Chromosome-centric Human Proteome Project (C-HPP)

HPP は、シドニーＨＵＰＯ国際会議（Sept. 2010）でアナウンスされ、ジュネーブでの国際会議で正式に承認された。 ゴールは、ヒト染色体上にコードされ、発現されるすべてのタンパク質に関する時間、場所、修飾、アイソフォーム、ＳＡＰなどすべて状態と機能に関する知識を統合整理して全く新しいデータベスを構築することにある。HPP は、２つの大きな柱を中心とする国際的な研究活動である。 １．C-HPP (chromosome-centric HPP) (Chair: Young-Ki Paik) と ２．B/D Proteomic Projects (biology and disease driven proteomics projects)(Chair: Ruedi Aebersol)である。 日本チームは染色体３番プロジェクト（PI: 西村俊秀 東京医科大学外科学第一講座）を担当し、産業技術総合研究所と共同で完全長 cDNA を基盤とする新規タンパク質データベース(H-Inv Extended Protein DB, H-EPD)を構築している。世界にはない完全長 cDNA（トランスクリプトームでのエビデンスがある）日本独自のタンパク質データベースを確立している。本データベースは、現在利用されているスイス、ＵＳのデータベースと共通の約18,000タンパク質群以外に、約17,000種のまったくアノテーションされていない独自タンパク質候補（融合蛋白質も含む）を有する。タンパク質同定に公に使用されているデータベースはゲノムデータベースをもととしているが、本新規タンパク質データベースはトランスクリプトーム（mRNA）をベースにしており世界で他に類を見ない独自性を持っている。

Human Proteome Project (HPP): Announced at the Sydney World Congress in September 2010 , launched at the Geneva World Congress in September 2011. The goal of the HPP is to construct integrated and structured Human Protein Encyclopedia in which entire proteins expressed in various forms (spatial and temporal information, modifications, and SAPs, and isoforms) coded on entire human chromosomes. Supporting scientific and technology resource pillars for the HPP is the wide array of 1) mass spectrometry platforms, 2) the antibody-based Human Protein, and 3) ProteomeXchange to integrate proteomics-based knowledge-bases (Fig. 1).

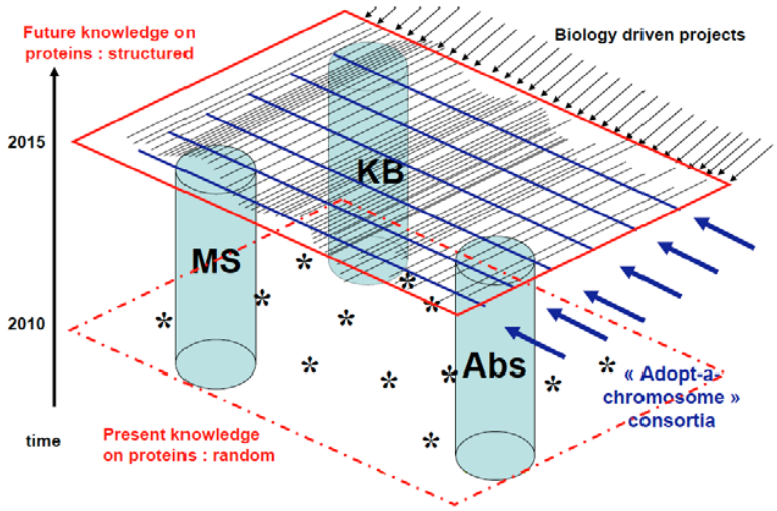
The HPP is organized into two investigative arms: a chromosome-centric HPP (C-HPP)(Fig. 2), with consortia so far, and a complementary effort to facilitate extensive biology and disease driven (B/D) proteomics projects. The Chairmans for C-HPP and B/D Proteomic Projects are Professors Young-Ki Paik and Ruedi Aebersold, respectively.

Prof. Toshihide Nishimura (Tokyo Medical Univ.) is leading Chromosome 3 Team (Headquarter: Prof. Hiromasa Tojo (Osaka Univ.); Bioinformatics: Prof. Tadashi Imanishi (AIST, Tokai Univ.)).Fig. 3 shows a current international collaborative contributors. Chromsome 3 Team has already published its activities in Journal of Proteome Res. 12(1): 62-66, 2013. Currently, Chromosome 3 Team has been constructing Unique H-Inv Extended Protein Database (H-EPD;

http://hinv.jp/hinv/h-epd/) has been developed by extending predicted protein data constructed from the full-length transcriptome-base H-Invitational Database (H-InvDB; http://hinv.jp/). (Fig. 4)

- Tadashi Imanishi, Yoko Nagai, Takuya Habara, Chisato Yamasaki, Jun-ichi Takeda, Sayaka Mikami, Yasuhiko Bando, Hiromasa Tojo, and Toshihide Nishimura Perspectives: Full-length transcriptome-based H-InvDB throws a new light on chromosome-centric proteomics. C-HPP Special Issue: 2013 Journal of Proteom Res. 12(1): 62-66.
- Hiromasa Tojo, Yoko Nagai, Sayaka Mikami, Masaharu Nomura, Kiyonaga Fujii, Yasuhiko Bando, Harubumi Kato, Norihiko Ikeda, György Marko-Varga, Tadashi Imanishi, Toshihide Nishimura Exploring Missing Proteins Encoded on Chromosome 3 by Re-Analysis of FFPE Lung Cancer Proteomic Data with Full-Length Transcriptome-based Database. Submitted to to JPR.
- Young-Ki Paik, Gilbert S. Omenn, Mathias Uhlen, Samir Hanash, György Marko-Varga, Ruedi Aebersold, Amos Bairoch, Tadashi Yamamoto, Pierre Legrain, Hyoung-Joo Lee, Keun Na, Seul-Ki Jeong, Fuchu He, Pierre-Alain Binz, Toshihide Nishimura, Paul Keown, Mark S. Baker, Jong Shin Yoo, Jerome Garin, Alexander Archakov, John Bergeron, Ghasem Hosseini Salekdeh, and William S. Hancock. Standard Guidelines for the Chromosome-Centric Human Proteome Project. 2012 J. Proteome Res., 11 (4), 2005–2013
- Hiromasa Tojo, Sayaka Mikami, Takeshi Kawamura, Yoko Nagai, Chisato Yamasaki, Yasuhiko Bando, Norihiko Ikeda, Harubumi Kato,György Marko-Varga, Tadashi Imanishi, Toshihide Nishimura. Is “Dark Matter” Really Expressed? : Hypothetical Short Proteins (20 - 79aa) Unique To H-inv Extended Protein Database (H-EPD). Proteome Forum Berlin 2013, March 17-22, 2013. (Abstract submitted.)

Fig. 1



Schematic representation of the HUPO Human Proteome Project. The three pillars represent: knowledge base (KB), mass spectrometry (MS), and protein capture (antibodies, Abs). The biology-driven projects are now captured in the B/D-HPP, and the adopt-a-chromosome consortia are the chromosome-specific C-HPP teams. (from Legrain et al., Mol Cell Proteomics, 2011 9.

Fig.2

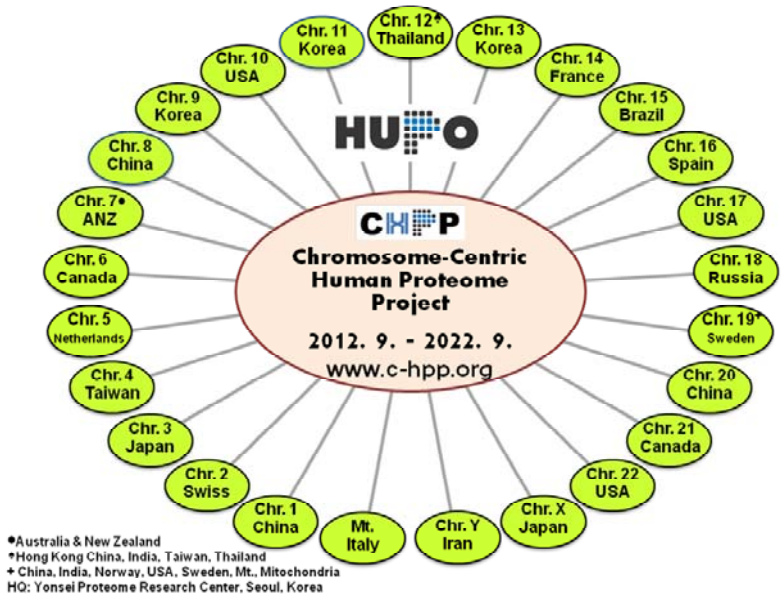
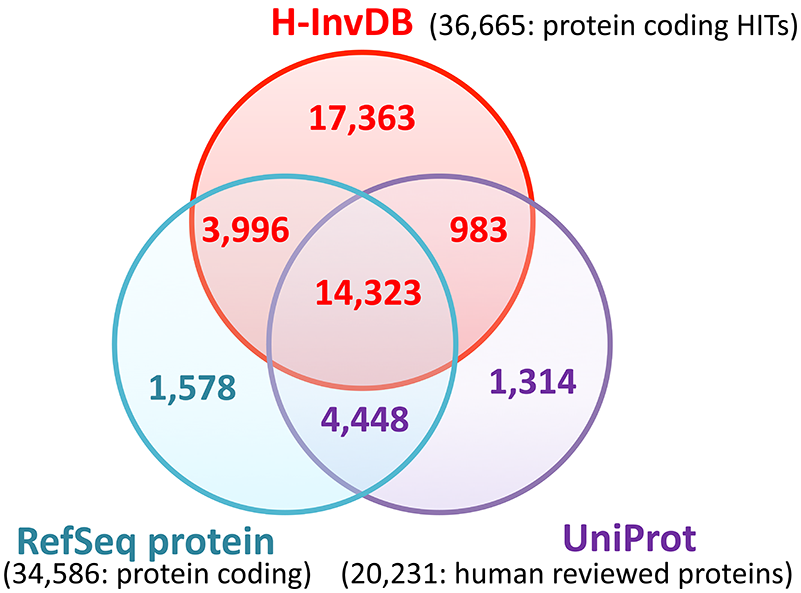


Fig. 3 Chromosome #3 Team

Fig. 4



Venn diagram of protein-coding genes/protein candidates of H-InvDB, UniProtKB/Swiss-Prot, and RefSeq protein. Values indicate numbers of protein entries for H-InvDB (red), UniProtKB/Swiss-Prot (purple), and RefSeq protein (blue). RefSeq protein includes AS isoforms, but H-InvDB and UniProtKB/Swiss-Prot do not. Numbers in the UniProtKB/Swiss-Prot circle do not sum up to the total number of UniProtKB/Swiss-Prot entries, because some UniProtKB/Swiss-Prot entries correspond to multiple H-InvDB entries for duplicated genes.