

Embargo until December 7, 16.00 CET

PRESS RELEASE

# THE HUMAN PROTEIN ATLAS

## AI-based efforts facilitate the mapping of the human building-blocks

**[December 7, 2022].** A new version of the open access Human Protein Atlas has been launched (version 22). Two new sections are introduced, both relying heavily on AI-based prediction modelling and machine learning. First, a new Human Disease Blood Atlas is launched using next generation blood profiling, initially aimed to facilitate research in the field of Cancer Prediction Medicine. Secondly, a new Structure resource section is launched using an AI-based prediction model (AlphaFold), to show the 3D structures for all human proteins. In addition, a major update of the Tissue Atlas section is released with detailed multiplex spatial profiling of proteins specific for single cell types in human testis or kidney. More data is also provided on single cell analysis of tissues and organs as well as data from an extensive catalogue of human cell lines.

The Human Protein Atlas consortium has today launched the version 22 of the open access resource for profiling of the human proteins ([www.proteinatlas.org](http://www.proteinatlas.org)). The resource explores different aspects of all proteins and contains 12 major sections, including two new sections; the Disease Blood Atlas section and the Protein 3D Structure section. In addition, all 5 million web pages in this open access resource have been updated.

### Disease Blood Atlas

This new section presents the result of a novel pan-cancer strategy to explore the proteome signature in blood from cancer patients. Plasma profiles of 1,463 proteins from more than 1,400 cancer patients representing 12 major cancer types were measured in minute amounts of blood plasma collected at the time of diagnosis and before treatment. Protein levels are based on analysis with proximity extension assay (Olink Explore) and targeted proteomics (mass spectrometry). We highlight proteins associated with each of the analyzed cancer types based on differential expression analysis as well as a machine-learning based disease prediction strategy. By combining the results from all cancer types, a panel of proteins suitable for the identification of individual cancer types based on a drop of blood is presented. Learn about in the new Disease Blood Atlas:

- comprehensive and accurate protein levels in blood covering 12 different cancer types
- the levels of proteins in blood using targeted proteomics and proximity extension assay
- proteins associated with each of the analyzed cancers

## **Structure section.**

This new section presents the three-dimensional structures of all human proteins based on the prediction model developed by the AlphaFold algorithm. AlphaFold is an artificial intelligence (AI) program developed by DeepMind, a subsidiary of Alphabet (Google), which performs prediction of protein structures. The program is designed as a deep learning system. In addition, experimentally verified structures from the Protein Data Bank (PDB) are also visualized. The binding site for the various antibodies used in the Human Protein Atlas can also be visualized for the individual proteins. Additionally, the amino acid positions of population variants and variants with known clinical relevance in are shown. Learn about in the new Structure section:

- the experimental and predicted 3D structure of human proteins
- the known missense variants with clinical significance
- the known missense variants in the population
- the antigen structure for a majority of antibodies

## **Tissue Atlas section**

The revised section has in this version implemented multiplex immunohistochemistry based on fluorescence high-resolution imaging, to study protein localization in single cell types of human testis or kidney. By combining deep annotation data and single-cell transcriptomics, three antibody panels have been developed to further characterize the spatial expression of proteins during testicular germ cell development (spermatogenesis), and highlight tubular and glomerular proteins in kidney. In the first version launched today, more than 500 protein targets have been analyzed in human testis or kidney. Learn about in the revised Tissue section:

- in-depth protein localization in testis and kidney based on multiplex profiling
- protein localization in tissues at a single-cell level
- a catalogue of genes enriched in a particular tissue (specificity)
- which genes have a similar expression profile across tissues (expression cluster)

## **Cell Line section**

The revised section now includes genome-wide data covering more than 1000 human cell lines analyzed either “in-house” or by the Cancer Cell Line Encyclopedia (CCLE). The data allows for researchers to select best cell lines for a particular application based on similarity to cancers, the presence or absence of various biological pathways and/or the presence of immunological molecules (cytokines). Learn about in the revised Cell Line section:

- a catalogue of genes enriched or lacking expression in a particular cell line (specificity)
- which cell line has the most consistent expression profile to its corresponding cancer tissue
- cancer-related pathway and cytokine activity of each cell line

Version 22 also contains a vast amount of new information within the various parts of the Human Protein Atlas, including revised summary pages for all human protein-coding genes and a new Methods Summary for the 12 sections with information how the data in each section have been

generated, analyzed and visualized. The strategies for dimensionality reduction and density-based clustering of co-expression patterns have been extended to explore the gene expression landscape and we present Expression UMAP clustering of all protein-coding genes.

“We believe that the new sections of the open access Human Protein Atlas with large amounts of novel data covering all human proteins provide new dimensions of valuable information for researchers interested in human biology and disease”, says Mathias Uhlen, Director of the Human Protein Atlas consortium. The work was funded by the Knut and Alice Wallenberg Foundation

Link to the new version of the Human Protein Atlas: [www.proteinatlas.org](http://www.proteinatlas.org)

**For more information, contact:** Mathias Uhlen (email: [mathias.uhlen@scilifelab.se](mailto:mathias.uhlen@scilifelab.se)) or Inger Åhlén (telephone +46-8-790 99 87)

## About

**Human Protein Atlas.** The Human Protein Atlas (HPA) is a program based at SciLifeLab (Science for Life Laboratory), Stockholm, that started in 2003 with the aim to map of all the human proteins in cells, tissues and organs using integration of various omics technologies, including antibody-based imaging, mass spectrometry-based proteomics, transcriptomics and systems biology. All the data in the knowledge resource are open access to allow scientists, both in academia and industry, to freely access the data for exploration of the human proteome. The Human Protein Atlas program has already contributed to several thousands of publications in the field of human biology and disease, and it has been selected by the organization ELIXIR ([www.elixir-europe.org](http://www.elixir-europe.org)) as a European core resource due to its fundamental importance for the wider life science community. The HPA consortium is funded by the Knut and Alice Wallenberg Foundation. For more information, see: [www.proteinatlas.org](http://www.proteinatlas.org)

**Knut and Alice Wallenberg Foundation.** The Knut and Alice Wallenberg Foundation is the largest private financier of research in Sweden and also one of Europe’s largest. The Foundation’s aim is to benefit Sweden by supporting basic research and education, mainly in medicine, technology, and the natural sciences. The Foundation can also initiate grants to strategic projects and scholarship programs. For more information, see: [kaw.wallenberg.org/en](http://kaw.wallenberg.org/en)

Figure 1. The home page of the version 22 of the Human Protein Atlas

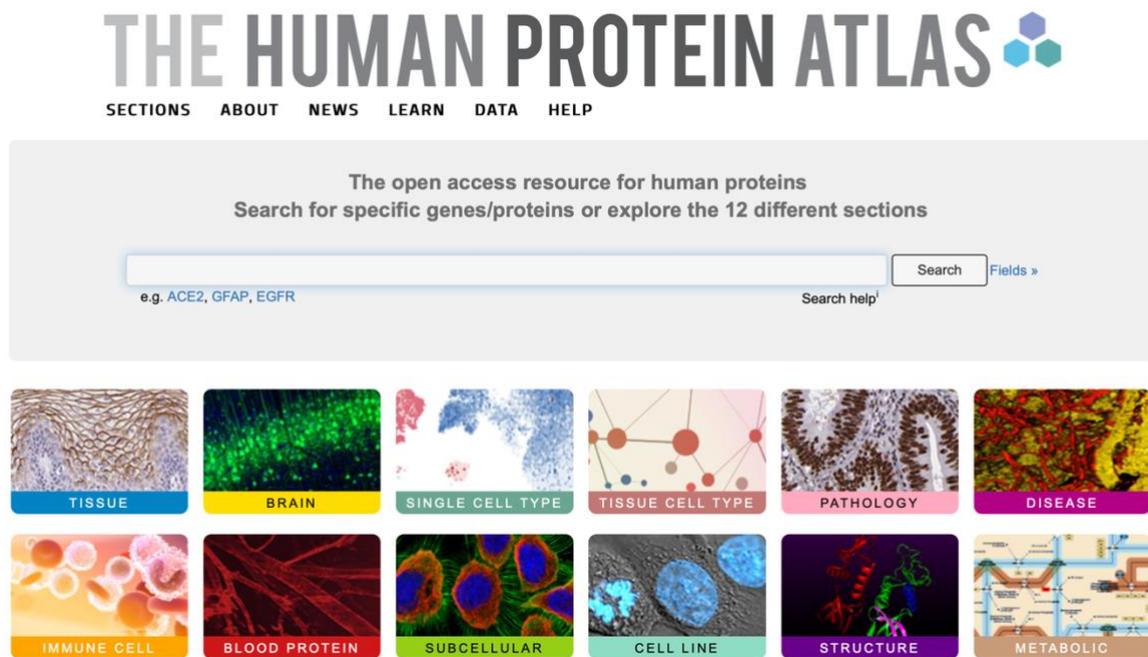


Figure 2. Multiplex single-cell protein profiling of human kidney tissue

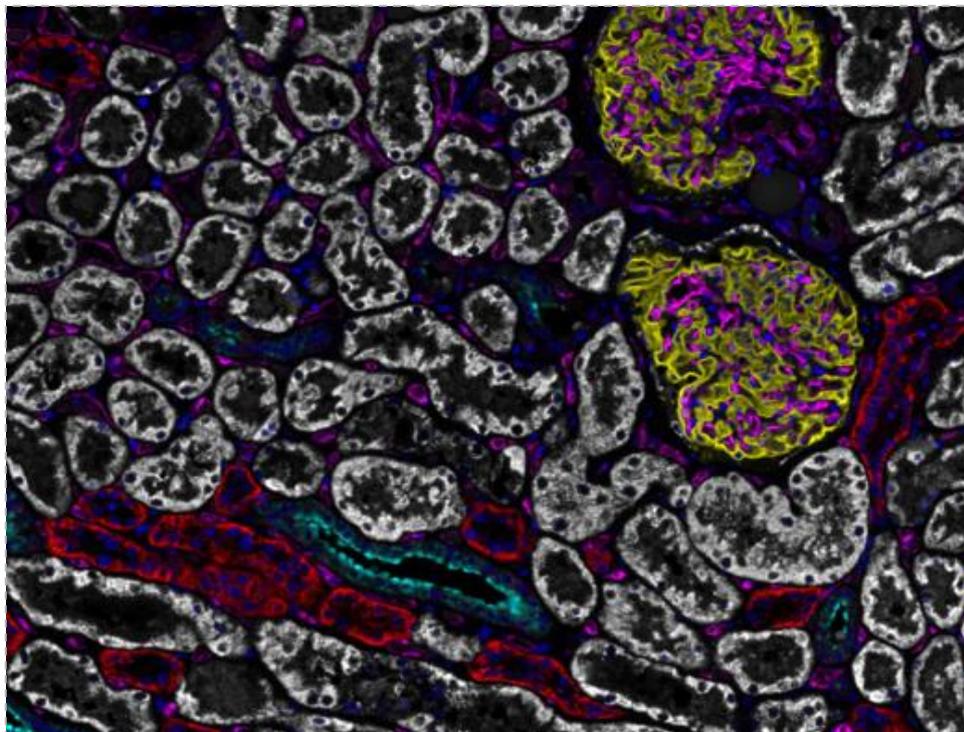


Figure 3. The new Structure section (AI-based predictions of 3D-structures). Example shows prediction of Her2 protein structure.

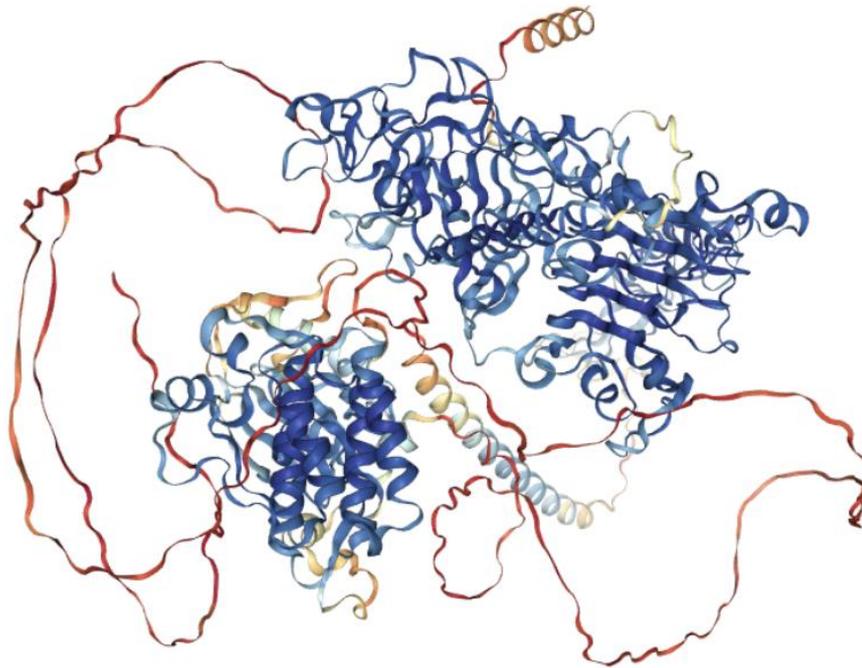


Figure 4. Towards next generation cancer prediction medicine. The example shows the elevated levels of the protein GFAP in the blood of patients with glioma.

