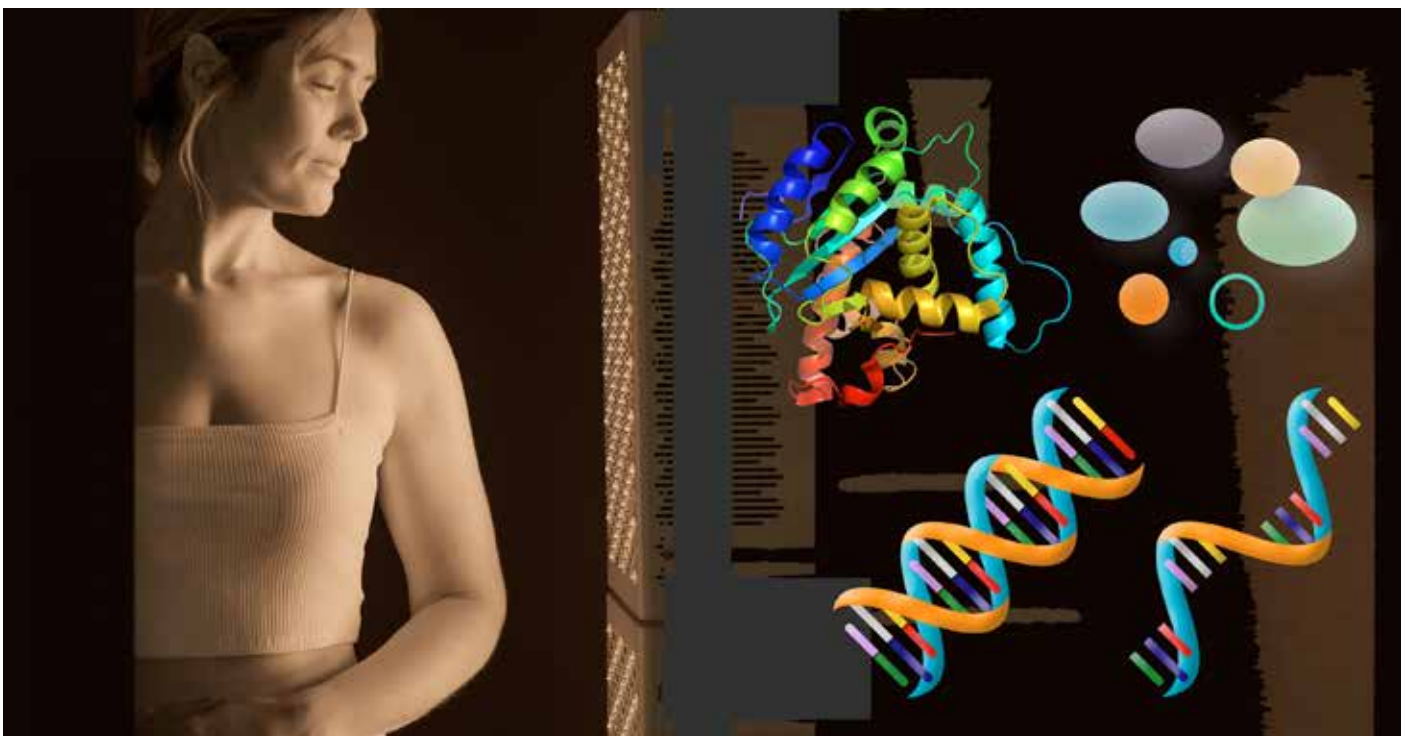


## Combining data from different sources for personalised treatment

The Bioinformatics Center at the University of Eastern Finland (UEF), led by Virpi Ahola, is developing new applications for analysing biomedical and multimodal data. These can be used to study cancers, metabolics, cardiovascular and neurodegenerative diseases.



Ahola has had a long career in bioinformatics. She was part of professor **Iikka Hanski's** metapopulation biology research group, which sequenced the whole genome of the Glanville fritillary butterfly that was the first reference genome solved in Finland. At the Karolinska Institute in Hong Kong, Ahola analysed gene function in different diseases at the single-cell level, and thereby studied how stem cells can be used to develop new drugs and treatments. She now heads the UEF Bioinformatics Center.

The Bioinformatics Center integrates different types of omics data (genomics, proteomics, transcriptomics) with clinical data and, in the future, possibly also imaging data.

"In addition to the usual omics analyses, we carry out multimodal data analysis for different research groups. This entails com-

binning the analysis of different types of data in order to provide more information than if they are analysed separately."

Analysis of the multimodal data varies depending on whether the data originates from different patients.

Omics is a research method that aims to analyse all genetically determined variables of a research subject simultaneously. Genomics analyses genetic variation and the function of genes, proteomics focuses on proteins, and epigenetics on the regulation of gene function and the storage of heritable information without changes in the DNA sequence. Metabolomics, for its part, analyses changes in metabolism caused by disease, diet or medication.

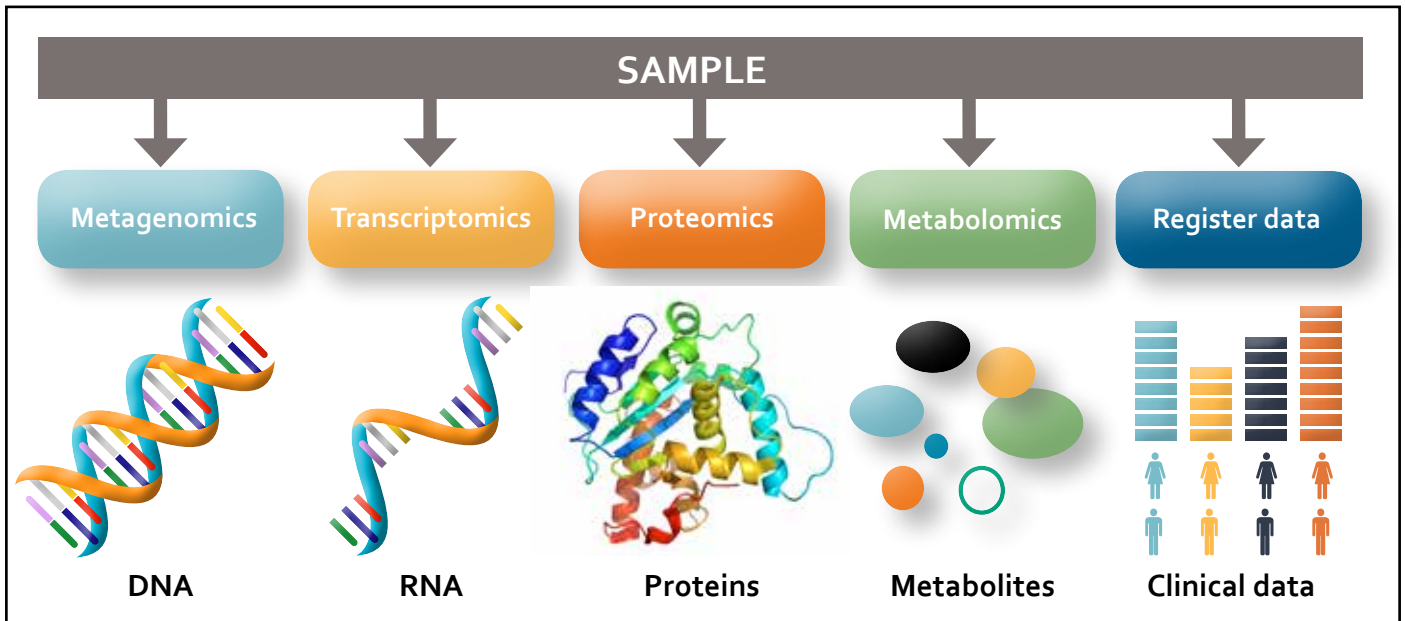
"We are developing bioinformatics services in collaboration with biomedical ex-

perts. One focus at the University of Eastern Finland is on understanding the molecular basis of key chronic diseases and improving their prevention and treatment," Ahola says.

### Translational medicine for personalised treatment

Translational medicine uses basic research in clinical trials, but also patient samples and disease models to identify disease mechanisms and drug targets. The research approach is interdisciplinary, which provides a good starting point for research but also improves treatments for patients.

"What is delaying the era of translational medicine is that we simply don't know enough. The idea behind combining several different data sources is to obtain more information. The integration is very much computational,



and requires CSC – IT Center for Science’s resources and infrastructures like ELIXIR.”

One example Ahola gives is single-cell technologies.

In transcription, the genetic code in DNA is copied into RNA. This is the first phase of protein synthesis. Transcriptomics provides precise information about the gene expression in an individual cell at a given moment.

“The use of single-cell transcriptomics is still expensive. The principles of open science exist, and therefore all data must be shared when it is published. This allows data to be reused, and different data sources to be combined.”

However, the challenge is that data is produced using different technologies.

“Different data sources may have different numbers of cells or different cell types. Which methods should be used to combine the different data? If this could be solved, we could analyse more effectively cell development and specialisation.”

### Computational methods for processing big data is a challenge for postgraduate students

Ahola’s aim is to provide more assistance in the use of computational methods.

The University of Eastern Finland’s Bioinformatics Center provides researchers computing capacity and helps researchers

in data pre-processing and analysis, and also assists in the use and installations of different computational methods and software.

“If there are no bioinformaticians in the same team or collaborative teams, researchers are expected to be proficient in computational methods and processing big data.”

Ahola admits that the requirements are tough, for example for postgraduate students.

Fortunately, the University of Eastern Finland has taken up this challenge by providing a Computational Biomedicine as an orientation option.

“One example of the reuse of data is Finnish biobanks, which contain the genomes of over half a million Finns. It’s not a simple matter to analyse the biobank data, because the amount of data is insane.”

Ahola is referring to the FinnGen research project, which was launched in autumn 2017. Its main goal is to increase understanding of the causes of diseases and promote their diagnosis, prevention and the development of treatment methods. FinnGen uses samples collected by all Finnish biobanks. By June 2023, more than 553,000 samples were collected for the FinnGen survey. The first phase of the research project lasted six years. There are only a few research projects of this scale in the world.

The research projects can combine genomic data with data from national health registries. Indeed, Finland has exceptionally good resources to carry out genetic research covering the entire population.

Clinical data from longitudinal studies combined with genetic data offers many opportunities. But there must be a lot of data.

“Data collections are needed because no single researcher can collect data from 10,000 or 100,000 individuals. If the dataset is smaller, it may not provide reliable information for studying genetically complex diseases.”

There are many research projects using different data sources underway at the University of Eastern Finland. A project on Alzheimer’s disease at the University of Eastern Finland and Kuopio University Hospital will combine clinical data collected from patient visits with FinnGen data. In this way, researchers are aiming to understand the biological mechanisms leading to the onset of Alzheimer’s disease.

“FinnGen’s biobank is a unique resource that could be used much more in research,” Ahola says.

“Another example of research on Alzheimer’s disease is a project with Rapta Therapeutics and UEF professors **Mikko Hiltunen** and **Annakaisa Haapasalo**. This project uses transgenic cell lines to study the

effect of different Alzheimer's treatments on protein function."

One interesting collaboration project is underway with Academy of Finland researcher **Kirsi Ketola**.

"The study investigates carboplatin treatment resistance mechanisms in prostate cancer. Carboplatin produces DNA cross-links, which lead to activation of a mechanism that repairs DNA and causes resistance, allowing cancer cells to divide again. The research uses single-cell techniques to measure both gene expression and chromatin changes at the single-cell level."

Chromosomes are located in the nucleus, in the form of long chromatin strands.

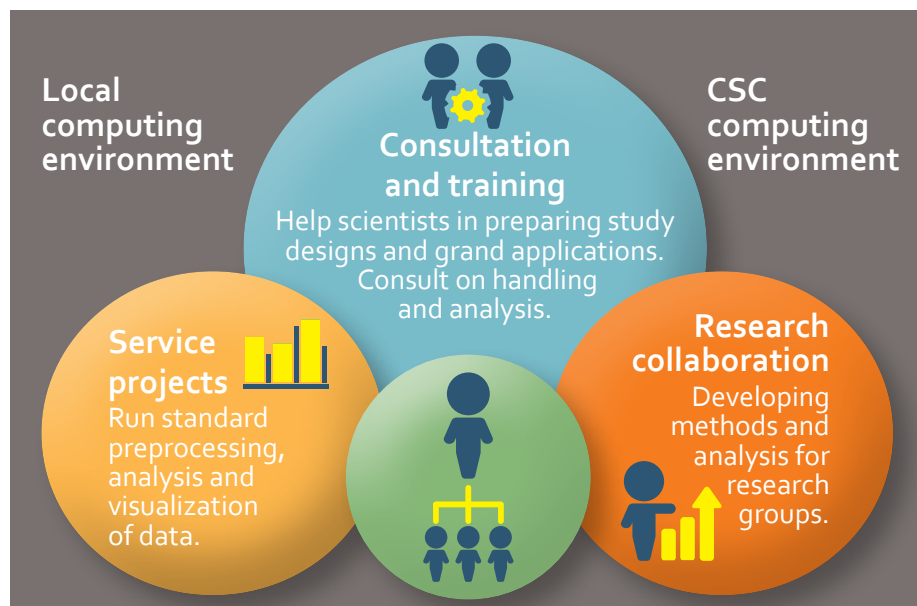
According to Ahola, careful data integration and analysis could promote development of personalised treatment.

### Data reuse requires know-how

Ahola is a tireless advocate for the openness and reuse of data, and for the development of methods and infrastructures that facilitate and encourage this. She cites the European Genome-phenome Archive (EGA) as an example. This is a data archive that makes it possible to share and, with permission, access biomedical data that has already been published.

"The archive contains human genomic data, combined with clinical and other metadata. Since in principle it may be possible to identify a person by genome and phenotype, data sharing is strictly regulated."

According to Ahola, the EGA allows for data sharing in the appropriate way. This makes it possible to reuse valuable biomed-



*Bioinformatics Core Facility, University of Eastern Finland.*

ical research data, for instance for creating or testing new research hypotheses.

"Existing data can be approached from a different perspective – for example, patients can be selected using different criteria than in an already published study, or data can be used as part of a larger data set."

Referring to Biocenter Finland, Ahola says that more should be done together. The centre brings together seven biocentres from different Finnish universities. It should not be impossible to increase collaboration between different biocentres and internationally, for example through the Finnish ELIXIR node CSC.

"ELIXIR is an avenue for us to network and learn from the experiences of other bi-

oinformatics core facilities, and to be part of discussions where research infrastructure issues are brought up and new initiatives are taken."

Because new technologies produce large and complex data sets, research infrastructures should also include data science experts, not research equipment alone.

"To make effective use of data, the computing capacity offered by CSC, for example, is not enough. Data processing and reuse also requires staff with expertise in the field. As I see it, better resourcing and systematic collaboration between biocentres could substantially facilitate and improve the processing, integration and reuse of large omics data sets."

23.8.2023 | **Ari Turunen**

### MORE INFORMATION:

#### Bioinformatics Center University of Eastern Finland

<https://uefconnect.uef.fi/en/group/bioinformatics-center/>

#### CSC – IT Center for Science

is a non-profit, state-owned company administered by the Ministry of Education and Culture. CSC maintains and develops the state-owned, centralised IT infrastructure.

<http://www.csc.fi>

<https://research.csc.fi/cloud-computing>

#### ELIXIR

builds infrastructure in support of the biological sector. It brings together the leading organisations of 21 European countries and the EMBL European Molecular Biology Laboratory to form a common infrastructure for biological information. CSC – IT Center for Science is the Finnish centre within this infrastructure.

<http://www.elixir-finland.org>

<http://www.elixir-europe.org>

#### ELIXIR FINLAND

Tel. +358 9 457 2821s • e-mail: [servicedesk@csc.fi](mailto:servicedesk@csc.fi)  
[www.elixir-europe.org/about-us/who-we-are/nodes/finland](http://www.elixir-europe.org/about-us/who-we-are/nodes/finland)

[www.elixir-finland.org](http://www.elixir-finland.org)

#### ELIXIR HEAD OFFICE

EMBL-European Bioinformatics Institute  
[www.elixir-europe.org](http://www.elixir-europe.org)