

Improving Chr-16 proteins coverage with high confidence and MIAPE compliant MS data



M. Sanchez del Pino^(1,10), J.A. Medina-Aunon^(2,10), L. Valero^(1,10), S.I. Gharbi^(2,10), C. Gonzalez-Tejedo^(2,10), K. Eloria^(3,10), J.M. Arizmendi^(3,10), F. Elortza^(4,10), M. Azkargorta^(4,10), M. Carrascal^(5,10), J. Abian^(5,10), I. Casal^(6,10), E. Olivera^(7,10), N. Colomé^(8,10), F. Canals^(8,10), F.J. Corrales^(9,10), J.P. Albar^(2,10)

(1) Valencia University (UV), Spain. (2) National Center of Biotechnology (CNB) - CSIC, Madrid, Spain. (3) University of the Vasque Country (UV-EHU), Bilbao, Spain, (4) CICBioGUNE, Bilbao, Spain, (5) LP-CSIC-UAB, Barcelona, Spain, (6) Biological Research Centre (CIB)-CSIC, Madrid, Spain, (7) Parc Científic de Barcelona (PCB), Barcelona, Spain, (8) Hospital Universitario Vall d'Hebron (HUVH), Barcelona, Spain, (9) Center for Applied Medical Research (CIMA), (10) ProteoRed-ISCIII.

A pilot shotgun study was conducted among 8 laboratories of the Chr-16 (Spanish) HPP consortium. The main goals of this experiment are: 1) to provide "in house", high confidence experimental MS data of the Chr16 proteins that could be used to develop SRM/MRM assays, 2) to define the standard operation procedures for further large scale experiments, 3) to submit generated data on the ProteomeXchange repository according to the HUPO PSI's guidelines and 4) to get information about the total proteome coverage of high confidence by integrating different mass spectrometry analysis platforms.

Four human cell lines were selected, MCF7 breast cancer epithelial cells, CCD18 colon fibroblasts, Jurkat T lymphocytes and RAMOS B lymphocytes, since a combination of fibroblasts, lymphoid and epithelial cells could provide theoretically up to 71% coverage of chromosome 16 proteins. This coverage was confirmed by transcriptomics experiments (J Proteome Res. 2013;12:112-22).

A robust pipeline to submit full MIAPE compliant data to PRIDE/ProteomeXchange repository were developed and 19 large scale shotgun 2D-LC-MS/MS experiments were performed in different MS platforms, giving a 45% Chr-16 protein coverage. The total coverage was improved by the combination of cell lines, extraction/fractionation procedures and analytical platforms more than accumulation of replicates from the same laboratory, underlying the importance of a multi-centric approach for this kind of projects.

Data Summary

Number of experiments:

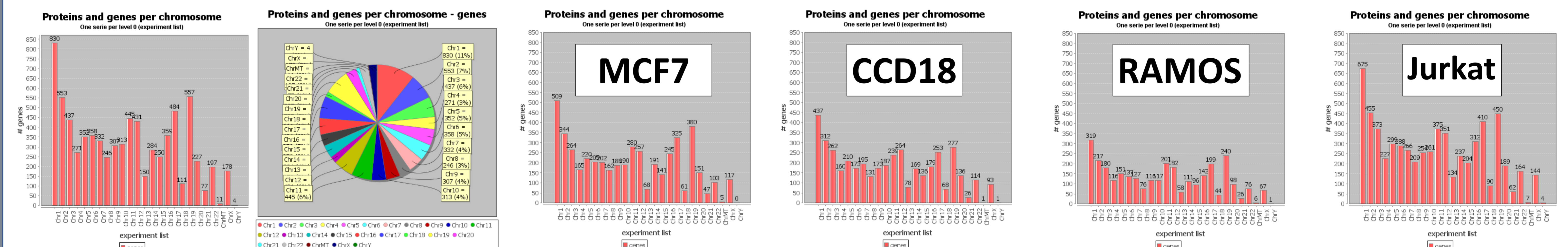
Cell line	separation			
	HPLC-RP	1D gel	offgel	
MCF7	4	3	1	8
CCD18	1	5		6
RAMOS	3			3
Jurkat	2			2
Total	10	8	1	19

Number of peptides/proteins/genes (1% FDR protein level):

Cell line	Total			Chr16		
	Genes	Proteins	Peptides (*)	Genes	Proteins	Peptides (*)
MCF7	5.161	5.273	37.602	272	264	1.693
CCD18	4.670	4.670	26.959	205	199	1.043
RAMOS	3.413	3.343	15.945	152	150	714
Jurkat	6.785	6.989	55.055	332	332	2.420
Total	8.131	8.766	77.805	383	383	3.345

(*) unique peptides

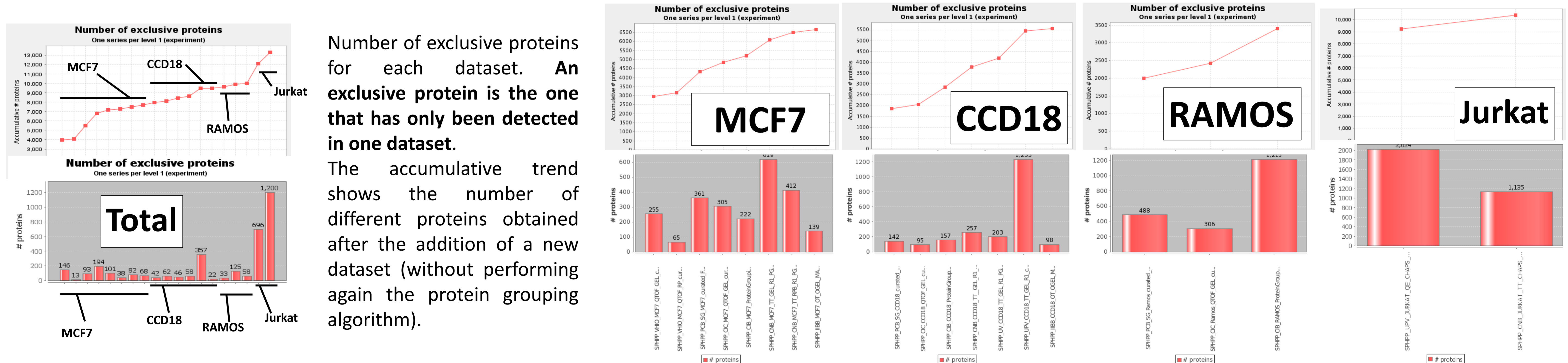
Chromosome coverages



Total

For each chromosome, here we show the number of protein coding genes that have been mapped to at least one detected protein

Exclusive proteins and accumulative data integration



Number of exclusive proteins for each dataset. An exclusive protein is the one that has only been detected in one dataset.

The accumulative trend shows the number of different proteins obtained after the addition of a new dataset (without performing again the protein grouping algorithm).

ProteomeXchange submission

Cell line	ProteomeXchange Submission Statistics				
	# experiments	# submitted files	Total file size	Submission time	PX accession
MCF7	8	538	60.8 Gb	3h 36 min	PXD 000442
CCD18	6	442	97.1 Gb	6h 36 min	PXD 000449
RAMOS	3	228	30.3 Gb	1h 56 min	PXD 000447
Jurkat	2	102	33.8 Gb	3h 01 min	PXD 000443
Total	19	1310	222 Gb	~ 14 h	

A fully MIAPE compliant ProteomeXchange data submission was performed using the ProteoRed MIAPE Extractor

Each fraction/band means:

- 1 MS raw file.
- 1 mgf peak file.
- 1 MSI mzIdentML file.
- 1 PRIDE result file.
- 2 MIAPE (MS and MSI) files.