THE HUMAN PLASMA PROTEOME: EMERGING RESULTS FROM THE HUPO PLASMA PROTEOME PROJECT

NIDDK/NHGRI Workshop on Standards for Proteomics Data and Analyses

4-5 January, 2005

Gilbert S. Omenn, Univ of Michigan

OUR GENETIC FUTURE

"Mapping the human genetic terrain may rank with the great expeditions of Lewis and Clark, Sir Edmund Hillary, and the Apollo Program."

--Francis Collins, Director

National Human Genome Research Institute, 1999

Next:

Understand the dynamic proteomic compartments.

FINANCIAL TIMES WEDNESDAY FEBRUARY 21 2001

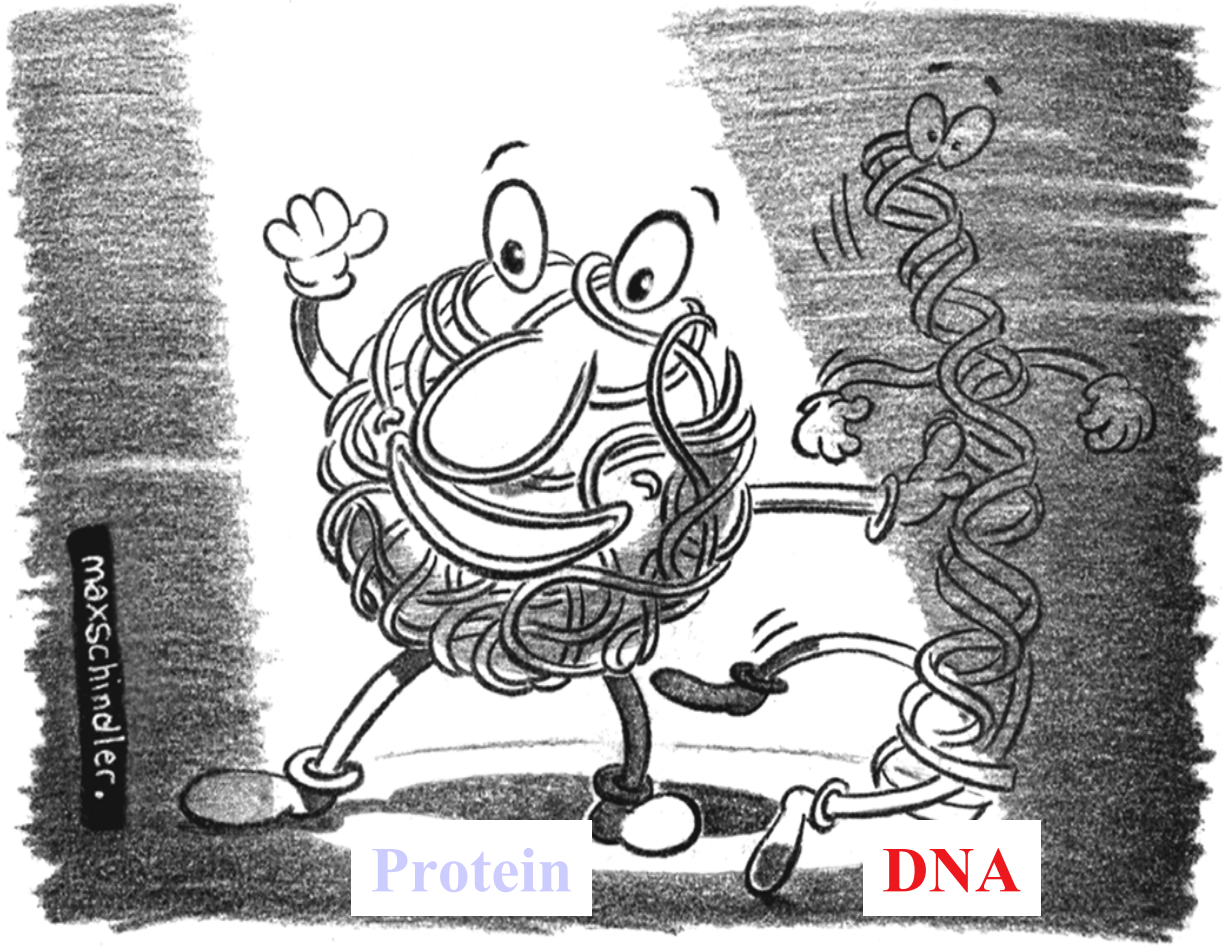
INSIDE TRACK

Strategy, Management, Technology & People

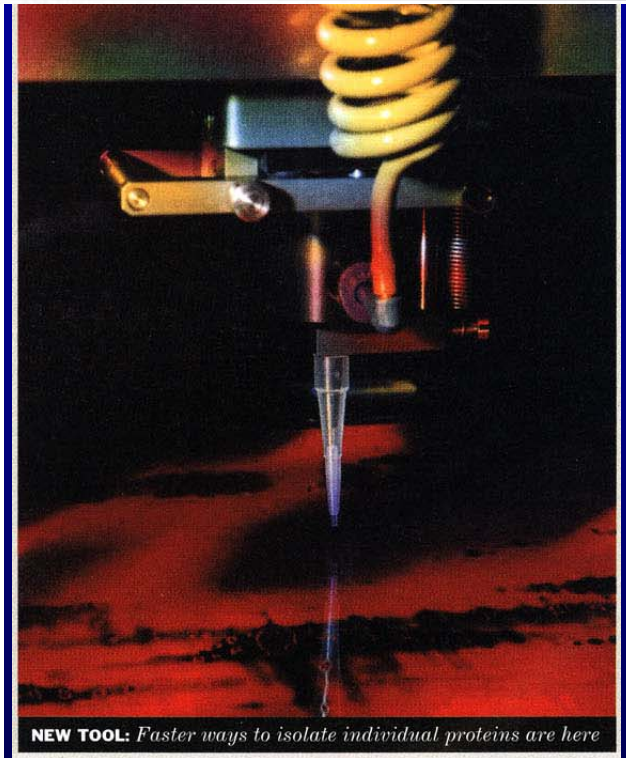
PROTEOMICS

Searching for the real stuff of life

The discovery that humans have fewer genes than expected has thrust proteins into the research spotlight, says Victoria Griffith



www.businessweek.com
BusinessWeek
APRIL 10, 2000
A PUBLICATION OF THE McGRAW-HILL COMPANIES



NEW TOOL: *Faster ways to isolate individual proteins are here*

BIOTECH'S NEXT HOLY GRAIL

Now, companies are racing to decipher the human protein set

Proteomics Standards Needed

1. Pre-analytical variables*
2. Fractionation of proteins
3. Tryptic digestion of proteins
4. Fractionation of peptides
5. Search engine parameters/filters*
6. Database variables*

* Highlighted in following slides

Corresponding needs for microarrays, direct MS-SELDI

Many open statistical questions

PRE-ANALYTICAL VARIABLES (1)

Patient/Specimen Donor: gender, age, diet, genetics, health history, lifestyle variables, fasting vs post-prandial vs random timing, medications

Venipuncture: needle gauge, collection tube set

Phlebotomy: tourniquet; position; tube order; venipuncture or existing line

Collection device: tube/bag; glass/plastic; gel/non-gel separator; protease inhibitors (peptides, small molecules); internal standards?

PRE-ANALYTICAL VARIBLES (2)

- **Blood processing: plasma vs serum; if plasma, EDTA v heparin v citrate; if serum, temp (platelet activation at 4C), duration, clot activator**
- **Separation of cells: centrifugation speed, duration, temp**
- **Aliquoting protocol; duration before analysis**
- **Storage: freezing method, materials; temp; thaw/re-freeze cycles permitted; expiration dating**

SEARCH ENGINE VARIABLES

- Choice of search engine (Sequest, Mascot, Sonar, Spectrum Mill, X!Tandem, Digger) : often embedded in the MS instrument; feasible to re-analyze and compare if have raw spectra or peaklists
- Number of MS runs, duplicates (sampling)
- Parameters/filters for peptide IDs, e.g. Sequest: Xcorr >= 1.9, 2.2, 3.75 for 1+, 2+, 3+; DCn >=0.1; Rsp <=4; fully tryptic
- Variability of "manual inspection" of spectra
- Probability of correct sequence: Mascot scores; PeptideProphet probability and error rate estimates

DATABASE MATCHING

- Choice of database: Swissprot, NCBI-nr, IPI,…
- Version of database: periodic updates
- Extent of annotation: proportion of “null”, “hypothetical”, and “similar to”
- Probability of correct match/estimates of error rates: species included (Homo sapiens, mammalian, broader); methods (Protein/Prophet, reversed sequence db, microbial sequence db)

RESPONSES TO WORKSHOP QUESTIONS

1. Current status: each investigator sets own criteria.

- o Reflects early stage of the field and complexity of analyses compared with individual proteins, some of which have Certified Reference Materials.
- o HUPO PPP uses reference specimens; IPI v 2.21 as database standard; recommended parameters for Sequest peptide IDs; protein concentration determinations, raw spectra, and peaklists for cross-lab analyses of submitted datasets and IDs
- o HUPO Protein Standards Initiative has issued consensus proposals for protein-protein interactions and for MS datasets
- o Carr et al provide guidelines for conduct of experiments and documentation for publication (MCP 2004, 3:531).

Responses to Workshop Questions (2)

2. Other fields: consensus efforts, stepwise refinement, exchange of materials for cross-analyses, improvement in S/N ratios, use of statistical criteria
3. Progress: develop, apply, and evaluate consensus guidelines; compare alternatives at every step from specimen collection to analyses
4. Barriers: too many open questions, evolving tech'y
5. Instrument manufacturers: proprietary
6. Concerns: encourage innovation as well as inter-operability
7. Integration/synergy: link with HUPO and journals

Responses to Workshop Questions (3)

8. "Proteomics Dictionary": good idea
9. Software tools: must be fully described; compared with alternatives
10. Access to data: criteria/filters for peptide and protein IDs; peptide sequences and associated confidence values; for discussion: raw spectra or peaklists
11. Comparisons of software tools: HUPO PSI, HUPO PPP, ISB
12. Journals: yes, enforce data guidelines
13. Data archives: huge undertaking; EBI, American Chemical Society, others have stated willingness
14. Follow-up working groups: YES.