Human Protein Atlas

- Mapping of the human proteome
- Generation of millions of immunohistochemical (IHC) images
  - tissues and cell lines
  - manual annotation, interpretation of IHC-positivity
- what can image analysis do for us and for PATHOLOGY?
Pathology
- study and diagnosis of disease

Ovarian cancer

Pathology addresses 4 components of disease:
- cause/etiology,
- mechanisms of development (pathogenesis),
- structural alterations of cells (morphologic changes),
- and the consequences of changes
Samples for diagnosis

1) **Histological samples (tissue)**
   - biopsies
   - surgical material
   - autopsy material

2) **Cytological samples (cells)**
   - gynecological cell samples
   - blood
Histological samples

Paraffin embedded tissue
+ preserved morphology
- takes a few days

Fresh-frozen tissue
+ quick
- poor morphology

Paraffin section

Frozen section

Fixation time
Ischemia
Sectioning...
Formalin-fixed paraffin embedded tissue

• Preserves tissue histology and morphological details.

• Causes fragmentation of nucleic acids
Weigert-Van Gieson

Cell nuclei: brown-black
Connective tissue: red
Muscle: yellow

Lung
Hematoxylin-Eosin

Cell nuclei: blue
Other: pink
Classification of tumors

Benign, Pre-malign, Malign

Baserad primarily on morphology

**Benign**
- Circumscribed (epithelial tumors)
- Differentiated
- No atypia
- Low proliferation
- No necrosis

**Malign**
- Invasive / Infiltrativ / Destructive
- Variable differentiation
- Cellular atypia
- Proliferative
- Necrosis

Morphology - the study of the form and structure of cells
Cellular atypia- morphological deviation from “normal”
- level of differentiation vs level of malignancy

Example: Bladder cancer

High level of differentiation = low grade of malignancy
Low level of differentiation = high grade of malignancy
Why immunohistochemistry?

Complement to histological staining

• Question:
  • tissue of origin?
  • prognosis?
  • treatment strategy?
Biopsy from liver cancer.

Question: does the tumor originate from bile-duct epithelium or hepatocytes?

Answer: tumor originates from bile duct epithelium

Cytokeratin 7 expressed in bile duct epithelium
Cytokeratin 18 expressed in hepatocytes
Do the metastases originate from the prostate?

Immunohistochemical staining for PSA=Prostate Specific Antigen
Proliferation assessment (prognostics)

Ki67-protein
Only fraction, intensity not considered
Immunohistochemistry

1. Antigen (target)
2. Specific antibody
3. Secondary antibody
4. DAB-reaction (brown)

Brown: Antibody binding
Blue: Hematoxylin counterstaining
Antibodies are proteins that are produced by the body to fight the intrusion of foreign molecules, such as toxins or other poisons. The antibodies are designed to bind very tightly to their target molecules (i.e., the antigens).

Polyclonal

Monoclonal

Inject antigen
Sample blood
Several weeks later

Production Of Monoclonal Antibodies
Human Protein Atlas

- Started 2003, estimated to be finished 2015
- Joint effort between Uppsala University and the Royal Institute of Technology (KTH, Mathias Uhlén) in total 90 persons
- Knut and Alice Wallenberg foundation
The Human Protein Atlas project

Three Main Goals

1. Generate and validate antibodies towards all human proteins (a major isoform for each protein encoding gene)

2. Create a map showing protein expression patterns in cells and in human tissues

3. Utilize reagents and protein expression data to identify and explore biomarkers of clinical relevance
Levels of protein expression profiling

Immunohistochemistry

Immunofluorescence

ORGANS

TISSUES

CELLS

ORGANELLES

THE HUMAN PROTEIN ATLAS

Knut och Alice Wallenbergs Stiftelse

UPPSALA UNIVERSITY
A publicly available Human Protein Atlas

www.proteinatlas.org
### How large is the human proteome?

<table>
<thead>
<tr>
<th>Category</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-redundant proteins</td>
<td>$\approx 22,000$</td>
<td>A representative protein from every gene locus</td>
</tr>
<tr>
<td>Protein variants</td>
<td>$&gt;200,000$</td>
<td>Different protein fragments (splice variants or proteolytic fragments)</td>
</tr>
<tr>
<td>Combinatorial variants</td>
<td>$&gt;10,000,000$</td>
<td>Different proteins due to somatic DNA rearrangements</td>
</tr>
<tr>
<td>Protein species</td>
<td>$&gt;100,000$</td>
<td>Proteins that differ in chemical composition</td>
</tr>
<tr>
<td>Protein alleles</td>
<td>$&gt;75,000$</td>
<td>Proteins that differ by genetic variation (4 coding SNPs per gene)</td>
</tr>
</tbody>
</table>
Immunohistochemistry - light microscopy

Determine tissue distribution and relative quantification
tisdag 9 oktober 12
The Workflow within HPA

PrEST design → Molecular biology → Sequencing → Protein factory → MS analysis → Immunization → Affinity purification

Cell profiling immunofluorescence → Tissue profiling immunohistochemistry → Bioinformatics & literature search → Western blot analysis → Protein array analysis

THE HUMAN PROTEIN ATLAS

THE HUMAN PROTEIN ATLAS
Workflow - Uppsala

1. Tissue microarray production
   - Test IHC: various tissues and cells
   - Tissues: 48 normal (3 individuals)
   - Disease: 20 cancer types (12 patients)
   - Cells: 46 cell lines
   - Fixation: Formalin
   - Staining: Immunohistochemistry (DAB)
   - Annotation: Pathology
   - Throughput: 708 images/antibody
                 60 antibodies/week

2. Immunohistochemistry

3. Image annotation

Tisdag 9 oktober 12
Tissue Microarrays

Donor tissue block  Recipient block  Microtome sectioning  Transfer to glass slide
Normal tissue profiling

48 different tissue types (triplicate samples)

Surgical specimens
Normal Morphology

Cancer profiling

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>12</td>
</tr>
<tr>
<td>Carcinoid</td>
<td>4</td>
</tr>
<tr>
<td>Cervical cancer</td>
<td>12</td>
</tr>
<tr>
<td>Colo-rectal cancer</td>
<td>12</td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td>12</td>
</tr>
<tr>
<td>Glioma</td>
<td>12</td>
</tr>
<tr>
<td>Head-Neck cancer</td>
<td>4</td>
</tr>
<tr>
<td>Kidney cancer</td>
<td>12</td>
</tr>
<tr>
<td>Liver cancer</td>
<td>12</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>12</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>12</td>
</tr>
<tr>
<td>Melanoma</td>
<td>12</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>12</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>12</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>12</td>
</tr>
<tr>
<td>Skin cancer</td>
<td>12</td>
</tr>
<tr>
<td>Stomach cancer</td>
<td>12</td>
</tr>
<tr>
<td>Testis cancer</td>
<td>12</td>
</tr>
<tr>
<td>Thyroid cancer</td>
<td>4</td>
</tr>
<tr>
<td>Urothelial cancer</td>
<td>12</td>
</tr>
</tbody>
</table>

20 cancer types

432 samples from 216 patients
Protein expression profiling

- Normal
- Cancer
- Cell lines
Imaging & Data

Scan 

20 x scanning of 144 spots
(one TMA section, 20x40 mm)
in 45 min

Extract

708 Images/Antibody

Transfer

Database

Web interface

tisdag 9 oktober 12
• ≈700 images/ab, today 17 300 abs adds up to 12 million images

• The images are handled by a number of data servers at Rudbeck laboratory and KTH /Stockholm.

• Back-up on tape.

• Barcode-ID for slides and antibodies

• Directly after scanning automated focus-check and separation of the TMA-cores is initiated.

• Manual approval of separation makes the images available for manual annotation.

• Highly automated work-flow, integrated into our Laboratory Information Management System (LIMS)

**Scanning facts**

* Resolution: 50 000 pxi/inch (20X) or 500 nm/pixl
  100 000 pxi/inch (40x) or 250nm/pixl

* 10-13 Gb of raw data/slide

* 25 Gb of processed data/antibody

* Max load/machine : 120 slides

* 1500 slides scanned /month (300 antibodies)

* At least 7.5 Tb of data is stored each month.
In-house developed software for separation of stained tissue cores.
The Mumbai HPA team
Dr. Sanjay Navani

High Quality Pathology Annotations
Fruitful scientific collaborations
• CELLTYPE
• Positivity
  • fraction
  • intensity
From TMA to expression profiling

Large number of samples in one TMA saves material

Digitalization of slides

Small amount of antibody is used to analyze more than hundred of samples simultaneously

Many sections with identical design

Staining profile for one antibody in a large number of tissues within the Human Protein Atlas
Production system with real data.

Welcome Anna Asplund

Automatic *logout* occurs after one hour of inactivity.
### Hematopoietic tumors

- **Myeloid**
  - K-562
  - HL-60
  - NB4
  - U.937
  - THP-1HEL
  - HMC-1

- **Lymphoid**
  - KM3BALL-1
  - Daudi
  - Ramos
  - U-698
  - B-CII
  - RPMI 8226
  - U-266-1970
  - U-266-1984
  - LP1
  - Karpas 707
  - HDML-2

### Solid tumors

#### Brain
- U-138 MG (glioblastoma)
- U-251 MG (glioma)
- SH-SY5Y (neuroblastoma)
- D431 Med (medulloblastoma)

#### Breast, female reproductive system
- SKBr-3 (breast carcinoma)
- MCF-7 (breast carcinoma)
- HeLa (cervical carcinoma)
- SiHa (cervical carcinoma)
- AN3 Ca (endometrial carcinoma)
- EFO-21 (ovarian carcinoma)
- BEWO (choriocarcinoma)

#### Urinary, male reproductive system
- HEK 293 (embryonal kidney carcinoma)
- RT-4 (transitional cell carcinoma)
- PC-3 (prostate carcinoma)
- NTERA-2D1 (testicular carcinoma)

#### Abdominal
- Hep G2 (liver carcinoma)
- CAPAN 2 (pancreatic carcinoma)
- CACO 2 (colon carcinoma)

#### Skin
- HaCaT (keratinocyte derived)
- A 431 (epidermoid carcinoma)
- SK-Mel 30 (malignant melanoma)
- WM 115 (malignant melanoma)

#### Lung
- A 549 (lung carcinoma)
- SCLC 21H (small cell lung carcinoma)

#### Sarcoma
- U 2 OS (osteosarcoma)
- RH 30 (rhabdomyosarcoma)
- U 2197 (malignant fibrous histiocytoma)

#### Miscellaneous
- TIME (endothelial cell derived)
- HTH 83 (anaplastic thyroid carcinoma)

---

**46 cell lines and 12 clinical cell samples**
Automated annotation

Manual annotation

<table>
<thead>
<tr>
<th>Gene data</th>
<th>Cell lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description: Antigen K-67.</td>
<td>HEL</td>
</tr>
<tr>
<td>Source: P44013 (UniProt)</td>
<td>HL-60</td>
</tr>
<tr>
<td>Chromosome: 10q26.2</td>
<td>HMC-1</td>
</tr>
<tr>
<td>Ensembl ID: ENSG00000148773</td>
<td>K-562</td>
</tr>
<tr>
<td>Splice variant 1: ENSP00000340355</td>
<td>NB-4</td>
</tr>
<tr>
<td>Splice variant 2: ENSP00000337642</td>
<td>TNP-1</td>
</tr>
<tr>
<td>Splice variant 3: ENSP00000368553</td>
<td>U-937</td>
</tr>
<tr>
<td>Splice variant 4: ENSP00000377994</td>
<td>Daudi</td>
</tr>
<tr>
<td>No. of exons</td>
<td>No. of exons</td>
</tr>
<tr>
<td>3956</td>
<td>2896</td>
</tr>
<tr>
<td>359 kDa</td>
<td>359 kDa</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>TM Region(s)</td>
<td>TM Region(s)</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Normal Tissues</th>
<th>Cell lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adrenal gland</td>
<td>HEL</td>
</tr>
<tr>
<td>Appendix</td>
<td>HL-60</td>
</tr>
<tr>
<td>Bone marrow</td>
<td>HMC-1</td>
</tr>
<tr>
<td>Breast</td>
<td>K-562</td>
</tr>
<tr>
<td>Bronchus</td>
<td>NB-4</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>TNP-1</td>
</tr>
<tr>
<td>Cerebral cortex</td>
<td>U-937</td>
</tr>
<tr>
<td>Colon</td>
<td>Daudi</td>
</tr>
<tr>
<td>Corpus uteri</td>
<td>HLCM-2</td>
</tr>
<tr>
<td>Endometrium</td>
<td>Karpas-797</td>
</tr>
<tr>
<td>Endometrium 2</td>
<td>KM3</td>
</tr>
<tr>
<td>Prostate</td>
<td>KP-1</td>
</tr>
<tr>
<td>Rectum</td>
<td>MOLT-4</td>
</tr>
<tr>
<td>Seminal vesicle</td>
<td>RPMI-8226</td>
</tr>
<tr>
<td>Skin</td>
<td>U-266/70</td>
</tr>
<tr>
<td>Glandular cells</td>
<td>U-266/84</td>
</tr>
<tr>
<td>Stomach</td>
<td>U-698</td>
</tr>
<tr>
<td>Glutathione</td>
<td>U-266/70</td>
</tr>
<tr>
<td>Glandular cells</td>
<td>U-266/84</td>
</tr>
<tr>
<td>Glandular cells</td>
<td>U-698</td>
</tr>
<tr>
<td>Glandular cells</td>
<td>D411 Med</td>
</tr>
<tr>
<td>Endothelial cells</td>
<td>U-138MG</td>
</tr>
<tr>
<td>Epithelial cells</td>
<td>U-251MG</td>
</tr>
<tr>
<td>Smooth muscle cells</td>
<td>U87MG (ATCC)</td>
</tr>
<tr>
<td>Glial cells</td>
<td>A-549</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cancer Tissues</th>
<th>Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast cancer</td>
<td>Leukemia, AML</td>
</tr>
<tr>
<td>Cervical cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>Leukemia, AML</td>
</tr>
<tr>
<td>Endometrial cancer</td>
<td>Leukemia, B-ALL</td>
</tr>
<tr>
<td>Head &amp; neck cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Kidney cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Liver cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Malignant glioma</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Malignant lymphoma</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Malignant melanoma</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Ovarian cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Pancreatic cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Skin cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Stomach cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Testis cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Thymus cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
<tr>
<td>Urinary bladder cancer</td>
<td>Leukemia, B-CLL</td>
</tr>
</tbody>
</table>
Confocal fluorescence microscopy

Determine subcellular localization
**Workflow**

1. Sample preparation
2. Image acquisition
3. Image annotation

**Cell lines:** A431, U2OS and U251mg

**Fixation:** Formaldehyde

**Staining:**
- HPR antibody
- Nucleus (DAPI)
- Micro-tubules
- ER

**Throughput:**
- 2 images/antibody and cell line
- 95 antibodies/week

Organelle probes

*tdsdag 9 oktober 12*
Probes:
Nucleus (DAPI)
Cytoskeleton (tubulin)
ER (calreticulin)

Subcellular localization

HPA → nucleus
A. Cytoplasm
B. Micro-tubules
C. Nucleus
D. Mitochondrion
E. Vesicles
F. Micro-filaments
G. Endoplasmatic reticulum
H. Golgi apparatus
I. Extra cellular matrix
SEARCH

Insulin

e.g. CD44, ELF3, KLK3, or use fields to search specific fields such as protein_class:Transcription factors or chromosome:X
<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene synonym</th>
<th>Gene description</th>
<th>Antibody</th>
<th>Antibody validation</th>
<th>Annotated expression (IH)</th>
<th>Tissue specificity (IH)</th>
<th>Subcellular location</th>
</tr>
</thead>
<tbody>
<tr>
<td>INS</td>
<td></td>
<td>Insulin</td>
<td>CAB000048</td>
<td></td>
<td></td>
<td>1 out of 66 cell types</td>
<td></td>
</tr>
<tr>
<td>SLC2A4</td>
<td>GLUT4</td>
<td>solute carrier family 2 (facilitated glucose transporter), member 4</td>
<td>CAB012098</td>
<td></td>
<td></td>
<td>17 out of 66 cell types</td>
<td></td>
</tr>
<tr>
<td>IDE</td>
<td></td>
<td>insulin-degrading enzyme</td>
<td>CAB012303</td>
<td></td>
<td></td>
<td>62 out of 63 cell types</td>
<td></td>
</tr>
<tr>
<td>INSIG1</td>
<td>CL-6</td>
<td>insulin induced gene 1</td>
<td>CAB016544</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INSIG1</td>
<td>MGC1405</td>
<td></td>
<td>CAB012303</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INSR</td>
<td>CD220</td>
<td>insulin receptor</td>
<td>CAB013113</td>
<td></td>
<td></td>
<td>23 out of 65 cell types</td>
<td></td>
</tr>
<tr>
<td>GCK</td>
<td>HK4</td>
<td>glucokinase (hexokinase 4)</td>
<td>HPA007034</td>
<td></td>
<td></td>
<td>55 out of 65 cell types</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td></td>
<td>MODY2</td>
<td></td>
<td>HPA007093</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGFBP1</td>
<td></td>
<td>insulin-like growth factor 1 (somatomedin C)</td>
<td>HPA007993</td>
<td></td>
<td></td>
<td>55 out of 66 cell types</td>
<td>Nucleus but not nucleoli</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CAB024999</td>
<td></td>
<td></td>
<td></td>
<td>Cytoplasm</td>
</tr>
<tr>
<td>IGFBP1</td>
<td>CD221</td>
<td>insulin-like growth factor 1 receptor</td>
<td>CAB010268</td>
<td></td>
<td></td>
<td>60 out of 60 cell types</td>
<td>Vesicles</td>
</tr>
<tr>
<td></td>
<td>IGFR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>JTK13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MGC18216</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGFBP2</td>
<td></td>
<td>insulin-like growth factor 2 (somatomedin A)</td>
<td>HPA007993</td>
<td></td>
<td></td>
<td>55 out of 66 cell types</td>
<td>Nucleus but not nucleoli</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CAB024999</td>
<td></td>
<td></td>
<td></td>
<td>Cytoplasm</td>
</tr>
<tr>
<td>INSIG1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INSIG1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INSL3</td>
<td></td>
<td>insulin-like 3 (Leydig cell)</td>
<td>HPA028615</td>
<td></td>
<td></td>
<td>5 out of 65 cell types</td>
<td>Nucleus but not nucleoli</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cytoplasm</td>
</tr>
<tr>
<td>INSL5</td>
<td></td>
<td>insulin-like 5</td>
<td>HPA030100</td>
<td></td>
<td></td>
<td>33 out of 65 cell types</td>
<td>Nucleus but not nucleoli</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CAB033849</td>
<td></td>
<td></td>
<td></td>
<td>Cytoplasm</td>
</tr>
<tr>
<td>INSL6</td>
<td></td>
<td>insulin-like 6</td>
<td>HPA021364</td>
<td></td>
<td></td>
<td>1 out of 63 cell types</td>
<td>Cytoplasm</td>
</tr>
<tr>
<td>SLC30A8</td>
<td></td>
<td>solute carrier family 30 (zinc transporter), member 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKT1</td>
<td></td>
<td>v-akt murine thymoma viral oncogene homolog 1</td>
<td>HPA002891</td>
<td></td>
<td></td>
<td>66 out of 66 cell types</td>
<td>Nucleus</td>
</tr>
<tr>
<td>PKB</td>
<td></td>
<td></td>
<td>CAB003765</td>
<td></td>
<td></td>
<td></td>
<td>Cytoskeleton (Microtubules)</td>
</tr>
<tr>
<td>PRKAB</td>
<td></td>
<td>v-akt murine thymoma viral oncogene homolog 2</td>
<td>CAB004204</td>
<td></td>
<td></td>
<td>55 out of 63 cell types</td>
<td></td>
</tr>
<tr>
<td>RAC</td>
<td></td>
<td>v-akt murine thymoma viral oncogene homolog 3 (protein kinase B, gamma)</td>
<td>CAB013090</td>
<td></td>
<td></td>
<td>59 out of 65 cell types</td>
<td>Plasma membrane</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HPA026441</td>
<td></td>
<td></td>
<td></td>
<td>Cytoplasm</td>
</tr>
<tr>
<td>AR</td>
<td>A1S</td>
<td>androgen receptor</td>
<td>CAB000001</td>
<td></td>
<td></td>
<td>13 out of 64 cell types</td>
<td>Nucleus</td>
</tr>
<tr>
<td></td>
<td>DHTR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HUMARA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NR3C4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SBMA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SMAX1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PKG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Gene and Protein Summary

<table>
<thead>
<tr>
<th>Gene name</th>
<th>INS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>insulin</td>
</tr>
<tr>
<td>Protein class</td>
<td>Candidate cancer biomarkers, Candidate cardiovascular disease genes, Mapped to UniProt SWISS-PROT, Plasma proteins, Potential transmembrane proteins</td>
</tr>
<tr>
<td>Entrez gene summary</td>
<td>After removal of the precursor signal peptide, proinsulin is post-translationally cleaved into three peptides: the B chain and A chain peptides, which are covalently linked via two disulfide bonds to form insulin, and C-peptide. Binding of insulin to the insulin receptor (INSR) stimulates glucose uptake. A multitude of mutant alleles with phenotypic effects have been identified. There is a read-through gene, INS-IGF2, which overlaps with this gene at the 5' region and with the IGF2 gene at the 3' region. Alternative splicing results in multiple transcript variants. [provided by RefSeq]</td>
</tr>
<tr>
<td>No of splice variants</td>
<td>5 in total</td>
</tr>
<tr>
<td></td>
<td>0 with predicted TM region</td>
</tr>
<tr>
<td></td>
<td>5 with predicted signal peptide</td>
</tr>
<tr>
<td>Subcellular Location</td>
<td>Tissue Type</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Central nervous system (Brain)</strong></td>
<td>Cerebral cortex</td>
</tr>
<tr>
<td></td>
<td>Cerebral cortex</td>
</tr>
<tr>
<td></td>
<td>Hippocampus</td>
</tr>
<tr>
<td></td>
<td>Hippocampus</td>
</tr>
<tr>
<td></td>
<td>Lateral ventricle</td>
</tr>
<tr>
<td></td>
<td>Lateral ventricle</td>
</tr>
<tr>
<td></td>
<td>Cerebellum</td>
</tr>
<tr>
<td></td>
<td>Cerebellum</td>
</tr>
<tr>
<td></td>
<td>Cerebellum</td>
</tr>
<tr>
<td><strong>Blood and immune system (Hematopoietic)</strong></td>
<td>Bone marrow</td>
</tr>
<tr>
<td></td>
<td>Lymph node</td>
</tr>
<tr>
<td></td>
<td>Lymph node</td>
</tr>
<tr>
<td></td>
<td>Tonsillar lymph node</td>
</tr>
<tr>
<td></td>
<td>Tonsil</td>
</tr>
<tr>
<td></td>
<td>Tonsil</td>
</tr>
<tr>
<td></td>
<td>Spleen</td>
</tr>
<tr>
<td></td>
<td>Spleen</td>
</tr>
<tr>
<td><strong>Liver and pancreas</strong></td>
<td>Liver</td>
</tr>
<tr>
<td></td>
<td>Liver</td>
</tr>
<tr>
<td></td>
<td>Gall bladder</td>
</tr>
<tr>
<td></td>
<td>Pancreas</td>
</tr>
<tr>
<td></td>
<td>Pancreas</td>
</tr>
<tr>
<td><strong>Digestive tract (GI-tract)</strong></td>
<td>Oral mucosa</td>
</tr>
<tr>
<td></td>
<td>Salivary gland</td>
</tr>
<tr>
<td></td>
<td>Esophagus</td>
</tr>
<tr>
<td></td>
<td>Stomach, upper</td>
</tr>
<tr>
<td></td>
<td>Stomach, lower</td>
</tr>
<tr>
<td></td>
<td>Duodenum</td>
</tr>
<tr>
<td></td>
<td>Small intestine</td>
</tr>
<tr>
<td></td>
<td>Appendix</td>
</tr>
<tr>
<td></td>
<td>Appendix</td>
</tr>
<tr>
<td></td>
<td>Colon</td>
</tr>
<tr>
<td></td>
<td>Rectum</td>
</tr>
</tbody>
</table>

**Respiratory system (Lung)**
- Nasopharynx: Respiratory epithelial cells
- Bronchus: Respiratory epithelial cells
- Lung: Pneumocytes
- Lung: Macrophages

**Cardiovascular system (Heart and blood vessels)**
- Heart muscle: Myocytes

**Breast and female reproductive system (Female tissues)**
- Breast: Glandular cells
- Vagina: Squamous epithelial cells
- Cervix, uterus: Squamous epithelial cells
- Cervix, uterus: Glandular cells
- Uterus, pre-menopause: Glandular cells
- Uterus, pre-menopause: Cells in endometrial stroma
- Uterus, post-menopause: Glandular cells
- Uterus, post-menopause: Cells in endometrial stroma
- Fallopian tube: Glandular cells
- Ovary: Follicle cells
- Ovary: Ovarian stromal cells

**Placenta**
- Placenta: Trophoblastic cells
- Placenta: Decidua cells

**Male reproductive system (Male tissues)**
- Testis: Cells in seminiferous ducts
- Testis: Leydig cells
- Epididymis: Glandular cells
- Prostate: Glandular cells
- Seminal vesicle: Glandular cells

**Urinary tract (Kidney and bladder)**
- Kidney: Cells in glomeruli
- Kidney: Cells in tubules
- Urinary bladder: Urothelial cells

**Skin and soft tissues**
- Skin: Epidermal cells
- Vulva/anal skin: Epidermal cells
- Skeletal muscle: Myocytes
- Smooth muscle: Smooth muscle cells

**Endocrine glands**
- Thyroid gland: Glandular cells
- Parathyroid gland: Glandular cells
- Adrenal gland: Glandular cells

---

Diabetes - Wikipedia

Tisdag 9 oktober 12
Validation of antibodies
Antibody quality assurance

1. Protein array
2. Western blot
3. Histochemistry (IHC)
4. Immunofluorescence

Reliability score:

- **High**: Two independent antibodies with same staining patterns
- **Medium**: Consistent with bioinformatics and literature (if available)
- **Low**: At least some evidence that supports the staining patterns
- **Very low**: No evidence that the antibody is correct
- **Failed**: The antibody is (most likely) not correct
Stress 70 protein - “mortalin” (HPA000898)

Mitochondria (supported by literature)

Protein array - specific

Mw according to gene prediction: 74K
Validation by two independent antibodies

PrEST 1 (K300007/ER)  PrEST 2 (K300009/ER)
Comparison IHC and ISH - validation using RNA

- satb2 colon 1
- satb2 colon 3
- krt17 tonsil 3
- krt17 tonsil 1
- krt17 colon 3
- krt17 colon 1

- pecam placenta 1
- pecam placenta 3
- mki67 colon 3
- brd esophagus 2
- jup skin 1
- vil1 colon

Tisdag 9 oktober 12
Histology dictionary
Histology of the bronchus

- Lumen
- Respiratory epithelium
- Smooth muscle layer
- Submucosa
- Peribronchial glands
- Cartilage

THE HUMAN PROTEIN ATLAS

tisdag 9 oktober 12
Histology of the bronchus
Histology of the bronchus
Histology of the bronchus
In silico biomarker discovery

Advanced queries

- Chromosome 14
- Kinases
- Expressed in skin

Advanced Search

Search for proteins expressed in:
- chromosome 14
- class: Kinases (Ki)
- skin

Search Results

Search results for advanced query: 3 hits

<table>
<thead>
<tr>
<th>#</th>
<th>Gene name</th>
<th>Description</th>
<th>Chr</th>
<th>Links</th>
<th>Class</th>
<th>Antibody ID</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AKT1</td>
<td>RAC-alpha serine/threonine-protein kinase (EC 2.7.11.1) (RAC-PK-alpha) (Protein kinase B) (PKB) (C-AKT).</td>
<td>14:q32.33</td>
<td>[Image]</td>
<td>Ch, Ez, Kg, Ki</td>
<td>HPA002891</td>
<td>PA, WB, IH, IF</td>
</tr>
<tr>
<td>2</td>
<td>RPS6KA5</td>
<td>Ribosomal protein S6 kinase alpha-5 (EC 2.7.11.1) (Nuclear mitogen- and stress-activated protein kinase 1) (90 kDa ribosomal protein S6 kinase 5) (RSK-like protein kinase) (RSK5)</td>
<td>14:q32.12</td>
<td>[Image]</td>
<td>Ez, Kg, Ki</td>
<td>HPA001274</td>
<td>PA, WB, IH, IF</td>
</tr>
<tr>
<td>3</td>
<td>VRK1</td>
<td>Serine/threonine-protein kinase VRK1 (EC 2.7.11.1) (Vaccinia-related kinase 1).</td>
<td>14:q32.2</td>
<td>[Image]</td>
<td>Ez, Ki, Kk</td>
<td>HPA000660</td>
<td>PA, WB, IH, IF</td>
</tr>
</tbody>
</table>
Validation of tissue biomarkers in clinical TMAs

- Colorectal cancer
- Prostate cancer
- Breast cancer
- Lung cancer
- Lymphoma
- Brain tumors
- Urothelial cancer
- Melanoma
- Ovarian cancer
- Renal cancer

Months from diagnosis

Overall Survival

p = 0.0003
SATB2
Colorectal cancer

- Transcription factor involved in development
- Approximately 80% of all colorectal cancers are positive
- SATB2 is a diagnostic marker for colorectal cancer
- SATB2 is a potential bad prognosis marker in advanced (stage III and IV) colorectal cancer

RBM3
Breast, ovarian, colon cancer & melanoma

- RNA binding protein involved in RNA metabolism and regulation
- Approximately 50% of colorectal, breast, ovarian cancers and melanoma are RBM positive (variable)
- RBM3 is a positive prognostic factor in these tumor types

HMGCR
Breast cancer

- Rate-limiting enzyme in cholesterol synthesis
- 75-80% of breast cancers are HMGCR positive (50% positive in pre-menopausal patients)
- HMGCR is a positive prognostic factor, stronger effect in ER positive tumors
- HMGCR positive tumors show better response to Tamoxifen treatment (irrespective ER status)

PODXL1
Lung cancer

- Sialomucin protein, involved in membrane integrity, stem cell factor
- Rare tumor cell populations positive in certain cancers
- PODXL1 is a bad prognosis marker in lung cancer and colorectal cancer
Biomarkers

RBM3 - example of a prognostic/predictive marker

RNA binding protein involved in RNA metabolism and regulation

Normal tissues

Breast cancer

Cancer tissues

Approximately 50% of breast cancers express RBM3 with variable intensity
Biomarkers

RBM3 - example of a prognostic/predictive marker

Manual assessment of IHC intensity
Biomarkers

RBM3 - example of a prognostic/predictive marker

Breast cancer

Consecutive cohort (512 patients)

Overall Survival

Time (Years)

>75%

0-75%

P = 0.001

tisdag 9 oktober 12
**Conclusions**

* RBM3 is a prognostic marker - increased expression of RBM3 protein correlates with better outcome

* Decreased RBM3 expression make cells less susceptible to cisplatin treatment in vitro indicating that RBM3 may be a useful prognostic and treatment predictive marker


Discovery phase

- Ki-67
- EGFR
- PSA
- ER
- HER2
- p53

tisdag 9 oktober 12
Image analysis of IHC staining in tissue

Staining complexity

Plasma membrane

Microvilli

Cilia

Extra cellular

Golgi

Endoplasmic ret.

Lysosomes

Peroxisomes

Mitochondria

Nuclei

Nucleoli

Nuclear membranes
Cancer is not one disease but many....
And even one cancertype can be very heterogenous
Examples of image analysis projects
Proteinkvantifiering

(% weak area + 2(% moderate area) + 3(% strong area))/# cells

=protein per cell score
RNA-protein correlation in cell lines

ENSG00000111716

Spearman: 0.84
Pearson: 0.87

log2 FPKM / log2 IHC

A549, CAC02, Daudi, HEK, Hel, HELA, HEPG2, HL60, K562, MCF7, PC3, RH30, RT4, SiHa, Time, U937
Modeling cell cycle as a continuous process

A. CLASSIFICATION (DISCRETE LABELS)

B. REGRESSION LINEAR

Garber K J Natl Cancer Inst 2001
Cyclin-B as surrogate marker for cell-cycle pattern recognition

G1: no staining at all
G2: Stronger staining in cytoplasm
S: weak staining in cytoplasm
M: strong staining in nucleus & cytoplasm
Pattern recognition of melanocyte in digital pathology data – Results

- 59 texture features selected to describe each cell in all data sets
- A SVM classifier model was trained and validated on the testing set

<table>
<thead>
<tr>
<th>Pattern-based classification</th>
<th>Other</th>
<th>Melanoma cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>14349</td>
<td>1977</td>
</tr>
<tr>
<td>Melanoma cells</td>
<td>1164</td>
<td>21578</td>
</tr>
</tbody>
</table>

Accuracy = 92.86 %
• Extra material- how do we make antibodies?
Protein Signature Epitope Tags (PrESTs)

100 to 150 amino acids selected from whole genome analysis (Ensembl-based)

Software to facilitate epitope selection

- Avoid homology (cross-reactivity)
- Avoid transmembrane spanning regions
- Allows “epitope” specific antibodies
**Molecular Biology**

**RT PCR**
- Amplification of PrEST regions from mRNA
- Analysis by agarose gel electrophoresis

**Cloning**
- Vector preparation
- Restriction of cDNA and insert in the vector
- Ligation

**Transformation**
- Transformation into E.coli
- Cultivation on agar plates
- Colony selection
- Production of glycerol stock
Protein Factory

Cultivation → Harvest

Protein purification
IMAC
denaturating conditions

His-tag for purification
ABP for immunopotentiation

Antigens
PrEST affinity matrix
Protein arrays

His ABP PrEST
His-tag for purification
ABP for immunopotentiation
Affinity-purification of the antibodies

Tag-specific depletion (His-column)

Tag-specific antibodies (discard)

Affinity capture (PrEST-specific column)

mono-specific antibodies (collect)

Non-specific antibodies (discard)