HUPO Human Proteome Project

www.thehpp.org

Maria M. Chan, Ph.D.

CDRH/OIR/DIHD

CDRH Public Workshop, 13 June, 2014

Acknowledgement

Gilbert S. Omenn, M.D., Ph.D.

Center for Computational Medicine and Bioinformatics

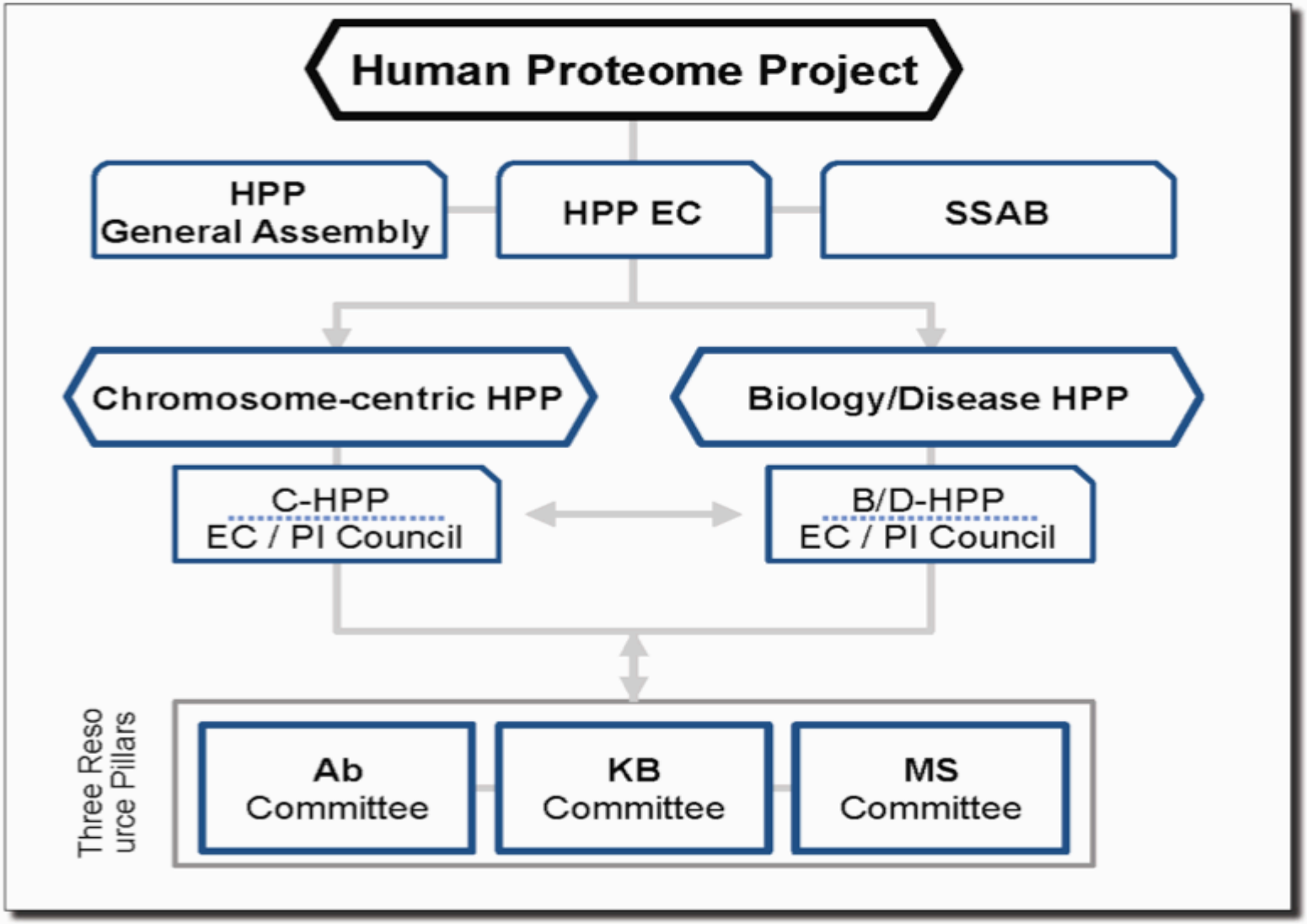
University of Michigan, Ann Arbor, MI

Chair, HPP Exec Committee

The Overall Goals of the HPP

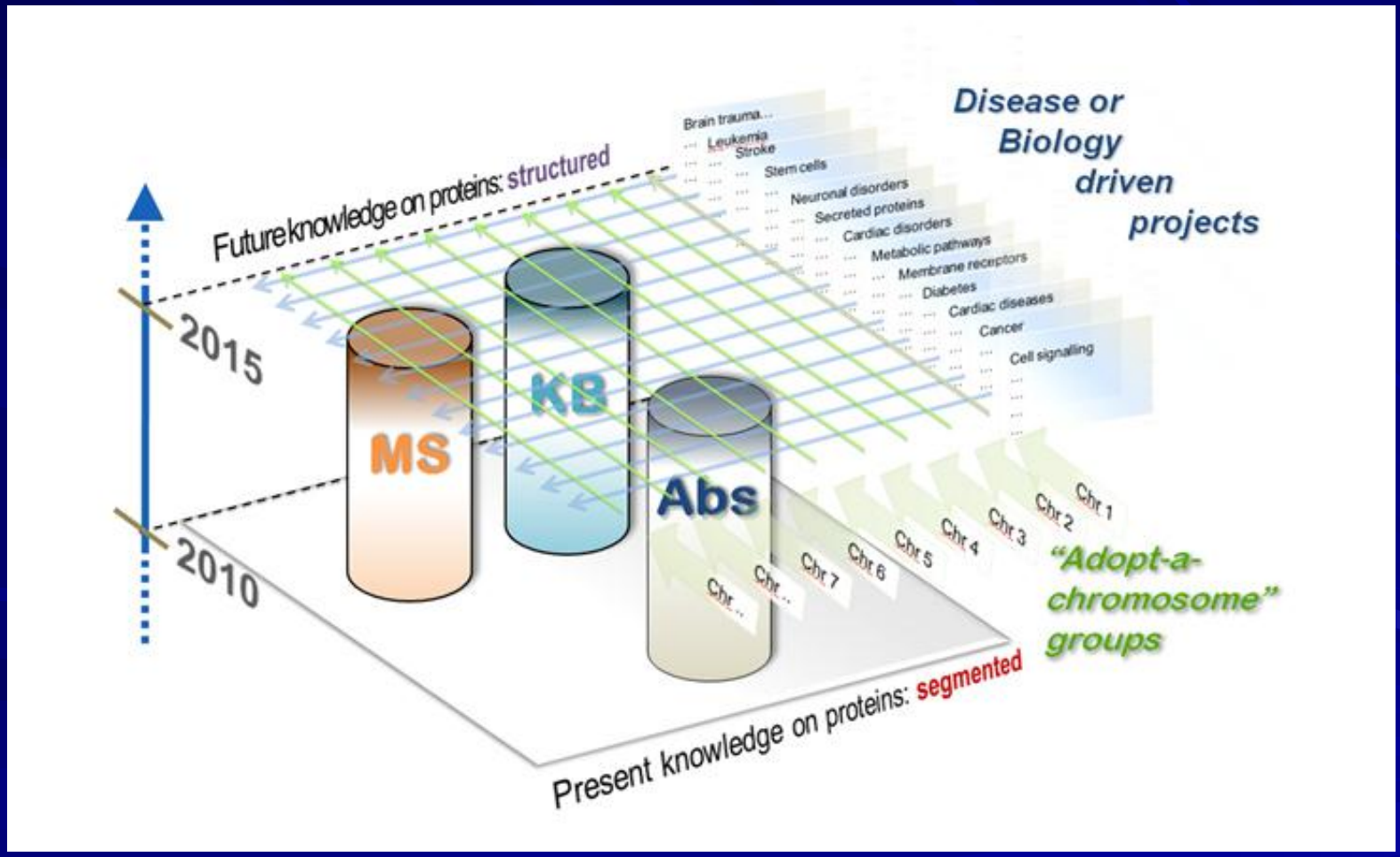
1. Make proteomics a full counterpart to genomics: Enhance the work of the entire biomedical research community with high-throughput instruments, reagents, specimens, pre-analytical preparation, and knowledge bases for identification, quantification, and characterization of proteins in network context.
2. Complete in stepwise fashion the Protein Parts List--identifying and characterizing at least one protein product and as many PTM, SNP, and splice variant isoforms as possible from all of the 20,300 human protein-coding genes that are expressed.

Human Proteome Project: Organization



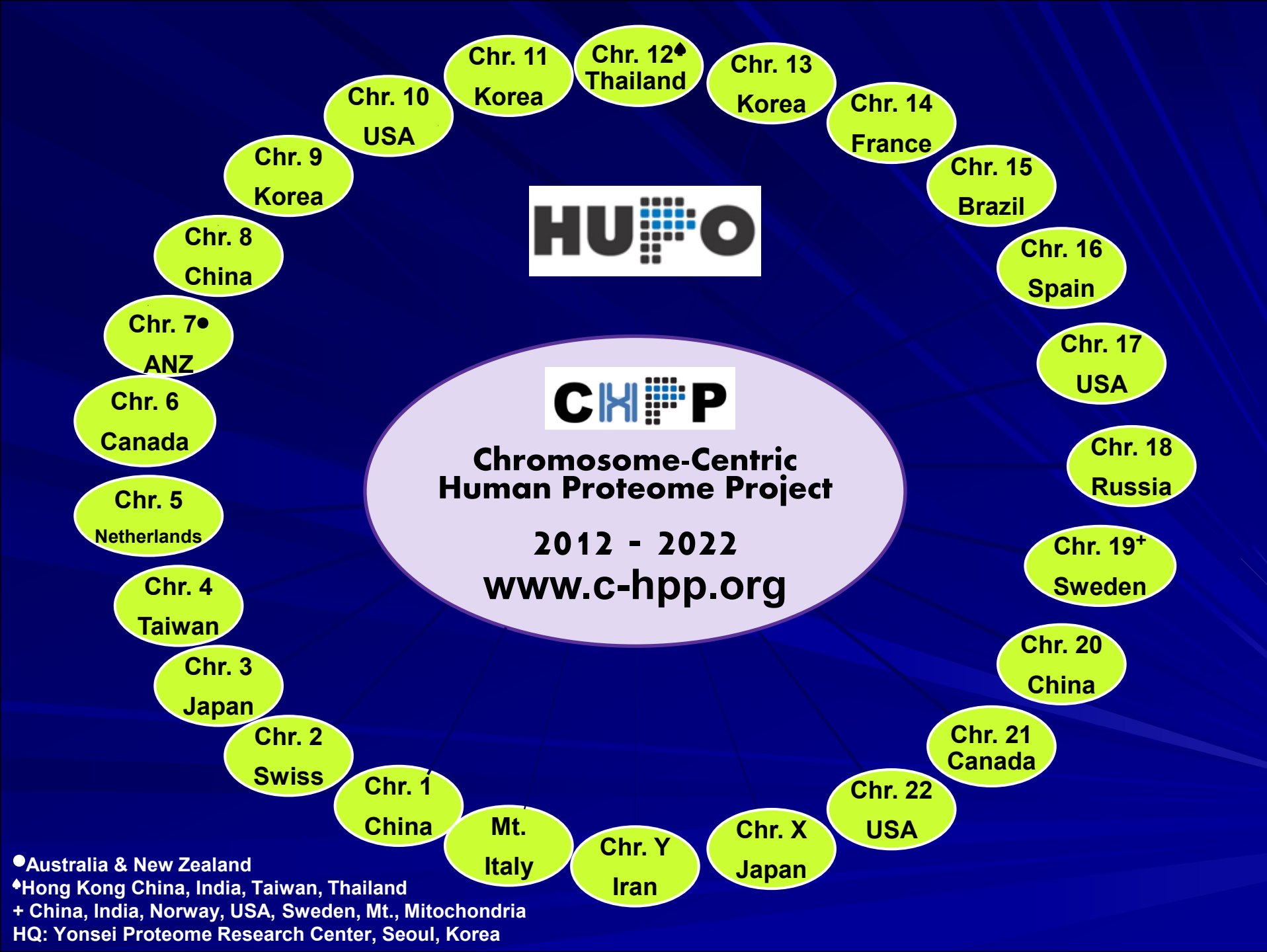
Human Proteome Project

Integration of human proteome information and resources



Chromosome-centric C-HPP

- Analogy to the Human Genome Project
- Goals
  - Map and annotate the entire human protein set encoded in each chromosome
  - Cooperation of 25-membered international consortium
  - 24 chromosomes and mitochondria



●Australia & New Zealand
♠Hong Kong China, India, Taiwan, Thailand
+ China, India, Norway, USA, Sweden, Mt., Mitochondria
HQ: Yonsei Proteome Research Center, Seoul, Korea

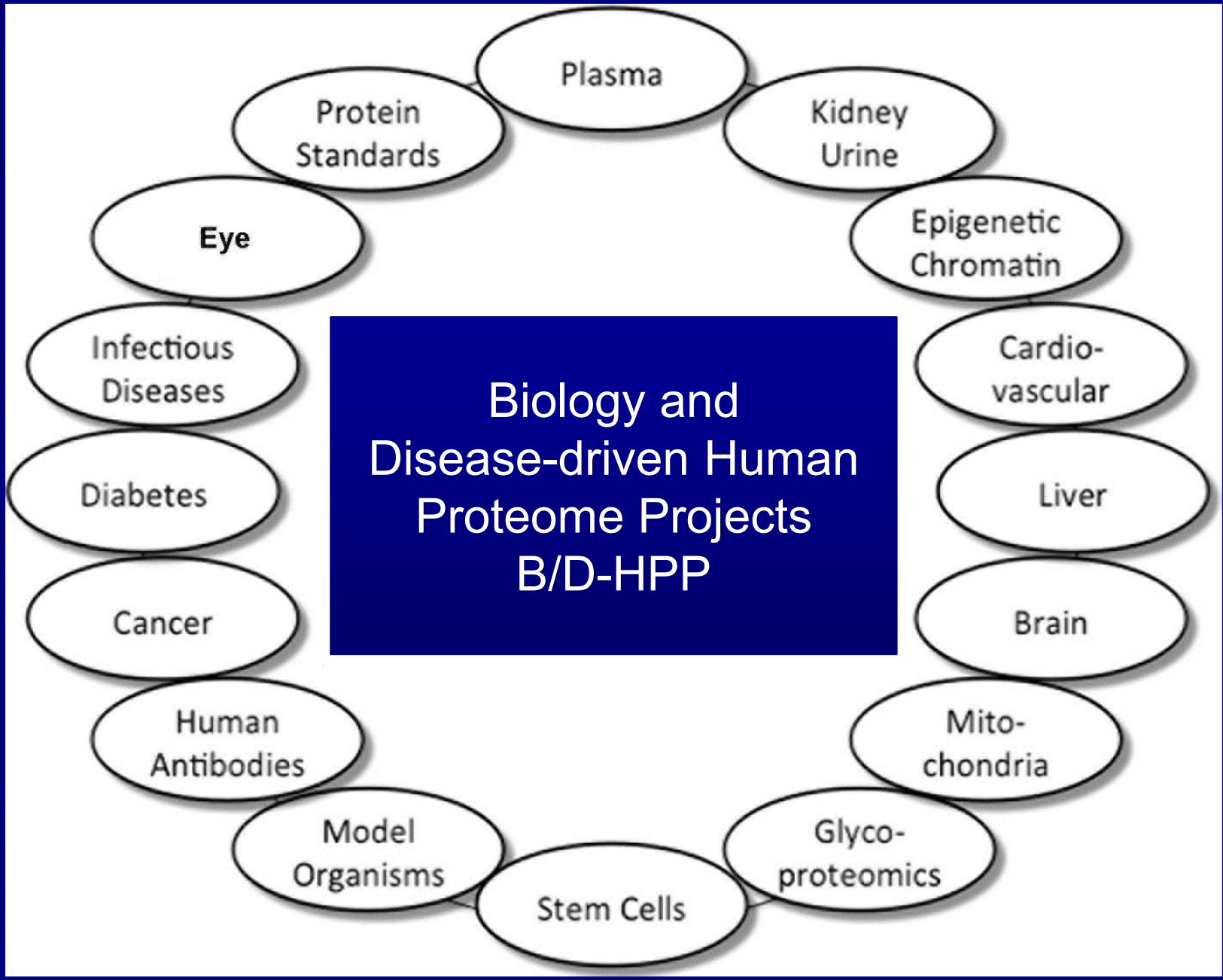
C-HPP Chromosome Centric Human Proteome Project

2013.1.1: First C-HPP Special Issue in Journal of Proteome Research

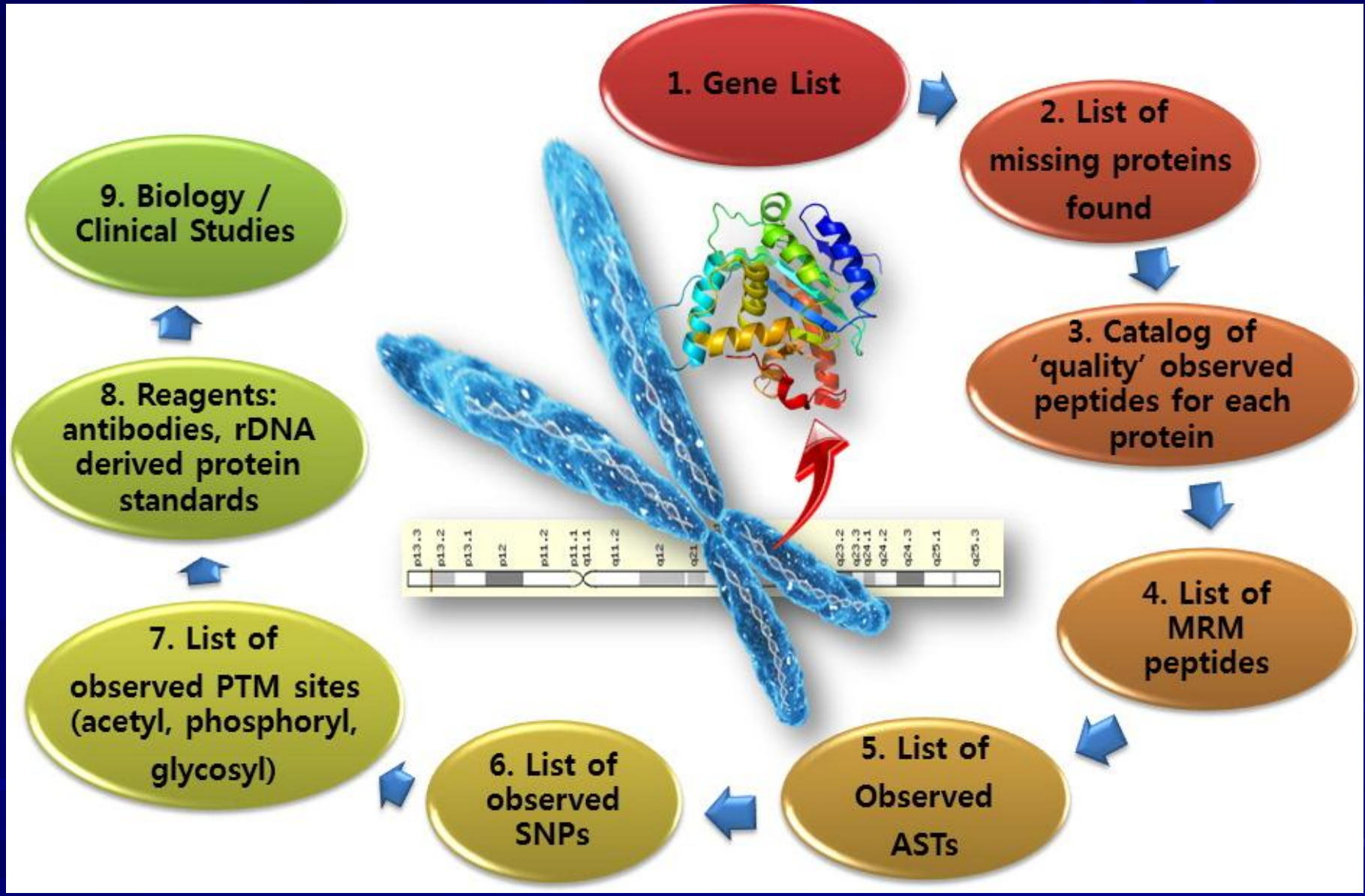
(Credited to Ghasem Hosseini Salekdeh)

Biology/Disease-HPP

- Expand understand of the human proteome with focus on biology research aspects and ongoing disease-focus research
- All active pre-existing initiatives elected to join the B/D-HPP (Biofluid and organ proteome etc.)
- New Initiatives: Diabetes, Cancers, Infectious Diseases, Auto-Immune Disorders, the Eye, Epigenetics, Mitochondria (shared with C-HPP).
- Latest team is Pedi-Ome (Pediatrics).



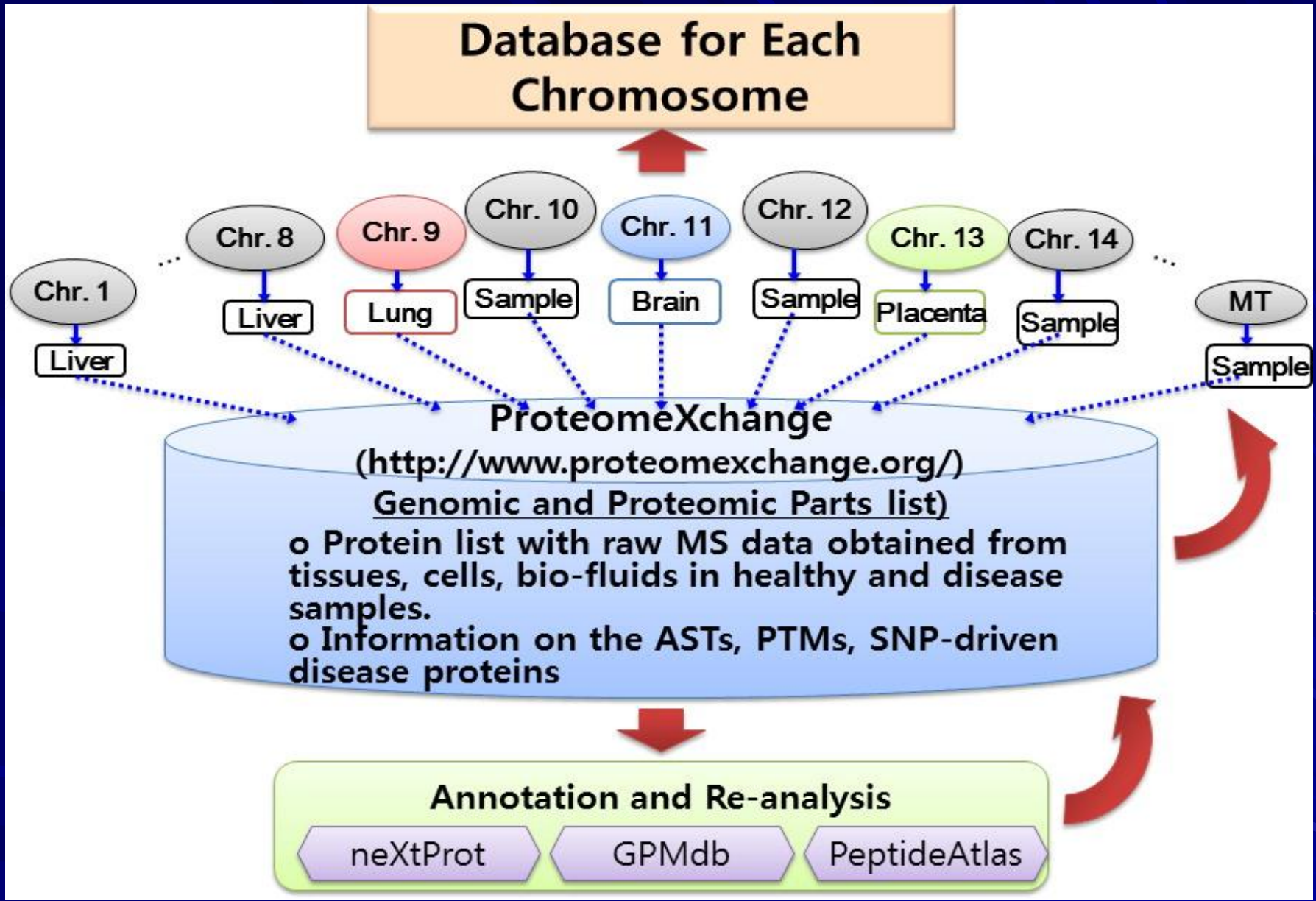
As of July 2013



Potential deliverables from the synergistic collaboration between the C-HPP and B/D-HPP

HPP Datasets

- Dataset Guidelines developed together with PSI and PX with HPP KnowledgeBase Pillar
- MIAPE compliance, PSI compliance and submitted to ProteomeXchange via PRIDE (MS/MS) or PASSEL/PeptideAtlas (MRM)
- PXD identifier must appear in last line of Abstract and be described in Methods
- PX links neXtProt/UniProt/SwissProt, PeptideAtlas, GPMDB, with anticipated addition of MassIVe and iProX



Data exchange and utilization through ProteomeXchange by the C-HPP consortium members

The HPP Master Table for the Parts List

- A starting point for the HPP: For how many of the protein-coding genes do we have confident evidence of a protein product?
- How many "missing" proteins must be found to complete the Human Proteome "parts list"?
- What are the attributes of the leading data repositories/knowledge bases for the launch and deliverables of the HPP?

The Data Resources for the HPP

- Ensembl (current version) provides the number of protein-coding genes.
- neXtProt, Peptide Atlas, and GPMdb provide the numbers of high-confidence proteins identified through mass spectrometry
- Human Protein Atlas provides the number of proteins for which polyclonal antibodies have detected protein expression by immuno-histochemistry (66 cell types, 48 tissues).

Table 1. Numbers of Highly Confident Protein Identifications in Each of the Major Data Resources as of Dec 2012 (Marko-Varga et al.[1]) and Sept 2013[a]

	Chr.	Ensembl protein-coding genes	neXtProt (gold)	human PeptideAtlas (canonical)	GPMdb (green)	human protein atlas evidence (high/medium; now "supportive")
Dec 2012	total	20 059	13 664	12 509	14 300	10 794
Sept 2013	total	20 123	15 646	14 012	14 869	10 976

[a]Inputs from Pascale Gaudet, Lydie Lane, Amos Bairoch—neXtProt; Terry Farrah, Eric Deutsch—Peptide Atlas; Ron Beavis—GPMdb; Emma Lundberg, Mathias Uhlen—Human Protein Atlas; and all C-HPP teams.

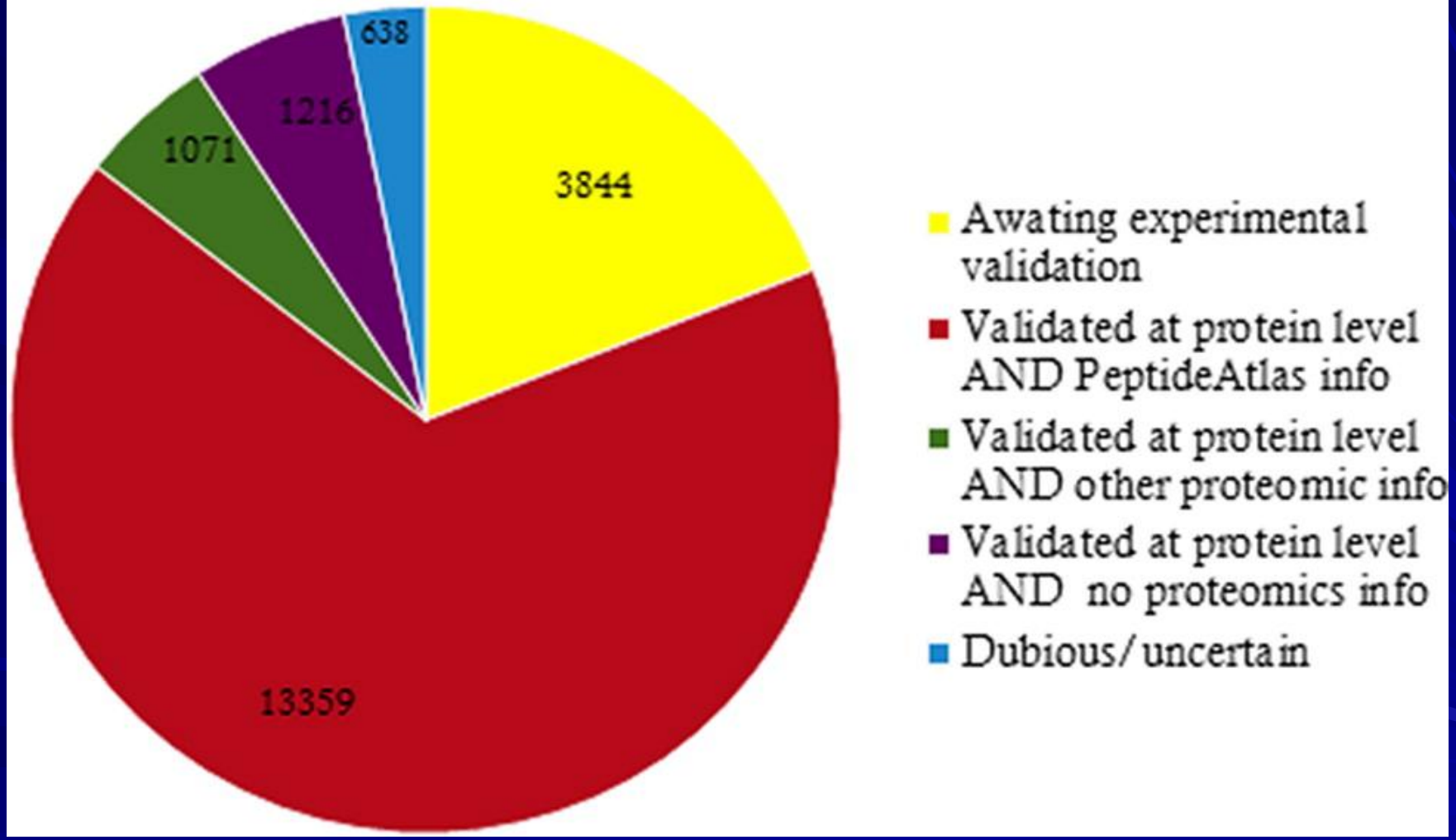
Table 2. Chromosome-by-Chromosome Updated HPP Metrics Table as Guide to Search for Missing Proteins

chromosome	Ensembl protein-coding genes v 72	neXtProt entries (Sept 2013)	neXtProt PE1 (Sept 2013)	PeptideAtlas (July 2013)	GPMdb green (Aug 2013)	HPA evidence supportive (Dec 2013)
1	2059	2061	1600	1415	1521	1119
2	1240	1239	1029	909	962	685
3	1071	1076	877	764	818	634
4	769	763	613	529	566	398
5	862	867	696	624	668	496
6	1094	1108	902	795	784	616
7	938	944	706	660	680	520
8	709	701	553	472	500	383
9	821	821	616	544	567	403
10	762	763	592	528	574	452
11	1314	1321	926	791	845	659
12	1026	1030	842	730	787	609
13	328	328	266	236	251	186
14	625	626	503	445	472	372
15	614	609	457	427	456	328
16	839	831	660	614	671	498
17	1167	1167	951	847	907	648
18	276	278	227	195	219	175
19	1414	1424	1018	1028	1076	716
20	547	552	442	384	398	276
21	254	254	167	142	148	122
22	458	464	356	337	339	254
X	866	827	613	554	626	398
Y	56	47	21	30	22	20
MT	14	14	13	12	12	
total	20123	20115[a]	15646	14012	14869	10967[b]
added since Oct 2012[1]			1982	1503	569	

[a]Total neXtProt entries (20128) include 15 not mapped to chromosome. [b]Human Protein Atlas version 12, released December 5, 2013.

dx.doi.org/10.1021/pr401144x | *J. Proteome Res.* 2014, 13, 15−20

Our Quest for the "Missing Proteins"



Lane, Bairoch, Beavis, Deutsch, Gaudet, Lundberg, Omenn, JPR 2014; 13:15-20

Journal of proteome research

Editorial

pubs.acs.org/jpr

A First Step Toward Completion of a Genome-Wide Characterization of the Human Proteome

Gyorgy Marko-Varga[†]
Gilbert S. Omenn[‡]
Young-Ki Paik[*,§]
William S. Hancock[*,∥]

Special Issue: Chromosome-centric Human Proteome Project

Published: December 20,

dx.

Journal of proteome research

Perspective

pubs.acs.org/jpr

The Biology/Disease-driven Human Proteome Project (B/D-HPP): Enabling Protein Research for the Life Sciences Community

Ruedi Aebersold,[†,‡,§] Gary D. Bader,[§,∥] Aled M. Edwards,[§,⊥,#] Jennifer E. van Eyk,[§,¶] Martin Kussmann,[§,×,□,●] Jun Qin,[§,△,■] and Gilbert S. Omenn[○,◆]

Special Issue: Chromosome-centric Human Proteome Project

Published: December 21, 2012

dx.doi.org/10.1021/pr301151m | *J. Proteome Res.* 2013, 12, 23−27

Additional Steps for the C-HPP

- Develop shared resource centers to prepare and distribute specific cell lines (link with ENCODE) and clinical specimens
- Perform proteomics on unusual cell types
- Perform top-down proteomics of intact proteins to better identify splice variants and SNVs
- Expand work on PTMs
- Develop ultrasensitive LC/MS for microsamples
- Reveal cis-regulatory phenomena

Next Steps for the B/D-HPP

- Make available/promote use of SRM Atlas, spectral libraries, directions for choice of transitions to monitor
- Create priority protein and SRM peptide lists for specific organs and specific diseases for widespread use (diabetes, ovarian cancers)
- Systematically capture data about PTMs, SNPs/SAAPs, splice isoforms, mutations/sequence variants; relate to biological networks and functions
- Plan B/D-HPP J Proteomics series of articles from Madrid HUPO-2014 Congress

Unmet Needs for PSI/MIAPE Standards

- Standards for various PTMs, including site, stoichiometry, cross-talk, functional effects
- Standards for splice isoforms, including splice junctions, mechanism of splicing, and functional consequences—recognizing differential expression of splice variants, sometimes with opposing actions, from the same gene

Unmet Needs for PX

- Standards for additional data resources to join the ProteomeXchange---recognizing that HPP teams, and others, are producing a variety of browsers and data resources
- As PX grows further, a sorting mechanism by species, organ, laboratory, and nature of study will add value for users
- Cooperation with journals to require inclusion of PXD identifiers and public release of data

Potential Applications of HPP to Regulatory World

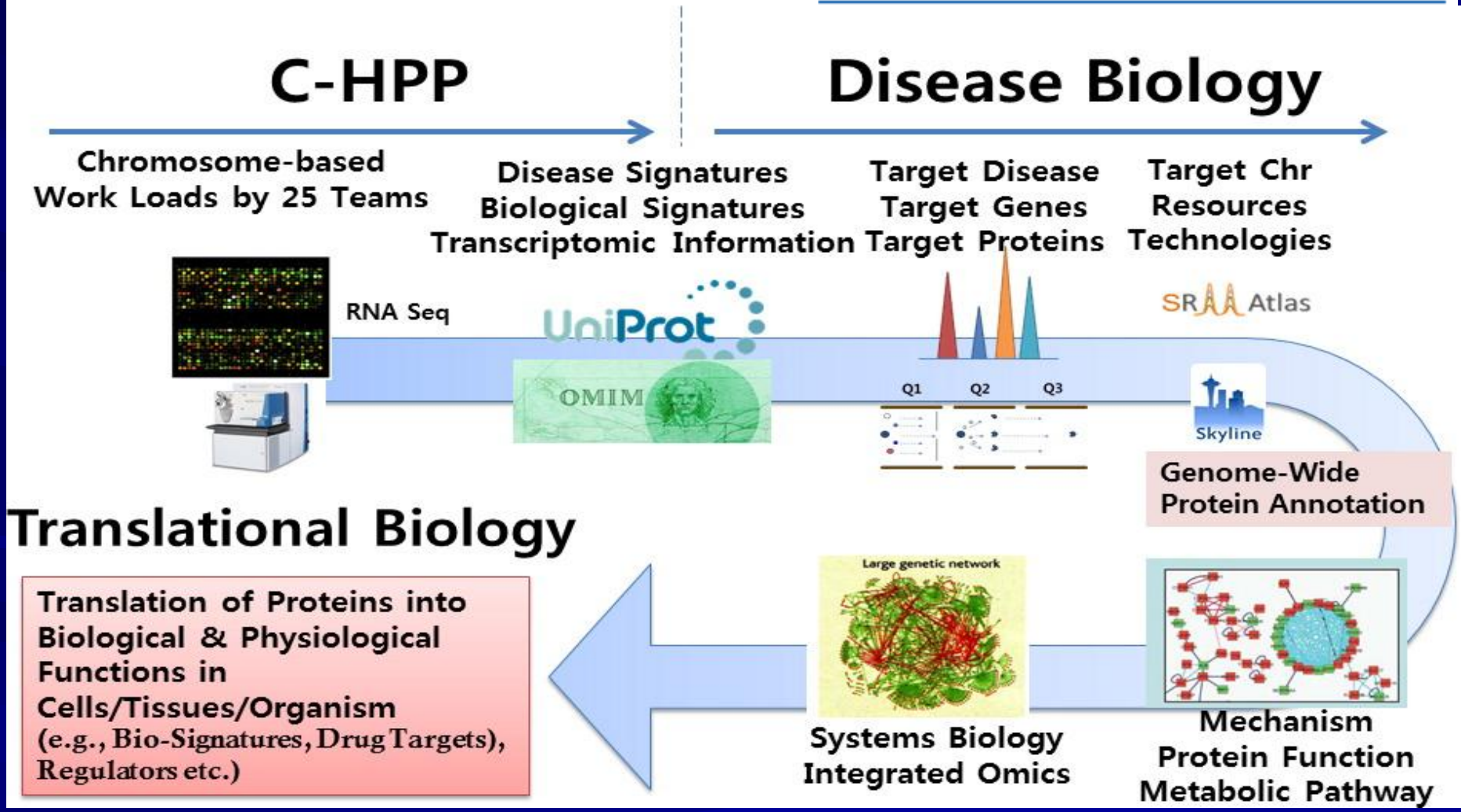
- The HPP results and careful curation and annotation in Peptide Atlas, GPMDB, neXtProt, and Human Protein Atlas provide a valuable resource for credibly identified proteins, peptides, and protein variants.
- Pathway and network analysis of these findings can point to potential drug targets, as well as off-target effects and resistance mechanisms.

Journal of proteome research

Editorial

pubs.acs.org/jpr

Genome-wide Proteomics, Chromosome-centric Human Proteome Project (C-HPP), Part II



www.thehpp.org