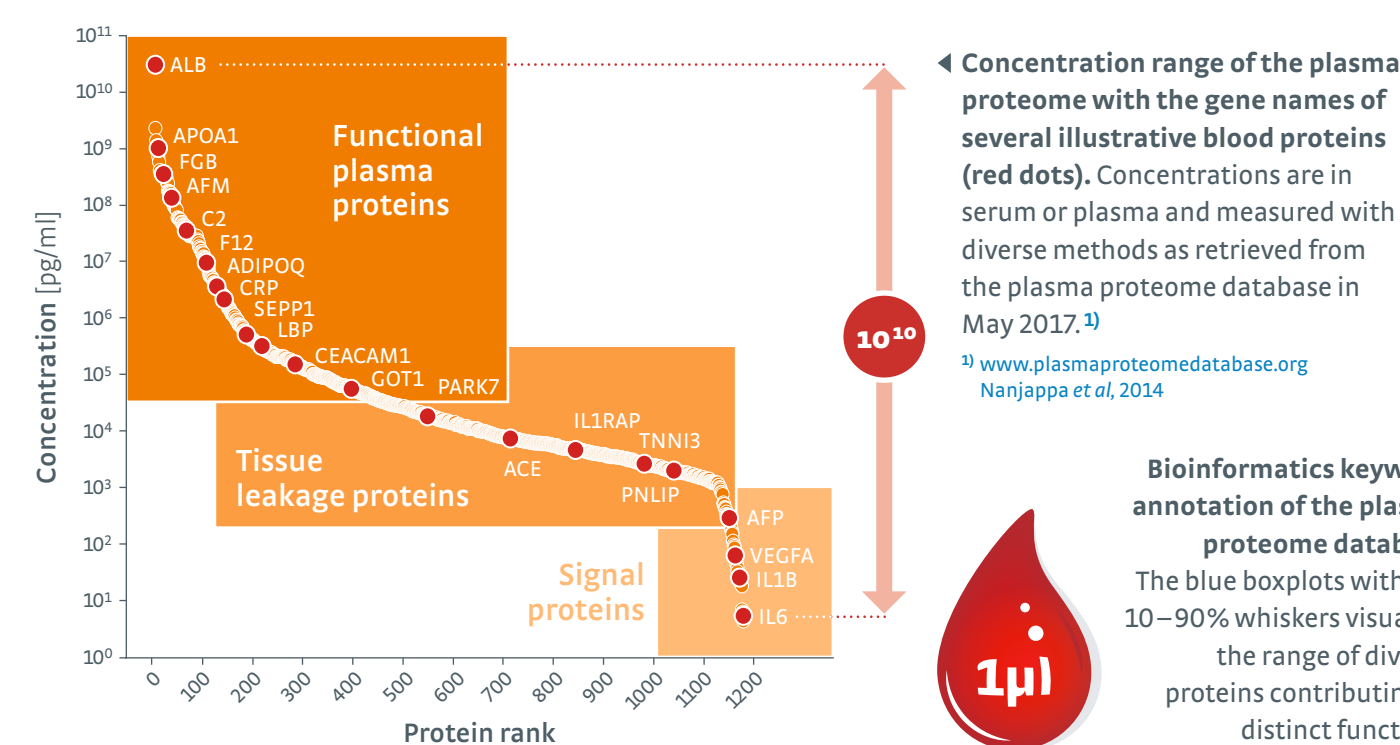# CLINICAL PROTEOMICS

By Philipp E Geyer, Lesca M Holdt, Daniel Teupser and Matthias Mann
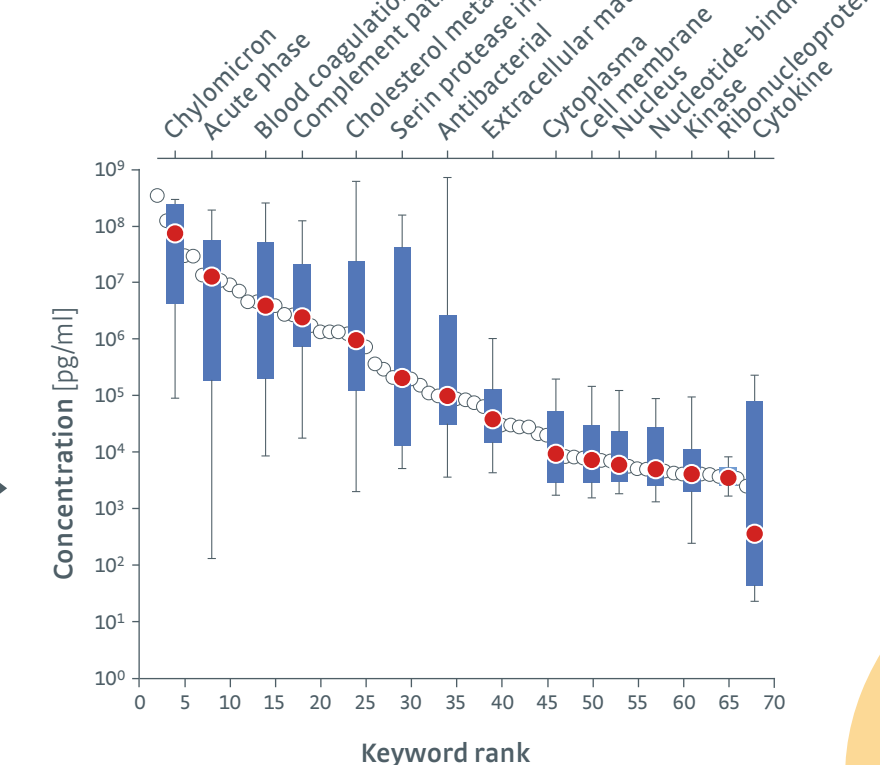
molecular systems biology

THE PLASMA PROTEOME

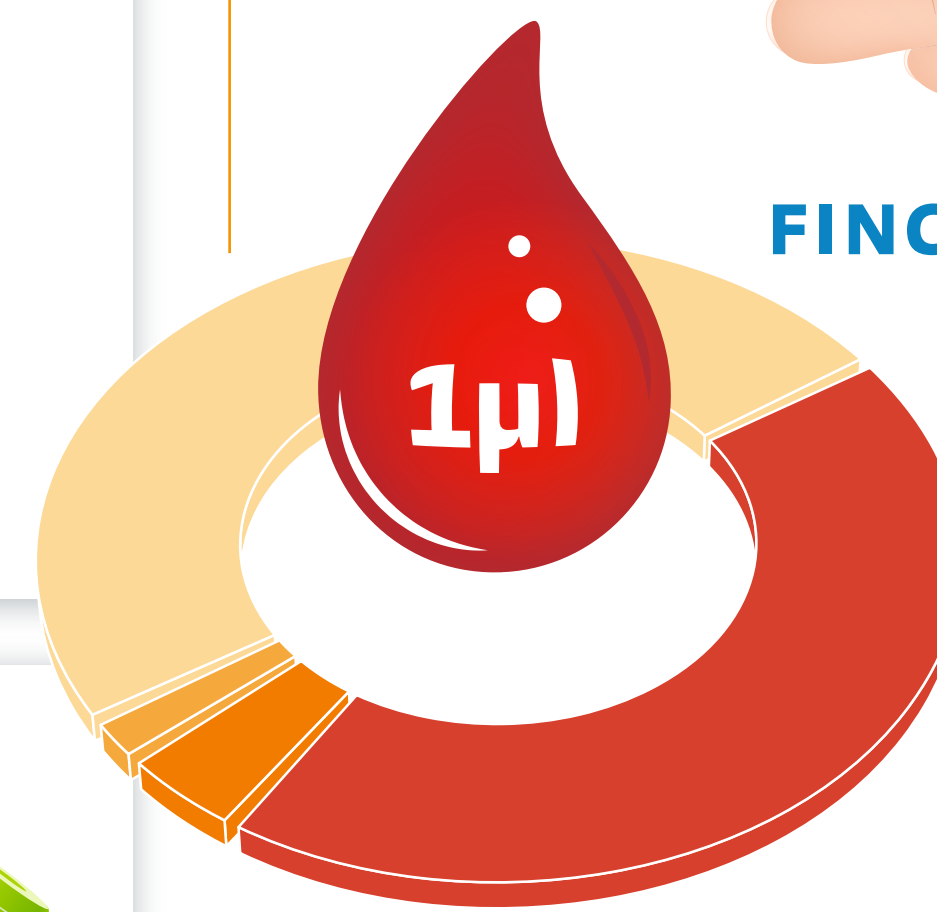### Proteins

ALB, APOA1, FCB, AFM, C3, F12, ADIPOQ, CRP, SEPP1, LBP, CFACAM1, GOT1, PARK7, IL1RAP, TNNI3, ACE, PNLIP, AFP, VEGFA, IL18, IL6

Functional plasma proteins
Tissue leakage proteins
Signal proteins

Concentration [pg/ml] — Protein rank

◀ Concentration range of the plasma proteome with the gene names of several illustrative blood proteins (red dots). Concentrations are in serum or plasma and measured with diverse methods as retrieved from the plasma proteome database in May 2017. [1]

[1] www.plasmaproteomedatabase.org Nanjappa et al, 2014

10^{10}

Bioinformatics keyword ▶ annotation of the plasma proteome database. The blue boxplots with the 10–90% whiskers visualize the range of diverse proteins contributing to distinct functions.

1µl

Functions

Chylomicron, Acute phase, Blood coagulation, Complement pathway, Cholesterol metabolism, Serum protease inhibitor, Antibacterial, Extracellular matrix, Cytoplasma, Nucleus, Cell membrane, Nucleotide-binding, Ribonucleoprotein, Cytokine

Concentration [pg/ml] — Keyword rank

TODAY'S BLOOD TESTS

In total 77 % of all clinical decisions in patients are based on laboratory testing. The numbers are based on 9 million tests performed in the year 2016 at the Institute of Laboratory Medicine, University Hospital Munich. The second figure illustrates the distribution of laboratory tests based on frequency of request.

Clinical decisions made...

...without lab testing 23%
...with lab testing 77%

Distribution of laboratory tests

Drugs 1.8%
Specific antibodies 1.5%
Nucleic acids 0.5%
Others 2.1%
Cells 16.9%
Proteins and enzymes 41.9%
Small molecules 35.3%

FINGER PRICK AND BLOOD COMPOSITION

56 % Plasma
90 % Water
3 % Small molecules
7 % Proteins

1µl

Blood is a suspension, consisting of a cellular (~44%) and a liquid component (~56%). Its cellular portion can be classified into thrombocytes, leucocytes and erythrocytes. The straw-colored liquid portion of blood is called plasma and is a mixture of water, small molecules like electrolytes, substrates or vitamins and an extensive diversity of all human proteins that are encoded by the 20,000 human genes.
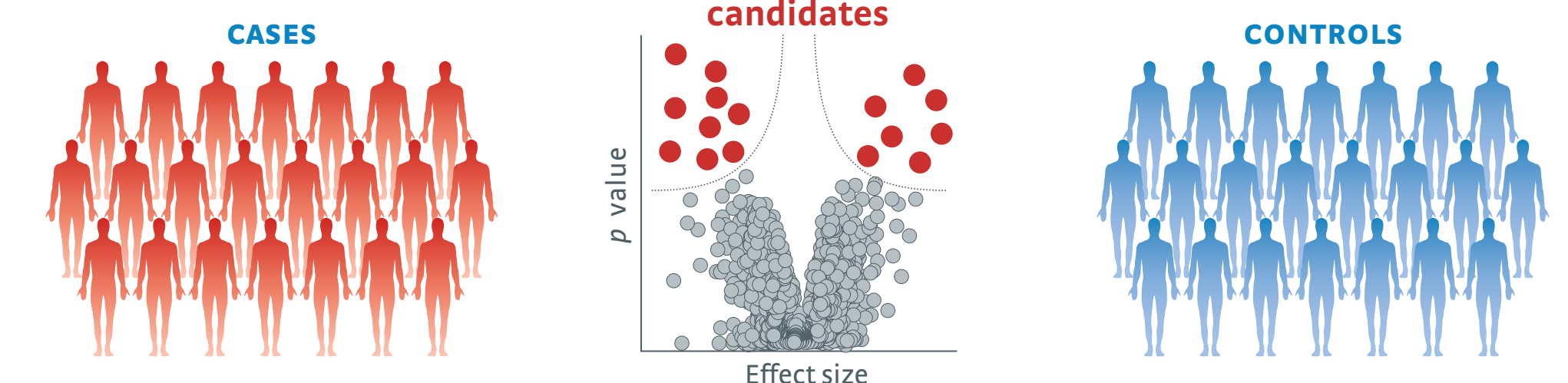
As the protein concentration in plasma is so high (50 µg/µl), a simple finger prick delivering just 1µl of plasma is enough to analyze the human plasma proteome.
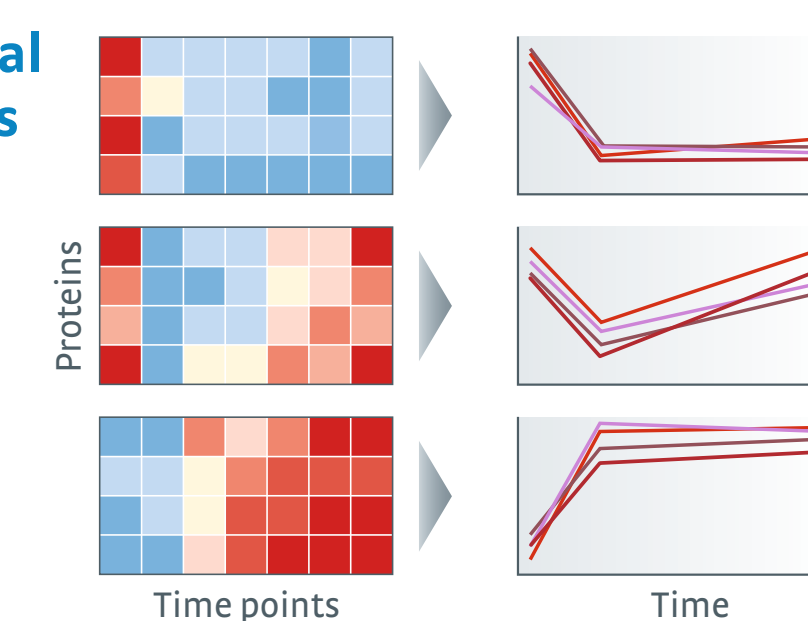
44 % Blood cells
5×10^6 Erythrocytes
$2-4 \times 10^5$ Thrombocytes
$5-10 \times 10^3$ Leucocytes

BIOMARKER DISCOVERY

Triangular strategy

Phases	Numbers of samples	Numbers of proteins
DISCOVERY Shotgun proteomics	10s	100s–1000s
VERIFICATION Targeted proteomics	10s–100s	10s
VALIDATION Immunoassays	100s–1000s	1–10

vs.

Rectangular strategy

Phases	Numbers of samples	Numbers of proteins
DISCOVERY Shotgun proteomics	Cohort 1	Not changed / Study specific / Validated biomarker/panel
VALIDATION Shotgun proteomics	Cohort 2 100s–1000s	Study specific / Not changed 100s–1000s

In the classical **'Triangular strategy'** a relatively small number of cases and controls are measured at great depth by hypothesis-free discovery proteomics, ideally leading to the quantification of thousands of proteins (top layer in the panel). This may yield dozens of differentially expressed candidates that are screened by targeted proteomics methods in cohorts of moderate size (middle layer). Finally, for one or a few of the remaining candidates, immunoassays are developed, which are then validated in large cohorts and applied in the clinic (bottom layer).

In the **'Rectangular strategy'** a large cohort is investigated in the discovery phase at the greatest proteome coverage possible. In the validation phase, another cohort is analyzed to confirm the biomarker candidates, utilizing the same technology and similar cohort size. Both cohorts can be analyzed in parallel, but only the proteins that are significantly different in both studies (orange as opposed to green circle in the right-hand part of the panel) are validated biomarkers.

POTENTIAL FOR NEW BIOMARKERS

Biomarker [%] — Abundance rank

POTENTIAL 280 Biomarkers 23%

70 Biomarkers ~23%
47 Biomarkers ~4%

The blue area illustrates the density of biomarkers as a function of increasing depth of the plasma proteome. Within the 300 most abundant proteins, 23% are already known biomarkers. The top of the yellow region extrapolates this proportion to the remainder of the plasma proteome. If the portion of biomarkers remained as high as it is in the 300 most abundant proteins, there are at least 233 potential biomarkers to be discovered (yellow area).

APPLICATIONS

Cases *versus* controls studies

CASES
Biomarker candidates
CONTROLS
p value — Effect size

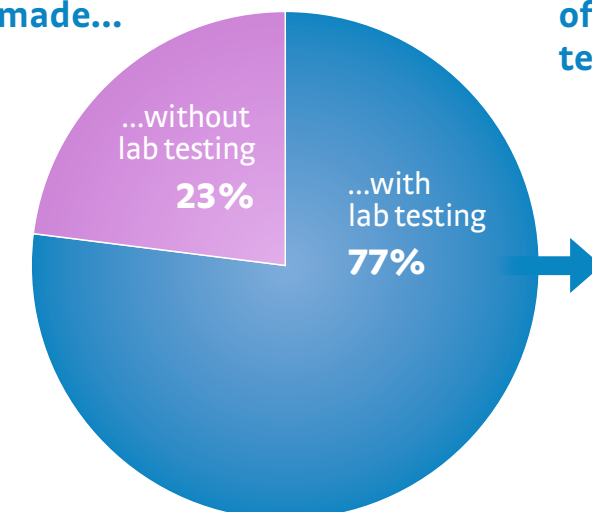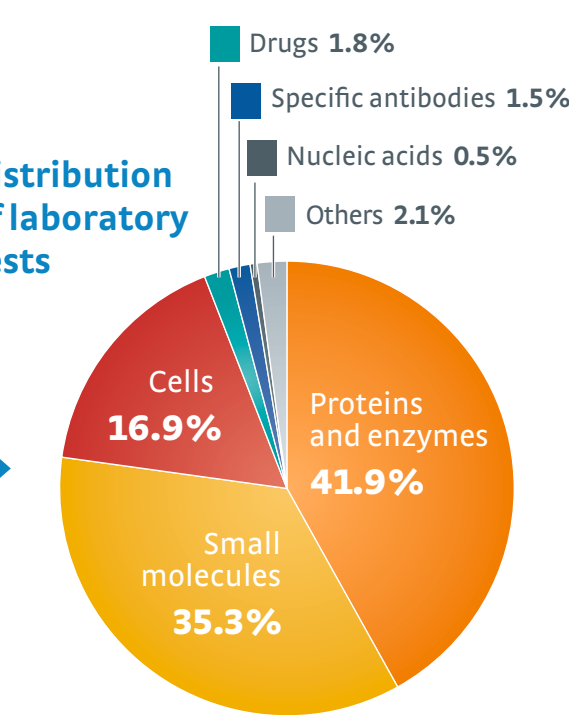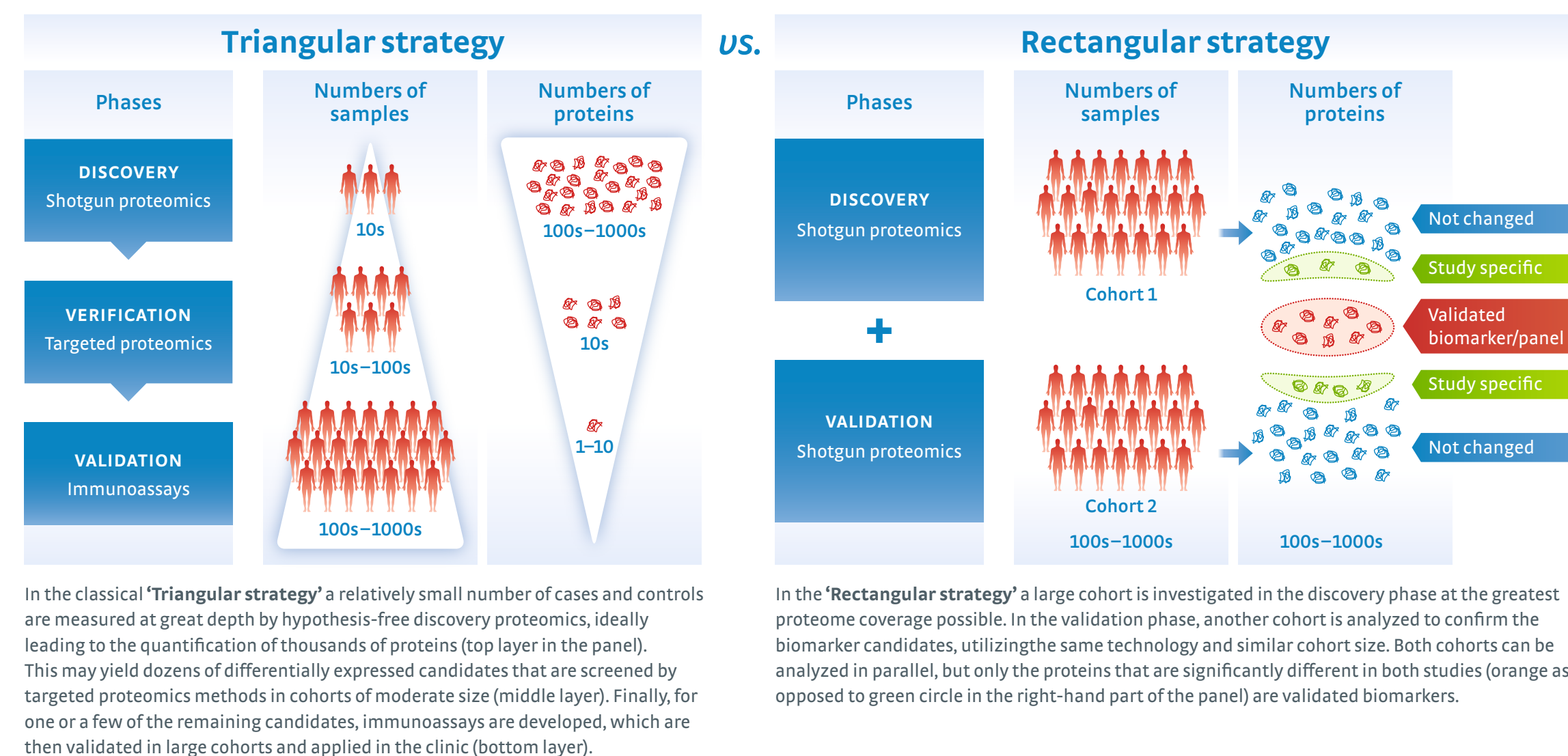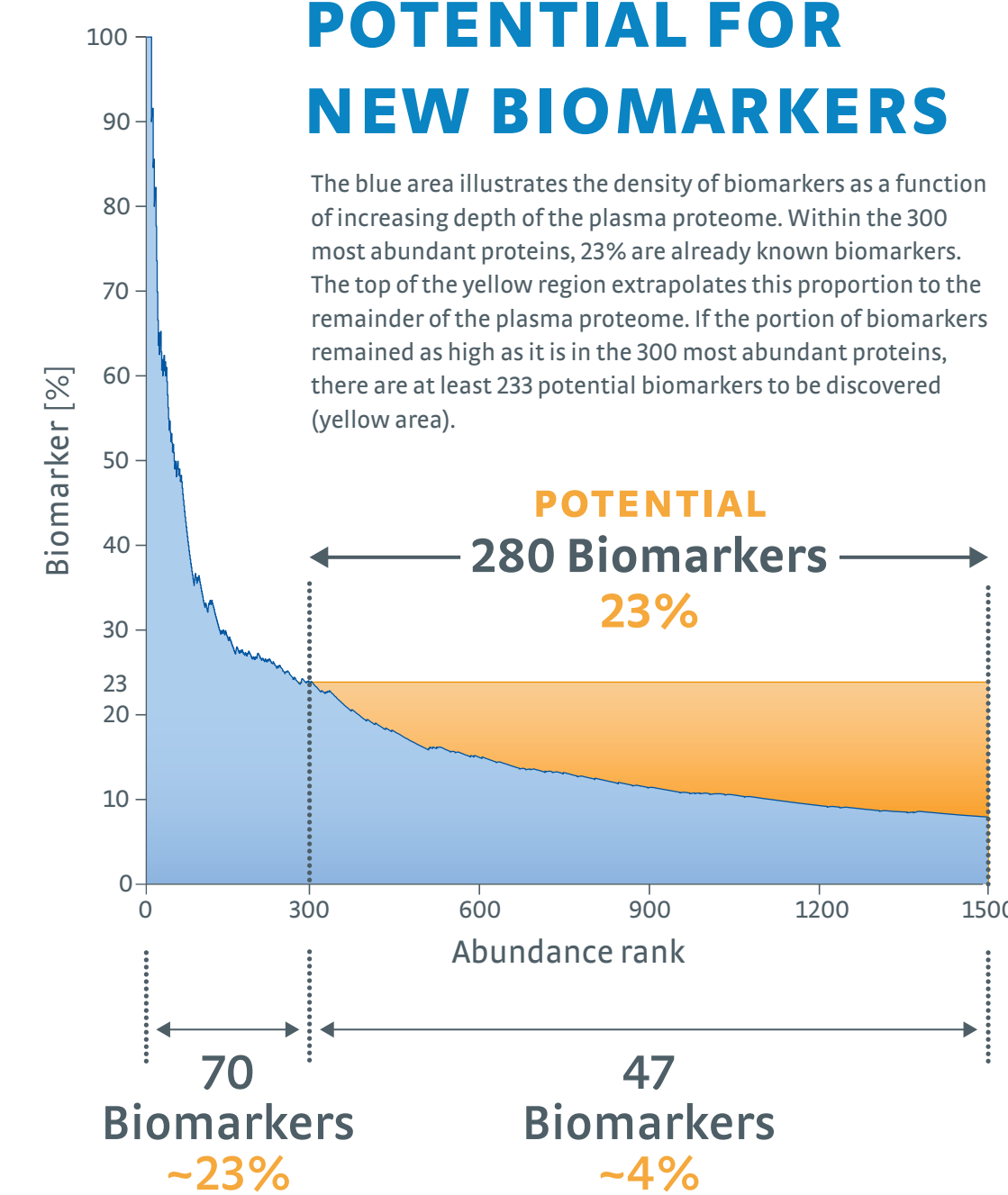### Longitudinal trajectories

Proteins — Time points — Time

Plasma proteome profiling can be further applied to compare case-control studies and for the investigation of longitudinal protein trajectories.

PLASMA PROTEOME PROFILING

Knowledge base
Disease | Risk | Treatment | Lifestyle

Individuals — Proteins

Plasma proteome profiling of diverse disease, risk, treatment, lifestyle or other relevant alterations will over time accrue a knowledge base that connects plasma protein changes to perturbations in a general manner.

MOLECULAR PHENOTYPING

The plasma proteome profile of a given individual can then be deconvoluted using the information and algorithms associated with the knowledge base.

1µl

Individual — Proteins — Plasma proteome profile