The Human Protein Atlas – mapping the building-blocks of humans

Mathias Uhlen
Science for Life Laboratory (KTH and KI)
Stockholm, Sweden
Disclaimer

- 20 start-up companies
- Atlas antibodies, Affibody, Abclon, ScandiBio Therapeutics
- AstraZeneca, GE Health

- Professor KTH, KI and DTU
- Director of Human Protein Atlas
- Founding Director of Science for Life Laboratory
- Member of the Royal Academy of Science (Sweden)
- Member of the National Academy of Engineering (USA)
- President of the European Federation of Biotechnology
<table>
<thead>
<tr>
<th>Century</th>
<th>Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>18th</td>
<td>Biology</td>
</tr>
<tr>
<td>19th</td>
<td>Chemistry</td>
</tr>
<tr>
<td>20th</td>
<td>Physics</td>
</tr>
<tr>
<td>21st</td>
<td>Medicine</td>
</tr>
</tbody>
</table>
Periodic Table (chemistry)

- Dimitri Mendeleev
- 150 year anniversary
- Prediction of missing elements
- 118 elements discovered (2019)
The human genome (2003)

- Few protein-coding genes (23,000)
- Blue-print for human biology and diseases
The human proteome – making a "periodic table" of the proteins

• Proteins – the building blocks of human biology
• Targets for all pharmaceutical drugs
• Targets for future precision medicine efforts

• What are the building-blocks of tissues and organs?
• What is the building-blocks of the cell?
• What are the targets for future drugs and diagnostics?
Content

1. The Human Protein Atlas – an update
2. The Human Proteome – an update
3. The human Secretome – a resource of secreted proteins
4. Precision medicine (wellness profiling)
5. Biologicals for drug treatment
The Human Protein Atlas – an update
The Human Protein Atlas

- Map of all human proteins in cells, tissues and organs (including cancer)
- Open knowledge resource for all researchers in academia and industry
- Started in 2003
- Funded by Wallenberg Foundation
Sweden and Asia

- SciLifeLab Solna (KTH)
- AlbaNova (KTH)
- Rudbeck (Uppsala)
- Neuoscience (Karolinska)
- Systems biology (Chalmers)

- South Korea (production of antibodies)
- China (transcriptomics and antibodies)
- India (pathologists for annotation)
Research factory

- 60,000 recombinant proteins (produced in E.coli and CHO cells)
- 55,000 antibodies (affinity-purified on the antigen)
- 21,000 validated antibodies for bioimaging of tissues and cells
- Integration with transcriptomics

Open access
More than 10 million images

Antibody-based bioimaging - "in-house" generation of 55,000 antibodies


More than 3000 citations (Google Scholar)

- Single cell resolution
- Context of neighboring cells
- In vivo analysis (tissues)

More than 400 citations (Google Scholar)

- Subcellular resolution (confocal)
- Single cell variation
A pathology atlas of the human cancer transcriptome

Matthew Uhl et al.

INTRODUCTION: Cancer is a leading cause of death worldwide, and there is great need to define the molecular mechanisms driving the development and progression of its distinct forms. The Pathology Atlas of Human Cancer Proteomes (Pathology Atlas) was developed to provide a systematic, high-quality, and comprehensive overview of the transcriptional and proteomic responses to tumors, including patient-to-patient and tumor-to-tumor variability. The atlas is designed to facilitate downstream analysis and interpretation of specific cancer-related transcripts.

RATIONAL: To address this need, we used a standardized approach to analyze the transcriptional profiles of five hundred cancers with respect to their relevant tissues and to generate a cancer-specific transcriptome database, enabling the discovery of novel biomarkers and therapeutic targets.

RESULTS: This study was made possible through the availability of large-scale cancer genomics and proteomics datasets, including the Cancer Genome Atlas (TCGA) and the Human Protein Atlas (HPA). Here, we used these datasets to generate a comprehensive transcriptome of 107 distinct cancer types, thereby facilitating the identification of novel biomarkers and therapeutic targets.

CONCLUSION: The Pathology Atlas has been created as part of the Human Protein Atlas project to explore the potential of high-throughput proteome analysis in cancer research and to identify novel therapeutic targets. The atlas provides a valuable resource for cancer research and clinical applications, allowing for a better understanding of cancer biology and the development of personalized therapies.

More than 400 citations (Google Scholar)
The Human Protein Atlas three separate parts

<table>
<thead>
<tr>
<th>Atlas</th>
<th>Description</th>
<th>Key publication</th>
</tr>
</thead>
</table>

Open access
Visitors from academia and industry

- 300,000 visitors per month
- 10 publications every day citing HPA
Human Protein Atlas – new additions

Blood Atlas – what proteins and cells are present in blood

Brain Atlas – what proteins are localized to different regions of the brain

Metabolic Atlas – what metabolic pathways are active in different tissues

Launched September 5, 2019
Map of human blood cell types

- 1,448 blood cell enriched genes
- 271 “specific” for a single cell type

Blood cell map contributors:
- Linn Fagerberg (KTH)
- Petter Brodin (Karolinska)
- Max Karlsson (KTH)
- Wen Zhong (KTH)
- Abdellah Tebani (KTH)
- Fredrik Edfors (KTH)
- Åsa Sivertsson (KTH)
- Jacob Odeberg (KI/KTH)
- Martin Zwahlen (KTH)
- Per Oksvold (KTH)
- Kalle von Feilitzen (KTH)

... and many others ...

Uhlen et al (2019A), in review
Blood specific genes

Categories based on expression (transcriptomics)
Regulatory T-cell enriched genes

CTLA4

FOXP3

CCR8

BFSP2
Molecular organization of the brain

- Organ level
- Region level
- Cellular level (cell types)
- Subcellular level (organelle)

- Proteomics (antibody-based)
- Transcriptomics

- Physiology
- Disease
A mammalian brain atlas

- Create a map of the gene expression of the mammalian brain
- Identify brain relevant genes for in-depth studies
- Identify species difference (human, pig and mouse)

**Brain Atlas contributors:**

- Jan Mulder (KI)
- Evelina Sjöstedt (KI)
- Tomas Hökfelt (KI)
- Yonglun Luo (BGI, China)
- Csaba Adori (KI)
- Linn Fagerberg (KTH)
- Wen Zhong (KTH)
- Martin Zwahlen (KTH)
- Per Oksvold (KTH)
- Kalle von Feilitzen (KTH)

...and many others...

Sjöqvist et al (2019), in review
RNA vs Protein mapping

Allen ISH

HPA IHC
Normalization of data from several sources and technology platforms

Human brain samples (n=1,047)
Comparisons of the brains in human, pig and mouse

Transcription factors

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>relative expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMX1</td>
<td></td>
</tr>
<tr>
<td>BHLHE22</td>
<td></td>
</tr>
<tr>
<td>ISL1</td>
<td></td>
</tr>
<tr>
<td>SIX6</td>
<td></td>
</tr>
<tr>
<td>PITX3</td>
<td></td>
</tr>
<tr>
<td>HOXB5</td>
<td></td>
</tr>
<tr>
<td>PAX2</td>
<td></td>
</tr>
<tr>
<td>NEUROD1</td>
<td></td>
</tr>
<tr>
<td>EOMES</td>
<td></td>
</tr>
</tbody>
</table>

Human | Pig | Mouse

Neurotransmitters

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>relative expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAT</td>
<td></td>
</tr>
<tr>
<td>DBH</td>
<td></td>
</tr>
<tr>
<td>TPH2</td>
<td></td>
</tr>
<tr>
<td>HTR5A</td>
<td></td>
</tr>
<tr>
<td>HTR5B</td>
<td></td>
</tr>
<tr>
<td>PENK</td>
<td></td>
</tr>
<tr>
<td>OPRD1</td>
<td></td>
</tr>
<tr>
<td>OPRM1</td>
<td></td>
</tr>
<tr>
<td>OPN4</td>
<td></td>
</tr>
</tbody>
</table>

Human | Pig | Mouse

Serotonin receptors

Opioid receptors
Single cell 3D-imaging of human brains

Tomas Hökfelt
Csaba Adori
Jan Mulder
Evelina Sjöstedt

HPA Neuro group
Karolinska Institutet
Tyrosine-hydroxylase in human locus coeruleus

Neurons involved in the regulation of mood, sleep and attention

HPA group, unpublished
Human Alzheimer brain (neocortex)

Green – pyramidal neurons

Red – beta-amyloid
The Metabolic Atlas

- 6,793 reactions
- 4,027 metabolites
- 3,316 genes (enzymes)

Jens Nielsen
Jon Robinson
Mihail Anton

Chalmers
Gothenburg, Sweden
The Brain Atlas explores the protein expression in the mammalian brain by visualization and integration of data from three mammalian species (human, pig, and mouse). Transcriptomic data combined with affinity-based protein in situ localization down to single cell detail is here available in a brain-centric sub atlas of the Human Protein Atlas. The data focuses on human genes and one to one orthologues in pig and mouse. Each gene is provided with a summary page, showing available expression data (mRNA) for summarized regions of the brain as well as protein location for selected targets. High resolution staining images as well as expression data for the individual sub regions are all available for exploring the most complex organ.

**THE BRAIN**

Gene classification based on expression in tissue types representing the whole human body enables the description of brain elevated proteins. Regional expression data is used for further brain - in depth classification, highlighting the complexity of the brain. Regional classification is performed in human, pig and mouse brain separately by comparing transcriptomic data summarized into 10 main regions of the brain. The regional classification in human brain is also compared to whole-body expression. The combination of transcriptomic data and antibody-based protein profiling is investigated on separate summary pages as a platform for further exploring the brain proteome.

**BRAIN REGIONS**

Brain samples are grouped into 10 anatomical regions, providing regional classification of ~9,000 genes based on RNA expression, indicating which proteins are elevated in one region compared to the other.

---

The Blood Atlas contains single cell type information of genome-wide RNA expression profiles of human protein-coding genes covering various B- and T-cells, monocytes, granulocytes and dendritic cells. The single cell transcriptomics analysis covers 18 cell types isolated with cell sorting followed by RNA-seq analysis. In addition, an analysis of the “human secretome” is presented including annotation of the genes predicted to be actively secreted to human blood, as well as the annotation of proteins predicted to be secreted to other parts of the human body, such as the gastric tract and local compartments. An analysis of the proteins detected in human blood are also presented with an estimation of the respective protein concentrations determined either with mass spectrometry-based proteomics or antibody-based immune assays.

---

**THE HUMAN BLOOD CELLS**

---

The Metabolic Atlas portion of the Tissue Atlas enables exploration of gene function and tissue specific gene expression in the context of the human metabolic network. For proteins involved in metabolism, a metabolic summary is provided that describes the metabolic subsystems/pathways, cellular compartments, and number of reactions associated with the protein. Over 120 manually curated metabolic pathway maps facilitate the visualization of each protein’s participation in different metabolic processes. Each pathway map is accompanied by a heatmap detailing the mRNA levels across 37 different tissue types for all proteins involved in the metabolic pathway.

**METABOLIC MAPS**

Maps are organized by individual pathways to facilitate visualization of metabolic areas of interest. Further detail and full cellular compartment maps are available at [metabolicatlas.org](http://metabolicatlas.org).

---

Table 1. A complete list of the pathways details in the Metabolic Atlas and number of enzymes involved in each pathway.

<table>
<thead>
<tr>
<th>Pathway</th>
<th># genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acyl-CoA hydrolisis</td>
<td>8</td>
</tr>
<tr>
<td>Acylglycerides metabolism</td>
<td>38</td>
</tr>
</tbody>
</table>
2.
The Human Proteome – an update
New classification of all human genes (tissue specificity)

4,482 genes (23%) are enriched in human tissues or organs

Only 586 genes (3%) are “specific” for one tissue (including insulin, troponin and PSA)

Hepataoma derived growth-like factor 1 (testis)

Uhlen et al, in review
Deep annotation of testis-specific proteins

- In-depth characterization of 500 testis elevated genes
- Detailed analysis of spatial protein localization in 8 testicular cell types

Dr. Cecilia Lindskog
Head Tissue Atlas
Human Protein Atlas

Dr. Charles Pineau
Inserm, France
Cells in seminiferous ducts (testis)

- **VCY1B**: All germ cells
- **TSPY1**: Spermatogonia and preleptotene spermatocytes
- **SHCBP1L**: Pachytene spermatocytes, round/early spermatids and elongated/late spermatids
- **SLCO6A1**: Sertoli cells, round/early spermatids and elongated/late spermatids
- **SYCP3**: Pachytene spermatocytes
- **SOX30**: Round/early spermatids
- **BPIFA3**: Elongated/late spermatids
- **TEX19**: Sertoli cells
Protein expression in premeiotic cells

57 proteins

11 proteins with unknown function

10 "missing proteins"

Involved in cell division and differentiation
Testis
Multiplex staining
Map of the tissue enriched genes

n=4,482

Uhlen et al, in review
How many proteins in humans?

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein-coding genes</td>
<td>19,670</td>
<td>The protein existence confirmed for 17,723 genes (90%)</td>
</tr>
<tr>
<td>Splice variants (isoforms)</td>
<td>82,271</td>
<td>So far, few examples of new functionalities (but interesting to explore)</td>
</tr>
<tr>
<td>Protein modifications</td>
<td>&gt;200,000</td>
<td>Modulate activity in enzymes and signal pathways</td>
</tr>
<tr>
<td>Somatic re-arrangements</td>
<td>&gt;20,000,000</td>
<td>The creation of immunological memory (IgG and T-cell receptors)</td>
</tr>
</tbody>
</table>

Status September 2019
Evidence for protein-coding genes

Altogether 19,670 predicted protein-coding genes (September 2019)

- 17,723 with evidence on the protein level (mainly antibody-based)
- 1,833 with evidence on transcriptional level
- 114 with no evidence (keratins, olfactory receptors and AC genes)
Evidence for protein existence – chromosome summary

Data from:
- HPA
- UniProt
- NextProt
- PeptideAtlas (MS)

www.proteinatlas.org
3.
The human secretome project
THE SECRETOME AND MEMBRANE PROTEOME

- 3000 secreted proteins
- 5500 membrane-bound proteins

Uhlen et al Science, 2015
The Human Secretome Project (HSP)

**Overall objective:**

- Production of all human secreted proteins
- High-throughput production in CHO cells
- Create a resource of reagents for drug discovery and development
3017 genes have been generated with synthetic biology

1600 bioactive proteins have been produced in CHO cells.

Phenotypic assay have been run in collaboration with AstraZeneca
How many secretome proteins in humans?

Åsa Sivertsson

Uhlen et al, submitted
The human secretome

Human secretome contributors:

- Åsa Sivertsson (KTH)
- Sophia Hober (KTH)
- Hanna Tegel (KTH)
- Fredrik Edfors (KTH)
- Andreas Hober (KTH)
- Jochen Schwenk (KTH)
- Adil Mardinoglu (KTH)
- Wen Zhong (KTH)
- Cheng Zhang (KTH)
- Peter Nilsson (KTH)
- Linn Fagerberg (KTH)
- Martin Zwahlen (KTH)
- Per Oksvold (KTH)
- Kalle von Feilitzen (KTH)

...and many others...

All input from community welcome

Uhlen et al (2019B), in review
Function of the human blood secretome (n=729)
Human plasma proteins (part of the Blood Atlas)

The estimated protein concentrations of proteins detected in human plasma based on:

1. AB-based immunoassays
2. Mass spectrometry-based proteomics
3. Ab-based Proximity Extension Assay
Objective: make assays to the whole blood proteome

Total: 729 blood proteins

139 not detected by any of the three platforms
4.

Precision medicine – an introduction
Precision medicine

Right treatment to right patient

More targeted treatment with less side-effects (biologicals)

Better diagnostic methods for analysis of health and disease
Diagnostic tools in hospitals and primary care

**Classical analysis**
- Blood sedimentation rate
- Blood pressure
- Puls
- EKG
- Oxygen levels (in blood)
- Spirometry
- Colonoscopy
- Ultrasound
- X-ray
- CRP (inflammation)
- Urine stick
- DNA-tests
- Troponin

**New analysis**
- Streptokock (quick tests)
- Sexual diseases (home kits)
- DNA-sequencing (nisch applications)
- Glucose – real time measurements (diabetes)
- Helicobacter (breath)
- Medical imaging
Probability of success in clinical phase transitions (n=9,985) with biomarkers involved in patient stratification

Source: Taiho pharma (unpublished)
Proteomics profiling of blood proteins

• Criteria:
  • More than 1000 targets
  • Precision - low technical variance (CV)
  • Specificity (low off target binding)
  • Multiplex (parallel) assays
  • Sensitivity (cytokine levels)

• Two competing platforms:
  • Olink (Uppsala, Sweden)
  • Somalogic (Boulder, Colorado, US)

• Not (yet) competitive:
  • Sandwich assays (ELISA etc)
  • Luminex
  • Proteomics (MS-based)
The Swedish SCAPIS SciLifeLab Wellness Profiling (S3WP) program

**Vision**

*To define the “wellness profile” of individuals using state-of-the art molecular analyses*

- Combine “classical” diagnostics, advanced imaging and new omics technologies
- Detect early signs of diseases
- Guide individualized treatments
Varying omics features by Subject - Wellness Cohort

Percentage of varying omics features with scaled value > 2SD

<table>
<thead>
<tr>
<th>Clinical Chemistry</th>
<th>Immune Cytome</th>
<th>Metabolome</th>
<th>Autoantibodies</th>
<th>Microbiome</th>
<th>Proteome</th>
<th>Transcriptome</th>
<th>Lipidome</th>
<th>Total</th>
</tr>
</thead>
</table>

Variance | Scale | Varying
Plasma protein profiling

- Olink multiplex panels based on proximity extension assay (PEA)
- 397 samples run in 11 panels with 92 proteins in each
- Longitudinal data for ~1000 proteins

Leptin

Folate Receptor 3

r=0.89

r=0.99
Wellness healthy cohort – protein examples
Longitudinal profiling of samples - UMAP

Proteome (n= 944)
Clinical chemistry (n= 67)
Autoantibodies (n= 1,456)
Metabolome (n= 133)

Microbiome (n= 1324)
Transcriptome (n= 944)
Immune cytome (n= 118)
Lipidome (n= 87)
Wellness Cohort – Mixed effect Modeling
Olink data – mixed effect modelling
Precision medicine effort

- Type 2 Diabetes
- Wellness healthy cohort
- Cardiovascular disease
- NAFLD (NASH)
- Preterm children

Preterm children
Type 2 Diabetes study

- Newly diagnosed T2D and treatment naïve at Visit 1
- Either elevated fasting glucose, elevated OGTT glucose, or both
- 52 subjects included (21 females and 31 males)
- 34/52 subjects were treated with metformin after Visit 1
- All were given lifestyle advice according to standard routine for T2D management
Healthy vs T2D – proteomics (Olink)

Wellness visit 1-6 (two years)

T2D visit 1-3 (three months)
Longitudinal Integrated Profiling of Preterm Children

Cohort
- 14 neonates from “Donna Mega” cohort from Queen Silvia Children's Hospital in Gothenburg, Sweden
- Samples collected 2013 – 2015
- Extremely preterm babies, gestational age 22-27 weeks
- Collected serum + feces

Blood samples:
#1: day 0 = cord blood
#2: day 1
#3: day 7 = 1 week
#4: day 14 = 2 weeks
#5: day 28 = 3 weeks
#6: gestation week 32
#7: gestation week 36
#8: gestation week 40
Profiling based on 459 proteins

PNA = Postnatal age
PMA = Postmenstrual age
Longitudinal Integrated Profiling of Preterm Children
Longitudinal Integrated Profiling of Preterm Children
Fat liver disease (NAFLD) – clinical trial

ScandiBio Therapeutics

Adil Mardonpulu
Jan Boren
Fat liver disease (NAFLD) – clinical trial

Functions as a major metabolic regulator. The protein stimulates the uptake of glucose in adipose tissue.

Moved to phase 2 clinical trials

Fibroblast growth factor 21
Take-home messages

• Protein profiling very important tool for precision medicine
• New tools for comprehensive and quantitative protein profiling
• Each healthy individual has a stable and unique protein profile
• Dramatic changes upon life style changes (and health changes)
• Dramatic changes upon drug treatment
• Dramatic changes in the pre-term babies
Multi-omics integration and wellness profiling

Mission: to perform integrative omics analysis based on precision medicine data as well as the Human Protein Atlas

Linn Fagerberg
Group leader

Abdellah Tebani
Post-doc

Wen Zhong
Post-doc

Max Karlsson
PhD student

Division of Systems Biology
Department of Protein Science
Science for Life Laboratory
KTH Royal Institute of Technology
5. Concluding remarks
## Mapping of human building-blocks

<table>
<thead>
<tr>
<th>Initiative</th>
<th>Funding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human Protein Atlas (Europe/Asia)</td>
<td>Wallenberg Foundation</td>
</tr>
<tr>
<td>Allen Brain and Cell Atlas (USA)</td>
<td>Paul Allen (Microsoft)</td>
</tr>
<tr>
<td>Human Cell Atlas (US and Europe)</td>
<td>Chan-Zuckerberg (Facebook)</td>
</tr>
<tr>
<td>Project Baseline - Verily (USA)</td>
<td>Google</td>
</tr>
<tr>
<td>Watson Health (USA)</td>
<td>IBM</td>
</tr>
</tbody>
</table>
Society-changing innovations

Technology:

• Integrated circuits (70:ies)
• Internet (90:ies)
• Smart phones (00:ies)
• Artificial intelligence
• Solar panels for electricity

Life science

• Gene technology (80:ies)
• Biological drugs (00:ies)
• “Next generation” precision medicine
• Mapping the building-blocks of humans
Will we (in the future) be able to construct a digital model of humans - covering cells, tissues, organs and diseases?

Will we (in this case) be able to construct a “digital twin” of everybody (patients) to test the efficacy of drugs digitally (on an individual basis) before decision on therapy?
Science for Life Laboratory

A national infrastructure for next-generation life science

**Global trends:**

- Need for major infrastructures
- Technology evolving rapidly
- Big data
Infrastructure resource for integrative omics

- Started in 2013
- 1,200 researchers
- More than 3,000 projects in 2018
Funding

- Wallenberg Foundation (Human Protein Atlas project)
- Novo Nordisk Foundation (Center for Biostainability)
- Erling Persson Foundation (Precision medicine)
- Heart and Lung Foundation (Biobank profiling)
- Chan Zuckerberg Foundation (Human Cell Atlas)
- ELIXIR – sharing of data resources