http://proteomics.cancer.gov

CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

Performance and Optimization of LC-MS/MS Platforms for Unbiased Discovery of Biomarker Candidates

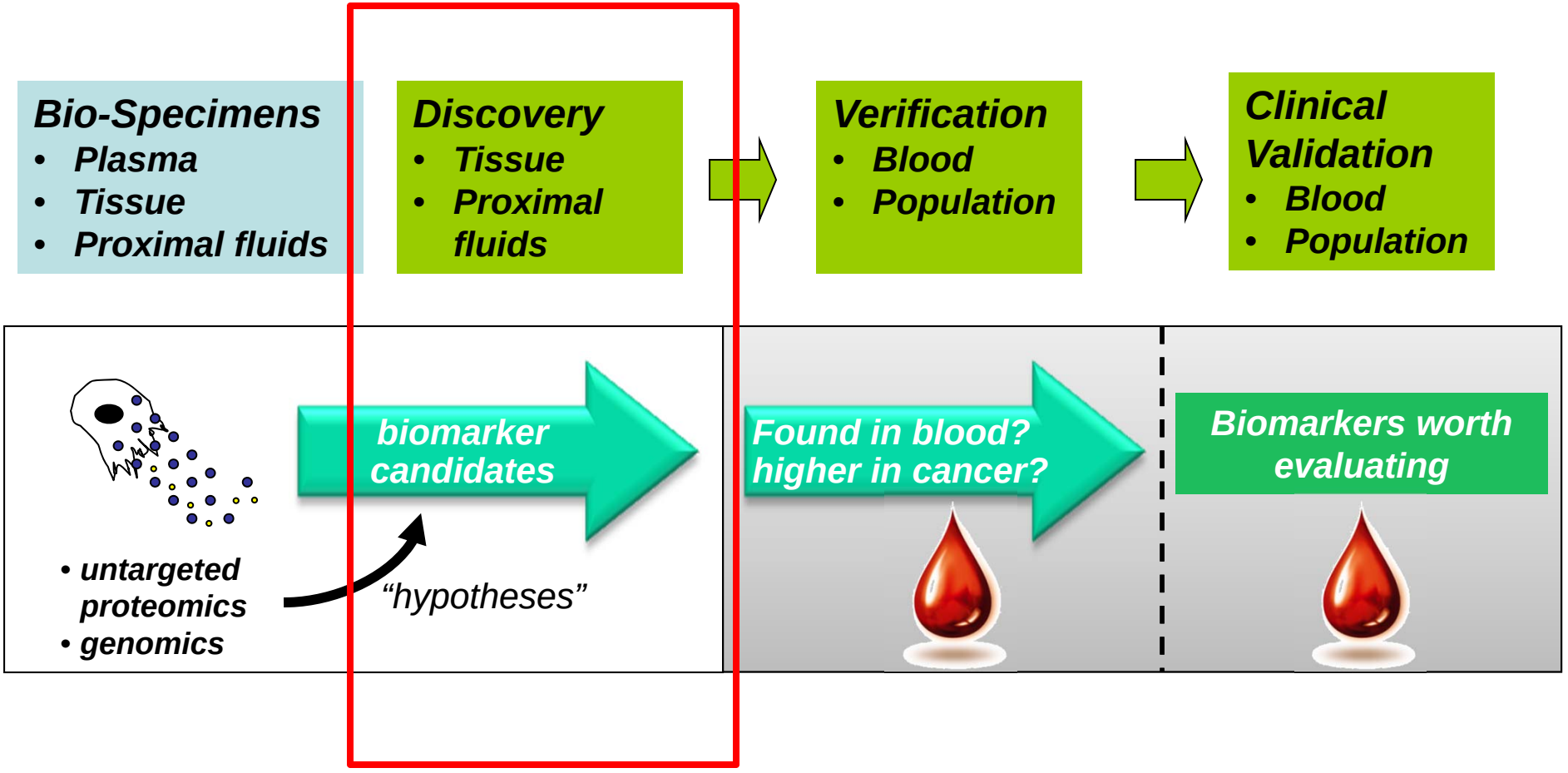
Daniel C. Liebler

Vanderbilt University School of Medicine
Vanderbilt, Tennessee

NATIONAL CANCER INSTITUTE

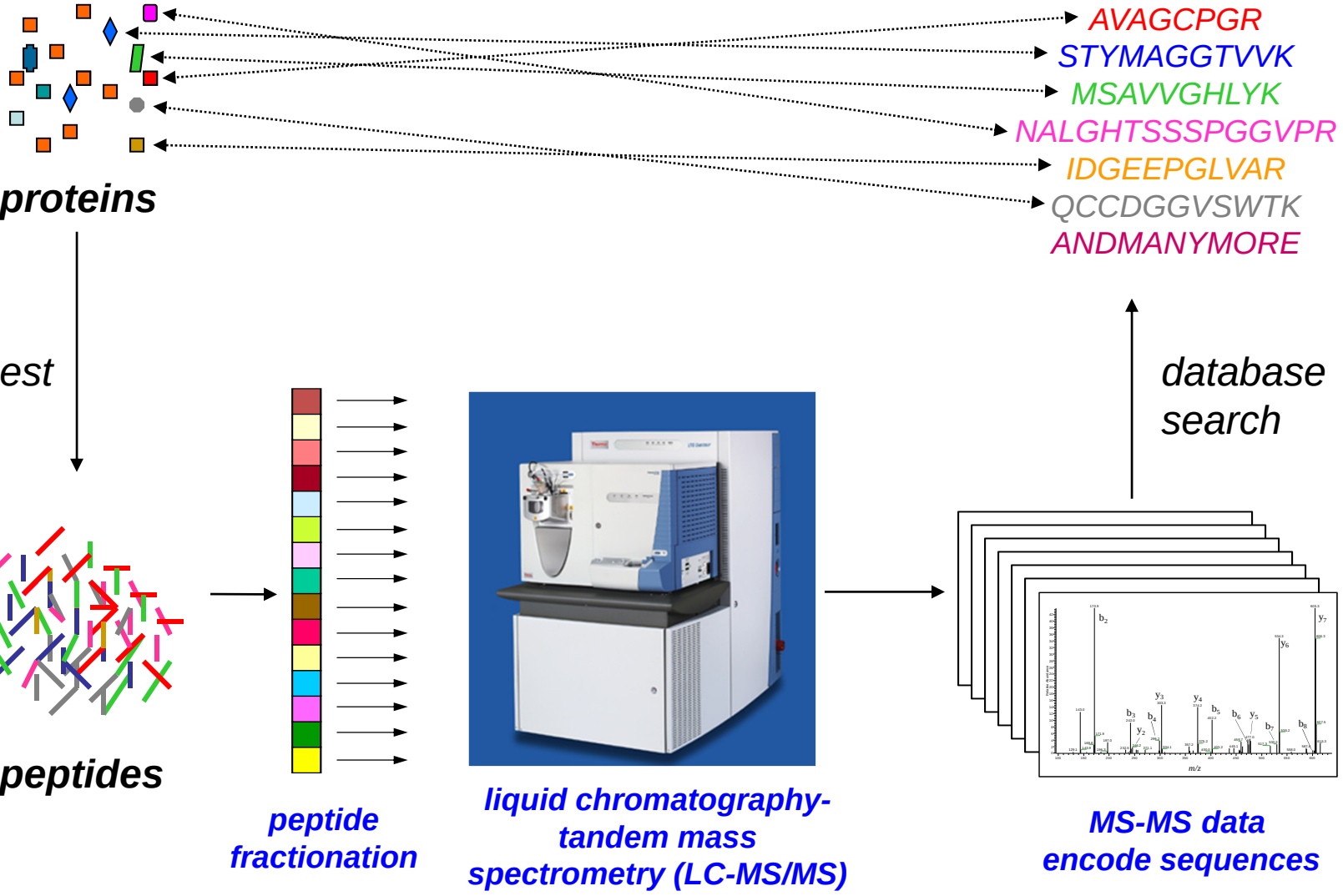
A Functioning Pipeline for Cancer Biomarker Development Requires Both Discovery and Directed Assay Components

CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

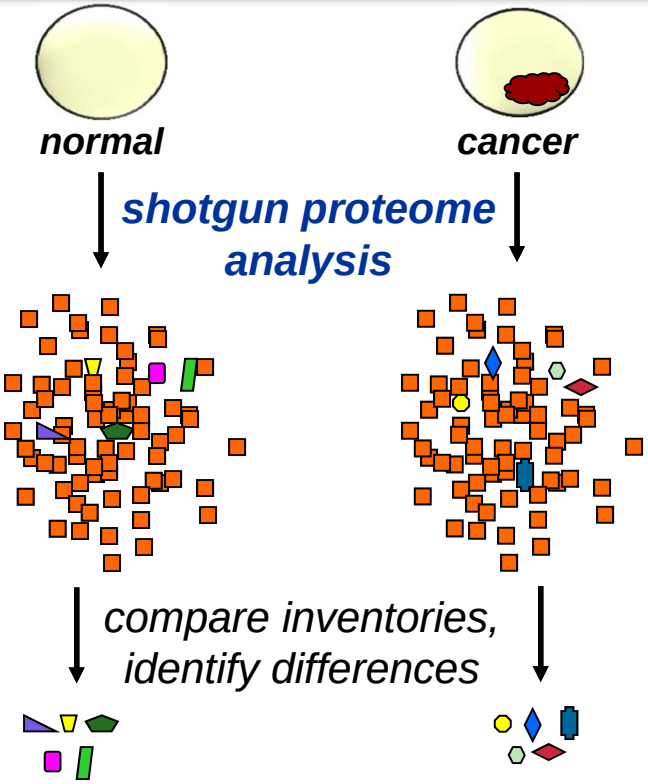


Overview of shotgun proteomics

CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER



Shotgun proteomics for biomarker discovery



- *few (<20) samples or sample pools*
- *low throughput*
- *Identify ~5,000+ proteins*
- *inventory differences ~50-500+ proteins*

Challenges

- **> 10^6 range of protein concentrations in tissues, biofluids**
- **no technology yet capable of complete sampling of proteomes**
- **multiple instrument systems used**
- **variability in detection of proteins**

No systematic studies of variation in shotgun proteomics

Impact of variation on performance for unbiased discovery unknown

HUPO Study

"A HUPO test sample study reveals common problems in mass spectrometry-based proteomics" *Bell et al. Nature Methods (2009)* ***6****: 423-430*

20 protein mixture distributed to 27 labs; no standardized methods or data analysis

- only 7 labs correctly ID all 20 proteins

Coaching → reanalyses → common bioinformatics

- all 27 labs correctly ID "most" proteins

Conclusions:

- Variable performance within and between labs
- better databases and search engines needed, as is training in their use

Key questions not addressed:

- ***Is the technology inherently variable?***
- ***What are the sources of variation?***
- ***How reproducible are analyses of complex biological proteomes?***

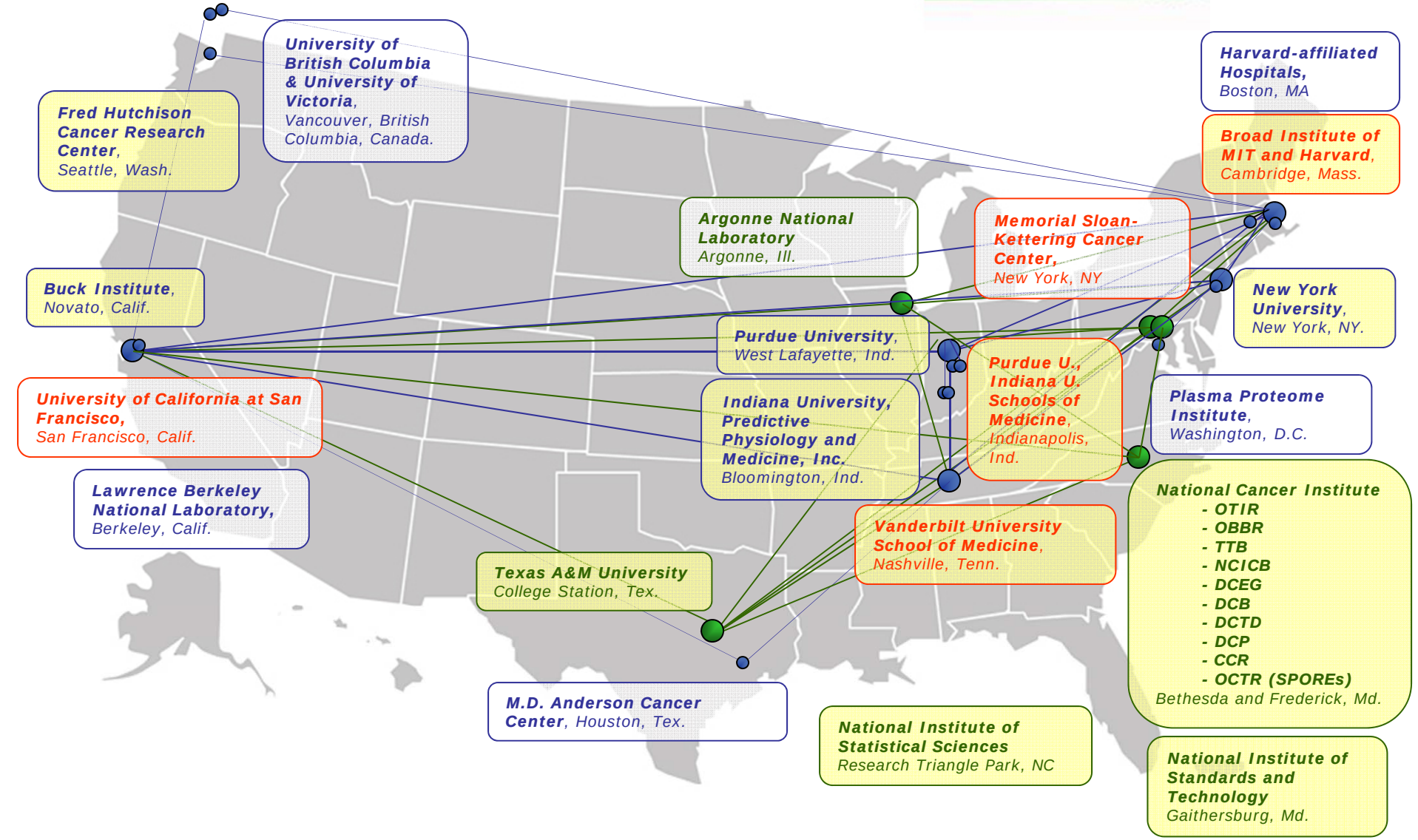
CPTAC Unbiased Discovery Workgroup: Goals

Evaluate and standardize performance of proteomic *discovery* platforms and standardize their use

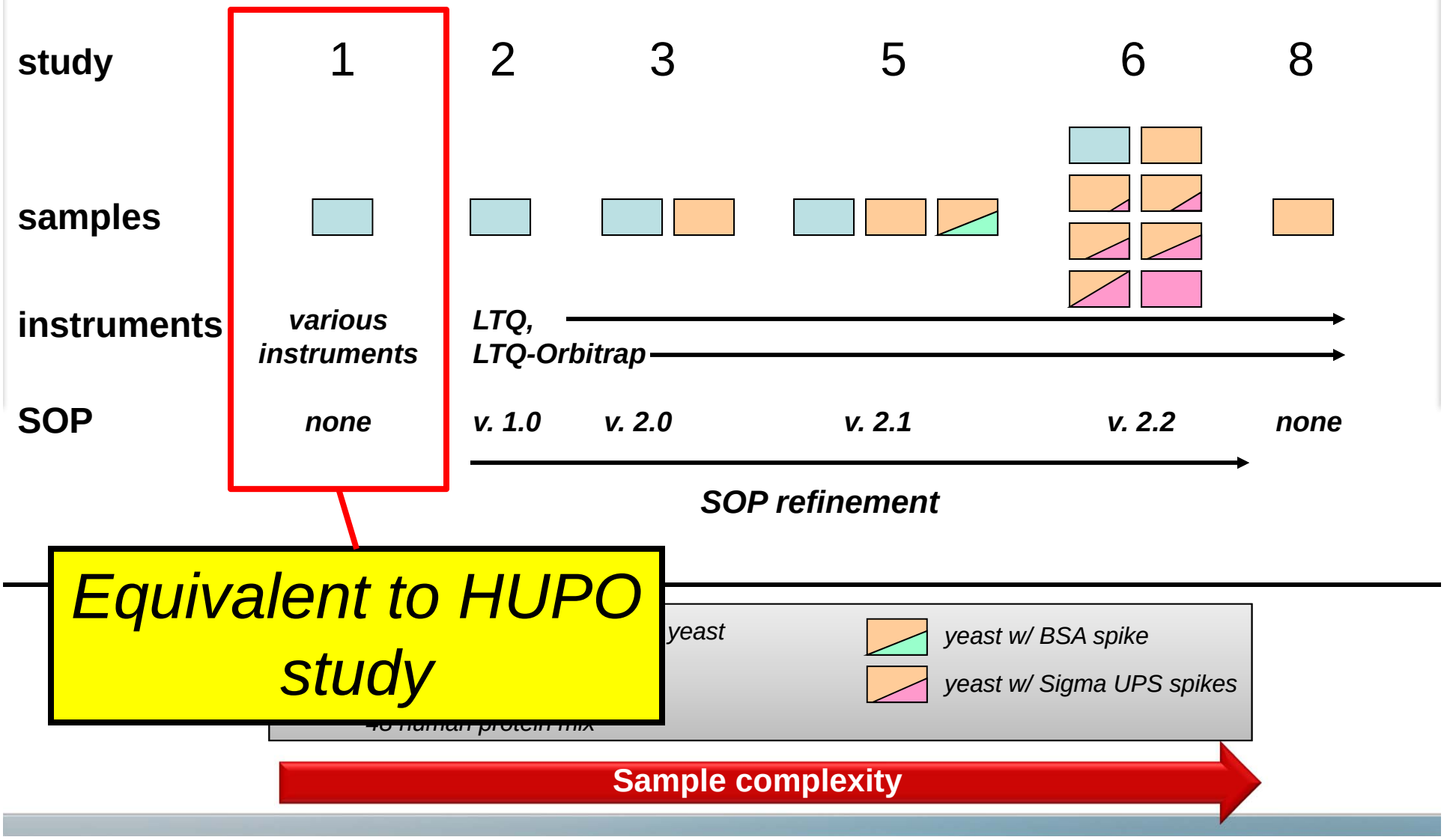
- **Identify and characterize sources of variation**
- **Develop quality control metrics**
- **Employ defined protein mixtures and biologically relevant proteomes**
- **Evaluate sensitivity for protein detection at defined levels of concentration**

Discovery WG participants

CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER



CPTAC Discovery WG studies: from simple to complex proteomes

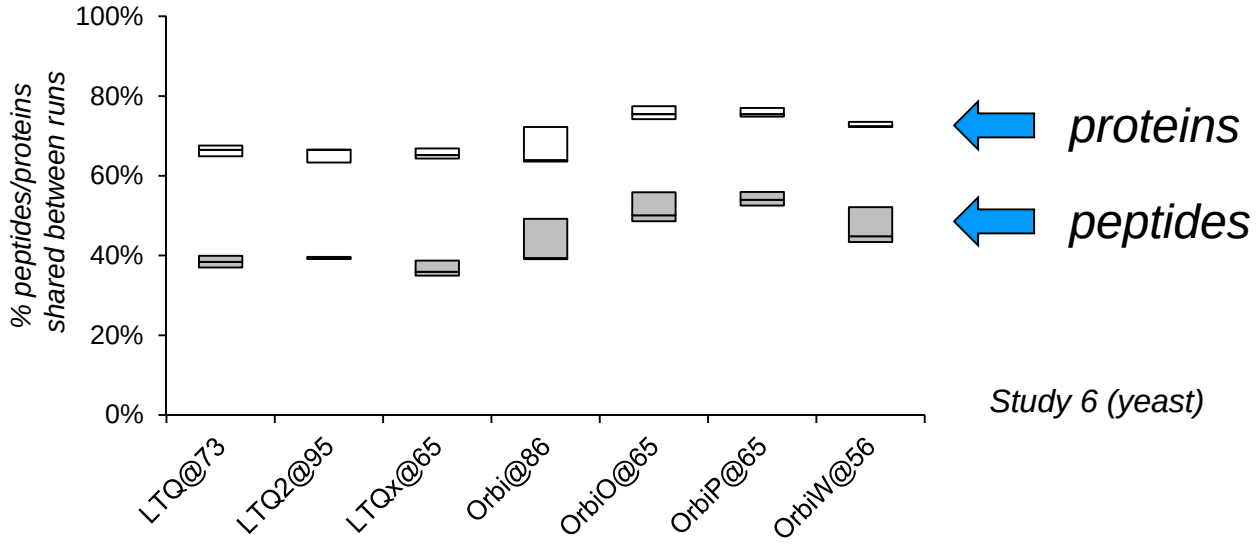


Repeatability and reproducibility are robust across laboratories

CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

Run-to-run repeatability in analyses of yeast across 7 instruments/sites

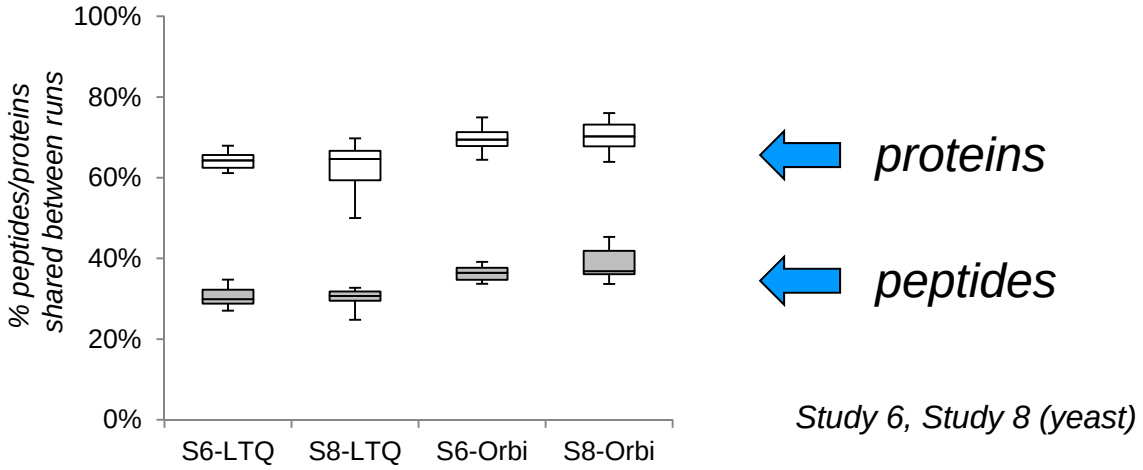
65-80% of protein IDs repeat between runs on individual instruments



Study 6 (yeast)

Interlaboratory reproducibility

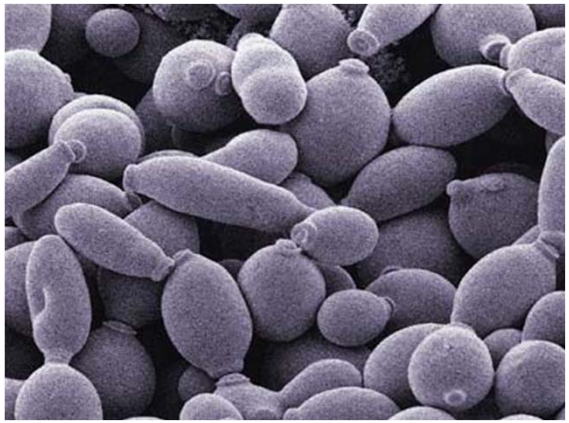
65-70% of protein IDs are reproduced across multiple instruments

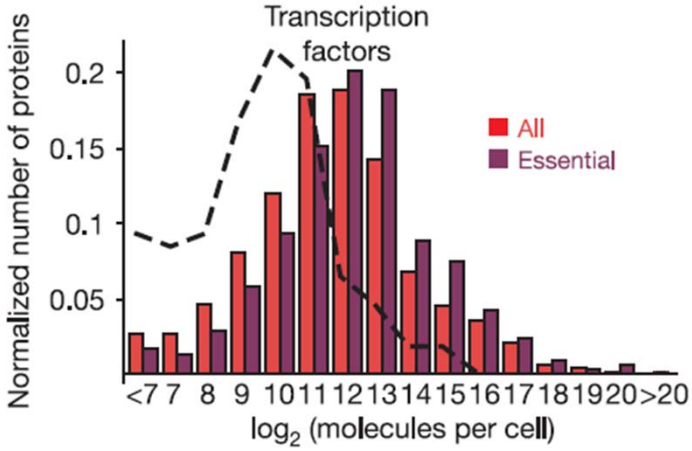


Study 6, Study 8 (yeast)

Saccharomyces cerevisiae proteome reference material

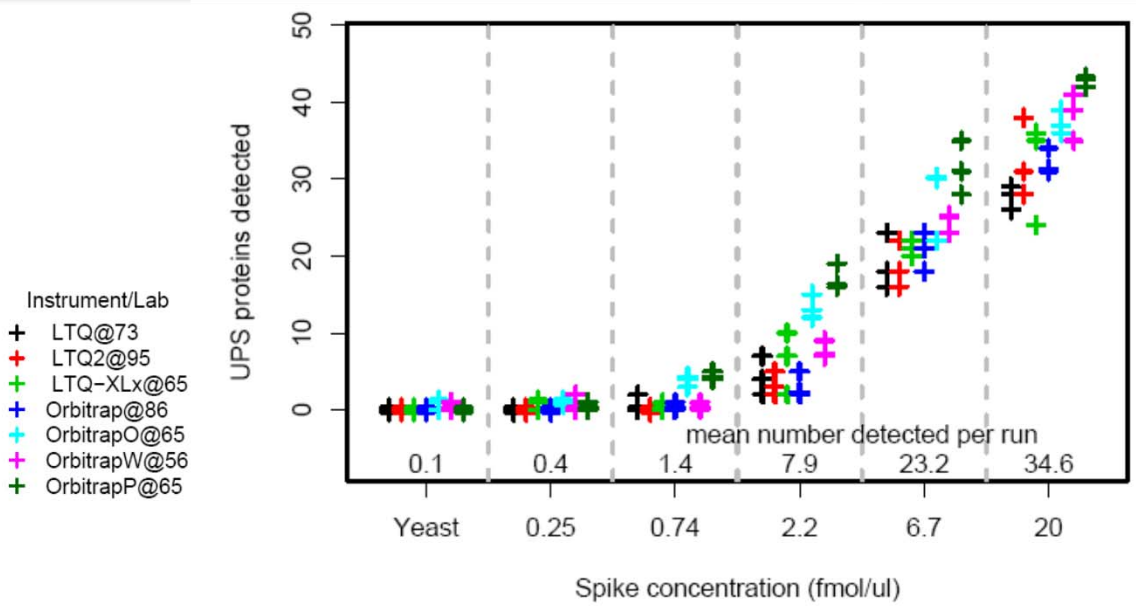
- **Complex protein matrix (6,000 ORFs)**
- **Reproducible preparation possible**
- **Quantitative TAP tag studies (Ghaemmaghami, S. et al. (2003) Nature 425, 737-741) provide calibration for expression levels**
- **Statistical power for modeling depth of coverage by proteomics platforms**
- **Use with human protein spikes enables modeling for biomarker discovery applications**
- **Made available to proteomics community (NIST)**





Ghaemmaghami, S. et al. (2003) Nature 425, 737-741

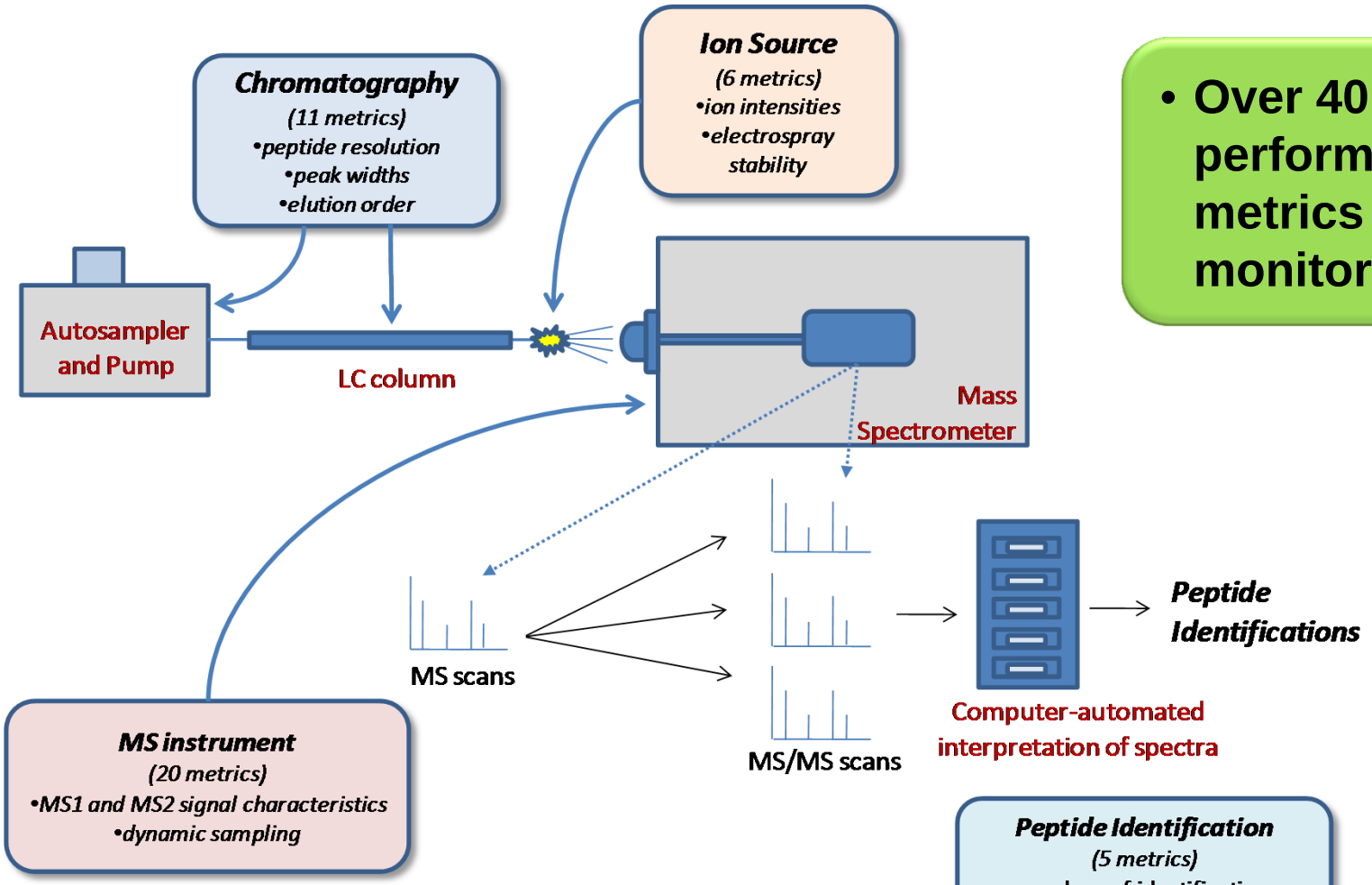
Modeling detection of human protein "biomarker" spikes in yeast matrix



- *48 human protein (UPS) spikes in yeast proteome background*
- *"simple" discovery system distinguishes differences equivalent to 0.5 – 1.3% of proteome*

Case Sample	# identified biomarkers[b]	# true positives[c]	# true negatives[d]	# false positives[e]	# false negatives[f]	Sensitivity[g]	Specificity[h]
Study 6 SOP (yeast+ 0.25 fmol / ul UPS)	0	0	2522	0	48	0.0	1.0
Study 6 SOP (yeast+ 0.74 fmol / ul UPS)	3	3	2522	0	45	0.1	1.0
Study 6 SOP (yeast+ 2.2 fmol / ul UPS)	17	17	2522	0	31	0.4	1.0
Study 6 SOP (yeast+ 6.7 fmol / ul UPS)	32	32	2522	0	16	0.7	1.0
Study 6 SOP (yeast+ 20 fmol / ul UPS)	43	41	2520	2	7	0.9	1.0

System map of LC-MS performance metrics



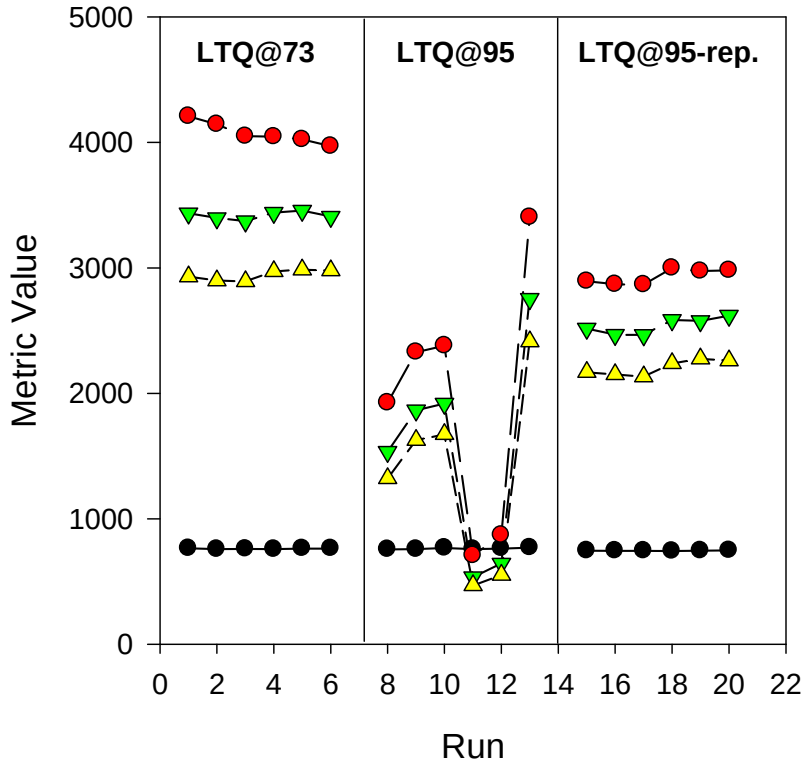
- **Over 40 performance metrics monitored**

Paul Rudnick
Steve Stein
(NIST)

Diagnosing and correcting system malfunction

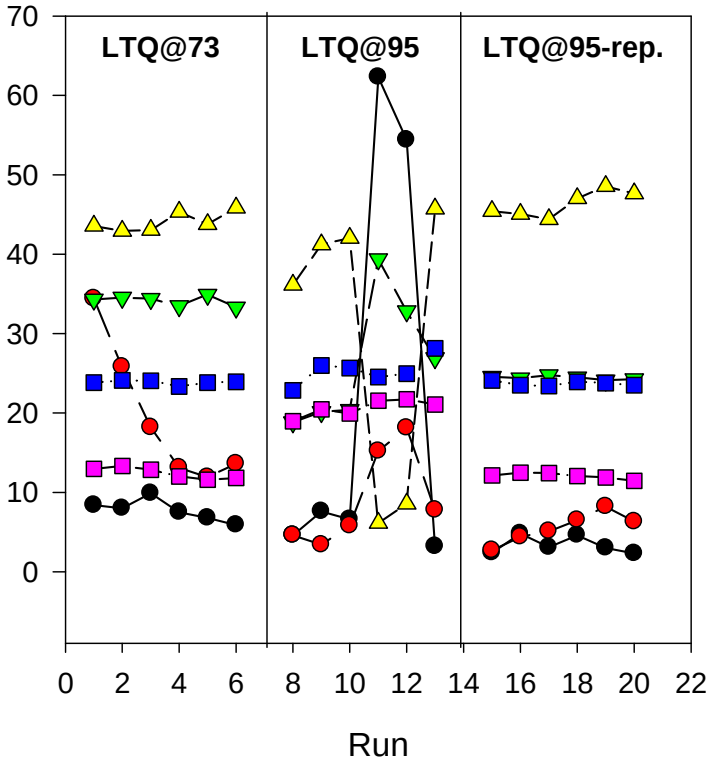
CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

Peptide Identification



- P-1 (med. f-value score for IDs) *1,000
- P-2A (total IDs)
- P-2B (unique ion IDs)
- P-2C (unique peptide IDs)

Chromatography



- C-1A ('bleed' -4 min.) %*100
- C-1B ('bleed' +4 min.) % *100
- C-2A middle 50% pep RT period
- C-2B rate (peptides/min.) over C-2A
- C-3A med. peak width
- C-3B (disperson for peak widths)

CPTAC study 5

CPTAC Unbiased Discovery Workgroup: Key Achievements

CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

1. First systematic, SOP-driven study of LC-MS/MS analytical systems across multiple laboratories
2. Quantitative assessment of repeatability and reproducibility in peptide vs. protein detection
3. Yeast reference proteome standard and accompanying datasets
4. Yeast reference proteome with spikes enables quantitative modeling of power to discover biomarker candidates
5. Performance metrics and software (“toolkit”) to monitor and troubleshoot system performance

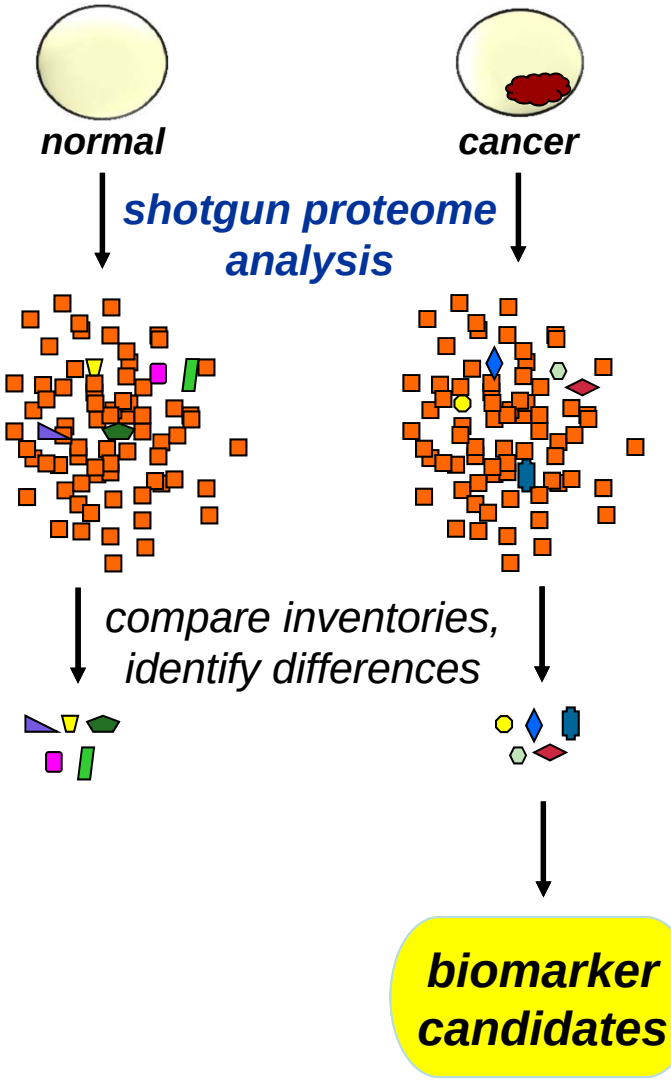
Next steps for Discovery WG

Evaluate performance of platforms to discriminate between cancer-relevant phenotypes

- Phase II Studies
  - Human breast cancer cell model; responses to TKI
  - Compare commonly employed quantitative methods for survey of differences
- Phase III studies
  - Human tumor tissue specimens corresponding to defined clinical phenotypes
    - Evaluate phenotype discrimination
    - Implement methods, metrics and approaches developed in Phase I, Phase II studies

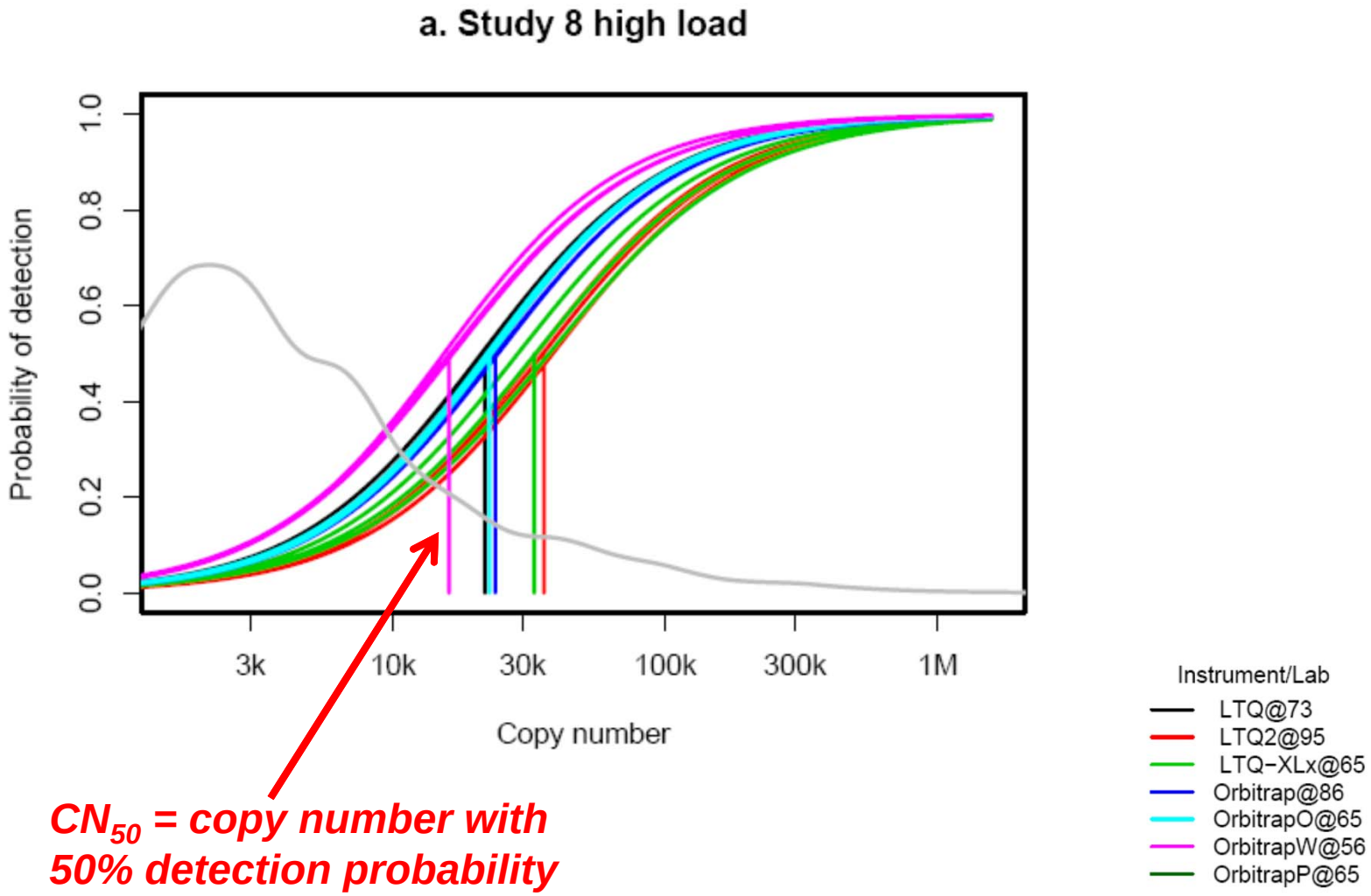
Backups

Why care about reproducibility in discovery proteomics?



1. Biomarker candidates come from comparing proteomes from different phenotypes
2. Need to know whether observed differences are due to biology or to variability in the analytical system.

Yeast proteome enables calibration and comparison of detection efficiency



CN_{50} = copy number with 50% detection probability

Metrics identify the greatest sources of variability

CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

CPTAC Study5 Intralaboratory Variability
3LTQs, 3 Orbitraps, 6 replicates each