

Embargo until October 22, 2024, 16.00 CET

PRESS RELEASE

THE HUMAN PROTEIN ATLAS

A new version 24 of the open access Human Protein Atlas resource has been released at the HUPO meeting in Dresden, Germany

[October 22, 2024]. A new version of the open access Human Protein Atlas has been launched (version 24) including many new features. The data is summarized in eight resources harboring information about all human protein-coding genes. Altogether 5 million web pages and over 10 million manually annotated high-resolution bioimages are presented, including 16 knowledge summaries in which various aspects of the human proteome is reported. A new Blood Disease Atlas containing open access next generation blood profiling data from 59 diseases, including patients with cancers, autoimmune, infectious, neurobiology and cardiovascular diseases is launched. New features include multiplex profiling of human tissues and a new section presenting spatial transcriptomics of the human brain (cerebral cortex). AlphaFold source code has been used to predict the 3D structures of more than 80,000 protein isoforms and variant data from AlphaMissense has been integrated. In the Interaction resource the protein interaction data from the IntAct database has now been complemented with data from BioGRID, Bioplex and OpenCell. 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 24 of the open access resource for profiling of the human proteins (www.proteinatlas.org). The resource explores different aspects of all proteins and contains eight major resources

The Blood resource

This new resource combines the data from the former Blood and Disease sections together with extensive novel information on blood profiles in patients with various diseases. The resource presents the result of a novel pan-disease approach to explore the proteome signatures in blood in patients from various major disease areas (Autoimmune, Infection, Metabolic, Cardiovascular, Psychiatric, and Pediatric). Plasma profiles of 1162 proteins, from more than 6,121 patients representing 59 diseases, were measured in minute amounts of blood plasma collected at the time of diagnosis and most often before treatment. Protein levels are based on analysis with proximity extension assay (Olink Explore) and targeted proteomics with spike in of stable isotope labeled standards (mass spectrometry). Based on differential expression analysis, we highlight proteins associated with each disease type analyzed. By combining the results from all cancer types, a panel of proteins suitable for identifying individual cancer types based on a drop of blood is presented.

Learn about:

- comprehensive and accurate protein levels in blood covering patients from 59 diseases
- the levels of proteins in blood using targeted proteomics and proximity extension assays
- proteins associated with elevated levels for each of the analyzed diseases

The Brain resource

The Brain resource is dedicated to the central nervous system with a focus on comparison between regions and cell-types of the brain. Over the years the Brain resource has increased completeness by providing the most extended overview of protein expression in 200 micro-dissected areas of the brain based on bulk RNA-sequencing data. In the current version, the first high-resolution spatial transcriptomics data of the human cerebral cortex is launched. With a capture resolution of 0.5 micrometer, a both cellular and spatial information on protein expression is provided for the first time, using tools to quantify, impute and visualize protein expression in the human cerebral cortex.

Learn about:

- proteins expression in different regions of the brain
- protein expression in different cell-types of the human cerebral cortex
- single cell analysis of cortex using spatial transcriptomics

The Subcellular resource

The Subcellular resource has in this version launched a major update in terms of a Cilia Atlas. The proteome and its subcellular organization have been systematically mapped in primary cilia across different cell lines, and in the flagellum of human sperm. With the help of high-throughput immunofluorescence imaging, we have been able to identify hundreds of proteins localized to these compartments and subcompartment localization.

The primary cilium, an antenna-like organelle with sensory capabilities, extending as a solitary unit from the surface of nearly all vertebrate cell types. We have identified 653 proteins within the primary cilium and its substructures. Intriguingly, the primary cilium stands out for exhibiting the highest degree of proteome heterogeneity among subcellular structures and organelles. This heterogeneity manifests in multiple features including cell type heterogeneity, single-cell heterogeneity, and multi-localization, indicating the fundamental role of the primary cilium in receiving and conveying cell-specific signals.

The sperm cell is the male reproductive cell, highly specialized in motility and fertilization of the female egg. To achieve this, sperm cells contain several unique subcellular structures, including a long motile cilium called the flagellum, promoting high motility and aiding egg penetration, and a sac containing enzymes crucial for penetrating the egg called the acrosome. We have identified 645 proteins to localize to specific substructures of the human sperm cell. Interestingly, we observe extensive single-cell heterogeneity for these proteins underscoring the importance of functional diversity within a sperm population.

Learn about:

- proteins localized to the primary cilia
- proteins localized to the flagellum and other sperm-specific subcellular compartments
- single-cell heterogeneity of the metabolic proteome

The Single cell resource

The new Single Cell Resource consolidates the previously separate sections—Single Cell Type, Tissue Cell Type, and Immune Cell—into a unified structure. Additionally, it now includes Single Nuclei Brain data, representing expression profiles of 2.5 million brain cells. This resource offers a comprehensive single-cell overview of all protein-coding genes. The Tissue Cell Type dataset has been expanded to include new cell type enrichment data for the spleen, salivary glands, adrenal glands, and pituitary gland. Furthermore, the resource now incorporates external cluster data from the Tabula Sapiens project, allowing users to compare their findings with the clustering and cell annotation provided by the Human Protein Atlas (HPA) pipeline. With the inclusion of the brain single nuclei data, users can explore cell clusters across 11 distinct brain regions and compare these results with the cell type specificity across other datasets within the resource.

Learn about:

- the Single cell type data, providing the expression profiles across 81 cell types from 31 human tissues, and cell type specificity based on gene expression.
- the Tissue cell type section, providing the predicted cell-type expression specificity based on bulk RNAseq
- the Single nuclei brain section, providing more details regarding cell type specificity within the brain, based on single nuclei RNAseq
- the Immune cell section, providing expression comparison between sorted immune cells

The Structure and interaction resource

This new resource contains data from the former Structure and Interaction sections. The Structure resource presents the three-dimensional structures of most human proteins, and now also their related isoforms, based on in-house predictions generated using the AlphaFold source code. The ProteinBrowser tool can be used to highlight features such as antigen sequences for HPA antibodies, InterPro domains and membrane regions in the structures. The positions of experimental and predicted clinical and population missense variants from Ensembl Variant and AlphaMissense can also be explored. The Interaction resource presents data on protein-protein interactions and metabolic networks. The interaction networks presented are now based on data from four different sources, IntAct, BioGRID, BioPlex and OpenCell, that have been integrated with data regarding protein expression, location and classification.

Learn about:

- the predicted 3D structure for more than 83,000 protein isoforms
- the structure of selected protein features and the position of missense variants
- the interaction partners of more than 15,000 proteins
- the expression, location and other features of all proteins in the interaction networks

- the pathways related to 2,900 metabolic genes and the expression profiles of the genes in each metabolic pathway

The Tissue resource

The Tissue resource has in this version launched a major update of multiplex tissue profiling using immunohistochemistry-based fluorescence high-resolution imaging. With the help of single-cell transcriptomics and an iterative staining-stripping method, seven antibody panels have been developed to study protein localization in single-cell types, cell states and subcellular structures in eight different human tissue types. Based on overlap of antibody staining with cell-specific markers, the dataset launched today shows the detailed spatial location of 612 proteins during germ cell development in testis, 178 proteins in different structures of motile ciliated cells, 162 proteins in different tubular and glomerular cells in kidney and 121 proteins in glandular and ductal cells in salivary gland.

Learn about:

- in-depth protein localization of 1,021 proteins using multiplex tissue profiling
- protein localization in tissues, single-cell types, cell states and subcellular structures
- a catalogue of genes enriched in a particular tissue (specificity)
- which genes have a similar expression profile across tissues (expression cluster)

The Cancer resource

The Cancer resource includes the analysis of the expression profiles of 6,918 patients across 21 cancer types using the gene annotations. The refined approach enabled us to offer an updated list of prognostic genes for several of the major human cancers. To strengthen the reliability of the cancer resource, data from 10 independent cancer cohorts were integrated, creating a cross-validated, reliable collection of prognostic genes. The updated resource lays the foundation for precision oncology and the development of personalized treatment strategies.

Learn about:

- the updated correlations between gene expression and survival outcomes using global gene expression profiling.
- independent datasets from 10 different cancer types were compiled to identify a robust set of confidence prognostic genes (CPGs).
- a substantial portion (53.6%) of protein-coding genes were expressed in all cancers analysed, while an additional 12.1% of genes were not detected in any of the cancer types.

The Cell Line resource

The resource now includes genome-wide data covering more than 1,200 human cell lines analyzed either “in-house” or through external resources such as the Cancer Cell Line Encyclopedia (CCLE). The data enables for researchers to identify the best cell lines for particular applications based on similarity to cancers, the presence or absence of various biological pathways and/or the presence of immune signalling molecules (cytokines). The new version complements the data with protein analysis from the CCLE covering 511 cell lines.

Learn about:

- 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

The Knowledge Summaries

Version 24 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 8 resources 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. In addition, we provide 15 knowledge summaries in which the HPA team has assembled data covering topics of high biological or medical interest:

1. The Disease Blood Atlas – explore the proteins profiles in blood from patients with various diseases
2. 3D-structures of proteins – explore the structures of human proteins isoforms
3. The cell and tissue specific proteome – explore all proteins elevated in tissues and cell types
4. The human secretome – explore all proteins predicted to be secreted from human cells
5. The human membrane proteome – explore all proteins predicted to be membrane-bound
6. The house-keeping proteome – explore the proteins essential for all cells
7. The human protein classes – explore the profiles of various protein classes
8. Evidence of the human protein-coding genes – explore the evidence for each of the proteins
9. The right cell line for your experiment – explore the protein profiles in human cell lines
10. The druggable proteome – explore the protein profiles of targets for human pharmaceuticals
11. The cancer proteome – explore the expression of proteins involved in human cancers
12. Transcription factor landscape – explore the cell specificity of all human transcription factors
13. Multiplex tissue profiling – explore the single-cell type-specific spatial location of proteins expressed in testis, kidney, salivary gland and ciliated cells
14. Spatial transcriptomics of the brain – explore the expression in the human cerebral cortex
15. Cilia and basal bodies – explore the detailed subcellular localization of proteins in these cells
16. Sperm and flagella - explore the detailed subcellular localization of proteins in these cells

“The HPA team is proud to launch a lot of new data in the new version of the open access Human Protein Atlas to provide extended knowledge about the building-blocks of humans; the proteins”, 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

For more information, contact: Mathias Uhlen (email: mathias.uhlen@scilifelab.se) or Åsa Sivertsson (email: asa.sivertsson@scilifelab.se) or Gustav Ceder (email: gustav.ceder@gmail.com).

Data availability

All data in the Human Protein Atlas is publicly available and to further increase and simplify access we have now summarized and structured the data available for each resource and included downloadable files with related meta data. The complete data is available in XML format and in collaboration with [BioImage Archive](#) also >10M raw images, corresponding to about 300TB of data, are now available and can be used as a resource for bioimage analysis.

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) as a European core resource due to its fundamental importance for the wider life science community and by the GCBR as a Global Core Biodata Resource. The HPA consortium is funded by the Knut and Alice Wallenberg Foundation. For more information, see: 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: <https://kaw.wallenberg.org/en>

Figure 1. The home page of the version 24 of the Human Protein Atlas.

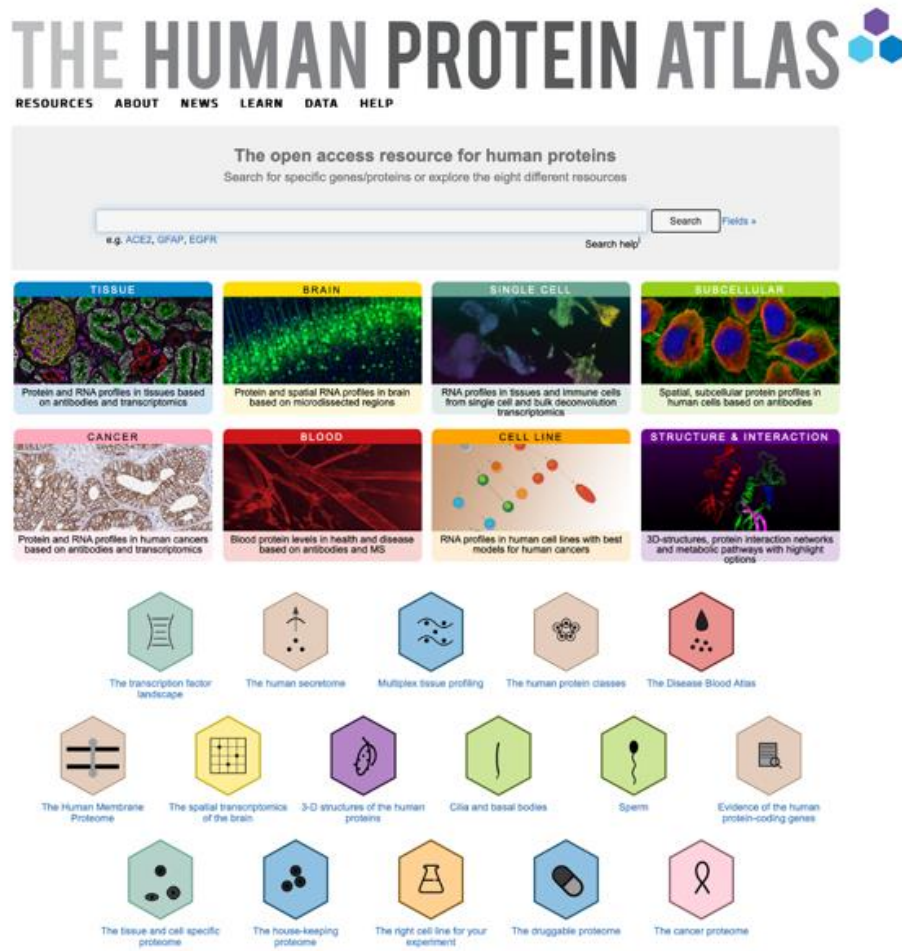


Figure 2. Multiplex single-cell protein profiling of human kidney.

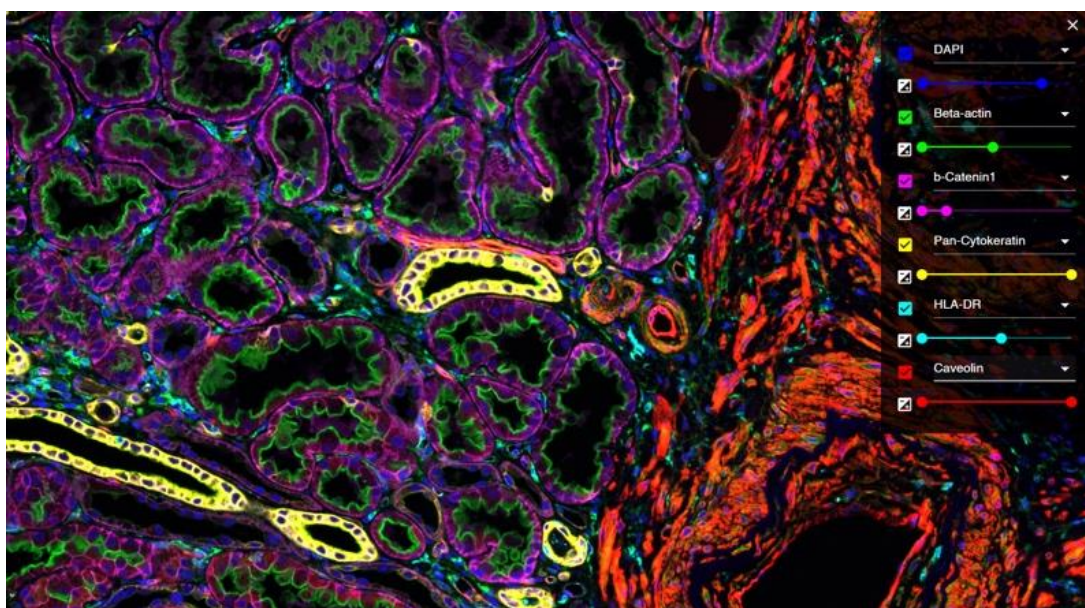


Figure 3. The new Structure section (AI-based predictions of 3D-structures). Example shows prediction of Her2 protein structure, main form and an example of a splice variant.

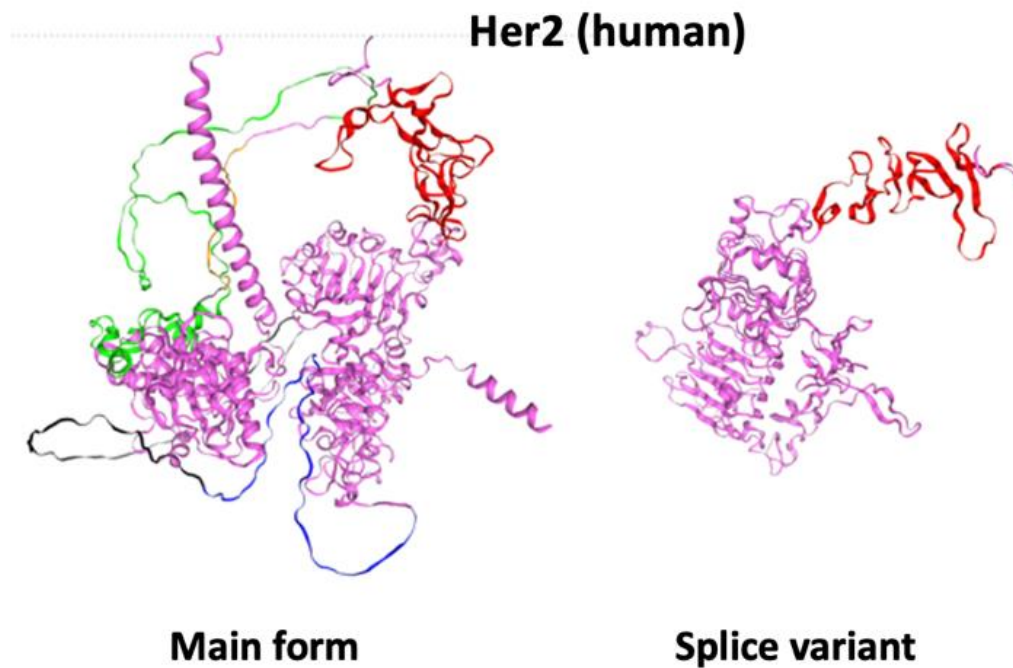


Figure 4. Towards next generation pan-disease prediction medicine. The example shows the blood profiles of the inflammation protein IL-6 across 59 human diseases and healthy controls.

