Sophgena

AI in Digital Biology and Genetics

Sophgena

sophgena_cz

Sophgena a.s.

www.sophgena.com

dna@sophgena.com

+420 734 55 22 55

Patients, healthcare providers, insurance companies and payers need solutions. Innovations that reduce the human and economic burdens of disease.

PROJECT SOPHGENA A.S. GLOBAL AND STRATEGIC VIEW

Sophgena's goal is to create a virtual clinic with the aim to provide physicians and end clients with personalized medicine solutions through end-to-end genomic analysis.

What: 1.) creating a functional ecosystem for personalized medicine based on genetic tests and information such as SNP, SNR, CNV, INDEL; 2.) running a genetics lab with top level technology and the best personnel; 3.) creating a database of clinically interpreted variants with AI running on top of them to refine diagnosis.

Why? The time is now ripe for **biology to move definitively from science to engineering**. It has become **less sporadic** and we can **be able to describe it, understand the language of biology, and describe chemical processes**. This means that **technology has reached a point where we can model at this enormously increasing rate of computational power, look for connections, use all the research projects that have been done so far, and the results to find connections about how amino acids behave, what proteins do, how they interact, what our variants in DNA do.**

=> THIS IS THE TIME WHEN WE CAN HARNESS THE POWER OF COMPUTING AND APPLY EXISTING KNOWLEDGE TO CLINICAL PRACTICE.

Patients, healthcare providers, and payers need solutions. Innovations that reduce the human and economic burden of diseases.

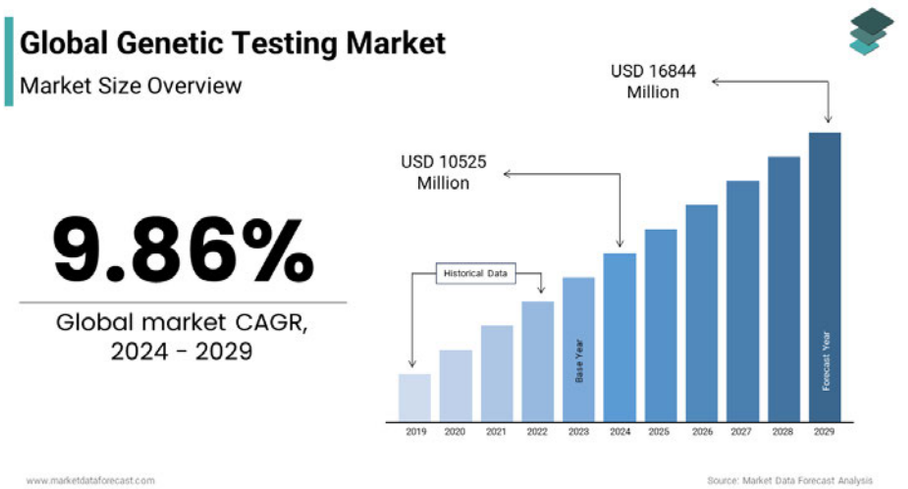
Sophgena's strategy is to take diagnostics and comprehensive healthcare to a new level to be faster, more successful, and more meaningful in what it delivers to clients. It also actively changes attitudes towards health.

PROJECT SOPHGENA A.S. GLOBAL AND STRATEGIC VIEW

The global genetic testing market is poised for significant growth this decade. The market size has been valued at approximately USD 15.9 billion in 2022. This market is expected to grow at a compound annual growth rate (CAGR) of approximately 10.2% from 2023 to 2032, reaching approximately USD 40.9 billion by 2032 (Market Data Forecast) (Allied Market Research).

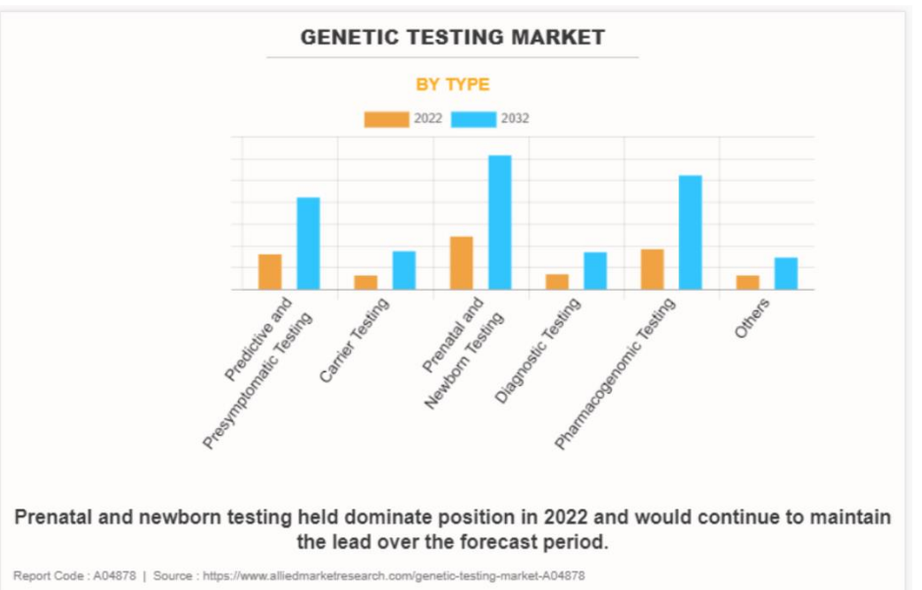
Another forecast states that the market could grow from $7.42 billion in 2023 to $36.4 billion by 2031, which would represent a significant 22% CAGR over the forecast period (SkyQuest Tech). The differences in market size estimates can be attributed to the different methodologies and market scopes considered in these analyses.

The European genetic testing market held the second largest share of the global market in 2023 and is expected to exhibit a remarkable CAGR during the forecast period.



Personalized medicine is used for various chronic diseases such as heart disease, cancer, diabetes, etc. The genetic testing of the individual receiving personalized medicine facilitates the drug developer's ability to efficiently and effectively treat the patient's disease.

The European genetic testing market accounted for the second largest share of the global market in 2023 and is expected to grow at a remarkable CAGR during the forecast period. Factors such as increasing demand for effective treatments, rising disease prevention awareness through ancestral DNA information, and growing government support in introducing reimbursement programs for the benefit of people are driving the genetic testing market in Europe.



11 REASONS TO USE SOPHGENA'S GENETIC SERVICE

Sophgena

1.Highly advanced technology and expert staff: Sophgena performs genomic testing in an accredited laboratory based on validated protocols using the latest technology; the quality of the tests is assured by top experts for test execution and genomic data analysis and interpretation.

2.Real personalized medicine: the Sophgena tests allow physicians to take a personalized approach to therapy that considers each patient's unique genetic differences. In the area of therapy selection, they allow physicians to precisely target treatment, be more effective, and minimize treatment side effects.

3. Disease prevention: The early detection of genetic risks using Sophgena tests enables the implementation of preventive measures before diseases develop. For example, if a genetic predisposition to certain types of cancer is detected, an individual form of monitoring can be chosen by the doctor, and lifestyle modifications can be recommended by specialists.

4.Genetic tests customized for the patient: with Sophgena tests, doctors can refine the selection of appropriate drugs based on genetic compatibility (pharmacogenomics), minimizing the risk of side effects and increasing the effectiveness of treatment.

5.Faster and more effective diagnosis: Sophgena tests use tools that make genetic analysis faster, more accurate, and more precise, allowing doctors to start effective targeted prevention or treatment sooner.

6.We are building a database of clinically interpreted variants: analysis of the large datasets generated by the whole-genome approach on which Sophgena tests are based, allows experts to better understand the genetic variations associated with different diseases, facilitating more accurate diagnosis and treatment.

7. Comprehensive range of genetic tests: Sophgena offers a wide range of targeted panel tests designed for specific health risks or diagnostic procedures such as KARDIO, ONKO, CARRIER, and newborn screening ...

8. The most comprehensive coverage, accuracy: For Sophgena tests, the technology of whole genome sequencing is used following international recommendations and standards in combination with evaluation criteria specified by professional entities (ACMG, Society of Czech Medical Genetics). The tests are performed following the EU rules for the performance of in vitro human diagnostics (IVDR regulation).

9. Maximizing data yield and participation in scientific research: Sophgena uses advanced informatics and bioinformatics methods, including AI methods, to maximize the yield of relevant data. The importance of aggregated and anonymized data obtained through Sophgena tests contributes, through collaboration between Sophgena and the medical and scientific community, to advancing faster and more accurate diagnosis and a personalized approach to prevention and treatment.

10. No investment for doctors: Hospitals and healthcare facilities can use Sophgena's services without investing in expensive technology and expert staff, providing a reliable solution for complex diagnoses at a great performance-to-price ratio.

11. Significant growth in the genetic diagnostics market: the use of genomic and genetic tests in human diagnostics and targeted personalized medicine is growing significantly globally, increasing the demand for expert services. Sophgena has the ambition to become one of the leaders in this market, providing comprehensive solutions based on the latest technologies, an expert team including laboratories, informatics, and interpretation of the data obtained.

EXPERIENCE OF THE FOUNDER AND MAIN SHAREHOLDER

Sophgena

I am an experienced executive chairman with strong focus on the healthcare industry in the CEE region.

With 16 years of expertise, I have combined healthcare within my own group, Innova Healthcare, FutureLife, specialized polyclinics, and private equity.

I am well-connected and capable of conducting business across the CEE region, with strong coverage in the Czech Republic, Slovakia, and Poland.

My influence and access within healthcare and medical organizations enable large-scale purchasing synergies, growth of unique patient bases, and high potential for M&A activity.

Motto: The goal is to help people understand that their health plays the most important role in their lives.

Up to 2001: Mathematics, physics, biology, chemistry, German, English

2001 - 2006: finance, credit derivatives, investments, banking at WU Wien, University of Cologne and VSB - TU Ostrava

2007 – 2013: private equity, negotiations, leading the healthcare projects, business development, and intensive executive management in IVD, outpatient clinics, and imaging diagnostics providers

2013 - 2015: co-founder of FutureLife, M&A, bringing the whole prepared project to investor with first phase hands-on execution

2016 - 2022: founder and major shareholder of Innova Healthcare (a group of medical facilities and healthcare providers that I built up consisting of an MRI clinic, an orthopaedic clinic, a gynaecology clinic, and a specialised immunology clinic), co-founder of Men's Health Clinic, M&A process, and integration of IVF clinics to the healthcare group.

2023 -> Sophgena - a new way of people accessing genetic testing and analysis, and the associated risk mitigating healthcare with clinical geneticists
Healthcare consolidation project

WE OFTEN TAKE PART IN WELL-KNOWN CONFERENCES, WHERE WE GIVE A POSTER OR LECTURE.

Sophgena

Sophgena a.s.
136 followers
3d •

✨ Naše talentovaná molekulární genetička Kristína Valovičová dnes prezentuje svůj poster na téma "Neinvazivní prenatální testování trizomie 21 a 18 s použitím digitální PCR" na prestižním kongresu ISPD 2024 v Bostonu! ✨

Kristína je momentálně ve finální fázi svého PhD studia v oboru Molekulární a buněčná biologie a genetika na MUNI. Zaměřuje se na molekulární metody v klinické genetice a jejím hlavním zájmem je neinvazivní prenatální testování (NIPT).

"Vybrala jsem si tuto oblast kvůli neustálému vývoji a posunu ve poznatcích. Chci přispět k tomu, aby se NIPT stalo dostupnějším pro více žen a abychom byli v budoucnu schopni testovat více nemocí neinvazivním způsobem," říká Kristýna.

Jsme na ni v Sophgena a.s. nesmírně hrdí a těšíme se na její budoucí úspěchy! 👏🔬🧬

#Genetika #PrenatalniPéče #Mo
#Výzkum #PhD

Juraj Bujnak • You
Healthcare, private equity, genetics, IVD, IVF
1mo •

Just arrived to ESHG 2024 Conference
Sophgena a.s.
#eshg2024

Sophgena a.s.
136 followers
1mo •

❓ Chcete mít kontrolu nad svým zdravím?
❓ Chcete učinit rozhodnutí, která zásadně ovlivní Váš dosavadní a budoucí život?
❓ Chcete znát odpovědi na otázky, na které ještě donedávna medicína neměla odpovědi?

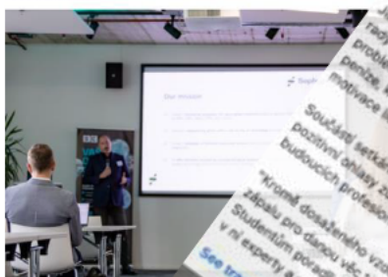
Pokud jste odpověděli třikrát ano, máme pro vás řešení.

🧬Sophgena je komplexní služba personalizované medicíny, která prostřednictvím genetické analýzy umožňuje vzít si váš osud a zdraví do vlastních rukou.

📩Pro více informací navštivte náš web, nebo nás kontaktujte prostřednictvím zprávy:
https://lnkd.in/ebUzUY-x

#vysetreni #testovani #genetics #zdravinaprvnimmiste #medicina #sophgena #dedicnenemoci #nextgenerationsequencing #genetika #analyzadna #NGS #epigenetics

See translation



Sophgena a.s.
136 followers
2mo •

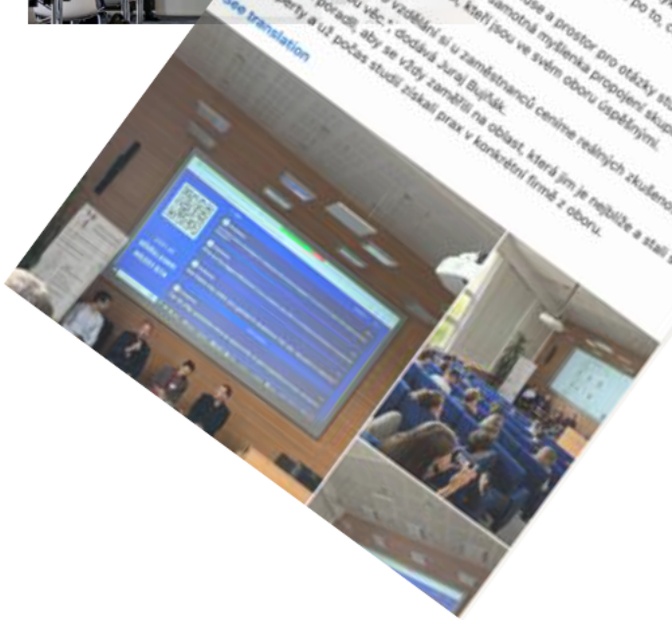
Zakladatel společnosti Sophgena a.s., pan Ing. Juraj Bujnak, se nedávno zúčastnil panelové diskuse s názvem "Lab to Market- Bridging Science and Industry" na půdě Akademie věd České republiky, kterou pořádala Skupina mladých imunologů (CYI).

Jako zástupce soukromé sféry předal studentům doktorského studia cenné rady a své zkušenosti z budování firmy se soukromým kapitálem. Nastínil problematiku, se kterou se firma v reálném životě potýká, od vize, cíle, vlastní peníze, které musíte dát jako první do projektu až po to, co je nejdůležitější - motivace lidí a zaměstnanců.

Součástí setkání byla i následná diskuse a prostor pro otázky studentů. Těší nás pozitivní ohlasy zúčastněných a samotná myšlenka propojení skupiny budoucích profesionálů a lidí, kteří jsou ve svém oboru úspěšnými.

"Kromě dostatečného vzdělání si u zaměstnanců cenímě reálných zkušeností a zápalu pro danou věc.", dodává Juraj Bujňák.
Studentům poradil, aby se vždy zaměřili na obšast, která jim je nejbliže a stáli se v ní experty a už počas studií získali prax v konkrétní firmě z oboru.

See translation

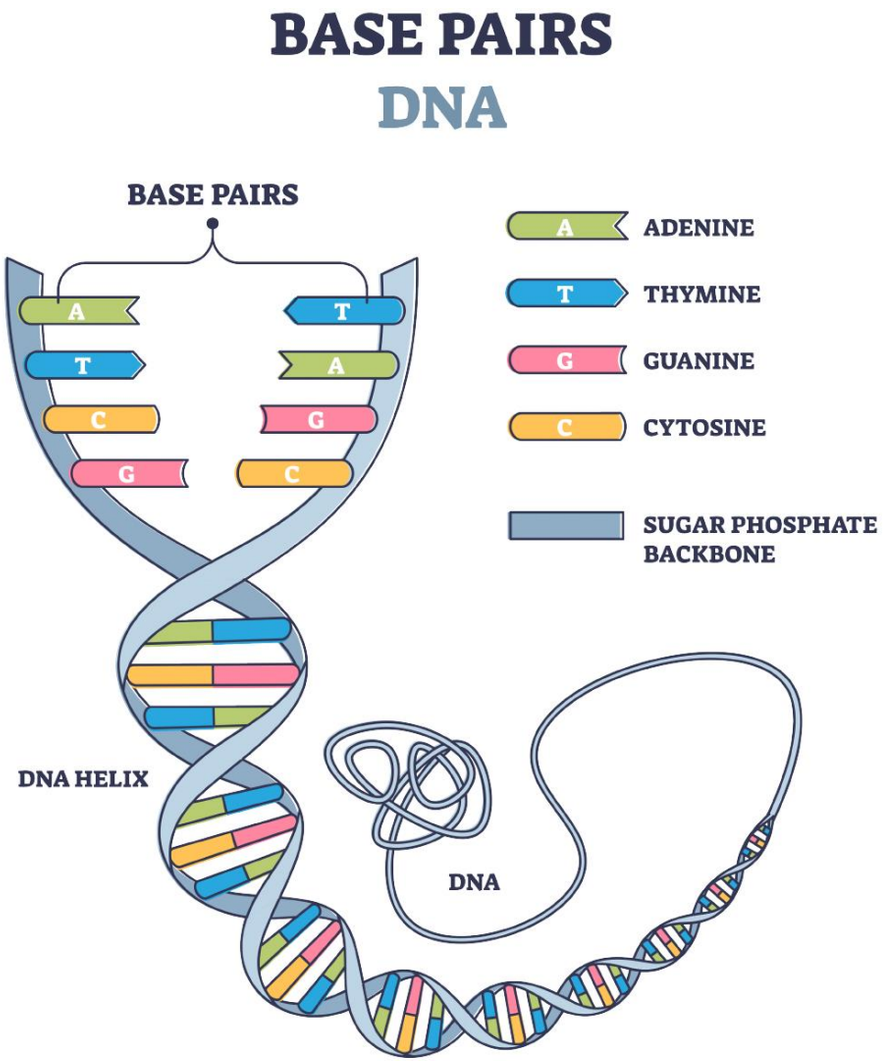


Sophgena

Genetics is the branch of biology that studies genes, genetic variation, and heredity in living organisms. It examines how traits are inherited through DNA, how genetic information shapes physical and biochemical traits, and how it changes over time due to mutations or environmental factors.

Genetic Information and DNA

- **DNA (Deoxyribonucleic Acid)** is the molecule that carries genetic instructions for all cellular processes and is fundamental to all life forms
- **DNA structure was resolved in 1953** by J. Watson, F. Crick, and R. Franklin
- DNA **contains the instructions for making proteins**, which perform a vast array of functions in organisms, from structural roles in cells to acting as enzymes that catalyze chemical reactions.
- These instructions are **encoded in the sequence of bases** (A, T, G, C), which are attached to a **sugar-phosphate backbone**. Each strand of DNA is held together by chemical interactions between these bases.

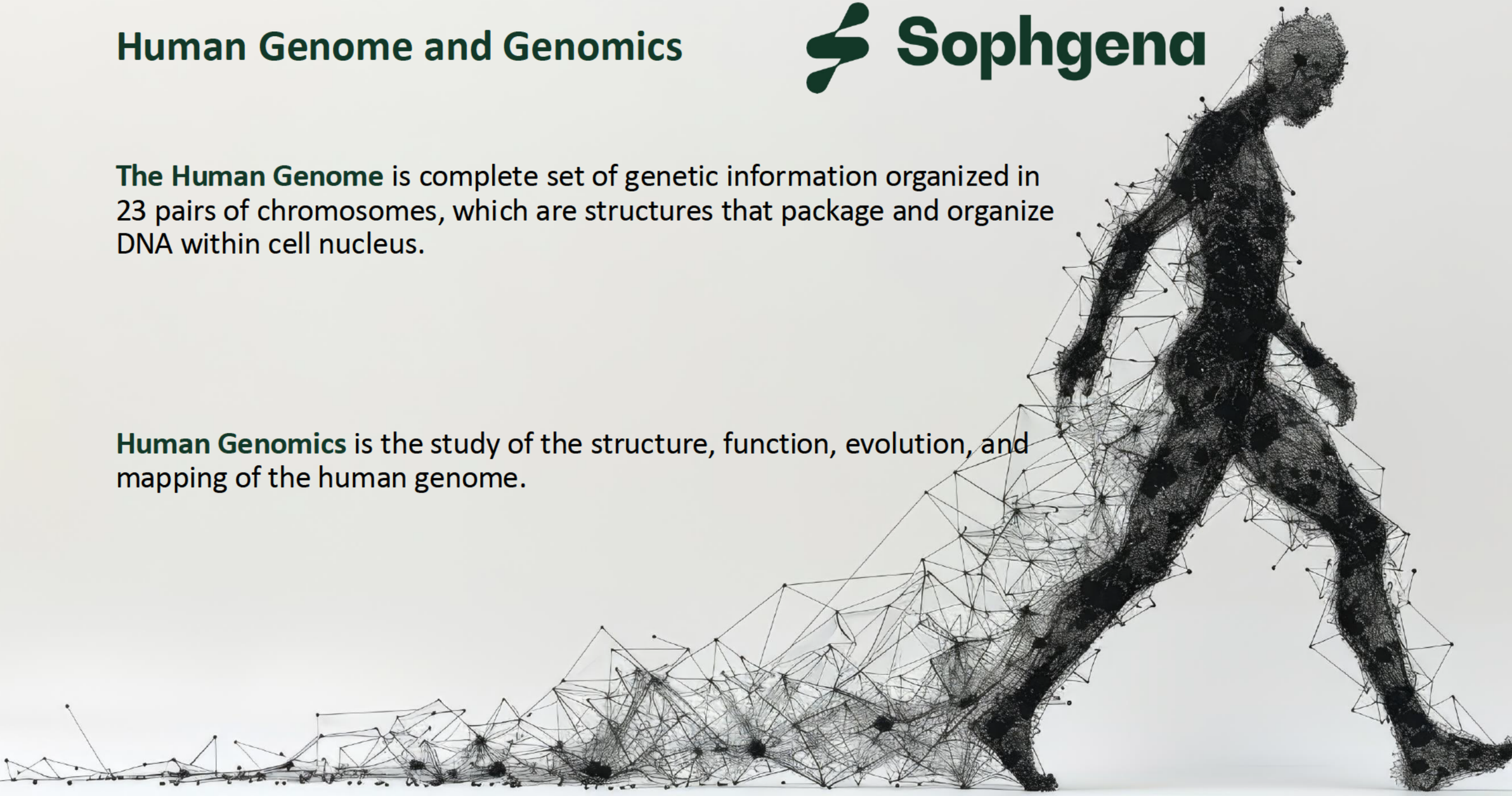


Human Genome and Genomics

Sophgena

The Human Genome is complete set of genetic information organized in 23 pairs of chromosomes, which are structures that package and organize DNA within cell nucleus.

Human Genomics is the study of the structure, function, evolution, and mapping of the human genome.

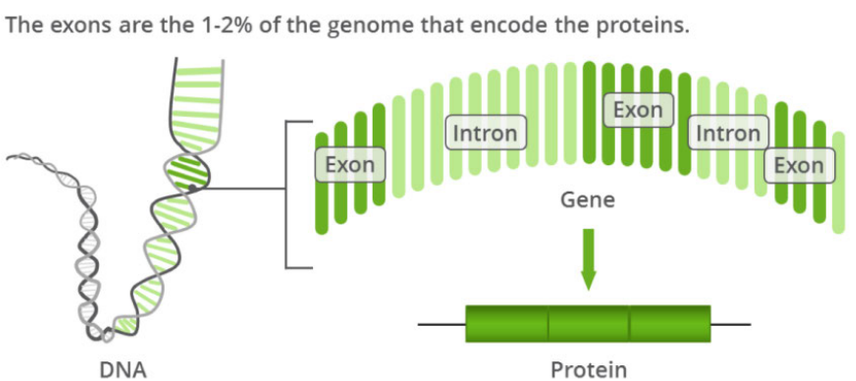


Key Features of Human Genome

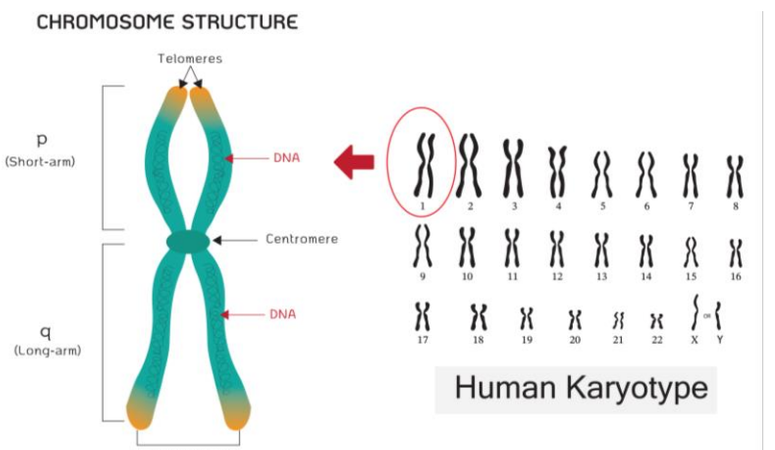
- The size of the Human Genome is about **3.2 Gigabases (Gb)** and it encodes approximatelly **20 000 to 25 000 genes.**

- **Only about 1-2% of the human genome directly codes for proteins (exons).** The rest consists of non-coding regions, some of which regulate gene expression or have structural functions.

- Although **human genomes are 99.9% identical among individuals, the 0.1% difference accounts for genetic diversity**, which contributes to differences in physical traits, susceptibility to diseases, and responses to environmental factors.



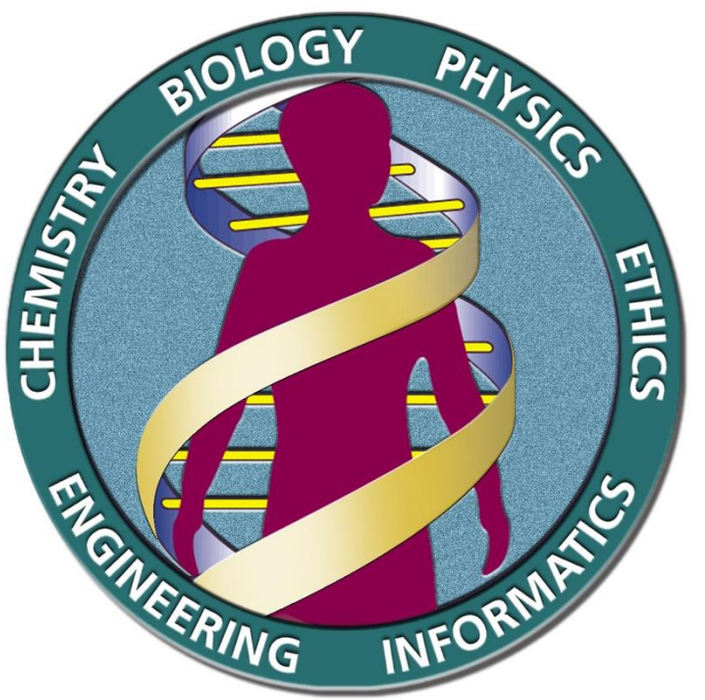
https://adntro.com/en/blog/salud/exome-complete/



The Human Genome Project (HGP)

- HGP was an **international research effort** to map and sequence the entire human genome. It was **launched in 1990** and **completed in 2003. The total costs** of HGP were estimated to **3 billion USD.**

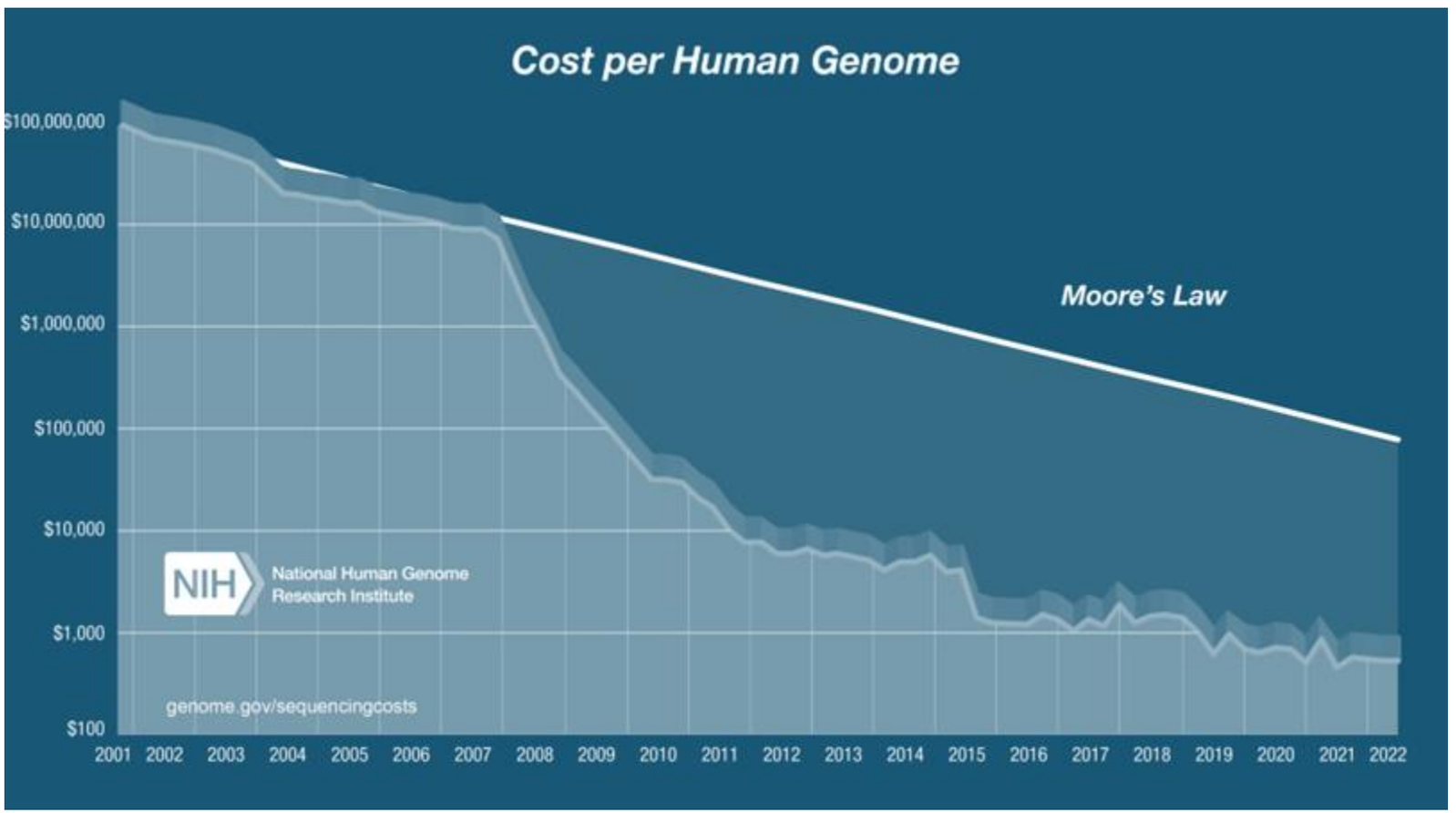
- **Aims and objectives:**
  - Sequence entire human genome consisting of 3.2 Gb
  - **Create a physical map** of the human genome
  - **Understand function** of genes and **identify all disease-causing** genes
  - **Make information available** for all researchers
  - **Develop tools** for processing and analysing genomic data



https://doe-humangenomeproject.ornl.gov/

Sophgena

Evolution of sequencing costs over last 20 years



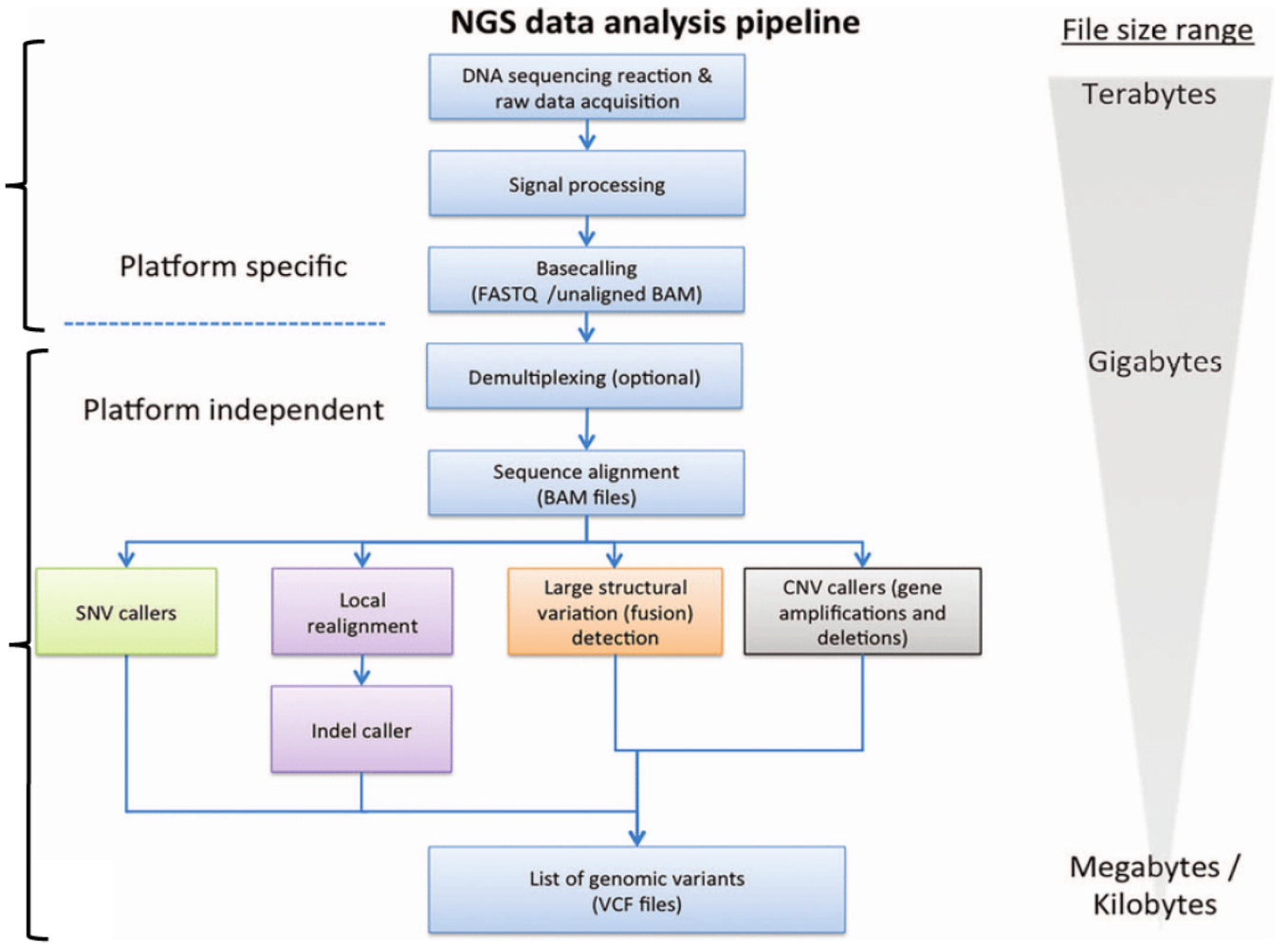
https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data

Sophgena

Next-generation sequencing analysis workflow

Primary Analysis

Secondary Analysis



https://www.researchgate.net/figure/General-schema-for-bioinformatics-workflow-for-next-generation-sequencing-NGS-testing_fig3_295684512

Secondary analysis – the task at hand

Actual sequencer output: ~1 billion ~100 basepair long DNA reads (30x coverage)

```
Read1: cttgggttgatattgtcttggaacatggaggttgtgtcaccgtaatggcacaggacaaacc
Read2: gatattgtcttggaacatggaggttgtgtcaccgtaatggcacaggacaaaccgactgtcg
Read3: tggaacatggaggttgtgtcaccgtaatggcacaggacaaaccgactgtcgacatagagct
Read4: ggttgtgtcaccgtaatggcacaggacaaaccgactgtcgacatagagctggttactgtcg
....
Read 1,000,000,0000: ....aactgtcgacatagagctggttactgtcgacatagagctggtt
```

True genome sequence: 3 billion bases in 23 contiguous chunks (chromosomes)

```
.......... cttgggttga tattgtcttg gaacatggag gttgtgtcac cgtaatggca
caggacaaac cgactgtcga catagagctg gttacaacaa cagtcagcaa catggcggag
gtaagatcct actgctatga ggcatcaata tcagacatgg cttcggacag ..........
```

Step 1 Align reads to a reference genome

Step 2 Infer the ***true*** genomic sequence(s)

Reads aligned to a reference genome

```
Reference: ...ttgtcttggaacatggaggttgtgtcaccgtaatggcacaggacaaacc...
Read1:     ...ttgtcttggaacatggaggttgtgtgaccgtaatggcacaggacaaacc
Read2:     ...ttgtcttggaacatggaggttgtgtgaccgtaatggcacaggacaaacc...
Read3:             tggaacatggaggttgtgtgaccgtaatggcacaggacaaacc...
```

Same as reference | Same as reference

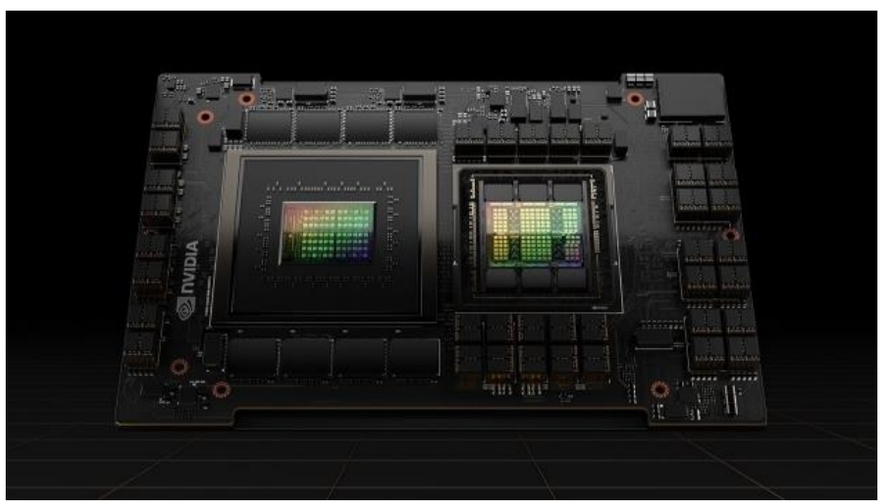
https://research.google/blog/deepvariant-highly-accurate-genomes-with-deep-neural-networks/

Sophgena

How fast can we perform secondary analysis today?

- Both primary and secondary analyses can be performed either **real-time** or **in a span of minutes** using today's advanced HW

- We performed benchmarking of NGS secondary analysis workflow using standard CPU computing server and **NVIDIA GPU computing server**



https://www.projectreylo.com/post/nvidia-unleashes-next-gen-ethernet-and-grace-hopper-superchip

Sophgena

Secondary analysis benchmarking – task and data

- **2 independent human DNA samples analysed using standardized whole-genome sequencing (WGS) protocol**
  - Illumina NovaSeq X Plus
  - 1.5 billion of short reads (2x151 nt PE) per sample accounting for genome coverage of ≥30x
  - Approximatelly 4.8 million single nucleotide variants (SNVs) and indels

- **map reads to hg38 reference genome (bwa mem, samtools)**
  - FASTQ -> BAM
- **mark or remove PCR duplicate reads (samtools, GATK)**
  - BAM -> clean BAM
- **call genetic variants (Google Deepvariant)**
  - BAM -> VCF

Sophgena

Secondary analysis benchmarking - hardware configuration

Standard CPU computing server

- **1x CPU Intel i7-14700KF 20-core/28-thread**
- 128 GB RAM
- 2x 14 TB HDD SATA RAID1
- Ubuntu 22 linux

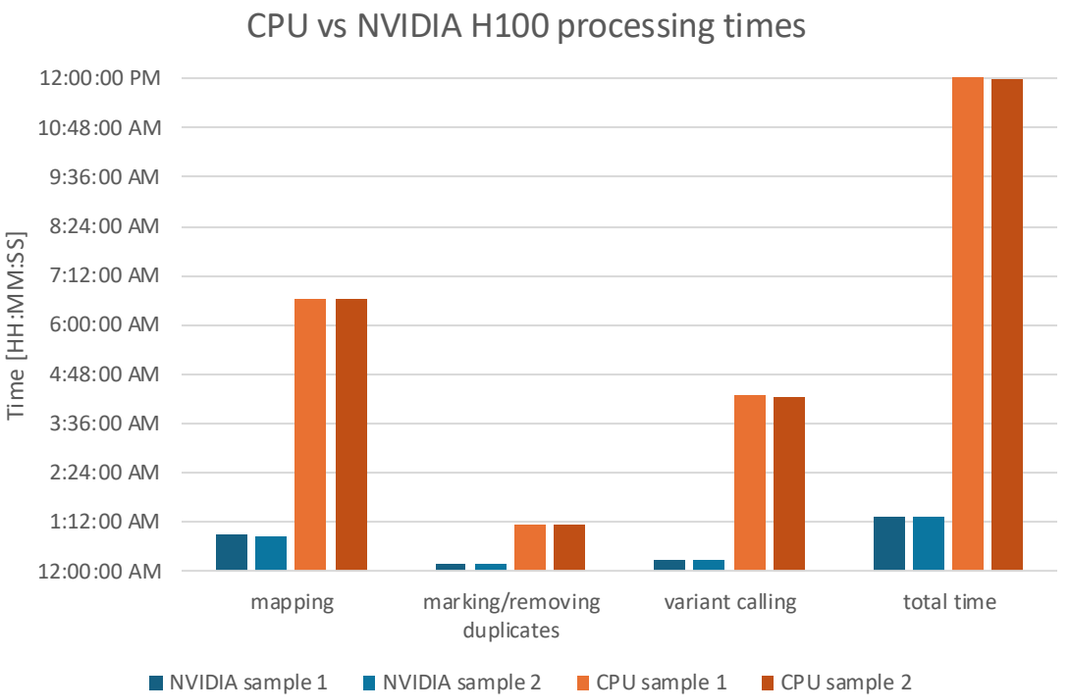
NVIDIA GPU computing server

- 1x CPU AMD EPYC 9334 32-Core
- 384 GB RAM
- 1x 3.84GB SSD SATA
- **1x NVIDIA H100 Tensor Core GPU**
- Ubuntu 22 linux
- NVIDIA Clara Parabricks 4.3.1-1

Secondary analysis benchmarking – results

H100 GPU processing is faster

- Alignment (mapping to reference) = 7.7x
- Marking and removing duplicate reads = 6.3x
- Variant calling = 15.2x
- Overall accelaration **= 9.1x**



Accelaration of secondary analysis in sequencing workflow is essential to applications such as **early pathogen diagnostics and wastewater monitoring,** as well as in fields like **oncology** and **newborn screening.**

Where is the bottleneck in data analysis today?

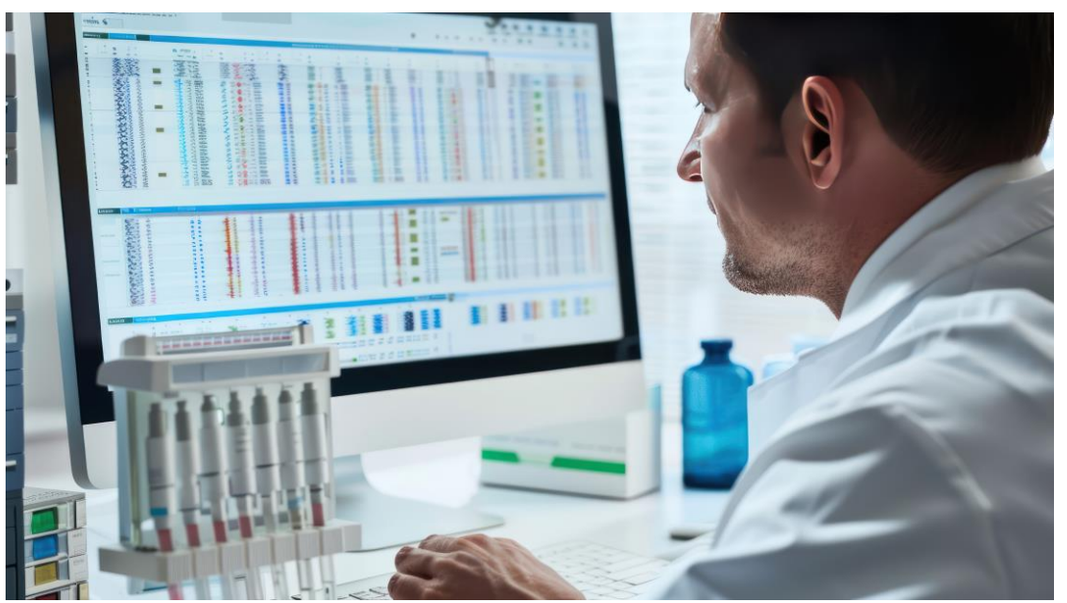
There is approximatelly **5 million variants** called in each WGS experiment which must be **reviewed and interpreted** in the context of patient clinical history.

In monogenic diseases, there is usually **single disease-causing variant,** making interperatation a *“needle-in-a-haystack”* problem

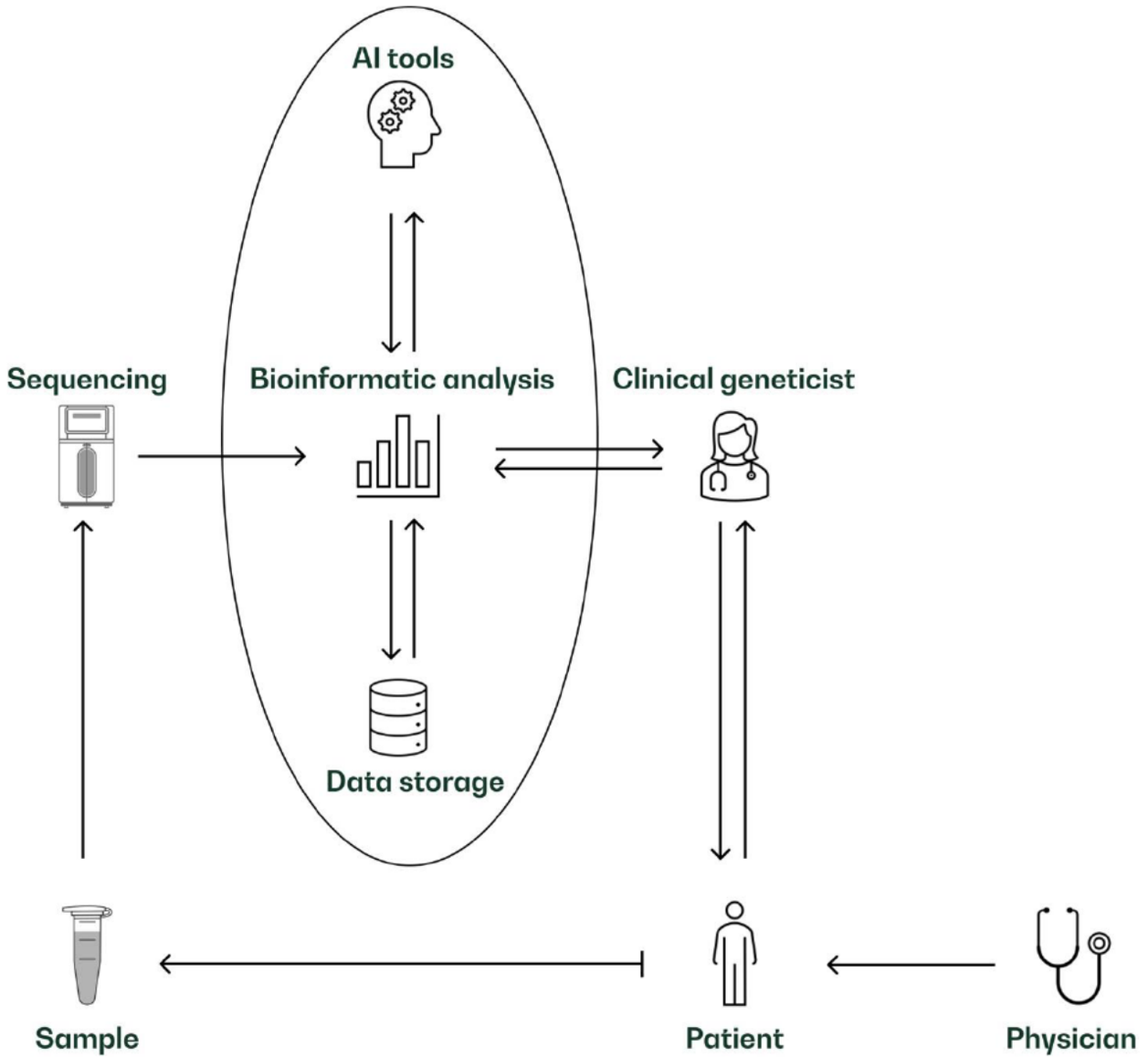
Data interpretation and reporting

=

Tertiary analysis



Considering future demands for reanalysis of genomic data



- **The amount of genomic data** produced in clinical diagnostics laboratories **has increased dramatically** due to the advent of NGS methods
- **New genetic variants are continuously added** to genomic databases, and **more powerful tools for data analysis are being developed**
- Genomic data need to be **securely stored** and **routinely re-analysed** to improve the diagnostic yield of genetic testing
- AI will be **essential in automating the proactive reanalysis** of genomic data

Use of Artificial Inteligence in Modern Genomics

- Artificial intelligence (AI) and machine-learning models (MLM) are essential in **database mining in genomics**, especially for **interpreting variants identified through sequencing**, however, the breadth of AI and MLM use also extends to:

1. **Variant classification** - AI and ML models automate the classification of genetic variants by comparing new variants to large databases like ClinVar or gnomAD.
2. **Functional impact prediction** - AI analyzes how variants might affect genes or proteins, predicting changes in structure or function.
3. **Annotation and prioritization** - AI annotates variants using multiple data sources, helping prioritize clinically relevant ones.
4. **Contextual analysis with phenotype integration** - Machine learning integrates patient data, like symptoms or family history, with genomic data to better interpret variants
5. **Prediction of variant-variant interactions -** AI detects complex interactions between multiple variants that contribute to disease.
6. **Data integration across multiple omics** - AI integrates data from genomics, transcriptomics, and proteomics to offer a comprehensive interpretation of how a variant affects biological processes, improving disease mechanism insights.
7. **Automation of workflows** - AI automates variant interpretation workflows, from variant identification to clinical reporting, speeding up the process and reducing human error.

Use of Artificial Inteligence in Modern Genomics

- AI and MLM are deployed to process vast amount of genomic datasets collected within **Genome-wide Association Studies** (GWAS)

- GWAS aims to identify **associations between genetic variants** (e.g., SNVs) and traits or diseases. AI and ML improve the process in the following ways:

  - **Feature selection and dimensionality reduction** – MLM simplify large genetic datasets from GWAS by reducing their complexity while retaining key genetic information
  - **Identification of complex interactions** - MLM can uncover complex, non-linear interactions between multiple genetic variants, revealing relationships that traditional GWAS methods might miss.
  - **Data integration** - AI models combine different types of biological data with GWAS results to uncover relevant pathways or regulatory mechanisms linked to genetic traits
  - **Improved prediction of phenotype** - AI techniques can predict disease risk or complex traits more accurately than traditional methods by analyzing multiple genetic variants at once.

Use of Artificial Inteligence in Modern genomics

- AI and MLM play a significant role in improving **Polygenic Risk Score** (PRS) models

- PRS **quantify an individual's genetic predisposition** to a disease based on **the cumulative effect of multiple genetic variants**. AI and MLM aids in PRS calculation as follows:

  - **Improved risk prediction**– AI models can improve PRS by considering complex relationships between SNPs and other factors (like age, environment, family history, lifestyle) for more accurate risk score assessment

  - **Handling population diversity**- AI helps make PRS models more accurate across different populations by adapting models trained on one group for use in others and adjusting for ancestry differences.

  - **Feature importance and interpretability** -AI methods like XGBoost or SHAP help identify which genetic variants contribute most to disease risk, improving our understanding of genetic influence.

Sophgena

Deep-learning model simulates a cell behavior

nature

Article | Published: 31 May 2023

Transfer learning enables predictions in network biology

Christina V. Theodoris, Ling Xiao, Anant Chopra, Mark D. Chaffin, Zeina R. Al Sayed, Matthew C. Hill, Helene Mantineo, Elizabeth M. Brydon, Zexian Zeng, X. Shirley Liu & Patrick T. Ellinor

Nature **618**, 616–624 (2023) | Cite this article

Deep-learning model learns to read human genome

nature machine intelligence

Article | Open access | Published: 23 July 2024

DNA language model GROVER learns sequence context in the human genome

Melissa Sanabria, Jonas Hirsch, Pierre M. Joubert & Anna R. Poetsch

Nature Machine Intelligence **6**, 911–923 (2024) | Cite this article

Sophgena

AI system developed by DeepMind enables prediction of protein stucture

nature

Article | Open access | Published: 15 July 2021

Highly accurate protein structure prediction with AlphaFold

John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, ... Demis Hassabis + Show authors

Nature **596**, 583–589 (2021) | Cite this article

1.78m Accesses | **3835** Altmetric | Metrics

https://www.nature.com/articles/s41586-021-03819-2

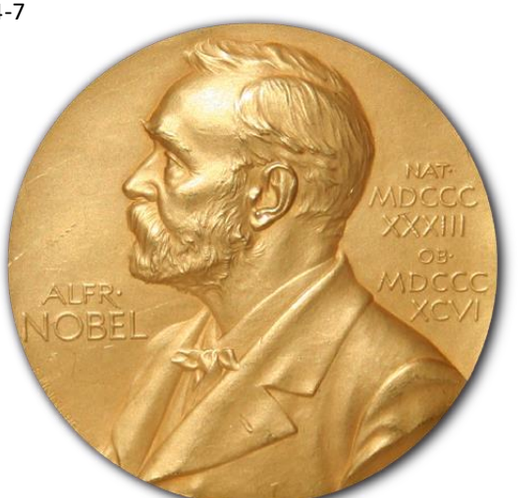
nature

NEWS | 09 October 2024

Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures

This year's prize celebrates computational tools that have transformed biology and have the potential to revolutionize drug discovery.

https://www.nature.com/articles/d41586-024-03214-7



https://en.wikipedia.org/wiki/Nobel_Prize_medal

Legal and Ethical Implications of AI use in Clinical Genetics

- **Data Privacy:** The use of AI requires large-scale genomic data, raising concerns about patient consent, data security, and potential misuse of sensitive genetic information.
- **Bias and Fairness:** AI models may perpetuate bias if trained on non-representative datasets, potentially leading to inaccurate genetic predictions for certain populations.
- **Transparency and Accountability:** The "black-box" nature of AI models in variant interpretation raises ethical questions about explainability, clinician responsibility, and ensuring that predictions are interpretable and reliable.
- **Regulation and Standards:** The integration of AI in clinical genetics demands clear regulatory frameworks to ensure the accuracy, safety, and ethical use of AI-driven decisions in healthcare.

Kazuistika - příklady pacientů s dětskou mozkovou obrnou a dystonií

- Pacientka nar. [redacted], od **1. roku** známky stagnace vývoje, **poruchy pohybu** a **motoriky**, od 5 let věku závislá na **inv. vozíku, porucha řeči, mikrocefalie, dystonie** (klonické křeče v různých částech těla)
- Oba rodiče bez klinické anamnézy nebo genetické zátěže v rodině
- 2020: jako diagnóza zvážená **kvadruspastická dětská mozková obrna**

https://patrondeti.cz/pribeh/rehabilitace-uvolni-kacence-ztuhle-a-bolave-svaly

https://youtu.be/lDRlv2ldTLo?t=8

Kazuistika – genetická vyšetření

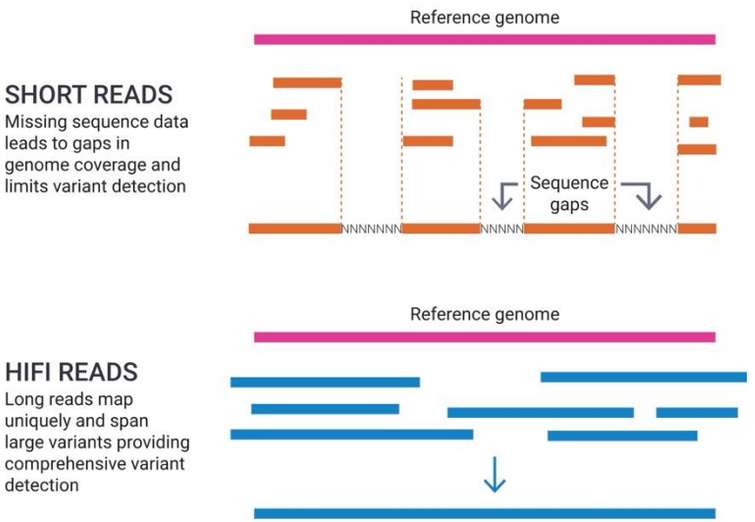
- standardní genetická diagnostika **bez patologického** nálezu (normální karyotyp 46, XX, mikročipové profilování: duplikace [redacted] (také u zdravého otce = **nemá patologický potenciál**)
- 2023: **specializované** vyšetření **fakultní nemocnice,** oba rodiče a dítě (trio) pomocí exomového sekvenování (WES), které teoreticky pokrývá všechny **protein kódující geny** v genomu člověka
- Nalezena genetická varianta v genu [redacted] – změna pořadí aminokyselin v proteinu vedoucí ke změně struktury proteinu [redacted], kóduje protein lehkého řetězce proteinu [redacted] - dlouhodobě spojován s časným nástupem dystonie
- *de novo* nález u pacientky, **pravděpodobně patogenní** v *in silico* modelech

Není přímý vztah s příznaky u pacientky – heterozygot (pouze jedna alela genu je zasažena), není publikována kazuistika s touto konkrétní variantou, nelze dohledat v online databázích = **bez uzavřené molekulární diagnózy**

Není vyloučena přítomnost **mutace**, která nebyla zjištěna použitými metodami = rodiče nemají jistotu, **že další dítě bude** v pořádku !!!

- Trio WES negativní = **vyčerpání** možností standardní péče, WES je top vyšetření hrazené ze ZP v ČR!
- Další možnosti: **samoplátci -** sekvenace celého genomu standardními metodami (po WES malá výtěžnost, ekonomicky nevýhodné
- experimentální metody - sekvenace celých molekul DNA (long-read sequencing) - [redacted]
- **Sophgena:** PacBio revio systém, sekvenování celého genomu čtením dlouhých molekul

Kazuistika – PacBio celogenomová sekvenace dlouhým čtením



https://www.pacb.com/human-genomics/

- **Pacientka přichází do Sophgena 06/24**, konzultace s klinickým genetikem – **indikace k LR-WGS**
- **09/24** – dokončení analýzy, **anotace dat pomocí SW s AI** využívající agregační online databáze genetických variant
- Potvrzeny předchozí nálezy **+ nalezena varianta v genu** [redacted] **– změna vlastností proteinu**
- [redacted] is **an inborn error of metabolism** characterized primarily by increased levels of 3-methylglutaconic acid (3-MGA) associated with variable neurologic deficits and neutropenia. The phenotype is highly variable: **most patients have infantile onset of a severe progressive encephalopathy with various movement abnormalities and delayed psychomotor development**. Other common variable features include seizures, recurrent infections due to neutropenia, anemia, and brain imaging abnormalities“ (OMIM, https://omim.org/entry/619835)
- **Nalezená varianta vysvětluje mnohem lépe fenotyp pacientky, probíhá validace výsledků ortogonální metodou**

Kazuistika – publikace

CLPB variants associated with autosomal-recessive mitochondrial disorder with cataract, neutropenia, epilepsy, and methylglutaconic aciduria.

Carol Saunders, Laurie Smith, Flemming Wibrand, Kirstine Ravn

Feb 5, 2015 in American journal of human genetics, Volume: 96, Issue: 2, Pages: 258-265
Linked by: CancerHotspots, GDC, ClinVar, Varsome AI, LOVD
52 citations • PubMed: 25597511 • DOI: 10.1016/j.ajhg.2014.12.020 • PDF

0

Abstract

3-methylglutaconic aciduria (3-MGA-uria) is a nonspecific finding associated with mitochondrial dysfunction, including defects of oxidative phosphorylation. 3-MGA-uria is classified into five groups, of which one, type IV, is genetically heterogeneous. Here we report five children with a form of type IV 3-MGA-uria characterized by cataracts, severe psychomotor regression during febrile episodes, epilepsy, neutropenia with frequent infections, and death in early childhood. Four of the individuals were of Greenlandic descent, and one was North American, of Northern European and Asian descent. Through a combination of homozygosity mapping in the Greenlandic individuals and exome sequencing in the North American, we identified biallelic variants in the caseinolytic peptidase B homolog (CLPB). The causative variants included one missense variant, c.803C>T (p.Thr268Met), and two nonsense variants, c.961A>T (p.Lys321*) and c.1249C>T (p.Arg417*). The level of CLPB protein was markedly decreased in fibroblasts and liver of affected individuals. CLPB is proposed to function as a mitochondrial chaperone involved in disaggregation of misfolded proteins, resulting from stress such as heat denaturation.

AI Tags

CLPB:c.803C>T CLPB:p.Thr268Met CLPB:c.961A>T CLPB:p.Lys321* CLPB:c.1249C>T CLPB:p.Arg417* CLPB mitochondrial disorder cataract neutropenia epilepsy 3-MGA-uria mitochondrial dysfunction febrile infection methylglutaconic aciduria 3-methylglutaconic aciduria liver recessive biallelic missense nonsense

https://www.cell.com/ajhg/pdf/S0002-9297(14)00526-6.pdf

ARTICLE

CLPB Mutations Cause 3-Methylglutaconic Aciduria, Progressive Brain Atrophy, Intellectual Disability, Congenital Neutropenia, Cataracts, Movement Disorder

Saskia B. Wortmann,[1,*] Szymon Ziętkiewicz,[2,25] Maria Kousi,[3,25] Radek Szklarczyk,[4,25] Tobias B. Haack,[5,6] Søren W. Gersting,[7] Ania C. Muntau,[8] Aleksandar Rakovic,[9] G. Herma Renkema,[1] Richard J. Rodenburg,[1] Tim M. Strom,[5,6] Thomas Meitinger,[5,6] M. Estela Rubio-Gozalbo,[10] Elzbieta Chrusciel,[2] Felix Distelmaier,[11] Christelle Golzio,[3] Joop H. Jansen,[12] Clara van Karnebeek,[13,14] Yolanda Lillquist,[13] Thomas Lücke,[15] Katrin Õunap,[16] Riina Zordania,[16] Joy Yaplito-Lee,[17] Hans van Bokhoven,[18] Johannes N. Spelbrink,[1,19] Frédéric M. Vaz,[20] Mia Pras-Raves,[20] Rafal Ploski,[21] Ewa Pronicka,[22] Christine Klein,[9] Michel A.A.P. Willemsen,[23] Arjan P.M. de Brouwer,[18] Holger Prokisch,[5,6] Nicholas Katsanis,[3] and Ron A. Wevers[24]

We studied a group of individuals with elevated urinary excretion of 3-methylglutaconic acid, neutropenia that can develop into leukemia, a neurological phenotype ranging from nonprogressive intellectual disability to a prenatal encephalopathy with progressive brain atrophy, movement disorder, cataracts, and early death. Exome sequencing of two unrelated individuals and subsequent Sanger sequencing of 16 individuals with an overlapping phenotype identified a total of 14 rare, predicted deleterious alleles in *CLPB* in 14 individuals from 9 unrelated families. *CLPB* encodes caseinolytic peptidase B homolog ClpB, a member of the AAA+ protein family. To evaluate the relevance of CLPB in the pathogenesis of this syndrome, we developed a zebrafish model and an in vitro assay to measure ATPase activity. Suppression of *clpb* in zebrafish embryos induced a central nervous system phenotype that was consistent with cerebellar and cerebral atrophy that could be rescued by wild-type, but not mutant, human *CLPB* mRNA. Consistent with these data, the loss-of-function effect of one of the identified variants (c.1222A>G [p.Arg408Gly]) was supported further by in vitro evidence with the mutant peptides abolishing ATPase function. Additionally, we show that CLPB interacts biochemically with ATP2A2, known to be involved in apoptotic processes in severe congenital neutropenia (SCN) 3 (Kostmann disease [caused by *HAX1* mutations]). Taken together, mutations in *CLPB* define a syndrome with intellectual disability, congenital neutropenia, progressive brain atrophy, movement disorder, cataracts, and 3-methylglutaconic aciduria.

https://www.cell.com/ajhg/pdf/S0002-9297(14)00519-9.pdf

Conclusion

”The application of AI technologies represents a critical opportunity for **dramatically improving the scalability**, **accuracy**, and **utility of clinical genomics**.” (Aradhya, S., 2023)

“**Variant prioritization computational tools** can complement and enhance manual analysis and reduce turnaround time in clinical diagnostic laboratory.” (O’Brien, T., 2022)

“It is important that AI methods **do not exacerbate existing socio-economic, racial,** and **ethnic disparities**.”(Walton, N., 2023)

“AI-based recommendations should **never entirely substitute for human judgements**. Until AI is judged to be extremely reliable, **accountability for the deployment of AI tools should tend to fall on practitioners**.” (Coghlan, S., 2023)

Sophgena

Bibliography

- Ada Lovelace Institute, Nuffield Council on Bioethics (2024). **Predicting: The future of health?** https://www.adalovelaceinstitute.org/report/predicting-the-future-of-health/
- Aradhya, S., Facio, F. M., Metz, H., **Applications of artificial intelligence in clinical laboratory genomics.** ***AJMJ Part C: Seminars in Medical Genetics*****, 193C:e32057 (2023)**. https://doi.org/10.1002/ajmg.c.32057
- Callaway, E., **Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures. Nature 634, 525-526 (2024).** https://doi.org/10.1038/d41586-024-03214-7
- Coghlan, S., Gyngell, C., Vears, D., **Ethics of artificial intelligence in prenatal and pediatric genomic medicine. J Community Genet 15, 13–24 (2024).** https://doi.org/10.1007/s12687-023-00678-4
- International Human Genome Sequencing Consortium. **Finishing the euchromatic sequence of the human genome**. Nature, 431, 931-945 (2004), https://doi.org/10.1038/nature03001
- Jumper, J., Evans, R., Pritzel, A. et al. **Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).** https://doi.org/10.1038/s41586-021-03819-2
- Sanabria, M., Hirsch, J., Joubert, P.M. et al. **DNA language model GROVER learns sequence context in the human genome. Nat Mach Intell 6, 911–923 (2024).** https://doi.org/10.1038/s42256-024-00872-0
- Theodoris, C.V., Xiao, L., Chopra, A. et al. **Transfer learning enables predictions in network biology. Nature 618, 616–624 (2023).** https://doi.org/10.1038/s41586-023-06139-9
- O'Brien, T., Campbell, N., Potter, A., et al. **Artificial intelligence (AI)-assisted exome reanalysis greatly aids in the identification of new positive cases and reduces analysis time in a clinical diagnostic laboratory.** Genetics in Medicine 24, 192-200 (2022). https://doi.org/10.1016/j.gim.2021.09.007.
- Walton, N., Nagarajan, R., Wang, C., **Enabling the clinical application of artificial intelligence in genomics: a perspective of the AMIA Genomics and Translational Bioinformatics Workgroup,** JAMIA **31, I536–541 (2024)**. https://doi.org/10.1093/jamia/ocad211
- Watson, J., Crick, F. **Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid**. Nature 171, 737–738 (1953). https://doi.org/10.1038/171737a0

Sophgena

Thank you for your attention

 Sophgena

 sophgena_cz

 Sophgena_a.s.

 www.sophgena.com

 dna@sophgena.com

 +420 734 55 22 55

Patients, healthcare providers, insurance companies and payers need solutions. Innovations that reduce the human and economic burdens of disease.