Chromosome 18 Transcriptome Profiling and Targeted Proteome Mapping in Depleted Plasma, Liver Tissue and HepG2 Cells

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · DECEMBER 2012

Impact Factor: 4.25 · DOI: 10.1021/pr300821n · Source: PubMed

Journal of proteome research

Article

pubs.acs.org/jpr

Chromosome 18 Transcriptome Profiling and Targeted Proteome Mapping in Depleted Plasma, Liver Tissue and HepG2 Cells

Victor G. Zgoda,[†] Arthur T. Kopylov,[†] Olga V. Tikhonova,[†] Alexander A. Moisa,[†] Nadezhda V. Pyndyk,[†] Tatyana E. Farafonova,[†] Svetlana E. Novikova,[†] Andrey V. Lisitsa,[†] Elena A. Ponomarenko,[†] Ekaterina V. Poverennaya,[†] Sergey P. Radko,[†] Svetlana A. Khmeleva,[†] Leonid K. Kurbatov,[†] Aleksey D. Filimonov,[†] Nadezhda A. Bogolyubova,[†] Ekaterina V. Ilgisonis,[†] Aleksey L. Chernobrovkin,[†] Alexis S. Ivanov,[†] Alexei E. Medvedev,[†] Yury V. Mezentsev,[†] Sergei A. Moshkovskii,[†] Stanislav N. Naryzhny,[†,‡] Elena N. Ilina,[§] Elena S. Kostrjukova,[§] Dmitry G. Alexeev,[§] Alexander V. Tyakht,[§] Vadim M. Govorun,[§] and Alexander I. Archakov*[,†]

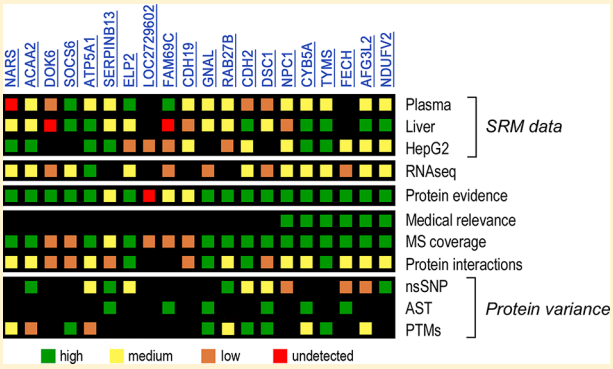
[†]Orekhovich Institute of Biomedical Chemistry of the Russian Academy of Medical Sciences, Russia

[‡]Petersburg Nuclear Physics Institute, Gatchina, Leningrad district, Russia

[§]Research Institute of Physical Chemical Medicine of the Federal Medical-Biological Agency of the Russian Federation, Russia

Supporting Information

ABSTRACT: The final goal of the Russian part of the Chromosome-centric Human Proteome Project (C-HPP) was established as the analysis of the chromosome 18 (Chr 18) protein complement in plasma, liver tissue and HepG2 cells with the sensitivity of 10^{-18} M. Using SRM, we have recently targeted 277 Chr 18 proteins in plasma, liver, and HepG2 cells. On the basis of the results of the survey, the SRM assays were drafted for 250 proteins: 41 proteins were found only in the liver tissue, 82 proteins were specifically detected in depleted plasma, and 127 proteins were mapped in both samples. The targeted analysis of HepG2 cells was carried out for 49 proteins; 41 of them were successfully registered using ordinary SRM and 5 additional proteins were registered using a combination of irreversible binding of proteins on CN-Br Sepharose 4B with SRM. Transcriptome profiling of HepG2 cells performed by RNAseq and RT-PCR has shown a significant correlation (r = 0.78) for 42 gene transcripts. A pilot affinity-based interactome analysis was performed for cytochrome b5 using analytical and preparative optical biosensor fishing followed by MS analysis of the fished proteins. All of the data on the proteome complement of the Chr 18 have been integrated into our gene-centric knowledgebase (www.kb18.ru).



KEYWORDS: *selected reaction monitoring (SRM), mass spectrometry, targeted proteomics, chromosome 18, human proteome project*

INTRODUCTION

As indicated in the roadmap of the Russian part of the Human Proteome Project (HPP), the Chr 18 is a target for gene-centric inventory of protein species.[1,2] The task accomplishment requires resolving the problem of concentration sensitivity in analytical means of proteomics providing detection of low- and ultralow-copied proteins in the biomaterial.[3−5] Only a portion of unmodified proteins, defined as master proteins, present in a high/medium-copied range is currently available for proteomic inventory. The master protein is the primary translation product of the coding sequence; it is at least one of the known protein forms, coded by the gene.[1]

Unlike the genome, the human proteome is tissue-specific: the tissue/organ-based dimension provides links between the gene centric HPP and pre-existing HUPO initiatives. Among HUPO initiatives, the plasma proteome project[6−8] (HPPP) and the liver proteome project[9,10] (HLPP) are the most profoundly developed.

We have chosen the liver because this biomaterial has been extensively studied by several Russian research groups using proteomic approaches.[11−14] The sufficient amount of mass-spectrometry data sets have been deposited into the public domain as the result of HLPP.[9,10] Presuming possible difficulties with an access to the liver samples, the hepatocellular carcinoma-derived cell line HepG2 was selected for the chromosome-centric HPP study. This cell line was reported as an appropriate model of hepatocytes,[15] exhibiting

Special Issue: Chromosome-centric Human Proteome Project

Received: August 30, 2012
Published: December 20, 2012

substantial similarity at the proteome level.[16] Recent deep proteome mining of HepG2 cells by tandem mass spectrometry (MS/MS) revealed 125 proteins originating from the Chr 18.[17]

In contrast to the liver, plasma is not an evident object for the investigation, as proteomes are generally specific for the cells, by which proteins are essentially produced. The rationale behind the chromosome-centric analysis of plasma is based on consideration of blood as a natural collector of variety of the protein species produced in the human body.[18] Therefore, targeted interrogation of plasma can be virtually successful for the products of every expressible gene provided that the appropriate analytical sensitivity of the measurements is achieved. Analysis of plasma proteome also has a practical impact as this type of biomaterial is most often used for medical diagnostics.

The final goal of the Russian part of C-HPP has been formulated as analysis of the Chr 18 protein complement in plasma, liver and HepG2 cells with a sensitivity of 10^{-18} M,[19] which corresponds to one protein molecule per 10^7 liver or HepG2 cells or 1 μL of plasma. The depth of proteome mining has been limited by the sensitivity threshold of 10^{-18} M, because using irreversible binding of proteins to CN-Br activated Sepharose 4B for protein enrichment, we have shown that 10^{-18} M is a minimal protein concentration detectable using the targeted SRM approach for BSA and BM3 proteins.[20] Using an ultra sensitive approach, the MS evidence could be collected for low-copied protein species, currently missing from the public databases.

In a broad sense, the scope of "missing" proteins is not restricted to unidentified proteins, which detection does not meet the gold threshold of neXtProt, the green threshold of GPMdb, the FDR of 1% at protein level in Peptide Atlas, or confident detection in Human ProteinAtlas. "Missing" proteins also includes those lacking accurate concentration measurement in a particular biological tissue.

The experimental part of this study is presented by the results of analysis of the transcripts level in the HepG2 cells, complemented by the SRM survey of master proteins of the total 277 genes of Chr 18 in plasma, liver tissue and HepG2 cells. We also report here the experimental design for gene-centric interactome analysis and describe the knowledgebase system for in-house data management.

Table 1 shows the gene-centric summary of Chr 18. The number of protein-coding genes equals 285 according to Ensembl (rel. 69, Oct 2012), whereas in the UniProt (rel. 2012_10 - Oct 31, 2012) and NextProt (rel.2012−10−07) databases this number is given as 277. The latter number of protein-coding genes was taken in our work for investigation of the Chr 18. Out of 277 genes, the expression of 240 transcripts is reported in RNaseqAtlas (Nov.2012). At the proteome level, the information on the identification of the products of 255 genes is currently available in PRIDE (v2.8.17). Limited quantitative data on the protein concentration in the biological material is available in emerging proteomic databases (MOPED,[21] MaxQB[22]). In view of that, in our work, the task of protein identification was accompanied by the task of quantitative analysis of Chr 18 proteins in blood plasma, liver tissue and HepG2 cell line.

Table 1. Gene-centric Summary of Chr 18 (Baseline Metrics)

Number of genes (Ensembl, rel.69, Oct.2012)	
Ensembl, protein-coding	285
Ensembl, noncoding	591
Ensembl, pseudogenes	227
UniProt, protein-coding (rel.2012_10)	277
NextProt, protein-coding (rel.2012−10−07)	277
Number of transcripts (RNaseq Atlas, Nov.2012)	240
Number of proteins confidently identified from MS data sets	
neXtProt(rel.2012−10−07, protein level, gold)	197
GPMdb (rel.2012−09−01) (green)	200
PeptideAtlas (human 2012_07 build, rel. Sep 2012, FDR of 1% at protein level)	217
PRIDE, v2.8.17	255
SRMAtlas, 2012_07	166
Antibody-based protein identification	
Human Protein Atlas (high and medium)	39

EXPERIMENTAL SECTION

Transcriptome analysis enables to evaluate the completeness of the proteome coverage[23] and to specify achievable targets for SRM-based research.[24] The RNA profiling has been started using HepG2 cells as the homogeneous and deeply mapped system compared with the liver tissue.

The RNA composition of the same sample transcripts has been independently analyzed by two methods. The RNaseq approach recently employed in a series of studies[23,25,26] was initially used for analysis of available transcripts and their abundance. More precisely, the expression level of transcripts has been determined by the quantitative Real-time PCR (RT-PCR)[27] and respective quantitative measures have been compared with transcriptome sequencing data.

RNaseq on HepG2 cells was performed by the 50-bp sequencing on the SOLiD4 platform (Supplementary Note 1, Supporting Information). Independent evaluation of gene expression was performed by RT-PCR of 45 genes (Supplementary Note 1). Genes were targeted to cover uniformly the range of RPKM values obtained by RNaseq. The measured quantities of transcripts were translated to copy number estimates based on the overall mRNA amount per HepG2 cell.[28]

A preliminary catalogue of proteins encoded by Chr 18 has been created by SRM scouting of biomaterial with subsequent abundance estimation by external calibration with highly purified (more than 95%) nonlabeled peptides (Supplementary Note 2, Supporting Information).

The results on the quantitative label-free analysis of proteins in biological samples were validated using the stable isotope-labeled peptides. For this purpose, synthesized Leu-$C^{13}N^{15}$ peptides were spiked into biological samples (human plasma, liver tissue and HepG2 cells) to the final concentrations corresponding to expected protein abundances, as previously estimated by the nonlabeled standards. Generation of the correlation curve for 16 light- and isotope-labeled peptides measured in plasma, liver and HepG2 cells illustrates applicability of such an approach (Figure 1, Supplementary Table 3, Supporting Information). The high correlation coefficient (R^2 = 0.96) between protein concentrations (double log) estimated by external calibration with light peptides and by isotope-labeled standards was observed. Therefore, we expect the external calibration with light peptides provides the suitable estimation for protein abundance.

The study of protein interactomics employed an experimental algorithm originally developed and tested for analysis of proteins, binding isatin (indoledione-2,3),[29,30] an endogenous regulator found in mammalian brain, peripheral tissues and

dx.doi.org/10.1021/pr300821n | J. Proteome Res. 2013, 12, 123−134

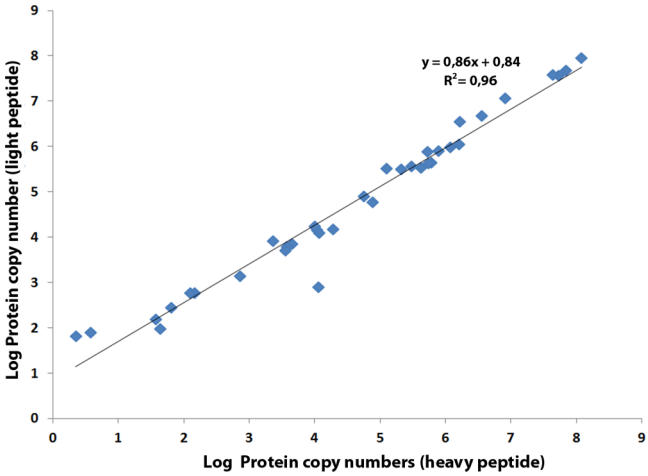


Figure 1. Correlation between protein concentrations estimated using external calibration with light peptides and spiked-in heavy (Leu-$C^{13}N^{15}$) isotope-labeled standards. Measurements were performed in depleted plasma, liver tissue and HepG2, respectively, for 16, 13, and 10 human proteins, respectively, encoded by Chr 18.

body fluids.[30−32] In these experiments we have used microsomal cytochrome b5, the protein encoded by Chr 18, as the model protein immobilized on: (a) the surface of molecular chip of an optical biosensor based on surface plasmon resonance (SPR)[33] and (b) cyanogen bromide activated Sepharose 4B (Supplementary Note 3, Supporting Information).

The chromosome-centric approach to data mining allowed characterizing each chromosome by the information about gene and protein properties. We used data available in Ensembl (rel.69, Oct.2012), NeXtProt (gold, rel. 2012−10−07), PubMed (07−2012), GPMdb (green, rel.2012−10−07), PRIDE (v.2.8.4), Peptide Atlas (rel.Sep.2012), ProteinAtlas (v.10.0) and GeneCards (v. 2.0) (Supplementary Table 1, Supporting Information). To characterize the involvement of the corresponding genes and their products into the scientific studies. We have queried the gene names (including synonyms) to the PubMed, and after filtering out overrepresented hits, we have recorded the number of genes cited in scientific articles. Information about heterogeneity of protein species was retrieved from NeXtProt and UniProt and expressed as number of single amino-acid polymorphisms (SAPs), alternative splicing (AS) and post-translational modifications (PTMs) per gene. For each chromosome, all of the values retrieved or data mining were normalized to the number of total (or affected, in case of protein variance) genes.

RESULTS AND DISCUSSION

Rational Focus on Chr 18

The distribution of the chromosomes among research teams has been completed (the list of the C-HPP Consortium Members, August 2012) in a more or less spontaneous way. Russia has selected Chr 18 in 2009. This choice was inspired by the intensive discussion occurring at the Moscow meeting,[2] while considering the criteria for ranking human chromosomes[1,34] according to their feasibility for the chromosome-centric HPP (C-HPP).

First, the chromosome investigated during C-HPP should contain a modest number of protein-coding genes due to economical considerations. The chromosome should be free (or carry just a few) of Ig-coding genes, as identification of Igs by proteomic methods is an intricate process due to the wide variability of this protein group.[35]

Second, the chromosome prioritization should rely on the medical relevance as a proportion of proteins associated with widespread diseases. In-depth examination of such proteins would be helpful for discovering new biomarkers and drug targets for early diagnosis and disease treatment. It should be also preferable, that proteins encoded by the chromosome have limited degree of variance introduced by amino-acid polymorphisms, splicing or post-translational modifications. Finally, in order to facilitate broader application of proteomic techniques, proteins encoded on the chromosome should be sufficiently explored beforehand. Addressing the relevant proteins can contribute to urgent fields of molecular biology, increasing usefulness of C-HPP deliverables.

All of the chromosomes are intrinsically the same in terms of their medical and scientific relevance and experimental coverage, with the exception of Chr Y, that falls out of the comparative analysis, probably due to minimal number of protein-coding genes.

In the ranking Chr 2, 4, 13, 18 occupied the top positions according medical relevance: ∼ 25% of their genes were reported in association with diseases. Chr 19 is much poorer from the clinical perspective, as only 14.2% of protein-coding genes are relevant to health distortions. Analysis of gene names' occurrences throughout PubMed has shown the high normalized values exceeding 75% of all proteins on each chromosome: this means, that most of the genes of all chromosomes are mentioned in scientific papers. For example, in Chr 6 the proportion of genes under investigation reaches 98%. Noteworthy, that same chromosome is characterized by a good coverage in MS experiments, having evidence for ∼82% of gene products (summarized data from GPMDb, PRIDE and PeptideAtlas).

The analysis of ProteinAtlas data showed approximately 15% coverage for Chr 17 (high and medium level of expression). Unexpectedly the low coverage (13%) was observed for Chr 21, despite this chromosome was targeted for piloting the antibody-based wing of C-HPP.[36]

The last section of the ranking spreadsheet (Supporting Information Table 1) refers to the variation of protein products expected due to single amino acid polymorphisms, splice isoforms and post-translational modifications. We assumed that a large degree of protein variability can possess difficulties in targeted SRM identification, as predicted proteotypic peptides will mismatch their polymorphic, isoform-specific or modified endogenous counterparts.

We found that on average, there are approximately 19.49 ± 3.42 (5.50 ± 1.26) SAPs, 2.96 ± 0.18 (1.95 ± 0.17) splice variants and 6.39 ± 1.0 (4.77 ± 0.82) PTMs per annotated protein according to NeXtProt, UniProt data are in parenthesis.

General analysis of chromosome ranking revealed Chr 13 and Chr 18 as the top hits.[34] Chr 18 consists of 76 Mbases and contains totally 492 annotated genes and 276 protein-coding genes.[37] Depending on the database this chromosome is assigned to 277−285 protein-coding genes (Supporting Information Table 1), and about 77% of them have been observed at the protein level (summarized data from GPMdb, PRIDE and PeptideAtlas).

The search in NeXtProt experimental data about the protein modifications (SAP, AS and PTM) revealed 132 protein-coding genes with 234 known alternatively spliced variants (on

average, 2.77 variants per on gene, considering one "canonical" sequence, see Table 2). Post-translational modifications lead to the appearance of 594 modified proteins per 1 protein-coding sequence, affecting 181 genes in total. The highest number of modified forms, 6120, appeared due to the emergence of single amino acid polymorphisms (SAPs), as the result of the presence of nonsynonymous nucleotide polymorphisms in the sequences of 265 protein-coding genes of Chr 18. Multiplying the average number of variations per gene, we have estimated the minimal total number of modified proteins (without combinatorial variants) as (1.77 + 1) × 5.94 × 23.09 × 277 = 126 740. Although this value reflects current situation (according to NeXtProt data, Table 2), the number of possible variants will definitely increase due to growth of the databases; therefore here we use the term "minimal number of proteins". Applying this calculation to all human genes, we could expect the minimal number of protein species ~9 500 000 or ~1 300 000 according to NeXtProt or UniProt data, respectively.

Table 2. Number of Protein Variants for 277 Genes on Chr 18 (neXtProt, rel.2012-10-07)

type	SAP	AS	PTM
Affected genes	265	132	181
Protein variants	6120	234	1076
Average per 1 gene	23.09	1.77	5.94
The minimal total number of modified proteins		126740	

Transcriptome Profiling

Analysis of Gene Expression in HepG2 Cells by RNaseq. In total, 63 million single-end cDNA reads were acquired for the whole genome exons. Reads were mapped onto the human reference genome (www.ncbi.nlm.nih.gov, v.36.1) corresponding to 16810 genes. Of that, 839 thousand reads specifically corresponding to the Chr 18 were mapped to 332 RefSeq transcripts and further resolved to 259 protein-coding genes (UniProt, 2012_07). The reads per kilobase of exon model per million mapped reads (RPKM) values representing estimations of transcript expression level were generated for the mapped genes of Chr 18 (Supporting Information Table 2).

The typical asymmetric bimodal distributions[23] were observed for the whole genome as well as for the chromosome-linked fraction of genes (Figure 2a). Approximately 20% of the transcripts were detected at the median level of 10−30 RPKM, while about 11% of transcripts were undetectable or had a low signal below 0.01 RPKM (Supporting Information Table 2). Excluding the transcripts expressed at low abundances (RPKM < 0.01), the number of target genes reduced to 205, comprising the 79% coverage of Chr 18. The resultant figure well corresponded to the total of 204 Chr 18 genes previously reported by RNaseq analysis[26] with an overlap of 183 transcripts.

We observed 10 transcripts for Chr 18 genes with RPKM ≥ 100 and 75 transcripts with RPKM ≤ 1 (less than one copy per cell[38]). The set of most abundant transcripts included products of the following Chr 18 genes: ATP5A1 (RPKM = 1393), TTR

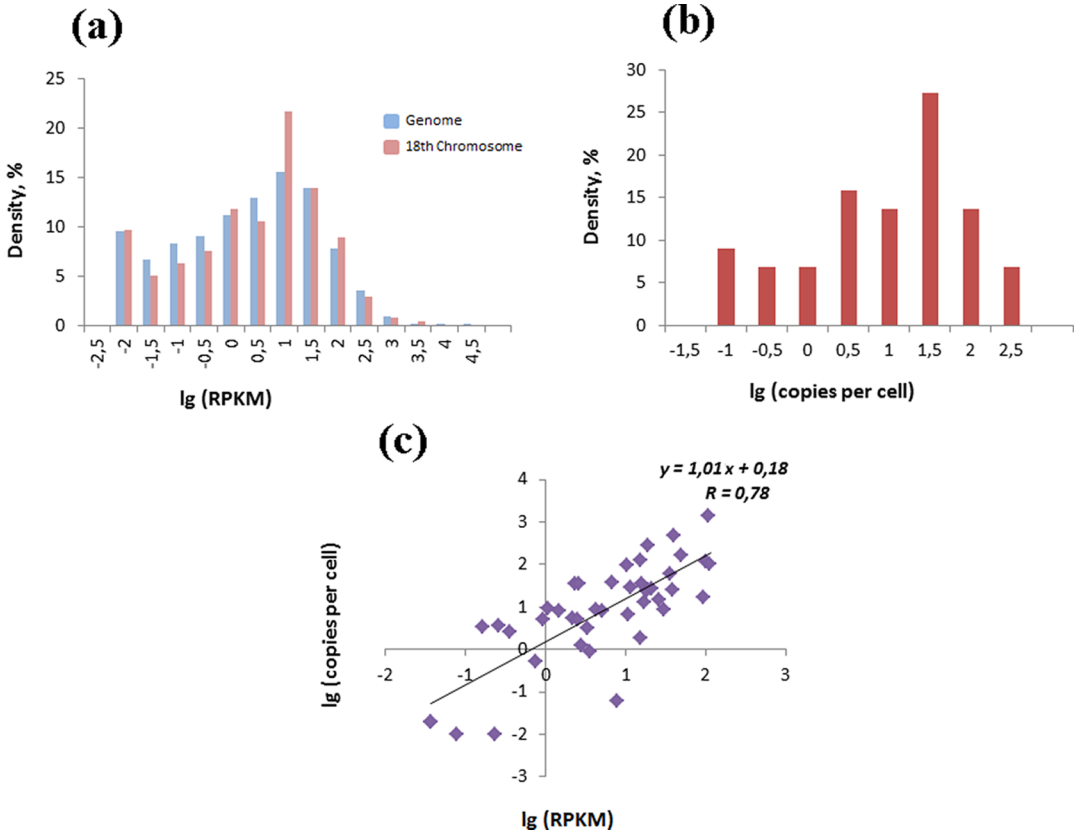


Figure 2. Transcriptome profiling of HepG2 cells by RNaseq and RT-PCR methods. (a) Distribution of lg RPKM values for all detected transcripts (blue) and a transcripts of the Chr 18 genes (red). (b) Distribution of lg transcript copies per single cell for 42 genes coded on Chr 18, detected by RT-PCR. (c) Correlation between RPKM and copies per cell obtained for 42 genes of Chr 18.

dx.doi.org/10.1021/pr300821n | J. Proteome Res. 2013, 12, 123−134

(RPKM = 495), MYL12A (RPKM = 381), NARS (RPKM = 277) and MYL12B (RPKM = 204). Low (but unambiguously detectable) abundances (0.1 < RPKM < 0.01) were measured for 35 protein-coding genes, including BCL2, AQP4, DSC3 and GALR1 (Supporting Information Table 2).

Quantitative PCR of Selected Genes in HepG2 Cells. In HepG2 cells the expression levels for 42 of 45 transcripts were measurable in our experiments (Supporting Information Table 2). It was not possible to detect any transcripts of three serpins (B2, B10 and B12) as no amplified PCR products of expected size were observed. We obtained low RPKM estimates for serpins B2 and B12, while serpin B10 was below the detection threshold of RNaseq. The low level of tissue expression of serpins was further confirmed by RNaseqAtlas data, in which serpin B2 was characterized by RPKM = 0.043 and two other serpins were reported as undetectable.

The distribution of transcript copy numbers, shown in Figure 2b, was generally more compressed than previously obtained distribution of RPKM values (Figure 2a). The dynamic range of RT-PCR data spanned 3,5 orders of magnitude from one transcript copy per 10 cells up to 10^3 copies per single HepG2 cell. The maximum peak of the distribution corresponded to 10−30 transcripts. The highest copy numbers were observed for proteasome assembly shaperone and ATP5A1 (114 and 110 copies per cell, respectively); whereas the least values of 4 copies per cell were detected for DSC1, OSBPL1A and RAB27B.

Comparison of RNaseq and RT-PCR Quantification Measures for HepG2 Cells. We have compared the different gene expression measurements using the data set of 42 genes. A significant Pearson correlation was observed (r = 0.78) as shown in Figure 2c, which is comparable to published results of RNaseq to RT-PCR comparisons.[39]

The RT-PCR quantitative data covered about 20% of the total number of transcripts measurable by RNaseq. The correlation between RNaseq and RT-PCR data enabled us to use the dependency in Figure 1c as the regression model for translation of the RPKM values to the transcript copy-numbers (Supporting Information Table 2). After translation, we identified the fraction of genes, characterized by the comparable level of expression estimated by both methods. For example, less than 25% difference in copy-number estimations by RNaseq and RT-PCR was obtained for ZNF236, SMAD2 and PSGM2 genes, present in 8, 18, and 115 copies per cell, respectively. On the other hand, some genes demonstrated significant differences in corresponding quantitative estimations. For example, the ATP5A1 abundance was found to be 110 copies per cell by RT-PCR, whereas estimation for the same gene by RNaseq was over 1 order of magnitude higher (2332 copies). That is the reason of such differences in respective quantitative measures is not clear yet.

Targeted Proteome Analysis

Detection and Quantitative Estimation of Target Proteins. The current stage of C-HPP is focused on constructing the chromosome-centric proteomic catalogues.[40] Representative proteins should be mapped in the tissues and quantitatively annotated. SRM is recommended for C-HPP as the method for targeted protein detection and quantification in complex samples due to its high sensitivity at the level of few protein copies per cell.[20] We applied SRM with the focus on a target set of proteotypic peptides of proteins encoded by genes of Chr 18. The currently adopted strategy of SRM assays generation implies the use of the libraries of crude synthetic peptides.[41] In our original strategy for C-HPP, the endogenously observed peptides were validated using the highly purified (more than 95%) synthetic (unlabeled) counterparts, which were used to generate calibration curves for quantitative estimation of protein abundances (see Experimental Section).

The targeted investigation of the Chr 18 proteome included scouting of the biomaterial, optimization of SRM-assay, calibration and quantitative estimation (Supplementary Note 2, Supporting Information). The preliminary lists of SRM transitions (fragmentation maps) were produced for 1277 proteotypic peptides of the target proteins by the manual inspection of fragmentation spectra in MS/MS repositories (PeptideAtlas, PRIDE and GPMDB). Transitions for the selected peptides were filtered according to the rules generally described by Sherwood et al.[42] (Supplementary Note 2, Supporting Information).

SRM measurements were performed using a triple-quadrupole MS (G6490 Agilent) in the nonscheduled mode to collect the transitions in triplicated SRM runs distinctly for each selected peptide. The transitions with sufficient signal-to-noise ratio (SNR) were assembled into protein-specific SRM assays, each consisting of 1−6 proteotypic peptides with corresponding chromatographic coordinates (time-scheduled SRM, Figure 3a). The presence of endogenous peptides was validated by time-scheduled SRM measurements of assays in plasma and liver.

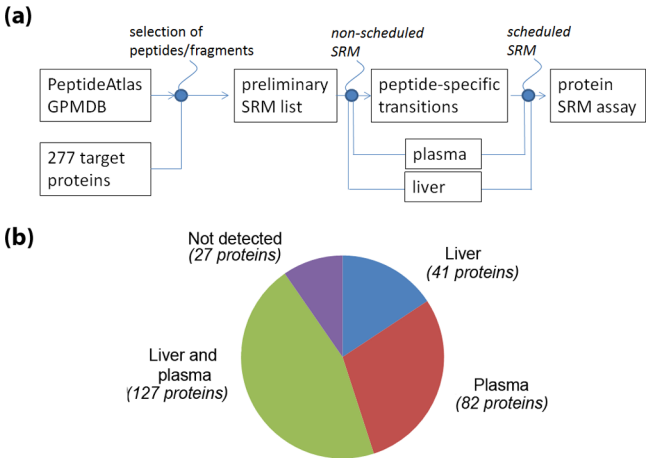


Figure 3. SRM-based scouting of depleted human plasma and liver tissue lysate. (a) Workflow applied to perform SRM measurements of 1277 proteotypic peptides of 277 proteins encoded by Chr 18. (b) Tissue distribution of 250 detected proteins.

The fragment ion maps for 277 proteins were assayed by SRM in depleted plasma and liver and evaluated as described in Supplementary Note 2 in the Supporting Information. Summarized SRM results for both types of biomaterial comprised in total 250 detected proteins (Supporting Information Table 3), which covered 90% of protein-coding genes of Chr 18. In the case of probed peptides for 27 proteins, we observed the absence of the SRM signal or poorly reproduced signal.

Analysis of tissue distribution (Figure 3b) demonstrated that a significant portion of detected proteins (51%) was observed both in human plasma and liver tissue. There were also proteins featured for plasma only (33%) or exclusively observed in the

dx.doi.org/10.1021/pr300821n | J. Proteome Res. 2013, 12, 123−134

Table 3. Copy numbers for the set of 29 proteins encoded by the human Chr 18 estimated by external calibration with non-labeled (light) synthetic peptides

AC Uniprot KB	protein name	protein copy numbers[a]		
		in 1 μL of human plasma, $\times 10^6$	per liver tissue cell, $\times 10^3$	per HepG2 cell, $\times 10^3$
P00167	Cytochrome b5	1.6	39	79
P25705	ATP synthase subunit alpha, mitochondrial	3.6	10	775
O00194	Ras-related protein Rab-27B	1.2	2.3	0.0003
O43396	Thioredoxin-like protein 1	1.5	0.2	n/a
O43776	Asparagine-tRNA ligase, cytoplasmic	0	1.2	19
Q10472	Polypeptide *N*-acetylgalactosaminyltransferase 1	68.7	4.5	19
Q96P63	Serpin B12	22.8	0.03	n/a
Q9Y4W6	AFG3-like protein 2	43.5	11	0.04
P19022	Cadherin-2	0.06	77.8	0.1
P19404	NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	53.7	535	3.7
Q08554	Desmocollin-1	0	0.9	n/a
P42765	3-ketoacyl-CoA thiolase, mitochondrial	0.3	1.1	11
Q9UIV8	Serpin B13	1.6	0.7	603
Q96FV9	THO complex subunit 1	0.4	0.01	n/a
Q96CS2	HAUS augmin-like complex subunit 1	119.0	3.6	0.1
Q6IA86	Elongator complex protein 2	291.0	3.5	0.0002
P38405	Guanine nucleotide-binding protein G(olf) subunit alpha	13.1	1.5	n/a
P04818	Thymidylate synthase (TS) (TSase) (EC 2.1.1.45)	8.2	12	547
O15118	Niemann-Pick C1 protein	0.2	0.004	4.5
Q32NC0	UPF0711 protein C18orf21 (HBV X-transactivated gene 13 protein) (HBV XAg-transactivated protein 13)	0.1	0.004	0.6
O14490	Disks large-associated protein 1 (DAP-1) (Guanylate kinase-associated protein) (hGKAP) (PSD-95/SAP90-binding protein 1) (SAP90/PSD-95-associated protein 1) (SAPAP1)	33.8	0.6	n/a
O14544	Suppressor of cytokine signaling 6 (SOCS-6) (Cytokine-inducible SH2 protein 4) (CIS-4) (Suppressor of cytokine signaling 4) (SOCS-4)	400.0	26	n/a
Q9H159	Cadherin-19	0.4	0.002	0.04
Q6PKX4	Docking protein 6 (Downstream of tyrosine kinase 6)	0.03	0	n/a
Q0P6D2	Protein FAM69C	277.6	0	0.002
A6NHP6	NPIP-like protein LOC729602	n/a	n/a	0.001
P22830	Ferrochelatase, mitochondrial	n/a	n/a	0.4
O95644	Nuclear factor of activated T-cells, cytoplasmic 1	147.8	n/a	n/a
P02686	Myelin basic protein	27.7	n/a	n/a

[a]n/a, measurements not performed; 0, SRM-response not detected.

liver (16%). Notably, we detected approximately two times less liver-specific proteins than that of plasma proteins. Such observation is consistent with the viewpoint that blood plays a role of the protein collector throughout all the tissues of the human body.[18]

The protein copy-numbers were estimated in the human plasma, liver tissue and HepG2 cells using the external calibration curves. As shown in Table 3, both in plasma and HepG2 cells the dynamic range of protein copies number spans more than 4 orders of magnitude, while in liver cells it covers 6 orders. In plasma the highest number 4×10^8 copies per 1 μL was observed for SOCS-4 protein, whereas the least copy number (3×10^4 per 1 μL) was estimated for Docking protein 6. In the liver tissue the most and least copy numbers were obtained for NADH dehydrogenase flavoprotein 2 (5.4×10^5 per cell) and cadherin-19 (2 per cell), respectively. In HepG2 cells the most abundant protein was ATP synthase subunit alpha and Elongator complex protein 2 was characterized by the lowest copy-number.

Undetected Proteins. As shown in Figure 3b, 27 proteins out of all master proteins coded by Chr 18 were undetected (Supporting Information Table 3) both in plasma and liver samples. Analysis of UniProt revealed that 18 of undetected products were reported at the protein level with information on tissue specificity available for 9 proteins (kidney, heart, skeletal muscle and fetal liver). Inspection of Gene Ontology[43] annotations of the undetected proteins revealed nuclear transcription factors (7 proteins), cell membrane receptors (7 proteins), signaling proteins, involved in cell division and apoptosis (7 proteins). Of 27 undetected proteins, the MS data on 17 proteins do exist in GPMDB (GPMdb, rel.2012−09−01, green) and also 17 proteins were observed on transcript level using RNaseq (Oct., 2012) in our experiments (Supporting Information Table 3).

The profile of functional annotations suggests that undetected proteins are present in biomaterial in ultralow abundances ($\leq 10^{-15}$ M). The main reason for the lack of positive SRM evidence may be attributed to the low abundance of these proteins or the absence of gene expression.

To enhance the detection limit of the SRM-based protein detection, we have included the pretreatment step to the sample preparation procedure to concentrate sample proteins by their irreversible binding to CN-Br activated Sepharose 4B. The initial step of the method consists in irreversible nonspecific binding of proteins present in a protein solution to CN-Br activated Sepharose 4B particles followed by blocking

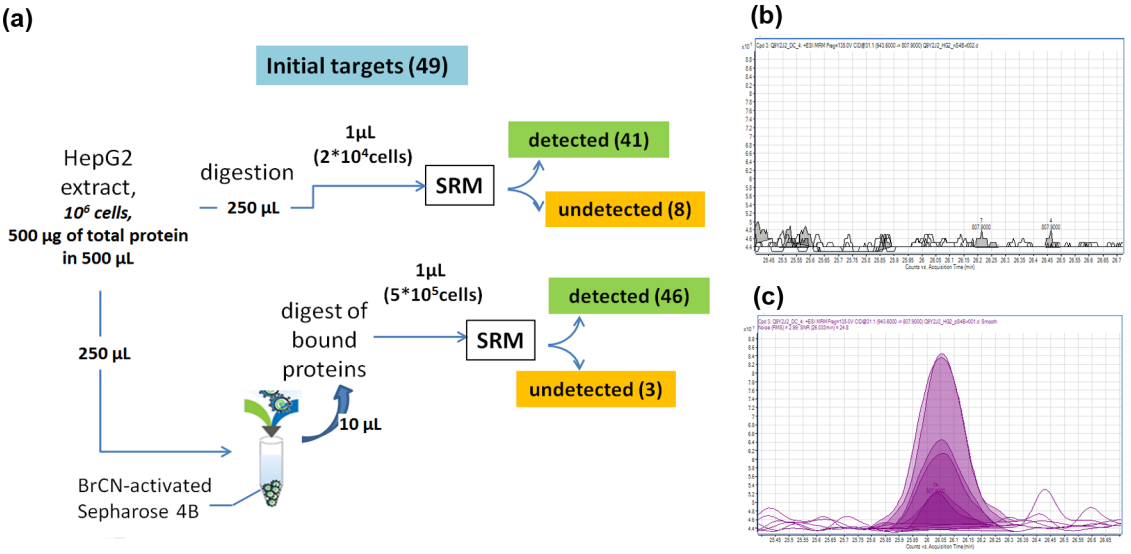


Figure 4. Approach to enhance the detection limit of SRM-based targeted proteome analysis of HepG2 cell line. (a) Experimental setup for the detection in extract of HepG2 cells of 49 target proteins, which were already successfully scouted in plasma and liver. Only 41 targets were initially detected in the HepG2 cell lysate, while additionally 5 proteins were picked up by SRM after increasing of sensitivity by factor 25 due to irreversible binding of proteins on CN-Br activated Sepharose 4B beads. Extracted ion chromatograms for IRPGEFEQFESTIGFK, m/z = 943.6 (2+) precursor ion (transition 943.6 → 807.9) (b) before and (c) after irreversible binding. See also Supporting Information Table 5.

of remaining reactive groups with a 10-fold excess of glycine. Sample protein concentration is achieved by sedimentation of Sepharose 4B and subsequent tryptic hydrolysis of covalently bound proteins. After hydrolysis these peptides were analyzed by SRM for the targeted detection of selected proteotypic peptides for the protein of interest. The detailed description of the method of irreversible binding in combination with the SRM technology is given by Kopylov et al.[20]

To evaluate the usefulness of sample pretreatment by irreversible binding, the preliminary study was performed as shown schematically in Figure 4. The extract of 10^6 HepG2 cells (500 μg of total protein) was split into two samples. The first sample (250 μg) was directly digested by the trypsin and the further 1 μg/1 μL (20 thousand cells) of peptide digest was used for detection of 49 target proteins in HepG2 cell by SRM of analysis of 78 peptides. These 49 proteins and their respective peptides were successfully scouted during the analysis of depleted plasma and liver cells, as described earlier.

Only 41 proteins were detected from the digested HepG2 lysate, whereas 8 proteins remained undetected. To reach these undetected proteins, we mixed the second portion of cell extract with CN-Br activated Sepharose 4B beads and concentrated proteins by factor 25 (from 250 to 10 μL). After incubation, the beads with covalently bound proteins were added with 1:100 (W/W) trypsin and 1 μL (500 thousand cells) of proteolytic peptides were analyzed by SRM (Supplement 4, Supporting Information). So, the resultant proteins mixture was concentrated on beads 25 times and thus we have revealed 5 additional proteins (O43310, O60346, P36952, Q9Y2J2, Q9Y4W6). Therefore, the increase for sensitivity by irreversible binding of proteins with activated beads followed by SRM detection enabled to retrieve additionally five proteins. It means that number of the detected proteins is a function of the sensitivity of the technology. In our opinion, other three proteins (Q9Y5X9, Q9Y5U9, Q9YBXX0) out of 49 investigated targets were undetected because of the insufficient sensitivity, especially taking into account, that the respective transcripts were observed by RNaseq (Supporting Information Table 2).

An Example of the Interactome Study

The study of human interactome provides new insights in our understanding of protein functions as integrated stable or temporally formed protein complexes, interactomes, and their roles in norm and various pathologies.[44] Direct experimental steps on realization of the Roadmap include development of an experimental scheme for identification and analysis of protein−protein interactions, which involved the proteins encoded on chromosome 18. Our studies of protein interactomics are based on the experimental algorithm originally developed and tested for analysis of isatin binding proteins (see Experimental Section).

LC−MS/MS analysis of the eluted fraction after fishing of human liver proteins performed using cytochrome b5 coupled to cyanogen bromide activated Sepharose 4B revealed 24 proteins, characterized by "gold" quality according to the NextProt (Supplementary Note 3, Supporting Information). None of these proteins was bound to control CN-Br activated Sepharose 4B, which was subjected to the same treatments as cytochrome b5-Sepharose 4B but without cytochrome b5.

Under these conditions, we failed to detect known protein partners of cytochrome b5. Reasons for this failure remain to be clarified. Although some of the identified proteins present in the Triton-X-100 extract of human liver homogenate can nonspecifically interact with cytochrome b5 due to their high abundances (e.g., hemoglobin), it should be noted that this experiment revealed some potentially interesting interactions.

Comparisons of STRING predictions with the list of proteins identified as partners of CYB5A by the molecular fishing (Supplementary Note 3, Supporting Information) revealed 5 coincidences with the medium confidence score ≤0.36. These included hemoglobins (HBG1, HBG2, HBB, HBA1) and ferritin light chain (FTL). The database search for the experimental interactions present in BioGRID, IntACT and

dx.doi.org/10.1021/pr300821n | J. Proteome Res. 2013, 12, 123−134

HPRD did not reveal any coincidences with our results obtained via molecular fishing.

Although functional importance of the reported interactions remains elusive, some of these interactions obviously reflect indirect interactions. The latter is especially true in the case of mitochondrial enzymes CPS1 and HMGCS2, which normally have different compartmentalization with the microsomal target protein, cytochrome b5. Other identified protein partners may potentially interact with microsomes and influence their functioning,[45,46] and thus, their interaction with cytochrome b5 cannot be ruled out.

Clinical Relevance of Chr 18

Medical diagnostics is mainly relied on detection of changes in the content of biomarkers, which correspond to the specific genome sequences. Although such markers are detected in blood of healthy individuals, development of certain diseases is accompanied by noticeable changes in their levels. Classical protein biomarkers and candidate biomarkers, as a rule, represent unmodified highly expressed protein products of corresponding genes (i.e., proteins without disease-associated modifications).[47] Since blood is a collector of proteins of various organs,[18] highly sensitive analytical methods may detect even several copies of any unmodified proteins in blood of healthy people.[5] Experimental approaches traditionally used for diagnostics of various diseases take into consideration protein concentration expressed either as molar concentration or as relative signal intensity (e.g., intensity of mass spectrometry peaks). Such signatures have been determined in early proteomic studies based on matrix assisted laser desorption ionization-time-of-flight (MALDI-TOF) profiling,[48,49] immunoanalysis,[50] and also in recent studies, where protein signatures have been obtained using SRM.[51,52]

Employment of a "digital" signature, in which aberrant protein biomarkers are probed either by the presence or absence of the analytical signal (by the 1/0 principle used, for example, in genomics and clinical relevance SNP analysis[53−55]), appears to be a next version of proteomic biomarker analysis.

Within the framework of the gene-centric approach to the human proteome the top priority should be given to highly sensitive (up to 10^{-18} M) and specific detection of those proteins in injured tissues and plasma, which are absent in healthy human body ("0") and typical exclusively for diseases ("1"). Alternatively, these would be a protein present in a healthy human body and absent at some pathology. The digital response capturing can be featured by protein modification events, including SAP, AS, PTM and also fused oncogenes due to chromosome aberrations.[56]

Classical examples of "digital" SAPs, which cause disease, include variants of hemoglobin. Thus, a single amino acid change from glutamate to valine in human hemoglobin beta position 7 leads to the threatening sickle-cell anemia.[57] Moreover, many other hereditary syndromes are known that caused by SAP. For example, changes from glutamate-693 in human amyloid beta A4 protein are of high clinical importance, which represent a molecular base for hereditary Alzheimer disease, type I.[58]

Applicability of digitalized diagnostics is supported by a series of studies demonstrating the possibility of proteomic identification and evaluation of association of modified proteins with diseases. For example, Su et al.[59] performed an SRM-based quantitative analysis of peptides with SAPs associated with development of diabetes mellitus and obesity. Omenn et al.[60] developed a strategy for discovery of new and well-known splice variants of proteins associated with the development of human pancreatic ductal adenocarcinoma and human Her2/neu-induced breast cancer in MS/MS data. Such an approach may be extended for targeted analysis of human plasma, other tissues and cell lines.[61]

Indicatively, the success of C-HPP depends on the new type of deliverables, which will be of use in the clinical diagnostics. To create the medical background for Chr 18, we have compiled information about 92 genes, reported in association with disease either in databases (GeneCards and OMIM (www.omim.org) or in the relevant literature. Each of the disease-associated genes has been queried to the UniProt to find out the number of alternatively spliced transcripts (ASTs), SAPs, and PTMs. Results of such survey were normalized to the number of genes and provided in Table 4 comparatively to the frequency of protein-affecting aberration over the whole Chr 18 (Supporting Information Table 1).

Table 4. Putative Variance for 92 Disease-associated Genes of Chr 18 (UniProt, 2012_11)

		disease-related genes	
aberration type	total number (per 92 genes)	number of affected genes	rate (per gene)[a]
SAPs	648	71 (77%)	7.04 ± 2.09
AS	133	44 (48%)	1.42 ± 0.21
PTMs	280	60 (65%)	3.04 ± 0.49
The minimal number of disease-related modified proteins		1061	

[a]Mean ± standard error of the mean.

Table 4 shows that more than half of Chr 18 genes can produce the products either affected by the amino acids variances or perturbed due to post-translational modifications. The comparison of the rate of aberrations reveals that the number of SAPs per disease-associated genes exceeds the background rate for Chr 18. The proportion of PTMs indicates an opposite case, so there are approximately 3 PTMs per clinically important gene. The number of protein forms per gene does not differ in disease-associated fraction of genes; however, UniProt annotations for alternative splicing (also of nsSNP) does not account for the vast amount of recently deposited RNaseq data. According to data about protein variances, we could expect the minimum number of disease-related protein forms for Chr 18 to be about 1000.

As the targeted SRM approach is limited by the analysis of selected ions, we have assessed the frequency of SAPs and PTMs occurrence within the specific proteotypic peptides. In 95% of cases, the putative aberration was located within the unique tryptic sequence of the protein; however, only 30% of these peptides in their unaffected form were observed by a shotgun MS/MS approach (according to GPMDB). It can be expected that the targeted SRM approach will uncover particular sequence variations and modifications that could be specific to the disease onset and progression.[59−61] If targeted SRM acquires the highest analytical sensitivity, the detection of SAPs, ASTs, and PTMs will become a basis for "digitalized" protein assays enabling the "yes/no" diagnostic decisions at the early stages of disease.

dx.doi.org/10.1021/pr300821n | J. Proteome Res. 2013, 12, 123−134

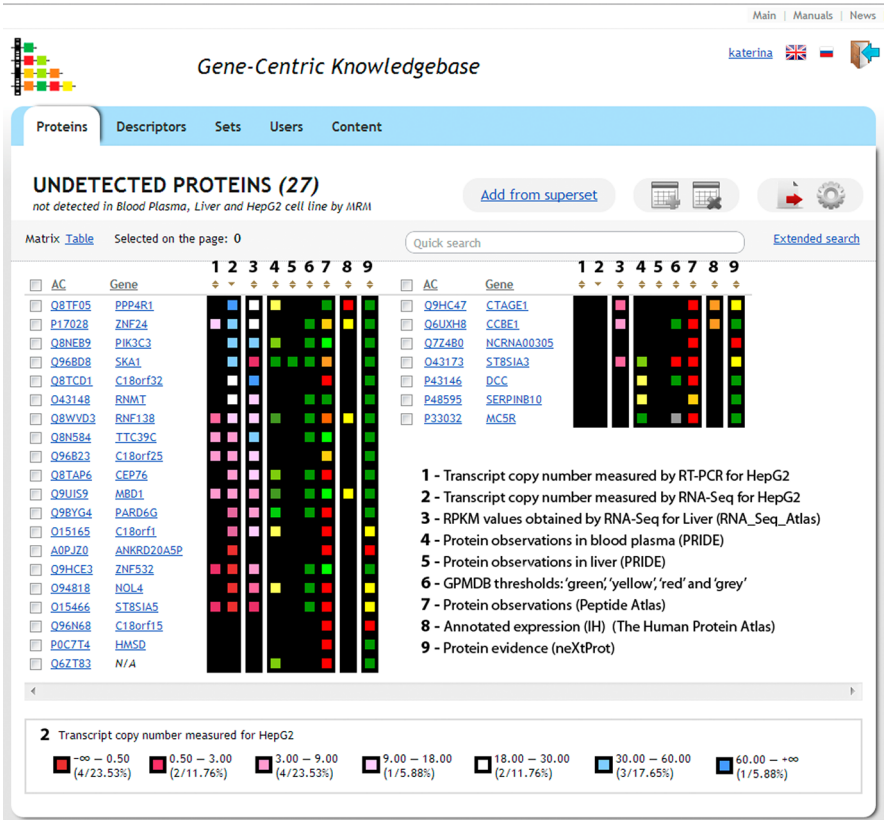


Figure 5. Screen shot of Gene-centric Knowledgebase for Chr 18. The subset for undetected proteins is loaded and decorated with 9 tracks on a heat map. Genes are sorted by transcript copy number, measured in HepG2 cells (2nd track).

CONCLUSION

To implement the chromosome-centric investigation, RNaseq, RT-PCR and SRM methods were applied to characterize the chromosome 18 transcriptome and proteome. The molecular composition of blood plasma, liver tissue samples and HepG2 cells has been investigated to analyze the feasibility of the experimental approaches for the C-HPP.

Using bioinformatics analysis, we observed that all of the chromosomes (except Chr Y) were quite comparable in terms of disease relevance, scientific impact and structural variance. In the case of chromosome 18, comprising 277 master-proteins, the amount of ~127 000 protein variants is expected due to the nonsynonymous polymorphism of coding regions, alternative splicing and post-translational modifications.

We have assayed 1277 peptide fragmentation maps to detect 699 parent ions, matching unique peptides of 250 proteins of Chr 18. In plasma the transition groups were observed for the peptides of 209 proteins, whereas using the liver tissue we have retrieved responses of 168 proteins, with over 50% overlap between results of plasma and liver survey. The focused set of 27 undetected proteins emerged as a negative result of the extensive SRM survey of plasma and liver (see Figure 3b). Assuming that insufficient sensitivity of measurements may comprise one of the main reasons for the failures in protein detection, we have shown that additional gain in number of detected proteins can be achieved by concentrating proteins with irreversible binding. Using the HepG2 cell line for a predefined set of 49 proteins (successfully scouted in plasma and liver cells) as an example, we have demonstrated that five additional protein species may be detected due to the sample pretreatment step (Figure 4).

To integrate all the data on the proteome complement of Chr 18, we have developed the gene-centric knowledgebase. Within the knowledgebase, information from 12 resources, including NextProt/UniProt, MS-repositories, ProteinAtlas, GeneCards and others, has been accumulated. The structure of the knowledgebase was organized into three sections: data mining of resources, original experimental data and predictions. The predictions were made using such Web-tools as STRING and GPMDB. Each entry of the knowledgebase corresponds to a specific NextProt record, while entries can be flexibly organized into the user-defined sets. Under guest login, three sets are currently accessible at http://www.kb18.ru, representing protein-coding genes detected in plasma, liver and HepG2 cells, respectively. The set is displayed as a heat map of color-coded features,[40] such as number of detected peptides, RNA expression level, associated clinical records, number of splice isoforms and parsed results of Web-queries to PRIDE, GPMDB, PeptideAtlas, ProteinAtlas and other repositories. The user can append/remove features' tracks from the heat map and also can create and manage the proprietary sets of target proteins. As an example of our knowledgebase, in Figure 5 we illustrate the list of 27 undetected proteins; for each protein, the color-coded values of 9 of the most popular descriptors are shown.

Conclusively, during the first years of C-HPP implementation, we addressed in practice the challenges in creating the chromosome-centric protein catalogue, embracing data mining, transriptome/proteome analysis, study of protein interactions and assessment of the clinical impact.

In our future work we plan to analyze the transcriptome of chromosome 18 in liver tissue and HepG2 cells using both

dx.doi.org/10.1021/pr300821n | J. Proteome Res. 2013, 12, 123−134

SOLiD and Illumina approaches and then to compare the obtained RNaseq data with the measurements by quantitative RT-PCR and DD (Droplet Digital)-PCR. The analysis of plasma, liver and HepG2 will be continued also at the proteome level to achieve the depth of coverage down to 10^{-18} M with an intention to create the transcripto-proteome of liver and HepG2. Next we expect to compete the prediction of protein−protein interactions and modifications (SAP, AS and PTMs) for the proteins encoded by Chr 18 and to proceed to the experimental validation of selected set of modified proteins based on our predictions.

■ ASSOCIATED CONTENT

Supporting Information

Supplemental notes and tables. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: alexander.archakov@ibmc.msk.ru; 2463731@gmail.com. Fax: +7 499 245-08-57.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was done in accordance with the Human Proteome Program of the Russian Academy of Medical Sciences. The work was funded in part by the Ministry of Education and Science of the Russian Federation (#16.522.12.2002 − targeted proteome analysis; #14.74011.0761 − quantitative analysis of proteins; #16.740.11.0372 − interactome study) and by the Russian Fund for Basic Research (grant 12-04-33109 − clinical relevance of Chr 18).

■ REFERENCES

(1) Archakov, A.; Aseev, A.; Bykov, V.; Grigoriev, A.; Govorun, V.; Ivanov, V.; Khlunov, A.; Lisitsa, A.; Mazurenko, S.; Makarov, A. A.; Ponomarenko, E.; Sagdeev, R.; Skryabin, K. Gene-centric view on the human proteome project: the example of the Russian roadmap for chromosome 18. *Proteomics* **2011**, *11*, 1853−6.

(2) Archakov, A.; Bergeron, J. J. M.; Khlunov, A.; Lisitsa, A.; Paik, Y.-K. The Moscow HUPO Human Proteome Project workshop. *Mol. Cell. Proteomics* **2009**, *8*, 2199−200.

(3) Ivanov, Y. D.; Govorun, V. M.; Bykov, V. A.; Archakov, A. I. Nanotechnologies in proteomics. *Proteomics* **2006**, *6*, 1399−414.

(4) Archakov, A. I.; Ivanov, Y. D.; Lisitsa, A. V.; Zgoda, V. G. AFM fishing nanotechnology is the way to reverse the Avogadro number in proteomics. *Proteomics* **2007**, *7*, 4−9.

(5) Archakov, A.; Ivanov, Y.; Lisitsa, A.; Zgoda, V. Biospecific irreversible fishing coupled with atomic force microscopy for detection of extremely low-abundant proteins. *Proteomics* **2009**, *9*, 1326−43.

(6) Omenn, G. S.; States, D. J.; Adamski, M.; Blackwell, T. W.; Menon, R.; Hermjakob, H.; Apweiler, R.; Haab, B. B.; Simpson, R. J.; Eddes, J. S.; Kapp, E. A.; Moritz, R. L.; Chan, D. W.; Rai, A. J.; Admon, A.; Aebersold, R.; Eng, J.; Hancock, W. S.; Hefta, S. A.; Meyer, H.; Paik, Y.-K.; Yoo, J.-S.; Ping, P.; Pounds, J.; Adkins, J.; Qian, X.; Wang, R.; Wasinger, V.; Wu, C. Y.; Zhao, X.; Zeng, R.; Archakov, A.; Tsugita, A.; Beer, I.; Pandey, A.; Pisano, M.; Andrews, P.; Tammen, H.; Speicher, D. W.; Hanash, S. M. Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. *Proteomics* **2005**, *5*, 3226−45.

(7) Omenn, G. S.; Baker, M. S.; Aebersold, R. Recent Workshops of the HUPO Human Plasma Proteome Project (HPPP): a bridge with the HUPO CardioVascular Initiative and the emergence of SRM targeted proteomics. *Proteomics* **2011**, *11*, 3439−43.

(8) States, D. J.; Omenn, G. S.; Blackwell, T. W.; Fermin, D.; Eng, J.; Speicher, D. W.; Hanash, S. M. Challenges in deriving high-confidence protein identifications from data gathered by a HUPO plasma proteome collaborative study. *Nature Biotechnol.* **2006**, *24*, 333−8.

(9) He, F. Human liver proteome project: plan, progress, and perspectives. *Mol. Cell. Proteomics* **2005**, *4*, 1841−8.

(10) Elortza, F.; Beretta, L.; Zhang, X.; He, F.; Mato, J. M. The 2011 Human Liver Proteome Project (HLPP) Workshop September 5, 2011, Geneva, Switzerland. *Proteomics* **2012**, *12*, 5−8.

(11) Zgoda, V. G.; Moshkovskii, S. A.; Ponomarenko, E. A.; Andreewski, T. V.; Kopylov, A. T.; Tikhonova, O. V.; Melnik, S. A.; Lisitsa, A. V.; Archakov, A. I. Proteomics of mouse liver microsomes: performance of different protein separation workflows for LC-MS/MS. *Proteomics* **2009**, *9*, 4102−5.

(12) Lisitsa, A. V.; Petushkova, N. A.; Thiele, H.; Moshkovskii, S. A.; Zgoda, V. G.; Karuzina, I. I.; Chernobrovkin, A. L.; Skipenko, O. G.; Archakov, A. I. Application of slicing of one-dimensional gels with subsequent slice-by-slice mass spectrometry for the proteomic profiling of human liver cytochromes P450. *J. Proteome Res.* **2010**, *9*, 95−103.

(13) Petushkova, N. A.; Kanaeva, I. P.; Lisitsa, A. V.; Sheremetyeva, G. F.; Zgoda, V. G.; Samenkova, N. F.; Karuzina, I. I.; Archakov, A. I. Characterization of human liver cytochromes P450 by combining the biochemical and proteomic approaches. *Toxicol. in Vitro* **2006**, *20*, 966−74.

(14) Zgoda, V.; Tikhonova, O.; Viglinskaya, A.; Serebriakova, M.; Lisitsa, A.; Archakov, A. Proteomic profiles of induced hepatotoxicity at the subcellular level. *Proteomics* **2006**, *6*, 4662−70.

(15) Mee, C. J.; Harris, H. J.; Farquhar, M. J.; Wilson, G.; Reynolds, G.; Davis, C.; van IJzendoorn, S. C. D.; Balfe, P.; McKeating, J. A. Polarization restricts hepatitis C virus entry into HepG2 hepatoma cells. *J. Virol.* **2009**, *83*, 6211−21.

(16) Slany, A.; Haudek, V. J.; Zwickl, H.; Gundacker, N. C.; Grusch, M.; Weiss, T. S.; Seir, K.; Rodgarkia-Dara, C.; Hellerbrand, C.; Gerner, C. Cell characterization by proteome profiling applied to primary hepatocytes and hepatocyte cell lines Hep-G2 and Hep-3B. *J. Proteome Res.* **2010**, *9*, 6−21.

(17) Geiger, T.; Wehner, A.; Schaab, C.; Cox, J.; Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics* **2012**, *11*, No. M111.014050.

(18) Anderson, N. L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R. S.; Conrads, T. P.; Veenstra, T. D.; Adkins, J. N.; Pounds, J. G.; Fagan, R.; Lobley, A. The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **2004**, *3*, 311−26.

(19) Archakov, A.; Zgoda, V.; Kopylov, A.; Naryzhny, S.; Chernobrovkin, A.; Ponomarenko, E.; Lisitsa, A. Chromosome-centric Approach to Overcoming Bottlenecks in the Human Proteome Project. *Expert Rev. Proteomics* **2012**.

(20) Kopylov, A. T.; Zgoda, V. G.; Lisitsa, A. V.; Archakov, A. I. Combining use of irreversible binding and multiple reactions monitoring technology for low- and ultralow copy-number proteins detection and quantitation. *Proteomics* **2012**, in press.

(21) Kolker, E.; Higdon, R.; Haynes, W.; Welch, D.; Broomall, W.; Lancet, D.; Stanberry, L.; et al. MOPED: Model Organism Protein Expression Database. *Nucleic Acids Res.* **2012**, *40*, 1093−9.

(22) Schaab, C.; Geiger, T.; Stoehr, G.; Cox, J.; Mann, M. Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol. Cell. Proteomics* **2012**, *11* (3), No. M111.014068.

(23) Nagaraj, N.; Wisniewski, J. R.; Geiger, T.; Cox, J.; Kircher, M.; Kelso, J.; Pääbo, S.; Mann, M. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **2011**, *7*, 548.

(24) Paik, Y.-K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Marko-Varga, G.; Aebersold, R.; Bairoch, A.; Yamamoto, T.; Legrain, P.; Lee, H.-J.;

Na, K.; Jeong, S.-K.; He, F.; Binz, P.-A.; Nishimura, T.; Keown, P.; Baker, M. S.; Yoo, J. S.; Garin, J.; Archakov, A.; Bergeron, J.; Salekdeh, G. H.; Hancock, W. S. Standard guidelines for the chromosome-centric human proteome project. *J. Proteome Res.* **2012**, *11*, 2005−13.
(25) Schwanhausser, B.; Busse, D.; Li, N.; Dittmar, G.; Schuchhardt, J.; Wolf, J.; Chen, W.; Selbach, M. Global quantification of mammalian gene expression control. *Nature* **2011**, *473*, 337−42.
(26) Lundberg, E.; Fagerberg, L.; Klevebring, D.; Matic, I.; Geiger, T.; Cox, J.; Algenas, C.; Lundeberg, J.; Mann, M.; Uhlen, M. Defining the transcriptome and proteome in three functionally different human cell lines. *Mol. Syst. Biol.* **2010**, *6*, 450.
(27) Nolan, T.; Hands, R. E.; Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* **2006**, *1*, 1559−82.
(28) Wilkening, S.; Stahl, F.; Bader, A. Comparison of primary human hepatocytes and hepatoma cell line Hepg2 with regard to their biotransformation properties. *Drug Metab. Dispos.* **2003**, *31*, 1035−42.
(29) Buneeva, O.; Gnedenko, O.; Zgoda, V.; Kopylov, A.; Glover, V.; Ivanov, A.; Medvedev, A.; Archakov, A. Isatin-binding proteins of rat and mouse brain: proteomic identification and optical biosensor validation. *Proteomics* **2010**, *10*, 23−37.
(30) Medvedev, A.; Kopylov, A.; Buneeva, O.; Zgoda, V.; Archakov, A. Affinity-based proteomic profiling: problems and achievements. *Proteomics* **2012**, *12*, 621−37.
(31) Medvedev, A.; Igosheva, N.; Crumeyrolle-Arias, M.; Glover, V. Isatin: role in stress and anxiety. *Stress (Amsterdam, Netherlands)* **2005**, *8*, 175−83.
(32) Medvedev, A.; Buneeva, O.; Glover, V. Biological targets for isatin and its analogues: Implications for therapy. *Biol.: Targets Ther.* **2007**, *1*, 151−62.
(33) Ivanov, A. S.; Gnedenko, O. V.; Molnar, A. A.; Mezentsev, Y. V.; Lisitsa, A. V.; Archakov, A. I. Protein-protein interactions as new targets for drug design: virtual and experimental approaches. *J. Bioinform. Comput. Biol.* **2007**, *5*, 579−92.
(34) Ponomarenko, E. A.; Poverennaya, E. V.; Pyatnitskiy, M. A.; Lisitsa, A. V.; Moshkovskii, S. A.; Ilgisonis, E. V.; Chernobrovkin, A. L.; Archakov, A. I. Comparative ranking of human chromosomes based on post-genomic data. *OMICS JIB* **2012**, *16* (11), 604−11.
(35) Lane, J.; Duroux, P.; Lefranc, M.-P. From IMGT-ONTOLOGY to IMGT/LIGMotif: the IMGT standardized approach for immunoglobulin and T cell receptor gene identification and description in large genomic sequences. *BMC Bioinform.* **2010**, *11*, 223.
(36) Uhlén, M.; Oksvold, P.; Älgenäs, C.; Hamsten, C.; Fagerberg, L.; Klevebring, D.; Lundberg, E.; Odeberg, J.; Pontén, F.; Kondo, T.; Sivertsson, Å. Antibody-based protein profiling of the human chromosome 21. *Mol. Cell. Proteomics* **2012**, *11*, No. M111.013458.
(37) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860−921.
(38) Hebenstreit, D.; Fang, M.; Gu, M.; Charoensawan, V.; van Oudenaarden, A.; Teichmann, S. A. RNA sequencing reveals two major classes of gene expression levels in metazoan cells. *Mol. Syst. Biol.* **2011**, *7*, 497.
(39) Chen, H.; Liu, Z.; Gong, S.; Wu, X.; Taylor, W. L.; Williams, R. W.; Matta, S. G.; Sharp, B. M. Genome-wide gene expression profiling of nucleus accumbens neurons projecting to ventral pallidum using both microarray and transcriptome sequencing. *Front. Neurosci.* **2011**, *5*, 98.
(40) Paik, Y.-K.; Jeong, S.-K.; Omenn, G. S.; Uhlen, M.; Hanash, S.; Cho, S. Y.; Lee, H.-J.; Na, K.; Choi, E.-Y.; Yan, F.; Zhang, F.; Zhang, Y.; Snyder, M.; Cheng, Y.; Chen, R.; Marko-Varga, G.; Deutsch, E. W.; Kim, H.; Kwon, J.-Y.; Aebersold, R.; Bairoch, A.; Taylor, A. D.; Kim, K. Y.; Lee, E.-Y.; Hochstrasser, D.; Legrain, P.; Hancock, W. S. The Chromosome-Centric Human Proteome Project for cataloging proteins encoded in the genome. *Nat. Biotechnol.* **2012**, *30*, 221−3.
(41) Picotti, P.; Rinner, O.; Stallmach, R.; Dautel, F.; Farrah, T.; Domon, B.; Wenschuh, H.; Aebersold, R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nat. Methods* **2010**, *7*, 43−6.
(42) Sherwood, C. A.; Eastham, A.; Lee, L. W.; Peterson, A.; Eng, J. K.; Shteynberg, D.; Mendoza, L.; Deutsch, E. W.; Risler, J.; Tasman, N.; Aebersold, R.; Lam, H.; Martin, D. B. MaRiMba: a software application for spectral library-based MRM transition list assembly. *J. Proteome Res.* **2009**, *8*, 4396−405.
(43) Ashburner, M.; Ball, C. A.; Blake, J. A.; Botstein, D.; Butler, H.; Cherry, J. M.; Davis, A. P.; Dolinski, K.; Dwight, S. S.; Eppig, J. T.; Harris, M. A.; Hill, D. P.; Issel-Tarver, L.; Kasarskis, A.; Lewis, S.; Matese, J. C.; Richardson, J. E.; Ringwald, M.; Rubin, G. M.; Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25−9.
(44) Ivanov, A. S.; Zgoda, V. G.; Archakov, A. I. Technologies of Protein Interactomics: A Review. *Russian Journal of Bioorganic Chemistry* **2011**, *37*, 4−16.
(45) Redondo, P. C.; Harper, M. T.; Rosado, J. A.; Sage, S. O. A role for cofilin in the activation of store-operated calcium entry by de novo conformational coupling in human platelets. *Blood* **2006**, *107* (3), 973−9.
(46) Hevi, S.; Chuck, S. L. Ferritins can regulate the secretion of apolipoprotein b. *J. Biol. Chem.* **2003**, *278*, 31924−9.
(47) Polanski, M.; Anderson, N. L. A list of candidate cancer biomarkers for targeted proteomics. *Biomarker Insights* **2007**, *1*, 1−48.
(48) Zhang, Z.; Bast, R. C.; Yu, Y.; Li, J.; Sokoll, L. J.; Rai, A. J.; Rosenzweig, J. M.; Cameron, B.; Wang, Y. Y.; Meng, X.-Y.; Berchuck, A.; Van Haaften-Day, C.; Hacker, N. F.; de Bruijn, H. W. A.; van der Zee, A. G. J.; Jacobs, I. J.; Fung, E. T.; Chan, D. W. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **2004**, *64*, 5882−90.
(49) Moshkovskii, S. A.; Vlasova, M. A.; Pyatnitskiy, M. A.; Tikhonova, O. V.; Safarova, M. R.; Makarov, O. V.; Archakov, A. I. Acute phase serum amyloid A in ovarian cancer as an important component of proteome diagnostic profiling. *Proteomics−Clin. Appl.* **2007**, *1*, 107−17.
(50) Yurkovetsky, Z.; Skates, S.; Lomakin, A.; Nolen, B.; Pulsipher, T.; Modugno, F.; Marks, J.; Godwin, A.; Gorelik, E.; Jacobs, I.; Menon, U.; Lu, K.; Badgwell, D.; Bast, R. C.; Lokshin, A. E. Development of a multimarker assay for early detection of ovarian cancer. *J. Clin. Oncol.* **2010**, *28*, 2159−66.
(51) Cima, I.; Schiess, R.; Wild, P.; Kaelin, M.; Schüffler, P.; Lange, V.; Picotti, P.; Ossola, R.; Templeton, A.; Schubert, O.; Fuchs, T.; Leippold, T.; Wyler, S.; Zehetner, J.; Jochum, W.; Buhmann, J.; Cerny, T.; Moch, H.; Gillessen, S.; Aebersold, R.; Krek, W. Cancer genetics-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancer. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3342−7.
(52) Huttenhain, R.; Soste, M.; Selevsek, N.; Rost, H.; Sethi, A.; Carapito, C.; Farrah, T.; Deutsch, E. W.; Kusebauch, U.; Moritz, R. L.; Nimeus-Malmstrom, E.; Rinner, O.; Aebersold, R. Reproducible Quantification of Cancer-Associated Proteins in Body Fluids Using Targeted Proteomics. *Sci. Transl. Med.* **2012**, *4*, 142ra94.
(53) Auffray, C.; Hood, L. Editorial: Systems biology and personalized medicine - the future is now. *Biotechnol. J.* **2012**, *7*, 938−9.
(54) Sachidanandam, R.; Weissman, D.; Schmidt, S. C.; Kakol, J. M.; Stein, L. D.; Marth, G.; Sherry, S.; Mullikin, J. C.; Mortimore, B. J.; Willey, D. L.; Hunt, S. E.; Cole, C. G.; Coggill, P. C.; Rice, C. M.; Ning, Z.; Rogers, J.; Bentley, D. R.; Kwok, P. Y.; Mardis, E. R.; Yeh, R. T.; Schultz, B.; Cook, L.; Davenport, R.; Dante, M.; Fulton, L.; Hillier, L.; Waterston, R. H.; McPherson, J. D.; Gilman, B.; Schaffner, S.; Van Etten, W. J.; Reich, D.; Higgins, J.; Daly, M. J.; Blumenstiel, B.; Baldwin, J.; Stange-Thomann, N.; Zody, M. C.; Linton, L.; Lander, E. S.; Altshuler, D. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **2001**, *409*, 928−33.
(55) Rajaraman, P.; Melin, B. S.; Wang, Z.; McKean-Cowdin, R.; Michaud, D. S.; Wang, S. S.; Bondy, M.; Houlston, R.; Jenkins, R. B.; Wrensch, M.; Yeager, M.; Ahlbom, A.; Albanes, D.; Andersson, U.; Freeman, L. E. B.; Buring, J. E.; Butler, M. A.; Braganza, M.; Carreon, T.; Feychting, M.; Fleming, S. J.; Gapstur, S. M.; Gaziano, J. M.; Giles,

G. G.; Hallmans, G.; Henriksson, R.; Hoffman-Bolton, J.; Inskip, P. D.; Johansen, C.; Kitahara, C. M.; Lathrop, M.; Liu, C.; Le Marchand, L.; Linet, M. S.; Lonn, S.; Peters, U.; Purdue, M. P.; Rothman, N.; Ruder, A. M.; Sanson, M.; Sesso, H. D.; Severi, G.; Shu, X.-O.; Simon, M.; Stampfer, M.; Stevens, V. L.; Visvanathan, K.; White, E.; Wolk, A.; Zeleniuch-Jacquotte, A.; Zheng, W.; Decker, P.; Enciso-Mora, V.; Fridley, B.; Gao, Y.-T.; Kosel, M.; Lachance, D. H.; Lau, C.; Rice, T.; Swerdlow, A.; Wiemels, J. L.; Wiencke, J. K.; Shete, S.; Xiang, Y.-B.; Xiao, Y.; Hoover, R. N.; Fraumeni, J. F.; Chatterjee, N.; Hartge, P.; Chanock, S. J. Genome-wide association study of glioma and meta-analysis. *Hum. Genet.* **2012**, *131* (12), 1877−88.

(56) Tomlins, S. A.; Rhodes, D. R.; Perner, S.; Dhanasekaran, S. M.; Mehra, R.; Sun, X. W.; Varambally, S.; Cao, X.; Tchinda, J.; Kuefer, R.; Lee, C.; Montie, J. E.; Shah, R. B.; Pienta, K. J.; Rubin, M. A.; Chinnaiyan, A. M. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **2005**, *310*, 644−8.

(57) Wishner, B. C.; Ward, K. B.; Lattman, E. E.; Love, W. E. Crystal structure of sickle-cell deoxyhemoglobin at 5 A resolution. *J. Mol. Biol.* **1975**, *98*, 179−94.

(58) Hardy, J. Framing beta-amyloid. *Nat. Genet.* **1992**, *4*, 233−4.

(59) Su, Z.-D.; Sun, L.; Yu, D.-X.; Li, R.-X.; Li, H.-X.; Yu, Z.-J.; Sheng, Q.-H.; Lin, X.; Zeng, R.; Wu, J.-R. Quantitative detection of single amino acid polymorphisms by targeted proteomics. *J. Mol. Cell Biol.* **2011**, *3*, 309−15.

(60) Omenn, G. S.; Yocum, A. K.; Menon, R. Alternative splice variants, a new class of protein cancer biomarker candidates: findings in pancreatic cancer and breast cancer with systems biology implications. *Dis. Markers* **2010**, *28*, 241−51.

(61) Blakeley, P.; Siepen, J. A.; Lawless, C.; Hubbard, S. J. Investigating protein isoforms via proteomics: a feasibility study. *Proteomics* **2010**, *10*, 1127−40.