/2018) sekä genomikeskustyöryhmän siihen liittyvät muistiot ovat parhaillaan lausuntokierroksella. Lain toteuttamiseen liittyy myös genomikeskuksen perustaminen. [109] Genomilain valmistelutekstissä mainitaan, että lain tarkoituksena on säätää genomitiedon asiantuntijaviranomaiseksi kansallinen genomikeskus, joka valvoo ja ohjaa genomitiedon vastuullisesta, yhdenvertaisesta ja tietoturvallisesta käytöstä. Työryhmän tehtävänä on tutustua ja ehdottaa mm. genomitiedon eettisestä käytöstä, kansallisen viite- ja variaatitietokannan perustamisesta, geenitestien tulkintapalveluiden luomisesta sekä rakenne- ja prosessiuudistuksista,

joita tarvitaan genomitiedon sisällyttämisestä terveydenhuoltoon ja tutkimukseen. Työryhmän toimikausi mainitaan päättyväksi vuoden 2019 loppuun mennessä.

Viimeaikaisin terveystietoja koskeva laki hyväksyttiin eduskunnan toimesta 13.3.2019, viisi päivää Juha Sipilän hallituksen eropäätöksestä. Laki sosiaali- ja terveystietojen toissijaisesta käytöstä (2019/552) [110] eli ns. toisiolaki purkaa mm. tieteellistä tutkimusta, tilastointia, kehittämis- ja innovaatiotoimintaa ja opetuskäyttöä varten hankittavien tietojen lupakäytäntöjä, kun kyseessä on sosiaali- ja terveydenhuollon toiminnassa syntyneitä asiakas- ja rekisteritietoja. Lain odotetaan parantavan kotimaisen lääketieteellisen tutkimuksen laajuutta ja tasoa, sekä edistävän potilastyötä, varsinkin kun eri tietolähteitä voi nyt yhdistellä vapaammin. Laissa säädetään myös tietolupaviranomaisesta, joka tulee toimimaan Terveyden ja hyvinvoinnin laitoksen (THL) yhteydessä kunhan STM:n asettaman väliaikaisen ohjausryhmän toimikausi päättyy kesäkuussa 2019.

#### **4.1.2. Potilasta koskeva lainsäädäntö**

Perimän tutkimukseen osallistuvan henkilön tai potilaan asemaa, oikeuksia ja tietoturva koskevia säädöksiä löytyy kuudesta eri laista. Näistä kattavin on nk. potilaslaki, jonka lisäksi yksilön oikeuksia terveystiedon käytöstä löytyy syrjintää, työelämää ja vakuutuksia koskevista laeista.

Laissa potilaan asemasta ja oikeuksista (1992/785) [111] mainitaan, että:

(3 §) Potilaalla on oikeus, ilman syrjintää, hyvään ja laadukkaaseen terveydenhoitoon. Kunnalla ja valtiolla on velvollisuus järjestää potilaalle tarvittava hoito niiden voimavarojen rajoissa, jotka kulloinkin ovat terveydenhuollon käytettävissä. Hoito on tapahduttava potilaan kanssa yhteisymmärryksessä. Potilaan yksilölliset tarpeet ovat mahdollisuuksien mukaan otettava hänen hoidossaan ja kohtelussaan huomioon.

(4 a §) Potilaan terveyden- tai sairaanhoitoa toteutettaessa on potilaan tai hänen laillisen edustajansa kanssa laadittava tarvittaessa tutkimus-, hoito-, kuntoutus- tai muu vastaava suunnitelma. Suunnitelmasta tulee ilmetä potilaan hoidon järjestäminen ja toteuttamisaikataulu.

(5 §) Potilaalla on oikeus saada selvitys hänen terveydentilastaan, hoidon merkityksestä, eri hoitovaihtoehdoista ja niiden vaikutuksesta. Tähän lasketaan myös hänen perimäänsä koskeva tieto. Samalla potilaalla on myös oikeus olla tietämättä edellä mainittuja tietoja. Selvitystä potilaan terveydentilasta ei tule myöskään antaa, kun on ilmeistä, että tiedosta aiheutuu potilaalle vakava hengen tai terveyden vaarantumisen uhka. Terveydenhuollon ammattihenkilön on annettava selvitys riittävän ymmärrettävästi, ja tarvittaessa turvaututtava tulkitsemisapuun.

(6 §) Potilasta on hoidettava yhteisymmärryksessä hänen kanssaan. Potilaalla on oikeus kieltäytyä tietyistä hoidosta tai toimenpiteistä, jolloin potilasta hoitavan tahon on mahdollisuuksien mukaan ja potilaan kanssa yhteisymmärryksessä jatkettava hoitoa muulla lääketieteellisesti hyväksyttävällä tavalla. Potilaan estyessä ilmaisemasta tahtoaan on hoidon toteutuksesta kuultava potilaan lähiomaista, laillista edustajaa tai muuta läheistä. Potilasta on hoidettava tavalla, jota voidaan pitää hänen henkilökohtaisen etunsa mukaisena, eritoten tilanteissa, joissa potilaan edustajien näkemys lääketieteellisestä hoidosta tai toimenpiteistä eroavat toisistaan.

(7 §) Alaikäisen potilaan mielipidettä hoidosta ja hoitotoimenpiteistä on kuultava silloin, kun se on hänen ikäänsä ja kehitystasoonsa nähden mahdollista. Alaikäisen potilaan ollessa kykenemätön päättämään hoidostaan on häntä hoidettava yhteisymmärryksessä hänen huoltajansa tai muun laillisen edustajansa kanssa.

(9 §) Tahtonsa ilmaisemiseen kykenemättömän potilaan lähiomainen, laillinen edustaja tai muu läheinen on oikeutettu saamaan tarpeelliset tiedot potilaan terveydentilasta, kun potilaan hoidon toteuttamisesta heitä kuullaan. Alaikäisellä potilaalla, jonka katsotaan olevan kykeneväinen päättämään hoidostaan, on oikeus kieltää terveydentilaansa ja hoitoansa koskevien tietojen antaminen huoltajalleen tai muulle lailliselle edustajalleen. Muutoin alaikäisen potilaan terveydentilaa koskeva selvitys annetaan hänen huoltajalleen tai muulle lailliselle edustajalle.

(12 §) Terveystieteiden ammattihenkilön tulee merkitä kaikki potilaan hoidon järjestämisen, suunnittelun, toteuttamisen ja seurannan kannalta olennaiset tiedot potilasasiakirjoihin. Asiakirjat, potilasnäytteet ja elinmallit tulee säilyttää edellä mainittujen käyttötarkoitusten lisäksi tieteellisen tutkimuksen ja mahdollisten korvausvaatimusten edellyttämä aika, ja hävittää välittämättömästi, kun niiden säilyttämiselle ei ole enää perustetta. Asiakirjojen, näytteiden ja mallien säilytysajasta säädetään tarkemmin sosiaali- ja terveysministeriön asetuksella 2009/298.

(13 §) Potilasasiakirjoihin sisältyvät tiedot ovat salassa pidettäviä, eikä tietoja saa ilman potilaan tai hänen laillisen edustajansa kirjallista suostumusta antaa sivulliselle. Salassapitovelvollisuus säilyy palvelussuhteen tai tehtävän päättymisen jälkeen. 13 §:n 3. momentissa on myös määritetty poikkeustilanteet potilasasiakirjojen tietojen antamisesta. Kohta 5 oikeuttaa kuolleen henkilön elinaikana laadittujen potilasasiakirjatietojen ja potilasnäytteiden käytön hänen omaistensa hyväksi silloin, kun perustellusti kirjallisella hakemuksella saatavat tiedot ovat omaisen etujen ja oikeuksien selvittämiseksi ja toteuttamiseksi välttämättömät.

Euroopan parlamentin ja neuvoston yleinen tietosuojasetus (EU) 2016/679, joka tunnetaan julkisuudessa nimellä GDPR-asetus, astui voimaan 25.5.2016 ja sitä alettiin soveltaa kahden vuoden siirtymäajan päätyttyä [112]. Asetus korvaa aiemman henkilötietodirektiivin 95/46/EY. Tietosuoja-

asetus on yleissäädös, joka koskee kaikenlaista henkilötietojen käsittelyä, eritoten rekisteröidyn henkilön oikeuksia tietojensa tarkistamisesta sekä rekisterinpitäjän velvollisuudesta osoittaa tietojen paikkansapitävyys ja oikeudenmukainen käyttö. Suomessa laitettiin vireille 1.3.2018 uuden tietosuojalain säätäminen [113], joka 1.1.2019 voimaan tullessaan korvasi vanhan henkilötietolain sekä lain tietosuojalautakunnasta ja tietosuojavaltuutetusta. Uusi tietosuojalaki (2018/1050) [114] täydentää EU:n yleistä tietosuoja-asetusta henkilötietojen käsittelystä sekä tietojen vapaasta liikkuvuudesta. Laki itsessään ei ole itsenäinen tai kattava kokonaisuus tietosuojasta ja henkilötiedoista, vaan sitä sovelletaan rinnakkain EU:n tietosuoja-asetuksen kanssa. Laissa myös määritetään valvontaviranomaiseksi nimetyn tietosuojavaltuutetun asema, kelpoisuusvaatimukset, toimikausi sekä tehtävät ja toimivaltuudet. Laissa on erikseen määritelty erityistilanteita, joissa henkilötietojen käsittelyssä voidaan poiketa tietosuoja-asetuksesta.

Yleisiä syrjintää rangaistuksen uhalla kieltäviä lakeja, joissa mainitaan yksilön terveystiedot yhtenä syrjinnän aiheena, on kaksi. Yhdenvertaisuuslain (2014/1325) [115] pykälässä 8 kielletään syrjintä ja vastatoimet mm. alkuperän ja terveydentilan perusteella. Rikoslain (1889/39) [116] pykälässä 11 kielletään ammatinharjoittajaa, virantoimittajaa tai muuta julkisessa tehtävässä olevaa syrjimästä ketään mm. rodun, kansallisen tai etnisen alkuperän, perimän, vammaisuuden tai terveydentilan perusteella sakko- tai vankeusrangaistuksen uhalla.

Laissa yksityisyyden suojasta työelämässä (2004/759) [117] pykälän 15 mukaan kielletään työnantajaa edellyttämästä työntekijäänsä osallistumasta geneettiseen tutkimukseen työhön otettaessa tai työsuhteen aikana. Työnantajalla ei myöskään ole oikeutta saada tietää työntekijälle jo tehdyistä geneettisistä tutkimuksista. Sen sijaan vakuutusopimuslain (1994/543) [118] tulkinta on vähemmän yksiselitteistä, sillä nimenomaan geenitesteistä ei kyseisessä laissa säädetä. Pykälän 10 mukaan vakuutuksen ehdot määräytyvät vakuutuksen hakemusajan terveydentilan perusteella silloin, kun hakemusasiakirjat on annettu vakuutuksenantajalle. Olemassa olevat tiedot ja testitulokset hakijan terveydestä on etukäteen kerrottava, ja diagnosoimattomia, piileviä sairauksia ei huomioida. Sen sijaan lakia voidaan tulkita niin, että jotkin diagnostiset, presymptomaattiset tai ennustavat geenitestit saatetaan rinnastaa muihin terveystietoihin, jotka on vakuutusta haettaessa kerrottava. Suomessa henkilövakuutukset ovat vapaaehtoisia ja normaalia, julkista terveydenhuoltoa täydentäviä. Potilaan tilanne on hyvinkin erilainen verrattuna niihin maihin, joissa terveydenhuolto rahoitetaan vakuutusten kautta.

Poliisin ja muiden esitutkintaviranomaisten toimintaa säätelevä pakkokeinolaki (2011/806) [119] rajoittaa yksilön oikeutta ruumiilliseen koskemattomuuteen tietyissä tilanteissa. Henkilönkatsauksessa (32 §) sanotaan, että yksilön oikeutta ruumiilliseen koskemattomuuteen voidaan rajoittaa, mikäli epäillään vakavaa rikosta ja henkilönkatsauksesta saadut näytteet, tässä tapauksessa DNA, voidaan tulkita olevan erittäin tärkeitä rikoksen selvittämiseksi. DNA-tunnisteet ja vastaavat tutkintatulokset on hävitettävä ja säilytetyt näytteet tuhottava viimeistään silloin, kun asia on saanut lainvoimaisen päätöksen. Pykälän 33 perusteella ainoastaan terveydenhuollon ammattihenkilö saa suorittaa lääketieteellistä asiantuntemusta vaativan tutkimuksen. Henkilönkatsauksessa viranomaisen ja tutkittavan tulee olla samaa sukupuolta, pois lukien verinäytteen, sylkinäytteen tai puhalluskokeen otossa sekä kliinisessä humalatutkimuksessa. Henkilönkatsauksesta ei saa aiheutua tutkittavalle mainittavaa terveydellistä haittaa.

Lain 9. luvussa *Eriyisiin tutkintakeinoihin liittyvät pakkokeinot* säädetään DNA-tunnisteiden määrittämisestä ja tallettamisesta pykälän 4 mukaisesti. Rikoksesta epäilylle saadaan tehdä DNA-tunnisteen määrittämistä varten tarpeellinen henkilöntarkastus. Vaatimuksena toimenpiteelle on epäily rikoksesta, josta säädetään vähintään kuuden kuukauden mittainen vankeusrangaistus. Vähintään kolmeksi vuodeksi vankeuteen tuomitulta henkilöltä saadaan tehdä DNA-tunnisteen määrittämistä varten tarpeellinen henkilönkatsaus sinä aikana, kun hän suorittaa vankeusrangaistusta. Samoin voidaan menetellä, mikäli henkilö määrätään mielentilansa vuoksi tahdosta riippumattomaan hoitoon eikä tuomita edellä mainittuun vankeusrangaistukseen. DNA-tunnisteen määrittäystä ei suoriteta, mikäli näin on jo tehty rikoksen esitutkinnassa. Poliisilla on oikeus tallettaa rajoitettua tietoa sisältävän DNA-tunnisteen henkilörekisteriin.

#### **4.1.3. Lääkäriä ja tutkimusta koskeva lainsäädäntö**

Lääkäriin asema, velvollisuudet ja oikeudet toimia työssään ovat määrätty lailla terveydenhuollon ammattihenkilöistä (1994/559) [120]. Laillistetulla lääkäriellä on oikeus päättää potilaan hoidosta, tutkimuksista ja diagnostiikasta, joiden kuitenkin täytyy perustua yleisesti hyväksytyihin sekä kokemusperäisesti perusteltuihin menettelytapoihin. Tämä antaa lääkärille tilanne- ja tapauskohtaisen päätösvallan geenitestien ja niiden tuloksin tarpeellisuudesta.

Lääketieteellistä tutkimusta ja diagnostiikkaa säätelevät kudoslaki, hedelmöityshoitolaki, biopankkilaki sekä laki lääketieteellisestä tutkimuksesta. Viimeksi mainittu on pitkälti yhtenevä Euroopan yleissopimuksen (2010/24) [121] kanssa, jossa säädetään ihmisoikeuksien ja ihmisarvon suojaami-



sesta biologian ja lääketieteen alalla. Sopimus tunnetaan myös nimellä biolääketiedesopimus, ja se onkin tärkein biologista ja lääketieteellistä tutkimusta ja hoitoa koskeva eurooppalainen säännöstö. Sen vuonna 2008 julkaistu lisäpöytäkirja geenitesteistä on myös Suomen osalta allekirjoitettu, mutta ei lakiin ratifioitu. Biolääketiedesopimuksessa kielletään selkeästi kaikenlainen perimään perustuva syrjintä sekä kielletään perimän testaus ja muokkaus muun kuin sairauden diagnosoiminnin, hoidon, ehkäisyyn tai tieteellisen tutkimuksen nimissä. Testaukseen tulee aina liittyä asianmukainen perinnöllisyysneuvonta. Henkilön iturataan, eli jälkeläisiin periytyviin ominaisuuksiin, kajoava geeniterapia tai -muokkaus on kielletty.

Laissa lääketieteellisestä tutkimuksesta (1999/488) [122] säädetään ihmistä koskevan tutkimuksen edellytykset sekä kiellot. Lääketieteelliseksi tutkimukseksi määritetään sellainen tutkimus, jossa puututaan ihmisen tai ihmisalkion koskemattomuuteen, ja jonka tavoitteena lisätä tietoa sairauksista, niiden oireista, diagnostiikasta, hoidosta, ehkäisystä ja terveydestä yleensä. Kaikenlaiseen lääketieteelliseen tutkimukseen on saatava eettisen toimikunnan myönteinen lausunto, ja tutkimuksesta saatava hyöty on oltava mahdollisia haittoja suurempi. Tutkittavan etu ja hyvinvointi on aina asetettava tieteen ja yhteiskunnan etujen edelle, eikä tutkimusta saa suorittaa ilman tutkittavan tai hänen laillisen edustajansa erityistä suostumusta. Pelkästään yhden henkilön hoitoon kohdistuvat lääketieteelliset tutkimustoimenpiteet, vaikka ne poikkeaisivat vakiintuneesta hoitokäytännöstä, eivät ole tutkimuslain mukaista lääketieteellistä tutkimusta. Sen sijaan lääketieteellisen laitteiston toimivuutta ja soveltuvuutta selvittävät, ihmiseen kajoavat toimenpiteet luetaan lain mukaiseksi tutkimukseksi. Perinnöllisyyslääketieteen osalta laissa on erikseen säädetty alkio- ja sikiötutkimusten edellytykset, ja hedelmöityshoitolaista poiketen ei tutkimuksen jälkeen alkioita saa siirtää ihmisen elimistöön.

Hedelmöityshoitolaissa (2006/1237) [123] säädetään perinnöllisyyslääketieteen toimintamallit. Lain mukaan alkiodiagnostiikka on sallittua ainoastaan, jos syntyvällä lapsella epäillänsä olevan vakava perinnöllisen sairauden vaara. Diagnostiikkaa voidaan tarjota myös vanhemmiksi tahtoville pareille, joilla ei ole hedelmöitysongelmaa, vaan sen sijaan korkea riski saada vakavasti sairas lapsi. Hedelmöityshoidoilla aikaan saatujen alkoiden soluista voidaan tulkita geneettinen sairausprofiili ja kohtuun siirtää alkio, jolla geenivirhettä ei ole. Vaikka lapsen sukupuolen etukäteen valinta on lailla kielletty, voidaan tästä poiketa siinä tapauksessa kun tiedetään geenivirheen olevan sukupuoleen leimautunut.

Syyskuussa 2013 voimaan tullut biopankkilaki (2012/688) [124] määrää tutkimuskäyttöön kerättävien biologisten näytteiden käsittelyä, hankintaa ja säilytystä koskevat yksityiskohdat. Biopankki-

toimintaa valvoo Valvira, ja uusien biopankkien perustaminen vaatii tutkimuseettisen toimikunnan hyväksynnän. Biopankkitoimintaan näytteen antanut kansalainen on oikeutettu saamaan tietoa jos ja kun hänen näytettään käytetään lääketieteelliseen tutkimukseen. Samalla näytteenantajalla on oikeus kieltää näytteidensä käytön missä tahansa tutkimuksen vaiheessa. Tutkijan kannalta biopankkilaki laajentaa tutkimuskäyttöön soveltuvien näytteiden käyttötarkoituksia – näytteitä voidaan luvanvaraisesti käyttää tutkimuksissa, joiden kaikki yksityiskohdat eivät ole vielä tiedossa.

#### 4.2. Laajojen geenitestien eettiset näkökannat

Geenitesteihin liittyvistä, hyväksi havaituista käytännöistä ja eettisistä näkökohdista on olemassa kokemusta jo 90-luvulta lähtien. Aiemmin geenien testaaminen on ollut harvojen asiaan erikoistuneiden tutkimuslaboratorioiden työtä, ja tulosten tulkinta perustunut jo itse tutkimusasetelmaan; löytyykö potilaalta juuri se geenimuutos, jota etsitään vai ei? Tutkimustiedon kasvun myötä myös mahdollisten geenitestien lukumäärä ja tulkintavaihtoehdot ovat lisääntyneet eksponentiaalisesti, mikä on lisännyt sekä diagnostisen genetiikan harjoittajien työmäärää että perinnöllisyysneuvonnan tarpeellisuutta. Ihmisen perimä pysyy koko eliniän kutakuinkin muuttumattomana, joten geenitestin tulos, tehdään se minä hetkenä hyvänsä, on peruuttamaton. Tämän vuoksi geneettisiä tutkimuksia tarjoavan ammattilaisen on annettava tietoon perustuvaa, ohjailematonta ja potilaan etujen mukaista neuvontaa geenitestien soveltuvuudesta ja niiden tuloksista. Muiksi eettisiksi kysymyksiksi on ajan mittaan noussut mm. lasten ja nuorten geenitestausta, sikiödiagnostiikka, geneettiset seulonnat sekä ennusteen antaminen vähäisen tai epävarman geenitiedon pohjalta. Toimintamalleja edellä mainittujen tilanteiden lisäksi muuhun ihmisen perimään liittyvään tutkimukseen asettavat Suomen laki sekä lääkäreiden työtä ohjaavat eettiset ohjeet. [125-126]

Uuden sukupolven sekvensointimenetelmät ovat luoneet uusia eettisiä ongelmia geenitiedon raportoimisessa ja jakamisessa. Kirjallisuudessa korostuu sivulöydösten rooli yleisenä eettisenä ongelmana, ja kuinka niistä tulisi potilaalle tai tutkimukseen osallistujalle kertoa. [127] Sivulöydöksiksi nimetään ne odottamattomat geneettiset muutokset, jotka ilmenivät tutkimuksen aikana, ja joilla on joko tunnettu, mahdollinen tai toistaiseksi epäselvä vaikutus tutkittavan terveyteen. American College of Medical Genetics and Genomics (ACMG) on laatinut luokitteluohjeen [128], jonka mukaisesti sekvensoinnissa havaitut geenivariantit tulisi jakaa viiteen eri ryhmään: ① patogeeninen, ② mahdollisesti patogeeninen, ③ vaikutus epäselvä, ④ mahdollisesti hyvänlaatuinen ja ⑤ hyvänlaatuinen. Näistä kahden ensimmäisen luokan kohdalla tulisi aina ilmoittaa tutkittavalle, ja varsinkin

kin jos löydöksen mahdollisesti aiheuttamiin terveysvaikutuksiin voidaan puuttua hoidolla, ehkäisyllä, varhaisella interventtiolla tai perinnöllisyysneuvonnalla. Vaikutukseltaan epäselvät variantit ovat tutkimuksissa havaittuja geenimuutoksia, jotka mahdollisesti vaikuttavat tutkittavan terveydentilaan tai sairastumisriskiin, mutta ovat toistaiseksi tutkimustiedon puutteen vuoksi luokittelemattomia. Epäselvien varianttien raportoinnin lisäksi tutkijan tulisi pohtia ehdottaako tutkittavalle lisätestejä löydösten merkityksen selvittämiseksi, ja sisällytetäänkö tutkittavan ydinperheen jäseniä tutkimukseen. Ero diagnostisesti tehtävän geenitestauksen ja lääketieteellisen tutkimuksen kesken on myös raportoinnissa; diagnostisen testauksen lähtökohtaisena ajatuksena on löytää potilaan oireille syy, jolloin sekä potilas että tutkija ovat jo valmistautuneet kuulemaan tulokset. Lääketieteellisessä tutkimuksessa sen sijaan tutkittava voi olla muutoin täysin terve, mutta kantaa vakavaakin perinnöllistä sairautta tietämättään. Potilaalla on aina oikeus ilmoittaa, haluaako tietää sivulöydöksistä ja niiden ilmi tuomista riskeistä. Ongelmana tässä tapauksessa ovat ne sairauteen liitetyt geenivariantit, jotka voivat vaikuttaa potilaan lisäksi hänen lähisukuunsa, ja kuinka löydöksistä tulisi raportoida potilaan ja hänen omaistensa tahtoa, yksityisyyttä ja hyvinvointia kunnioittaen. [125]

ACMG on laatinut listan 59 sivulöydöksenä todetusta geenivariantista, jotka tulisi aina ilmoittaa tutkittavalle kun suoritetaan kliinisiä sekvensointitutkimuksia [129]. Listalle on valittu työryhmän konsensuksen mukaiset variantit, jotka ovat yhdistettävissä yleisesti tunnettuihin monogeenisiin sairauksiin, joille on olemassa ehkäisevää tai oireenmukaista hoitoa ja jotka voivat olla pitkään oireettomia. Kriteereinä raportoinnille ovat listan varianttien löytyminen laajojen geenitestien yhteydessä, oli tutkimuksen indikaatio mikä tahansa. Tutkijan tai tutkimuksen pyytäjän tulee myös varmistaa, että potilas saa tarvittavaa neuvontaa tutkimusta edeltäen ja tulokset saatuaan. Suositus ei ota kantaa perinatologiaa (ennen ja jälkeen raskausaikaa ja syntymää) koskeviin tai terveiden lasten ja aikuisten sekvensointitutkimuksiin. Työryhmä ei myöskään näe hyötyä epäselvien varianttien raportoinnista.

Toisena kirjallisuudessa esiintyvänä eettisenä ongelmana mainitaan genomitiedon tuottamisesta ja käsittelystä syntyvät tietoturvariskit. Kliinisesti tehtävien, diagnostisten geenitestien tuottaman tiedon käyttö ja säilytys ovat kattavasti säädellyt lainsäädännön ja eettisten toimintaohjeiden myötä, mutta muussa tutkimuksessa, tuotekehityksessä ja toissijaisessa tarkoituksessa käytettävä genomitieto voi olla riskialttiimpaa. Arkaluontoisten, ja siten salassa pidettävien terveystietojen käytöstä tutkimuksessa on esitetty mielipiteitä, joissa vahvistetaan tutkimukseen osallistumisen periaatteet, eli vapaaehtoisuus ja tietoon perustuva suostumus, [130] ja että toistaiseksi ei tietoon ole tullut merkittäviä tietojen väärinkäytöksiä [131]. Geneettinen tutkimustyö on kuitenkin riippuvainen korkea-

laatuisesta ja kansainvälisesti jaettavasta anonymisoidusta geenitiedosta. Lisäksi esim. tutkittavan ikä, sukupuoli, maantieteellinen sijainti ja yleiskuvaus terveydentilasta voivat olla tutkimuksissa tarvittavia metatietoja, joita yhdistelemällä muihin julkisiin rekistereihin ja sosiaaliseen mediaan voidaan tietyissä tilanteissa paljastaa tutkittavien henkilöllisyyksiä. [132] Riskin lievittämiseksi useat julkiset tietokannat ovat siirtäneet osan tiedoista valvotun käyttöoikeuden taakse.

Geneettisen tiedon käsittely EU:n ulkopuolella voi heikentää testattavan oikeusturvaa. Tietoon perustuvan suostumuksen ohittaminen internetin välityksellä markkinoitavissa geenitesteissä voi olla yksinkertaista esim. raksi ruutuun -menetelmällä [68], jolloin asiakas ei voi olla täysin varma mihin on suostumassa. Myöskin internetin välityksellä tarjottavat testitulokset voivat olla alttiina tietovarauksille, ja mitä tapahtuu terveystiedoille jos yritys ajautuu konkurssiin tai fuusioituu osaksi suurempaa konsernia? [66] Vääriin käsiin joutunut genomitieto, joka on yhdistettävissä yksilön tunnistetietoihin tai pahemmassa tapauksessa muihin terveystietoihin, voi asettaa kyseisen henkilön heikompaan asemaan esimerkiksi vakuutusta, lainaa tai työtä hakiessa. Syrjinnän esto lainsäädännöllä ei ole tae hakemusten asianmukaisesta käsittelystä, varsinkaan jos hakija ei tiedä tietojensa joutuneen vääriin käsiin. Huoli tietojen väärinkäytöstä on suurempaa niissä maissa, joissa yksilön taloudellinen tilanne määrittelee saatavan terveydenhuollon tason. [133]

## 5. POHDINTA

Genomilääketiede on kokenut nopeaa kasvua 2010-luvulle saavuttaessa, ja nyt kun uuden vuosikymmenen alku lähestyy, ollaan varsinkin Suomessa geenitiedon hyödyntämisen suhteen varovaisen optimistisia. Sosiaali- ja terveystieteiden ministeriön käynnistämä strategia tähtää genomitiedon arkipäiväistämiseen ja perimästä saatavan tiedon liittämistä osaksi normaalia terveydenhuoltoa. Kansallisen genomikeskuksen perustus oli suunniteltu vuodelle 2018, mutta lausuntokierrokset ovat vielä kesken ja työryhmän toimikausi voimassa vielä tämän vuoden loppuun saakka. Keskus ja sitä tukeva lainsäädäntö tulevat asettamaan toimintamallit sekä lääketieteelliselle tutkimukselle että yhä laajeneville diagnostisille testeille. Vaikka geenitiedolle asetetaan suuria odotuksia, on pidettävä mielessä kaikki tiedon tuottamiseen, tulkitsemiseen ja käyttöönnottoon liittyvät rajoitukset. Kaikki eivät hyödy perimän testauksesta, eikä esimerkiksi luunmurtuman diagnosoimiseksi tarvita geenipaneelitutkimuksia. Vasta niissä tapauksissa, joissa sairaudelle ei osata asettaa diagnoosia muiden tutkimusten ja löydösten perusteella, diagnoosi tarvitsee varmistuksen tai jo anamneesin oton yhteydessä herää vahva epäily perinnöllisestä sairaudesta, on geenien tarkempi tutkimus kohdallaan. Monet perinnölliset sairaudet ovat harvinaisia, ja juuri harvinaissairauksien diagnosoimiseksi laajat geenitestit ovat jo osoittautuneet käyttökelpoisiksi. Suurimmat vaikutukset kansanterveyteen tosin saadaan yleisiä pitkäaikais- ja syöpäsairauksia hoitamalla, joista monet ovat monitekijäisiä ja siten hankalasti tulkittavia. Sattuman osuus sairastumiseen tulisi myös ottaa huomioon perimän ja ympäristön lisäksi. Näin kommentoi geenitutkija Juha Kere Lääkärilehdessä 10/2018, jonka lisäksi hän suhtautuu varauksella genomistrategiassa mainittuun geenien seulontaan.

Katsauksessa käsitellyistä toisen sukupolven sekvensointilaitteista Illuminan CRT-menetelmää käyttävät laitteet ovat yhä markkinoiden kärjessä. Laitteiston toimintaperiaate on säilynyt ennallaan tuotekehityksen ja teknologisten edistysten rinnalla. Käyttökokemuksen tuomasta varmuudesta huolimatta on alustalla edelleen sille ominaisia tyyppivirheitä. Matala-asteiset indel-virheet pystytään korjaamaan lukusyvyyttä lisäämällä, mutta lyhyistä fragmenteista koottavat, referenssiin rinnastettavat lukujaksot voivat jäädä virhekorjausalgoritmien suodattamiksi varsinkin toistojaksoalueilla. Tämä tekee osia perimästä sekvensointiin kelpaamattomiksi, minkä vuoksi tehokkaimmillakin CRT-laitteilla ei saada täyttä kuvaa koko genomista. Lisäksi saadakseen täyden hyödyn tehokkaasta laitteistosta tulisi sitä käyttää aina maksimikapasiteetissa, sillä esimerkiksi yksittäisten geenien tai suppeiden geenipaneelien sekvensointi olisi resurssien hukkausta. Skaalautuvuuden puutteen vuoksi laboratoriot saattavat joutua hankkimaan useita eri kapasiteetin laitteistoja. Valtavan suorituskyvyn genomisekvensointilaitteet ovat yhä suuria investointeja, kun pelkän laitteiston lisäksi täytyy ylläpi-

tää bioinformatiikkaohjelmistoja ajavaa infrastruktuuria, kouluttaa henkilökuntaa ja pitää asianmukaiset toimitukset voimassa.

Kolmannen sukupolven sekvensointialustat näyttävät jo ainakin tekniseltä kannalta ratkaisevan edellä mainittuja puutteita. Ion Torrentin PGM-alusta onkin nopeasti noussut markkinoilla Illuminan rinnalle. 454-pyrosekvensointialustalta lainaamien näytekirjastojen valmistelun ja polymeerasi-reaktion sivutuotteiden havainnoinnin pohjalta rakennettu alusta sekvensoi pidempiä fragmentteja ja on tarpeen mukaan skaalautuva. Kohtuuhintaisten laitteistojen ja reagenssien vuoksi puolijohdesekvensointi näyttääkin houkuttelevalta vaihtoehdolta pienemmän kokoluokan laboratorioille. Näytekirjaston fragmenttikoon äärimmäisyyksiin vievät yhden molekyylin sekvensointialustat taas edustavat katsauksen uusinta teknologiaa. Lähes reaaliajassa DNA:ta sekvensoivat PacBion RS II ja Oxford Nanopore Technologies:n alustat pyrkivät ratkaisemaan lyhyiden fragmenttien aiheuttamia rajoituksia. Varsinkin nanohuokosia käyttävä taskukokoinen MinION on saanut huomiota tiedejulkaisuissa. Matalaan hintaan markkinoitavaa laitetta suositellaan opetuskäytön lisäksi kenttätyöskentelyyn, sillä kaksijuosteisen DNA:n sekvensointiin soveltuvaa alustaa voi nähtävästi käyttää pieneliöiden genomien kartoitukseen.

Kliiniseltä kannalta geenitestaus on suunnannut kohti yhä laajentuvia, kokonaisia elinjärjestelmiä ja useita diagnoosivaihtoehtoja kattavia tutkimuksia. Aikaisempien kyllä tai ei –koeasetelmien sijaan katsotaan potilaan ehdokasgeenien variaatiota, jota sitten verrataan aikaisempaan tutkimustietoon. Täten erilaisia testivaihtoehtoja voidaan mainita olevan vähintään yhtä paljon kuin ihmisellä on genejä. Toisaalta uuden sukupolven sekvensointialustojen mahdollistama eksomisekvensointi kattaa jopa 85 % tunnetuista geneettisistä sairauksista, mutta tuo mukanaan aika-, resurssi- ja sivulöydös-ongelmia. Koko genomien sekvensointia ei voi vielä kutsua realistiseksi vaihtoehdoksi diagnostiikassa, varsinkin kun läheskään kaikkia perimän variaatioita ei ole mahdollista tietää. NGS-alustojen muita sovelluksia, kuten RNA-sekvensointia ja epigenomin kartoitusta käytetään jo tutkimuksessa, minkä ansiosta jatkuvasti löydetään uusia kohteita täsmähoidoille ja -lääkkeille. Yksi genomilääketieteen odotetuimmista sovelluksista onkin yksilöllistetty lääketiede (personalized medicine, PM), jonka periaatteena on lääketieteellisen diagnostiikan ja hoidon räätälöiminen yksilön ominaisuuksien ja tarpeiden mukaan. Toimintamallin tavoitteena on ennakoida potilaan terveydentilan muutoksia jo ennen oireiden ilmenemistä sekä poissulkea turhia ja tehottomia hoitovaihtoehtoja ennen kuin niitä edes kokeillaan. Nykyisin yksilöllistetyn lääketieteen näkyvin osa-alue on farmakogenetiikka, jonka avulla voidaan selvittää potilaan elimistön tapaa käsitellä lääkkeitä. Tunnettuja farmakogeenettisiä ominaisuuksia ovat mm. aihiolääkkeiden vaarallisen nopea metabolia, lääkkeiden liian

nopea eliminaatio sekä hitaasti toimivien entsyymien aiheuttamat lääkeaineen kertymät elimistöön. Käytännössä farmakogeenejä tulisi testata aina ennen kuin potilaalle määrätään uutena jokin tunnetusti riskejä sisältävä lääke. Yksilöllistetyn lääketieteen lisäksi uusia, genetiikkaan pohjautuvia kliinisiä käyttöaiheita joudutaan kuitenkin vielä odottamaan kattavien tapaustutkimusten puutteen vuoksi.

Geeniteknologian ja bioinformatiikan kehitystä on alusta asti tukenut tutkimustiedon jakaminen. Kaupallisten ohjelmistojen lisäksi avoimen lähdekoodin julkaisut muodostavat nykyisin jopa liialti kattavan valikoiman erilaisia tietoteknisiä työkaluja, joista valtaosa on räätälöity yhden ainoan tehtävän ratkaisemiseksi. Alati kasvavat tietokannat perimän variaatioista, sekvenssikuvauksista ja genotyyppi-fenotyyppihavainnoista yhdistettynä genomitason sekvensoinnista saatavaan raakadataan ovat kääntäneet geenitutkimuksen perusongelman, tiedonpuutteen, päälaelleen. Lääketieteellisesti tärkeän tiedon poimiminen tuhansista tai jopa miljoonista muuttujista vaatii vielä genomitiedon manuaalista suodattamista ja analyysiohjelmistojen hienosäätöä. Tiedon tallentamiseen ja analyysiin on kirjallisuudessa ehdotettu muun muassa pilvipalvelimia ja laajoja serverifarmeja, mikä vaatii infrastruktuurin uusimista, jos ja kun genomitietoa suunnitellaan sovellettavaksi lääkärin arkeen.

Suomessa geeni- ja terveystietoa koskeva lainsäädäntö on hajautettua. EU:n asetuksia ja kansainvälisiä sopimuksia täydentävä sekä hoito- ja tutkimustapahtumiin osallistuvien henkilöiden asemaa säädellään lukuisilla osittain toisiinsa viittaavilla lailla. Uusimmat lakimuutokset ovat siirtymäajallaan ja tulevat voimaan muutamassa vuodessa, näistä geeniteknologian kannalta tärkeimpinä mainittakoon lääkinnällisiä laitteita koskevat asetukset. Yksilön tietosuojaa on pyritty parantamaan GDPR-asetusta täydentävällä tietosuojalaille ja lääketieteellistä tutkimusta biopankkilaille sekä terveystietojen toisiokäyttölailla, joka astuu voimaan kokonaisuudessaan vuonna 2021. Genomilaki on vielä suunnitteluasteella, joten sen vaikuttavuutta terveydenhuoltoon ja muuhun terveystietojen käyttöön ei voida vielä arvioida.

Kuluttajille markkinoitavat geenitestit ovat kokeneet paradigmanmuutoksen nyt 2010-luvulla. Yhdysvalloissa FDA puuttui diagnostisia testejä markkinoivien bioteknologiayritysten toimintaan, joista toistaiseksi yksi on saanut genomipalveluilleen markkinointiluvan. Suuri enemmistö nykyisin toimivista DTC-yrityksistä tarjoaa suku- ja väestöhistoriatestejä, joiden tietoarvo on enemmänkin viihdeellistä. Vanhemmuutta, perinnöllistä taitotasoa tai ravitsemuksen genetiikkaa testaavan tulisi myös ymmärtää, ettei monessakaan tilaamassaan palvelussa ole vahvaa tieteellistä näyttöä. Ulkomaisen yrityksen kanssa vieraalla kielellä asioiva ei myöskään välttämättä ymmärrä mitä palvelua on tilaa-

massa ja mihin kaikkeen hänen henkilötietojaan käytetään. Asiantuntijoiden mukaan kuluttajien tulisi harkita tarkkaan mitä testejä teettää, sillä tulosten tulkintaa ei välttämättä osata perusterveydenhuollon vastaanotollakaan.

Huoli terveystietojen vuotamisesta ulkopuolisten nähtäville sekä laajojen geenitestien mukana kulkeva riski sivulöydöksistä ovat tämänhetkisen genomiikan eettisiä ongelmakohtia. Oletuksena pidetään, että terveydenhuollon tilaama ja säilyttämä perimätieto pysyy turvassa. Terveysorganisaatioiden ylläpitämät biopankit sekä väestötutkimuksia tekevät tutkimuslaitokset ovat velvoitetut poistamaan tunnistetietoja tutkimukseen käytettävistä näytteistä, mutta olemassa on jo tapauskertomuksia, joissa yksilön geneettiseen profiiliin on onnistettu yhdistämään nimi, ikä ja asuinalue. Tietovuodosta voi aiheutua haittaa yksilön sekä sosiaaliselle että ekonomiselle asemalle. Genomisekvensoinnin yhteydessä ilmenevien sivulöydösten suhteen löytyy jo valmiiksi toimintaohjeita. ACMG:n suositukset näyttävät olevan yleisesti hyväksytyt, ja diagnostisissa geenitesteissä ilmoitusvelvollisuus on hoitavalla lääkäriellä. Arviolta yhdellä tuhannesta voi löytyä sairauteen viittaava geenimuutos, jonka riskiarviota voi olla vaikea selvittää.

Yhteenvedona mainittakoon, että suurista edistysaskelista huolimatta on genomien laajuisten tutkimusten kliininen käyttö yhä epävarmaa. Epätietoisuus löydösten merkityksestä, varsinkin kun kyse on odottamattomista sivulöydöksistä, ei välttämättä edes aja potilaan etua. Tarvitaan siis yhä enemmän näyttöön perustuvaa tutkimustietoa, mutta massadatan asettamat haasteet tiedon käsittelystä ja säilyttämisestä kaipaavat vielä kestäviä ratkaisuja. Geenipaneelitutkimukset ovat nykyaikaa, ja niiden asianmukaisen käytön tulisi hiljalleen levitä terveydenhuollon arkeen. Isojen yliopistosairaaloitten laboratoriot ovat jo tarttuneet tilaisuuteen, ja lähitulevaisuudessa genomitietoa käytetään entistä laajemmin. Ainakin jos genomistrategiaa on uskominen.



## LÄHTEET

1. Tripp S, Grueber M. Economic Impact of the Human Genome Project. Battelle Memorial Institute 2011 (luettu 1.2.2018). [www.battelle.org/docs/default-source/misc/battelle-2011-misc-economic-impact-human-genome-project.pdf](http://www.battelle.org/docs/default-source/misc/battelle-2011-misc-economic-impact-human-genome-project.pdf)
2. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431: 931-945.
3. National Human Genome Research Institute. The Human Genome Project FAQ: Who participated? (päivitetty 12.11.2018). [www.genome.gov/human-genome-project/Completion-FAQ](http://www.genome.gov/human-genome-project/Completion-FAQ)
4. Ezkurdia I ym. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum. Mol. Genet.* 2014; 23 (22): 5866-5878.
5. Sharp AJ ym. Segmental Duplications and Copy-Number Variation in the Human Genome. *Am. J. Hum. Genet.* 2005; 77 (1): 78-88.
6. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* 2006; 7 (7): 552-564.
7. Schloss JA. How to get genomes at one ten-thousandth the cost. *Nature Biotechnology* 2008; 26: 1113-1115.
8. National Human Genome Research Institute. Genome Technology Program: Advanced Sequencing Technology Awards (päivitetty 25.2.2019). [www.genome.gov/Funded-Programs-Projects/Genome-Technology-Program#al-6](http://www.genome.gov/Funded-Programs-Projects/Genome-Technology-Program#al-6)
9. National Human Genome Research Institute. The Cost of Sequencing a Human Genome (päivitetty 6.7.2016). [www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost](http://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost)
10. Lohmann K, Klein C. Next Generation Sequencing and the Future of Genetic Diagnosis. *Neurotherapeutics* 2014; 11: 699-707.
11. Watson JD, Crick FH. A structure for deoxyribose nucleic acid. *Nature* 1953; 171 (4356): 737-738.
12. Ticono L Jr, Bustamante C. How RNA folds. *J. Mol. Biol.* 1999; 293 (2): 271-281.
13. Rojansky R, Cha M-Y, Chan DC. Elimination of paternal mitochondria in mouse embryos occurs through autophagic degradation dependent on PARKIN and MUL1. *ELife* 2016; 5: e17896.

14. Kere J, Knuutila S. Mitä lääkärin tulisi tietää kromosomeista, DNA:sta ja geenisäätelystä. Kirjassa: Aittomäki K, Moilanen J, Perola M. Lääketieteellinen genetiikka, 1. painos. Helsinki: Kustannus oy Duodecim 2016. ISBN: 978-9516564671.
15. Beadle GW, Tatum EL. Genetic control of biochemical reactions in *Neurospora*. Proc. Natl. Acad. Sci. USA 1941; 27 (11): 499-506.
16. Campbell AM, Heyer LJ. Kirjassa: Discovering genomics, proteomics and bioinformatics, 2. painos. San Francisco, USA: Benjamin Cummings 2006. ISBN: 978-0805382198.
17. Durmaz AA ym. Evolution of Genetic Techniques: Past, Present, and Beyond. Biomed Res Int. 2015; 2015: 461524.
18. Genohub Inc. NGS Library Preparation Applications and Kits (luettu 15.5.2019). <https://genohub.com/ngs-library-preparation-kit-guide/>
19. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 1977; 74 (12): 5463-5467.
20. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J. Mol. Biol. 1975; 94 (3): 441-448.
21. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA. 1977; 74 (2): 560-564.
22. Sanger F ym. Nucleotide sequence of bacteriophage  $\phi$ X174 DNA. Nature 1977; 265 (5596): 687-695.
23. Saiki RK ym. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 1985; 230 (4732): 1350-1354.
24. Murray V. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res. 1989; 17(21): 8889.
25. Anderson S. Shotgun DNA sequencing using cloned DNase I-generated fragments. Nucleic Acids Res. 1981; 9 (13): 3015-3027.
26. Smith LM ym. Fluorescence detection in automated DNA sequence analysis. Nature 1986; 321 (6071): 674-679.
27. Tabor S, Richardson CC. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proc. Natl. Acad. Sci. USA. 1995; 92 (14): 6339-6343.
28. Marsh M ym. High-throughput DNA sequencing on a capillary array electrophoresis system. J. Capillary Electrophor. 1997; 4 (2): 83-89.
29. Mu W ym. Sanger confirmation is required to achieve optimal sensitivity and specificity in next-generation sequencing panel testing. J. Mol. Diagn. 2016; 18 (6): 923-932.

30. Pareek CS. An Overview of Next-generation Genome Sequencing Platforms. Kirjassa: Xu J. Next-generation sequencing: Current Technologies and Applications. Norfolk, UK: Caister Academic Press 2014. ISBN: 9781908230331.
31. Margulies M ym. Genome sequencing in open microfabricated high density picoliter reactors. *Nature* 2005; 437 (7057): 376-380.
32. Ronaghi M ym. A sequencing method based on real-time pyrophosphate. *Science* 1998; 281 (5375): 363-365.
33. Dressman D ym. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl. Acad. Sci. USA.* 2003; 100 (15): 8817-8822.
34. Park S-J ym. Advances, practice, and clinical perspectives in high-throughput sequencing. *Oral Diseases* 2016; 22 (5): 353-364.
35. QUIAGEN. Pyrosequencing (luettu 20.9.2018). [www.qiagen.com/fi/resources/technologies/pyrosequencing-resource-center/](http://www.qiagen.com/fi/resources/technologies/pyrosequencing-resource-center/)
36. Fedurco M ym. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* 2006; 34 (3): e22.
37. Bentley DR ym. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008; 456 (7218): 53-59.
38. Schirmer M ym. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res.* 2015; 43 (6): e37.
39. Shendure J ym. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 2005; 309 (5741): 1728-1732.
40. Mardis ER. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 2008; 9: 387-402.
41. Reuter JA ym. High-throughput sequencing technologies. *Mol. Cell.* 2015; 58 (4): 286-597.
42. Harris DT ym. Single-molecule DNA sequencing of a viral genome. *Science* 2008; 320 (5872): 106-109.
43. Eid J ym. Real-time DNA sequencing from single polymerase molecules. *Science* 2009; 323 (5910): 133-138.
44. Travers KJ ym. A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Res.* 2010; 38 (15): e159.
45. Rothberg JM ym. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011; 475 (7356): 348-352.

46. Feng Y ym. Nanopore-based fourth-generation DNA sequencing technology. *Genomics Proteomics Bioinformatics* 2015; 13 (1): 4-16.
47. Kasianowicz JJ ym. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl. Acad. Sci. USA.* 1996; 93 (24): 13770-13773.
48. Schneider GF, Dekker C. DNA sequencing with nanopores. *Nat. Biotechnol.* 2012; 30 (4): 326-328.
49. Steinbock LJ, Radenovic A. The emergence of nanopores in next-generation sequencing. *Nanotechnology* 2015; 26 (7): 074003.
50. Jain M ym. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 2016; 17 (1): 239.
51. Oxford Nanopore Technologies. GridION X5 (luettu 20.9.2018).  
<https://nanoporetech.com/products/gridion>
52. Oxford Nanopore Technologies. About PromethION (luettu 20.9.2018).  
<https://nanoporetech.com/products/promethion>
53. Ip CLC ym. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. Version 1. *F1000Res.* Online 2015; 4: 1075.
54. Jain M ym. MinION Analysis and Reference Consortium: Phase 2 data release and analysis of R9.0 chemistry. Version 1. *F1000Res.* Online 2017; 6: 760.
55. Lightbody G ym. Review of applications of high-throughput sequencing in personalized medicine: barriers and facilitators of future progress in research and clinical application. *Briefings in Bioinformatics* 2018; 1-17.
56. Myllykangas S, Koskenvuo JW, Alastalo T-P. Uuden sukupolven sekvensointimenetelmät geenidiagnostiikassa. *Duodecim* 2013; 129 (2): 141-148.
57. Shevchenko Y, Bale S. Clinical Versus Research Sequencing. *Cold Spring Harb. Perspect. Med.* 2016; 6 (11): a025809.
58. Di Resta C ym. Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities. *EJIFCC.* 2018; 29(1): 4-14.
59. Need AC ym. Clinical application of exome sequencing in undiagnosed genetic conditions. *J. Med. Genet.* 2012; 49: 353-361.
60. Ku C-S, Cooper DN, Patrinos GP. The Rise and Rise of Exome Sequencing. *Public Health Genomics* 2016; 19: 315-324.
61. Sims D ym. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics* 2014; 15: 121-132.

62. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015; 526: 68-74.
63. Kalokairinou L, Howard HC, Borry P. Direct-to-consumer genetic testing. E-julkaisu 15.9.2014 (luettu 30.5.2019) eLS. <https://doi.org/10.1002/9780470015902.a0025181>
64. Vayena E. Direct-to-consumer genomics on the scales of autonomy. *J. Med. Ethics* 2015; 41: 310-314.
65. Burton A. Are we ready for direct-to-consumer genetic testing? *Lancet Neurology* 2015; 14: 138-139.
66. Strachan T, Goodship J, Chinney P. Genetic testing for complex disease and direct-to-consumer genetic testing (s. 474). Kirjassa: *Genetics and genomics in medicine*. New York, USA: Garland Science 2015. ISBN: 978-0815344803.
67. U. S. Food and Drug Administration. FDA allows marketing for first direct-to-consumer tests that provide genetic risk information for certain conditions. FDA News Release 6.4.2017 (luettu 4.4.2019).  
[www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm551185.htm](http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm551185.htm)
68. Phillips AM. Only a click away — DTC genetics for ancestry, health, love...and more: A view of the business and regulatory landscape. *Applied & Translational Genomics* 2016; 8: 16–22.
69. Regalado A. 2017 was the year consumer DNA testing blew up. *MIT Technology Review* (luettu 30.5.2019). E-julkaisu 12.2.2018. [www.technologyreview.com/s/610233/2017-was-the-year-consumer-dna-testing-blew-up/](http://www.technologyreview.com/s/610233/2017-was-the-year-consumer-dna-testing-blew-up/)
70. Regalado A. More than 26 million people have taken an at-home ancestry test. *MIT Technology Review* (luettu 30.5.2019). E-julkaisu 11.2.2019.  
[www.technologyreview.com/s/612880/more-than-26-million-people-have-taken-an-at-home-ancestry-test/](http://www.technologyreview.com/s/612880/more-than-26-million-people-have-taken-an-at-home-ancestry-test/)
71. Blueprint Genetics. Accreditations & Certifications (luettu 30.5.2019).  
<https://blueprintgenetics.com/certifications/>
72. HUSLAB tutkimusohjekirja. Erikoisalakohtaiset tutkimushakemistot - Genetiikka (luettu 30.5.2019). [https://huslab.fi/ohjekirja/genetiikka\\_hakemisto.html](https://huslab.fi/ohjekirja/genetiikka_hakemisto.html)
73. Kalokairinou L ym. Legislation of direct-to-consumer genetic testing in Europe: a fragmented regulatory landscape. *J. Community Genet.* 2018; 9 (2): 117-132.
74. Valtakunnallinen sosiaali- ja terveystieteiden neuvottelukunta ETENE. Kannanotto: Eettisiä huomioita kaupallisten geenitestien tarjoamisesta kuluttajille. E-julkaisu 15.12.2011 (luettu 6.2.2018). [https://etene.fi/lausunnot\\_ja\\_kannanotot\\_2011](https://etene.fi/lausunnot_ja_kannanotot_2011)

75. Sitra: Kuluttajien asenteet geenitutkimuksia kohtaan. Taloustutkimus Oy, 12/2013 (luettu 31.5.2019). E-julkaisu 2.2.2014: [www.slideshare.net/SitraHyvinvointi/sitra-kuluttajien-asetet-geenitutkimuksia](http://www.slideshare.net/SitraHyvinvointi/sitra-kuluttajien-asetet-geenitutkimuksia)
76. Dayhoff MO, Ledley RS. Comprotein: a computer program to aid primary protein structure determination. Julkaisussa: Proceedings of the December 4-6, 1962, Fall Joint Computer Conference. New York: ACM 1962: 262-274.
77. Paulig L, Zuckerkandl E. Chemical paleogenetics: molecular "restoration studies" of extinct forms of life. Acta Chem. Scand. 1963; 17: S9-16.
78. Sievers F, Higgins DG. Clustal Omega, accurate alignment of very large number of sequences. Methods Mol. Biol. 2014; 1079: 105-116.
79. Crick FH. The origin of the genetic code. J. Mol. Biol. 1968; 38: 367-379.
80. Staden R. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. 1979; 6: 2601-2610.
81. Devereux J, Haeblerli P, Smithies O. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 1984; 12: 387-395.
82. Malthiery B ym. Apple II PASCAL programs for molecular biologists. Nucleic Acids Res. 1984; 12: 569-579.
83. Johnsen M. JINN, an integrated software package for molecular biologists. Nucleic Acids Res. 1984; 12: 657-664.
84. Queen C, Korn LJ. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. 1984; 12: 581-599.
85. Hamm GH, Cameron GN. The EMBL data library. Nucleic Acids Res. 1986; 14: 5-9.
86. Bilofsky HS, Christian B. The GenBank® genetic sequence data bank. Nucleic Acids Res. 1988; 16: 1861-1863.
87. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. 1988; 85: 2444-2448.
88. Altshul SF ym. Basic local alignment search tool. J. Mol. Biol. 1990; 215: 403-410.
89. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. Genome Res. 1998; 8: 195-202.
90. Sutton GG ym. TIGR assembler: a new tool for assembling large shotgun sequencing projects. Genome Sci. Technol. 1995; 1: 9-19.
91. Pevzner PA, Tang H, Waterman MS. An Eulerian path approach to DNA fragment assembly. Proc. Natl. Acad. Sci. USA 2001; 98: 9748-9753.

92. Glowniak J. History, structure, and function of the internet. *Telenuclear Med.* 1998; 28: 135-144.
93. Gauthier J ym. A brief history of bioinformatics. *Briefings in Bioinformatics* 2018; 1-16.
94. Levin C ym. A data-supported history of bioinformatics tools. E-julkaisu (18.7.2018). Cornell University arXiv<sup>®</sup>. <https://arxiv.org/abs/1807.06808>
95. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 2000; 16: 276-277.
96. Schmidt B, Hildebrandt A. Next-generation sequencing: big data meets high performance computing. *Drug Discov. Today* 2017; 22: 712-717.
97. Oliver GR, Hart SN, Klee EW. Bioinformatics for Clinical Next Generation Sequencing. *Clinical Chemistry* 2015; 61: 124-135.
98. Goldfeder RL ym. Human Genome Sequencing at the Population Scale: A Primer on High-Throughput DNA Sequencing and Analysis. *Am. J. Epidemiol.* 2017; 186: 1000-1009.
99. Genome Reference Consortium. GRC Human Build 38 patch release 13 (luettu 9.5.2019). [www.ncbi.nlm.nih.gov/assembly/GCA\\_000001405.28](http://www.ncbi.nlm.nih.gov/assembly/GCA_000001405.28)
100. Lek M ym. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536: 285-291.
101. Online Mendelian Inheritance in Man, OMIM<sup>®</sup>. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD) (luettu 9.5.2019). [www.omim.org/](http://www.omim.org/)
102. Landrum MJ ym. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 2016; 44: D862-D868.
103. McLaren W ym. The Ensembl Variant Effect Predictor. *Genome Biology* 2016; 17: 122.
104. Laki terveydenhuollon laitteista ja tarvikkeista annetun lain muuttamisesta 2017/936. Annettu Helsingissä 19.12.2017. <http://finlex.fi/fi/laki/alkup/2017/20170936>
105. EUR-Lex. Euroopan parlamentin ja neuvoston asetus (EU) 2017/745 lääkinnällisistä laitteista (MD-asetus) (luettu 25.3.2019). <https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=CELEX:32017R0745>
106. EUR-Lex. Euroopan parlamentin ja neuvoston asetus (EU) 2017/746 in vitro -diagnostiikkaan tarkoitetuista lääkinnällisistä laitteista (IVD-asetus) (luettu 25.3.2019). <https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=CELEX:32017R0746>
107. Laki terveydenhuollon laitteista ja tarvikkeista 2010/629. Annettu Naantalissa 24.6.2010. <http://finlex.fi/fi/laki/alkup/2010/20100629>

108. Sosiaali- ja terveysministeriö. Parempaa terveyttä genomitiedon avulla. Kansallinen genomistrategia. Työryhmän ehdotus. Julkaisussa: Raportteja ja muistioita (STM): 2015:24. <http://urn.fi/URN:ISBN:978-952-00-3586-0>
109. Sosiaali- ja terveysministeriö. Genomikeskus – genomitiedon käsittelyn asiantuntijaviranomainen (luettu 30.5.2019). <https://stm.fi/genomikeskus>
110. Laki sosiaali- ja terveystietojen toissijaisesta käytöstä 2019/552. Annettu Helsingissä 26.4.2019. <http://finlex.fi/fi/laki/alkup/2019/20190552>
111. Laki potilaan asemasta ja oikeuksista 1992/785. Annettu Helsingissä 17.8.1992. <http://finlex.fi/fi/laki/ajantasa/1992/19920785>
112. EUR-Lex. Euroopan parlamentin ja neuvoston asetus (EU) 2016/679 (yleinen tietosuoja-asetus) (luettu 25.3.2019). <https://eur-lex.europa.eu/legal-content/FI/TXT/?uri=CELEX:32016R0679>
113. Eduskunnan kirjasto, lakihankkeiden tietopaketti (LATI). EU:n tietosuojauudistuksen kansallinen täytäntöönpano (päivitetty 21.12.2018). [www.eduskunta.fi/FI/tietoaeduskunnasta/kirjasto/aineistot/kotimainen\\_oikeus/LATI/Sivut/EUn-tietosuojauudistus.aspx](http://www.eduskunta.fi/FI/tietoaeduskunnasta/kirjasto/aineistot/kotimainen_oikeus/LATI/Sivut/EUn-tietosuojauudistus.aspx)
114. Tietosuoja laki 2018/1050. Annettu Helsingissä 5.12.2018. <http://finlex.fi/fi/laki/ajantasa/2018/20181050>
115. Yhdenvertaisuuslaki 2014/1325. Annettu Helsingissä 30.12.2014. <http://finlex.fi/fi/laki/ajantasa/2014/20141325>
116. Rikoslaki 1889/39. Annettu Helsingissä 19.12.1889. <http://finlex.fi/fi/laki/ajantasa/1889/18890039001>
117. Laki yksityisyyden suojasta työelämässä 2004/759. Annettu Helsingissä 13.8.2004. <http://finlex.fi/fi/laki/ajantasa/2004/20040759>
118. Vakuutuslakilaki 1994/543. Annettu Naantalissa 28.6.1994. <http://finlex.fi/fi/laki/ajantasa/1994/19940543>
119. Pakkokeinolaki 2011/806. Annettu Naantalissa 22.7.2011. <http://finlex.fi/fi/laki/ajantasa/2011/20110806>
120. Laki terveydenhuollon ammattihenkilöistä. Annettu Naantalissa 28.6.1994. <http://finlex.fi/fi/laki/ajantasa/1994/19940559>
121. Tasavallan presidentin asetus ihmisoikeuksien ja ihmisarvon suojaamiseksi biologian ja lääketieteen alalla tehdyn yleissopimuksen sekä siihen liittyvien ihmisten toisintamisen kieltämisestä ja ihmisalkuperää olevien elinten ja kudosten siirroista tehtyjen lisäpöytäkirjojen lainsäädännön alaan kuuluvien määräysten voimaansaattamisesta annetun



- lain voimaantulosta 2010/24. Annettu Helsingissä 19.2.2010.  
[www.finlex.fi/fi/sopimukset/sopsteksti/2010/20100024/20100024\\_1](http://www.finlex.fi/fi/sopimukset/sopsteksti/2010/20100024/20100024_1)
122. Laki lääketieteellisestä tutkimuksesta 1999/488. Annettu Helsingissä 9.4.1999.  
<http://finlex.fi/fi/laki/ajantasa/1999/19990488>
123. Laki hedelmöityshoidoista 2006/1237. Annettu Helsingissä 22.12.2006.  
<http://finlex.fi/fi/laki/ajantasa/2006/20061237>
124. Biopankkilaki 2012/688. Annettu Helsingissä 30.11.2012.  
<http://finlex.fi/fi/laki/ajantasa/2012/20120688>
125. Clarke A, Wallgren-Pettersson C. Etiikka ja sairauksien geneettiset aspektit. Kirjassa: Aittomäki K, Moilanen J, Perola M. Lääketieteellinen genetiikka, 1. painos. Helsinki: Kustannus oy Duodecim 2016. ISBN: 978-9516564671.
126. Saarni S, Kattelus M, Nummi V. Kirjassa: Lääkärin etiikka, 7. painos. Helsinki: Suomen Lääkäriliitto 2013. ISBN: 978-9519433646.
127. Hallowell N ym. Revealing the results of whole-genome sequencing and whole-exome sequencing in research and clinical investigations: some ethical issues. *J. Med. Ethics* 2015; 41: 317-321.
128. Richards S ym. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 2015; 15 (5): 405-424.
129. Kalia SS ym. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet. Med.* 2017; 19 (2): 249-255.
130. Lehtonen L, Saxén H. Geenitieto on salassapidettävää terveydentilatietoa. *Suomen Lääkärilehti* 2018; 73: 1044-1045.
131. Kääriäinen H. Genomitiedon käytön eettiset näkökohdat. *Duodecim* 2017; 133: 769-770.
132. Shi X, Wu X. An overview of human genetic privacy. *Ann. N. Y. Acad. Sci.* 2017; 1387 (1): 61-72.
133. Clayton E ym. A systematic literature review of individuals' perspectives on privacy and genetic information in the United States. E-julkaisu: *PloS One* 2018; 13 (10): p.e0204417  
<https://doi.org/10.1371/journal.pone.0204417>