Journal of proteome research

Article

pubs.acs.org/jpr

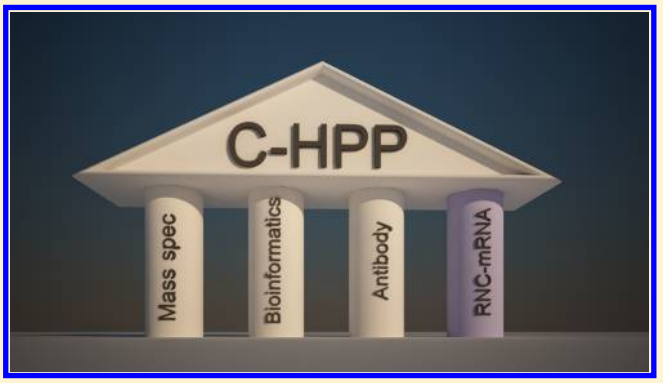
Resolving Chromosome-Centric Human Proteome with Translating mRNA Analysis: A Strategic Demonstration

Jiayong Zhong,† Yizhi Cui,† Jiahui Guo,† Zhipeng Chen, Lijuan Yang, Qing-Yu He,* Gong Zhang,* and Tong Wang*

Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes, Institute of Life and Health Engineering, College of Life Science and Technology, Jinan University, 601 Huangpu Avenue West, Guangzhou 510632, China

Supporting Information

ABSTRACT: Chromosome-centric human proteome project (C-HPP) aims at differentiating chromosome-based and tissue-specific protein compositions in terms of protein expression, quantification, and modification. We previously found that the analysis of translating mRNA (mRNA attached to ribosome-nascent chain complex, RNC-mRNA) can explain over 94% of mRNA-protein abundance. Therefore, we propose here to use full-length RNC-mRNA information to illustrate protein expression both qualitatively and quantitatively. We performed RNA-seq on RNC-mRNA (RNC-seq) and detected 12 758 and 14 113 translating genes in human normal bronchial epithelial (HBE) cells and human colorectal adenocarcinoma Caco-2 cells, respectively. We found that most of these genes were mapped with >80% of coding sequence coverage. In Caco-2 cells, we provided translating evidence on 4180 significant single-nucleotide variations. While using RNC-mRNA data as a standard for proteomic data integration, both translating and protein evidence of 7876 genes can be acquired from four interlaboratory data sets with different MS platforms. In addition, we detected 1397 noncoding mRNAs that were attached to ribosomes, suggesting a potential source of new protein explorations. By comparing the two cell lines, a total of 677 differentially translated genes were found to be nonevenly distributed across chromosomes. In addition, 2105 genes in Caco-2 and 750 genes in HBE cells are expressed in a cell-specific manner. These genes are significantly and specifically clustered on multiple chromosomes, such as chromosome 19. We conclude that HPP/C-HPP investigations can be considerably improved by integrating RNC-mRNA analysis with MS, bioinformatics, and antibody-based verifications.



KEYWORDS: *C-HPP, strategy, RNC-seq, translating mRNA, RNC-mRNA, chromosome*

INTRODUCTION

In view of systems biology, a complete understanding of the flow of genetic information in different human cell types is fundamentally dependent on an all-inclusive knowledge-base at the gene, transcription, translation, and protein levels, both qualitatively and quantitatively. Human genome project (HGP)[1] and intensive transcriptome investigations (reviewed in ref 2) addressed the first part of this question by characterizing approximately 20 300 human protein coding genes that can be transcribed into mRNAs. In parallel, chromosome-centric human proteome project (C-HPP) is to find the protein evidence for all of these coding genes and to characterize their tissue/cell/subcellular distributions (reviewed in refs 3−5). The ultimate impact of C-HPP will be emphasized by cooperating with the biology- and disease-driven projects (B/D-HPP) to provide a normal proteome baseline of different human body compartments.[6] With this reference, disease-oriented investigations of both the systems biology and reductionism will be fundamentally improved.[7]

As the first resource pillar of C-HPP, the state-of-the-art mass spectrometry (MS) allows the identification of ∼12 000 proteins in human cells at steady-state and the high-throughput peptide verification by selected/multiple reaction monitoring (SRM/MRM).[8,9] In contrast, transcripts of more than 14 000 genes are typically detectable in transcriptome of various human cell lines.[8,10−15] These investigations significantly contributed to the correlation and annotation landscape of transcriptome versus proteome. Hence, transcriptome information has an indicative value for protein expression, especially with regard to discover the "missing proteins" that lack MS evidence or Ab detection[16−19](reviewed in ref 3), the primary goal of C-HPP. For example, Liu et al.[20] found that 59 out of 1169 chromosome 17 genes products are "missing proteins" based on exploration of knowledgebases proposed by the C-HPP guidelines.[4] The second pillar of C-HPP, antibody-based validation[21] (reviewed in refs 3 and 4), may assist to unravel this discrepancy between transcriptome and proteome in terms of identification numbers.

Special Issue: Chromosome-centric Human Proteome Project

Received: July 17, 2013
Published: November 7, 2013

Regarding this, it has been recognized by Human Proteome Organization (HUPO) that when using Human Protein Atlas for the exploration of antibody evidence it is necessary to distinguish "high or medium immunostaining" from the lower staining that is less reliable. It should be noted that a proportion of mRNAs in the transcriptome is not translated to proteins.[22−25] We previously reported that this proportion is ∼5% in multiple cell lines.[15] This implicates that using transcriptome as a reference to lead peptide evidence searching and antibody verification has a risk of futilely focusing on nontranslated mRNAs.

As the third resource pillar of C-HPP, bioinformatics has significant contributions to its scientific questions, including providing clues from knowledge bases to match evidence of proteins and transcripts, but proteomic data integration is still a challenge. The primary reason is that the inconsistency of data acquisition in different laboratories makes it difficult to directly compare protein identifications and quantifications among various samples and via probing numerous databases (reviewed in ref 3). These interlaboratory variations include sample normalization, peptide modifications, contaminations, MS platforms, experiment settings, and algorithms.[26] As an example, in a simple comparison, only 7 out of 27 laboratories successfully reported all proteins in a highly purified and equimolar 20-protein mixture.[27] Thus, a uniform reference basis of the identifications and quantifications is necessary.

In addition, C-HPP involves in complex aspects of protein, including isoforms, derived from alternative splicing transcripts (ASTs) and single-nucleotide variations (SNVs) as well as post-translational modifications (PMTs) (reviewed in refs 3−5). Especially with regard to ASTs and SNVs, the low sequence coverage of identified peptides to a protein is another limitation and obstacle of current MS. Working with bioinformatics resources has been shown to partially overcome this drawback by increasing the detection rates of AST products[28−30] (reviewed in ref 31); however, the low peptide sequence coverage is still a bottleneck for its wide application.

We here propose the analysis of translating mRNA (the mRNA attached to ribosome-nascent chain complex, RNC-mRNA) by the next-generation sequencing (NGS) can be another resource pillar of HPP/C-HPP. The primary rationale of this strategic improvement is that translating mRNA analysis excludes the nontranslated mRNAs from the beginning, bypassing the low peptide sequence coverage through providing translational evidence with high sequence coverage. Moreover, as shown in our previous report, a quantitative rule of the tight correlation of translating mRNA-protein relative abundance has been discerned.[15] This makes it possible to predict the protein abundance by the RNC-mRNA abundance on a genome-wide scale, serving as a theoretical basis of our rationale for a novel addition to current C-HPP strategy.

In this study, we developed our proposed strategy by analyzing normal human bronchial epithelial (HBE) cells and differentiated human colorectal adenocarcinoma Caco-2 cells. We demonstrated a solution to resolve chromosome-centric human proteome by full-length translating mRNA sequencing analysis (RNC-seq) that is useful to overcome the obstacles of C-HPP mentioned above.

MATERIALS AND METHODS

Cell Culture

Human Caco-2 cells were obtained from American Type Culture Collections (ATCC, Rockville, MD). Cells were cultured in complete DMEM media (Invitrogen, Carlsbad, CA), supplemented with 10% fetal bovine serum (FBS) (PAA Australia, Weike Biochemical Reagent, Shanghai, China), 1% penicillin/streptomycin, and 10 μg/mL ciprofloxacin. Cells were harvested when the culture reached 100% confluence.

Sequencing of mRNA and RNC-mRNA

Total RNA and RNC-RNA of Caco-2 cells were extracted as previously described.[15] In brief, the polyA+mRNA and RNC-mRNA were selected by RNA Purification Beads (Illumina, San Diego, CA). Equal amounts of total mRNA and RNC-mRNA isolated from three independent cultures were, respectively, pooled for subsequent library construction and RNA-seq analysis. The library was constructed by using the Illumina TruSeq RNA sample Prep Kit v2 and sequenced by the Illumina HiSeq-2000 for 50 cycles. High-quality reads that passed the Illumina quality filters were kept for the sequence analysis. All sequencing data sets of Caco-2 cells are available at Gene Expression Omnibus database (accession number GSE48603). The RNA-seq data for mRNA and RNC-mRNA of HBE cells were reported in our previous study,[15] available at Gene Expression Omnibus database (accession number GSE42006).

Methodologically, as a common method to extract RNC-mRNA from cells under physiological conditions, sucrose cushion removes the nonribosome bound proteins and mRNAs in one step.[15] This type of method results in high specificity for RNC purification in both bacteria and higher mammalian cells.[32−37] As a mild treatment, the high-quality full-length translating mRNA can be obtained and sequenced for subsequent analysis.

Sequence Analysis

The sequencing reads were mapped to the RefSeq-RNA reference sequence (downloaded from http://hgdownload.cse.ucsc.edu/downloads, accessed on Jan 21, 2013) using FANSe 2 algorithm[38] (http://bioinformatics.jnu.edu.cn/software/fanse2/) with the parameters −L55 −E5 −U0 −S10 (for Caco-2 data sets) and −L80 −E7 −U0 −S10 (for HBE data sets), respectively. The mismapping rate of this algorithm has been shown to be as low as 1×10^{-5}.[38] Alternative splice variants were merged.[15] The genes with at least 10 mapped reads were considered as reliably detected and quantified genes.[39] The translating genes were quantified using rpkM method.[40] The RNC-mRNA reads were imported into edgeR software package to calculate the up-/down-regulation of translating mRNAs comparing HBE and Caco-2 cells.[41]

RNC-mRNA sequence coverage was calculated by dividing the nucleotides covered by sequencing reads and the full mRNA length or coding sequence (CDS). Peptide sequence coverage was calculated by dividing the amino acids covered by MS-detected peptides and the full protein length.

Splice variants detection was performed as previously described.[15,42] SNVs were calculated by Fisher's exact test for the predominant nucleotide (>50% occurrence) at a single position that differs from the RefSeq sequence. Translation ratio (TR) that is a parameter reflecting translation initiation was calculated based on our published method.[15]

Statistics

Data were analyzed for statistical significance by the Fisher's exact test with MATLAB R2012a software package (MathWorks, Natick, MA).

RESULTS AND DISCUSSION

Full-Length Translating mRNA Analysis Achieved High Sequence Coverage

With current MS technology, it is very difficult to reach high sequence coverage for a certain proteome, which is required for

dx.doi.org/10.1021/pr4007409 | J. Proteome Res. 2014, 13, 50−59

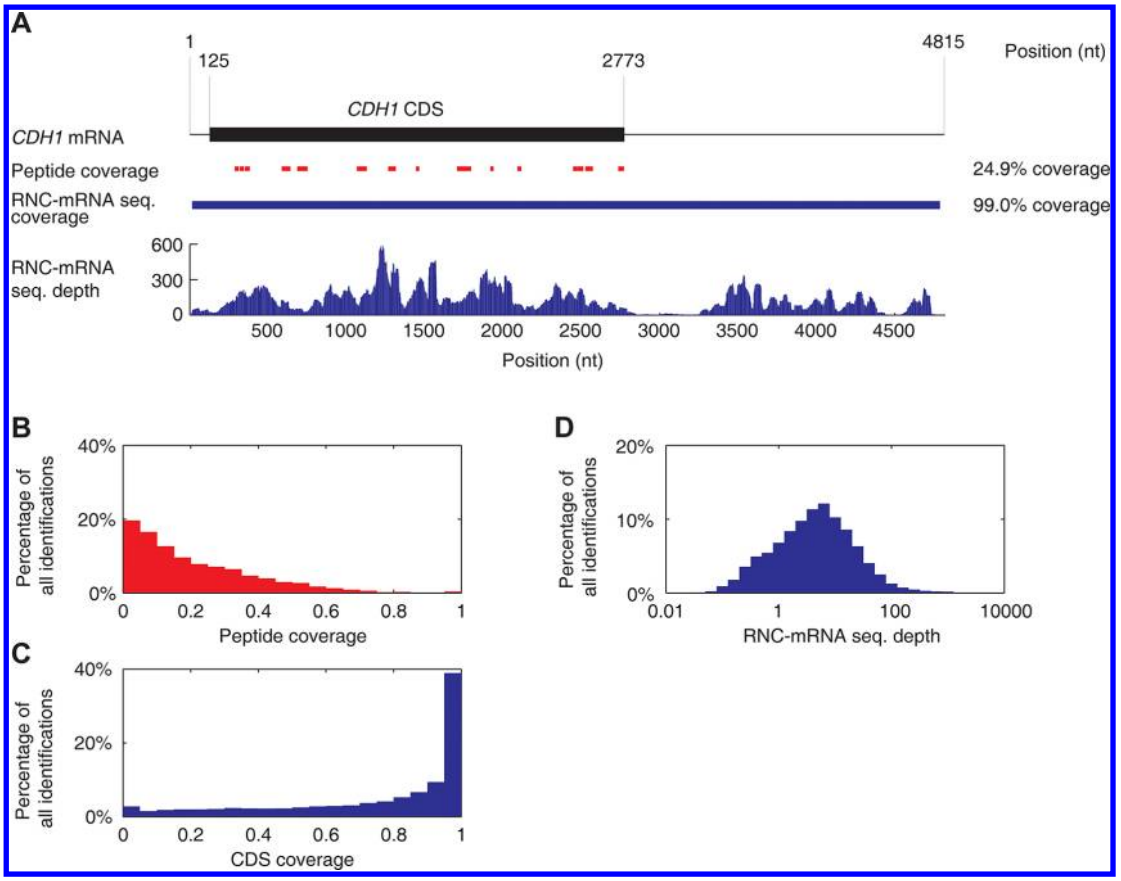


Figure 1. Sequence coverage of translating mRNA sequencing in comparison of mass spectrometry. (A) Peptide sequence coverage of E-cadherin and RNC-mRNA sequence coverage of its coding gene *CDH1* of Caco-2 cells. MS-identified peptides (red) and RNA-seq-acquired RNC-mRNA sequence (blue) are aligned to the nucleotide position bar of *CDH1* mRNA with a black box indicating the coding sequencing (CDS) region. The frequency of detection of each nucleotide (seq. depth) is shown with a histogram below. (B–D) Genome-scale comparison of sequence coverage in Caco-2 cells. Genes were classified into categories in a stepwise manner based on the sequence coverage of peptide (B) and RNC-mRNA (C) as well as the RNC-mRNA seq. depth (D), respectively. Each bar indicates the percentage ratio of the number of genes in this category to the number of all identifications.

the AST and SNV exploration.[43,44] In contrast, translating mRNA analysis with NGS can deliver millions of reads and potentially much higher sequence coverage. To demonstrate this argument, we used Caco-2 cells and performed RNA-seq on its full-length RNC-mRNAs. For comparison, we obtained the label-free MS data of Caco-2 cells from Wisniewski et al.[45]

As an example, out of ~14.8 million total 50-nt reads acquired by RNA-seq, 12 195 were mapped onto *CDH1* gene, a moderately expressed gene encoding E-cadherin. Each nucleotide in the full-length mRNA was covered for 126.6 times on average by the short sequencing reads, and 99.0% of the *CDH1* full-length mRNA was covered by at least one read, while its CDS was 100% covered by the sequencing reads (Figure 1A). In comparison, only 24.9% of the amino acids in the full-length protein were detected in peptides by MS. On a genome-wide scale, merely ~1% of the proteins were well-covered (coverage >80%) by MS identified peptides, while a majority (58.5%) of the proteins were detected with the peptide sequence coverage <20% (Figure 1B). In contrast, 60.1% of the RNC-mRNA sequencing detected genes were highly covered by 50 nt sequencing reads (CDS coverage >80%) (Figure 1C), and 79.9% of the mRNA reached the sequencing depth more than 1; namely, each nucleotide of the full-length mRNA was sequenced at least once on average (Figure 1D).

To be noted, we used a similar RNA-seq throughput to our previous work for better comparison.[15] The mRNA sequence depth can be boosted by increasing the read length with the widely used 2 × 100bp pair-end sequencing that provides four times greater sequencing depth. For example, an entire Illumina HiSeq-2000 lane for such sequencing can deliver over 240 million reads,[46] sixteen times more than that in this study. Hence, the high-throughput RNC-seq is an economic and reliable method to provide translating evidence of gene expression with high sequence coverage. As a standard operation for C-HPP, highly confident proteins with protein false discovery rate (FDR) <1% are mandated for data inclusion in numerous proteome databases, such as PeptideAtlas.[47] This stringent criterion is very helpful to rule out the noise in protein identifications. However, it also leads to a reduction of the peptide sequence coverage, which is crucial for mapping of PMTs and de novo sequencing.[48] Our results shown above helped to overcome this obstacle by remarkably increasing the CDS coverage while providing genome-wide translating evidence and the RNC-mRNA abundance information for predicting the protein abundance based on our published model.[15]

Detection of Sequence Variations with Full-Length Translating mRNA Analysis

According to the high sequence coverage and sequencing depth shown above, we believe that the accurate identification of SNVs and ASTs is feasible by analyzing the RNC-mRNA sequencing reads.

dx.doi.org/10.1021/pr4007409 | J. Proteome Res. 2014, 13, 50–59

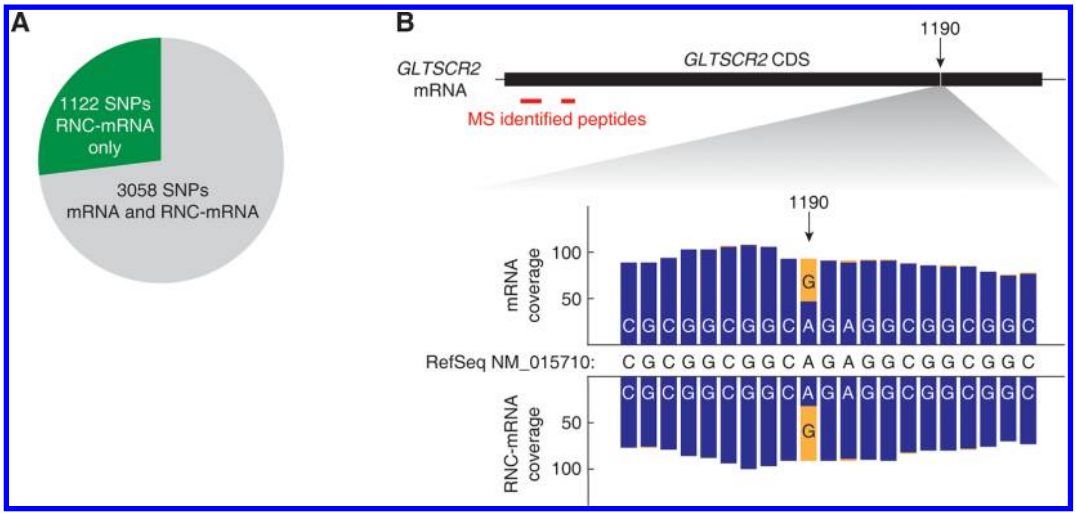


Figure 2. Single nucleotide polymorphisms (SNVs) identified from RNC-mRNA sequencing reads. (A) Fraction of all 4180 SNVs identified in RNC-mRNA was not significant in mRNA. (B) *GLTSCR2* (NM_015710) as an example. The two MS-identified peptides are marked as red. The region near the position 1190 in the mRNA is enlarged. The bar charts (upper chart for mRNA and lower chart for RNC-mRNA) represent the nucleotides covered at this position: the nucleotides in the reads are same (blue) or different (orange) to the reference sequence.

In this regard, we identified 4180 significant SNVs in 2373 Caco-2 translating mRNA (Supplementary Table 1 in the Supporting Information). However, 1122 of these SNVs were significantly detected only in the translating mRNA fraction but not in the mRNA fraction ($P < 0.05$, Fisher's exact test; Figure 2A). Although the sampling error cannot be ruled out completely, a number of highly confident SNVs that were only significant in translating mRNA fraction were observed. As an example, the gene *GLTSCR2* (NM_015710) is considerably expressed with 258.2 and 194.6 rpkM in mRNA and RNC-mRNA, respectively. According to the RefSeq-RNA reference sequence, the nucleotide at position 1190 is an adenosine. This is consistent with our results that more than half of the reads covered this position (47 out of 93) in the mRNA fraction were adenosines. However, 59 out of 91 reads covered this position in RNC-mRNA were guanosine, showing an SNV A1190G ($P = 0.015$, Fisher's exact test) that causes an amino acid change Q389R at protein level (Figure 2B). Such amino acid change was not previously searched by Wisniewski et al. and could be only detected with a specific query in the error-tolerant mode.[45] Similar cases can be found in multiple genes, for example, *CYC1* (NM_001916) T270C, *PSMD8* (NM_002812) C877T, *SCD* (NM_005063) C480G, A2459G, and so on. This indicates that some SNVs can be translated with much higher efficiency than the others, showing a genome-wide and unproportional modulation of SNVs at the translation level. To be noted, this is the downstream modulation of transcriptional information and thus independent from the sources of SNVs (e.g., genetic mutations, transcriptional errors and RNA editing, etc.). We found numerous SNVs in the CDS (e.g., *GLTSCR2* A1190G), 5′-UTR (e.g., *SCD* C480G), and 3′-UTR (e.g., *SCD* A2469G) regions, respectively. It is known that the SNVs in the 5′-UTR and 3′-UTR may alter the translational regulation by changing the binding sites of regulatory RNAs (e.g., miRNAs) and proteins, leading to functional consequences.[49−51] This type of information can only be obtained by sequencing full-length RNC-mRNA. The high-throughput and parallel feature of NGS promotes the genome-wide SNV detection to a quantitative level, which is major technology compensation to current MS. In addition, we have shown that the transmission efficiency of various ASTs from mRNA to RNC-mRNA and protein levels can be precisely investigated.[15] Here we also successfully detected three splice variants of the gene *BDP1* in Caco-2 cells, including *H5C7152.2*, *H5C7152.5*, and *H5C7152.9* with translating mRNA analysis. Actually, working with proteomic information resources is another efficient way to identify specific known and novel splice variants in tissue and plasma specimens[28−30,52] (reviewed in ref 31). These technologies are complementary and beneficial to each other by the data integration strategy shown in this study.

Translating mRNA Provides a Reference Standard for MS-Data Integration

As previously shown, full-length RNC-mRNA sequencing can generate high-quality translating evidence with optimal sequence coverage and gene identifications. Hence, translating mRNA information should be potentially able to serve as an independent reference standard for the integration of MS data from different laboratories and with various MS technologies.

To address this question, we identified 14 113 genes in Caco-2 RNC-mRNA data set (Supplementary Table 2 in the Supporting Information), which is close to the number of translating genes that were identified in human lung cancer A549 and H1299 cells.[15] These numbers can serve as the estimated numbers of protein species of each respective cell line based on the translating evidence. For data integration analysis, we downloaded four MS data sets on Caco-2 cell proteome identified by ICAT,[53] SILAC,[54] label-free,[45] and ICPL[55] techniques, respectively. Duplicated identifications were merged. The confident proteins that are derived from the merged data set across the four different platforms can be annotated to 7876 genes. Among these genes, 6688 were identified by RNC-mRNA NGS (Figure 3). In this regard, there were ~7000 translating genes remained unidentified by MS. We also noted that there were 595 MS-identified proteins that were not included in the 14 113 genes identified by RNC-mRNA sequencing. This discrepancy can have multiple origins, including system limitation of both MS and RNC-seq technologies, database mapping inconsistency, contaminations, as well as extreme conditions, such as highly discordant degradation rates of RNC-mRNA and proteins (reviewed in ref 3). For example, our RNC-mRNA sequencing

dx.doi.org/10.1021/pr4007409 | J. Proteome Res. 2014, 13, 50−59

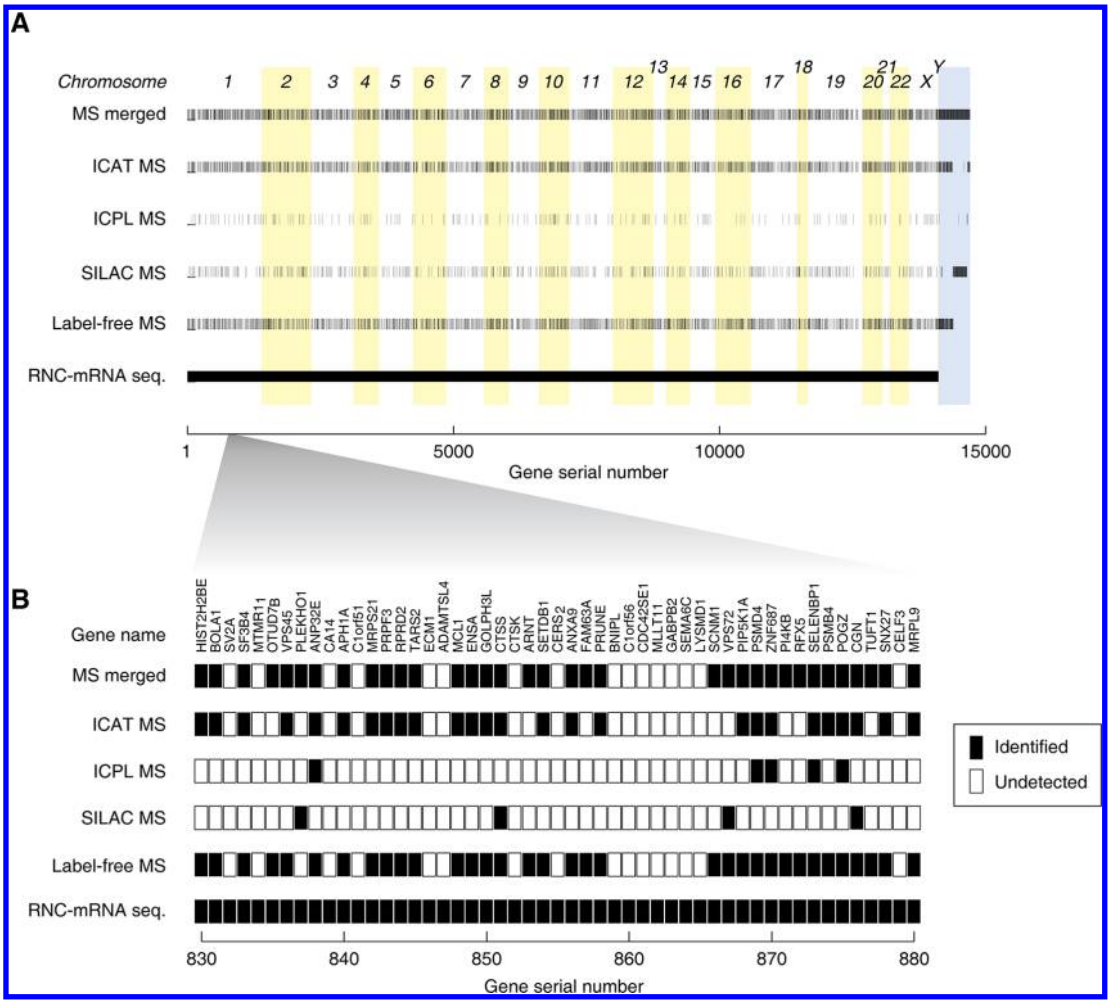


Figure 3. Data integration of mass spectrometry across different platforms and laboratories by full-length translating mRNA sequencing. (A) Proteins of Caco-2 cells that were identified by the ICAT, iTRAQ, SILAC, and label-free based MS, together with their merged identifications, are respectively aligned to the RNA-seq identified translating genes in a chromosome-by-chromosome manner. The genes that were identified by MS, but not by RNA-seq are shaded in a blue box. Each black vertical line represents an identified protein/gene. The scale bar below indicates the gene serial numbers of RNA-seq identifications, consistent with those included in Supplementary Table 1 in the Supporting Information. (B) Local zoom-in of a segment in panel A. Identified proteins/genes are indicated by black boxes.

reads were mapped to NCBI RefSeq-RNA database, whereas the MS data sets were searched using either IPI or UniProt databases. It is known that a number of entries of these databases are not mappable, which can be improved by the collective efforts of the community to standardize the data collection and deposition (reviewed in ref 3).

With the previous results, we demonstrated that full-length translating mRNA analysis can provide translation evidence and RNC-mRNA abundance information for the majority of proteins, thus serving as a useful baseline of the MS data integration in C-HPP. Because of biochemical and biophysical features, a number of proteins are very resistant to digestion and thus not suitable for MS identification (reviewed in ref 3). These proteins, together with those with extremely low abundances, represent the majority of "missing proteins". As integrated in PeptideAtlas (protein FDR < 1%), at least one highly confident peptide can be mapped to each of the ~12 500 Swiss-Prot entries, and ~7500 coding genes have no confident protein evidence.[47] In agreement, an MS data integration assay on 13 normal and cancer cell lines discovered protein evidence for 12 101 Swiss-Prot entries (protein FDR < 1%).[9] It can be foreseen that the translating evidence and RNC-mRNA abundance information acquired by this study, plus the large scale RNC-seq on various human cell/tissue/organ in collaboration with the scientific community, will generate a comprehensive map of human-translating mRNA to assist on the HPP/C-HPP goals. In addition, we observed that among the Caco-2 translating genes, 1397 were categorized as noncoding genes in RefSeq (Supplementary Table 2 in the Supporting Information). With RNC-seq, we previously detected and validated novel translating genes that have no proteomic evidence to date, such as *HMGB3P1*.[15] This capacity of RNC-seq may confer great impact

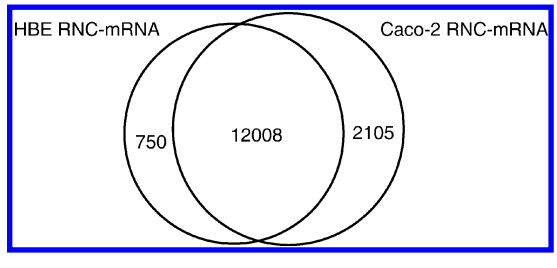


Figure 4. Translatome comparison of HBE and Caco-2 cells.

dx.doi.org/10.1021/pr4007409 | J. Proteome Res. 2014, 13, 50−59

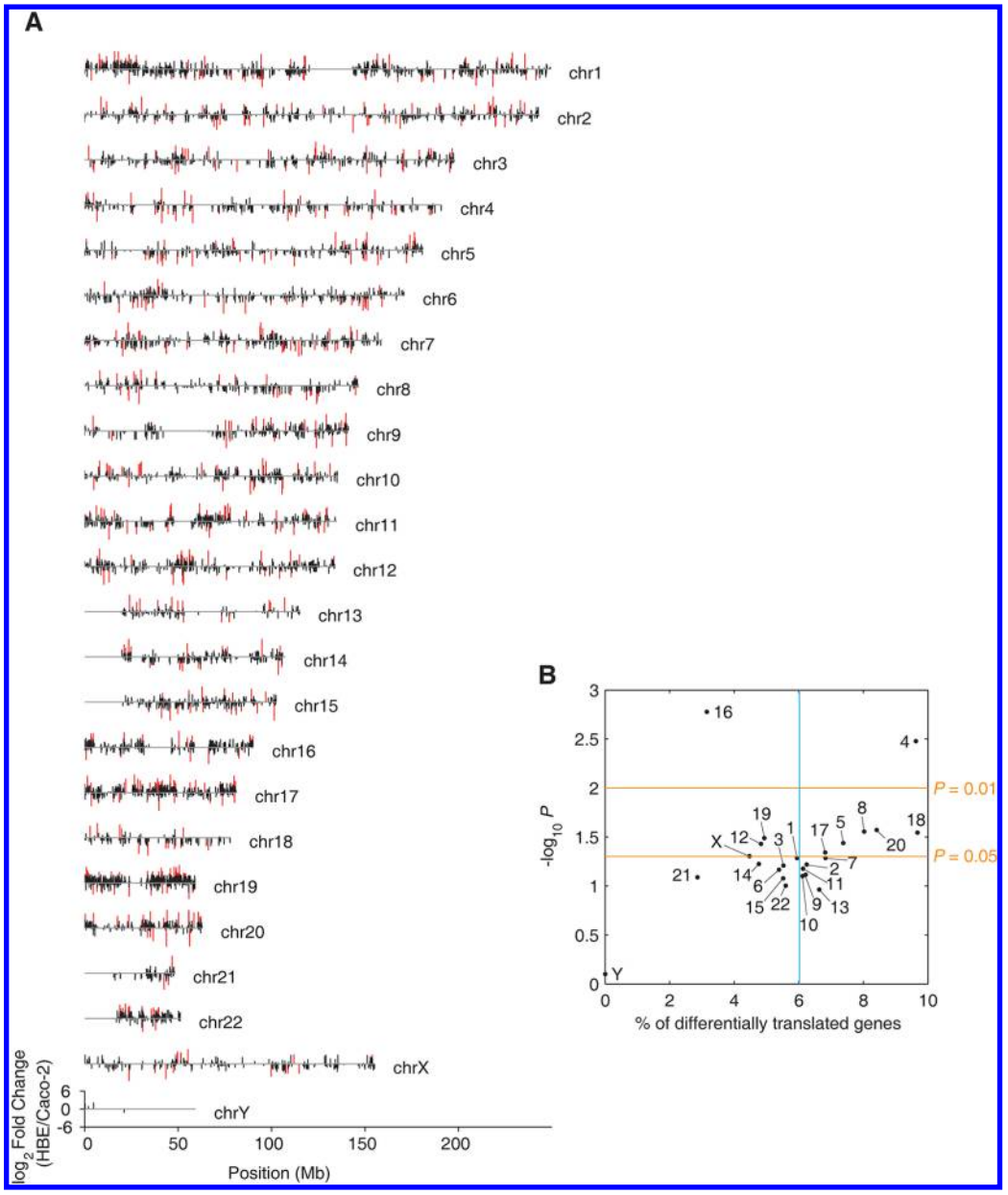


Figure 5. Cell-specific and chromosome-centric analysis of translational control. (A) Differentially translated genes and their chromosomal distributions. The fold changes of the up- and down-regulated genes (comparing HBE with Caco-2 cells) were shown above and below the baseline of each chromosome, respectively. The position scale bar illustrates the chromosomal position of each gene, with the chromosome Y as an example. Statistically significant changes ($P < 0.001$) were shown in red, otherwise in black. (B) Enrichment analysis of translational alterations. Individual chromosome is plotted in the graph according to the percentage of DTGs and the P value determined by Fisher's exact test. The α sizes of 0.05 and 0.01 were plotted by the orange lines, respectively. The average percentage of DTGs was indicated by the blue line.

on both C-HPP and diverse biologies, allowing for discovery of new proteins even in nonmodel species with no available proteome knowledgebases. Favorable to this notion, we detected that 141 reads (4.05 rpkM) acquired from the RNC-seq on Caco-2 cells were mapped to the *ESRG* gene (NR_027122.1) that is still marked as a nonprotein coding gene in NCBI RefSeq database (Supplementary Table 2 and Supplementary Figure S1 in the Supporting Information). However, recent studies revealed that this gene is actually the coding gene of the HESRG protein.[56,57]

Differentially Translated Genes Are Unevenly Distributed Across Chromosomes

We showed that RNC-seq is complementary to C-HPP, especially regarding its primary goal to define at least one protein for each coding gene of human genome. We next tried to address the equally important question of how to use translating mRNA information to predict the protein abundance in different cell types. Hence, we compared the relative abundance of RNC-mRNA between HBE and Caco-2 cells, followed by the analysis in a chromosome-by-chromosome manner.

In this study, we quantified 14 113 translating genes in Caco-2 cells compared with 12 758 translating genes in HBE cells as reported previously.[15] Among them, 12 008 genes were detected in both cell lines, while 750 genes in HBE cells and 2105 genes in Caco-2 cells, respectively, are cell-specific translating genes (CSTGs) (Figure 4). We employed edgeR, which uses the trimmed mean of M-values method based on the negative binomial distribution to reliably detect the differential

dx.doi.org/10.1021/pr4007409 | J. Proteome Res. 2014, 13, 50−59

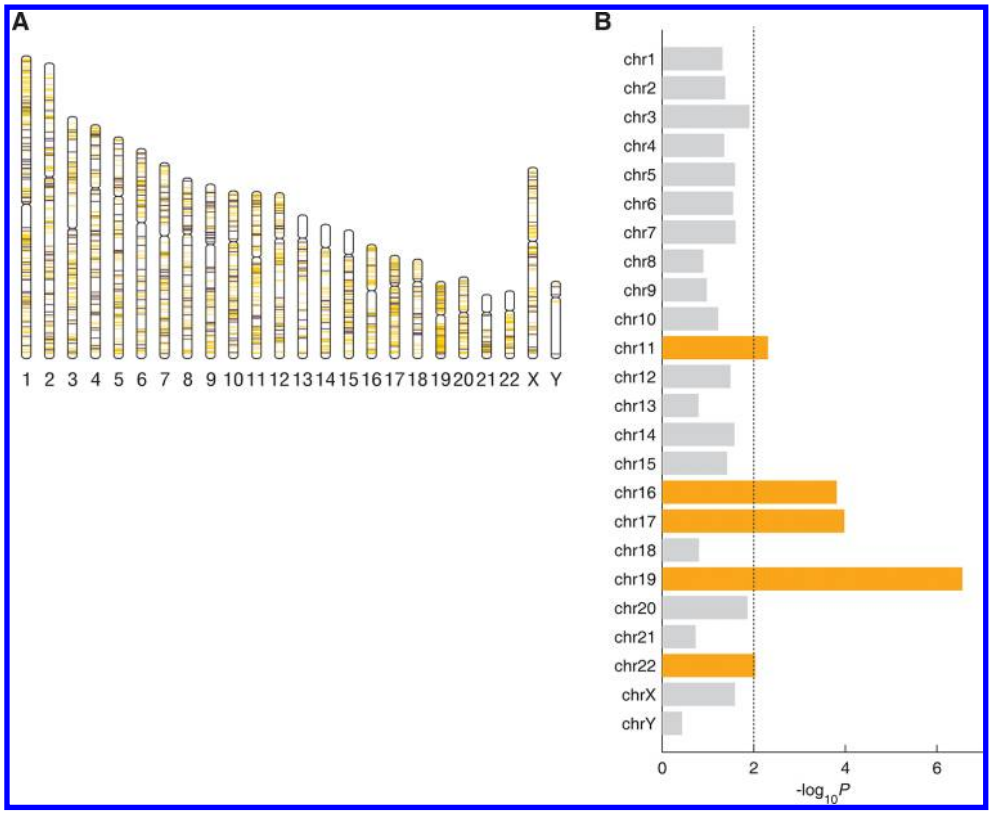


Figure 6. Chromosome-centric analysis of cell-specific translating genes. (A) Chromosomal distribution of cell-specific translating genes. Genes that were translated in only HBE (purple bands), and only Caco-2 (yellow bands) cells were indicated in a chromosome-by-chromosome manner, respectively. (B) Statistical comparison of the chromosomal distribution of cell-specific translating genes. Fisher's exact test was employed to test whether the translating genes detected in only HBE and Caco-2 cells are evenly distributed in each chromosome. The α size of 0.01 was plotted by a dashed line as a threshold. The orange bars indicate the chromosomes that have the significant enrichment of Caco-2 cell-specific translating genes, while the gray bars show the chromosomes with nonsignificant enrichment in either cell line.

expression.[41] When focusing on the genes identified in both cell lines, a total of 682 differentially translated genes (DTGs, Fisher exact test, $P < 0.001$ calculated by edgeR), with approximately greater than eight-fold up-/down-regulation of RNC-mRNA abundance,[15] were observed (Figure 5A). They exhibited differential enrichment pattern among chromosomes (Figure 5A). Visually, genes in chromosomes 4 and 18 were more actively translated in Caco-2 cells, while those in chromosome 20 were more translated in HBE cells (Figure 5A). Statistically, the DTGs between the two cell lines were significantly enriched on chromosomes 4, 5, 8 17, 18, and 20 (Fisher exact test, $P < 0.05$), while these DTGs tend not to distribute on chromosomes 12, 16, 19, and X (Fisher exact test, $P < 0.05$) (Figure 5B). In addition, we found 185 genes with TR difference of more than four-fold between the two cell lines. These genes are not evenly distributed in chromosomes either and are significantly enriched on chromosomes 3 and 19 (Fisher exact test, $P < 0.05$) (Supplementary Figure S2 in the Supporting Information), suggesting that the genes on these two chromosomes are more actively involved in translational control.

It is known that mRNA correlates poorly to protein; namely, only <40% of mRNA-protein bivariate correlation in their abundances exists (reviewed in ref 2). Theoretically, the significance of the strategy shown in this study is based on our previous discovery on a linear multivariate model among the relative abundances of RNC-mRNA and protein as well as the mRNA length.[15] This model allowed us to calculate >94% of protein relative abundance based on the RNC-mRNA information. To be noted, we have shown that the relative abundance of total mRNA (information at transcriptome level) cannot be used to replace its counterpart of RNC-mRNA in this model as the stepwise regression is nonsignificant.[15] This argues the advantage and necessity of focusing on translating mRNA per C-HPP rationale. In addition to MS data integration for profiling, quantification is an equally important issue for C-HPP. In this regard, the RNC-mRNA information has a unique advantage in predicting protein abundances, inheriting the accurate sequence coverage and improved gene identifications of transcriptome, while providing translating evidence for each of these genes.

Cell-Specific Translating Genes Are Unevenly Distributed Across Chromosomes

We then tried to answer whether the genes that are specifically translated in Caco-2 and HBE cells are evenly distributed across chromosomes. Addressing this question is to demonstrate an approach for probing "missing proteins" in different cell types and their distribution characteristics, which is expandable to human proteome level by further data integration.

We observed remarkably uneven distribution of CSTGs across chromosomes (Figure 6A). Statistically, we employed Fisher's exact test to test the null hypothesis that there is no biased distribution of the CSTGs on each chromosome. Our results overruled this null hypothesis by detecting that the CSTGs of Caco-2 cells are significantly concentrated on chromosomes 11, 16, 17, 19, and 22 ($P < 0.01$) (Figure 6B). To be noted as an example, Caco-2 cells have an active translation-on behavior on the chromosome 19 (Figure 6A,B). This implies that the

dx.doi.org/10.1021/pr4007409 | J. Proteome Res. 2014, 13, 50−59

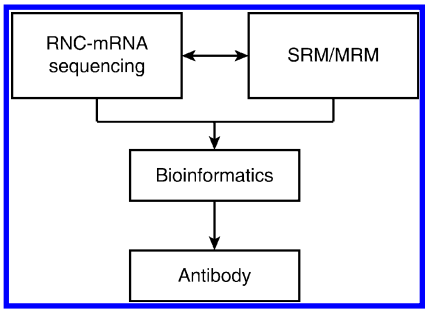


Figure 7. Strategic addition of translating mRNA analysis for HPP/C-HPP.

chromosomal structure variations may cause greater impact on genes of chromosome 19. In favor of this notion, it was reported that the long-arm deletion of chromosome 19 correlates to the differentiation of glioma cells as well as various cancers, such as ovarian cancer[58] and neuroblastomas.[59] These findings are comparable to previous reports from us[15] and other groups,[60,61] showing that chromosome 19 is highly involved in cellular development and carcinogenesis. In another aspect, the chromosome imbalance has been observed in many cancer cells (reviewed in ref 62). Concerning Caco-2 cells, chromosomal imbalances were observed on chromosomes 1, 4, 8, 9, 10, 11, 12, 16, 17, 18, and 20.[63] This does not match the chromosome-centric pattern of CSTGs, especially those on the chromosome 19, detected in this study. This suggests that the analysis of translating mRNA by RNA-seq can provide critical information on chromosome-centric expression patterns at translational level. Therefore, our results represent a novel chromosomal view of translation on and off behaviors on a genome-wide scale that is beneficial for the C-HPP goal of finding "missing proteins".

■ CONCLUSIONS

We proposed here to include translating mRNA analysis as the fourth pillar of HPP/C-HPP, in addition to SRM/MRM, bioinformatics, and antibody-based verifications (Figure 7). For the first phase of C-HPP,[3,5] the work flow of this strategy starts with the data integration and reciprocal validation with protein and translating evidence to explore the expression of the coding genes of each human chromosomes. This procedure will predominantly rely on bioinformatics knowledgebases and tools. As a result, a comprehensive list of "missing proteins", all with translating evidence and RNC-mRNA abundance information (Figure 3), will be created and subjected to antibody-based verifications. In the future, this strategy will potentially contribute to the B/D-HPP in terms of predicting tissue/disease-specific proteome and protein abundance differences, SNVs, as well as ASTs by translating mRNA analysis.

■ ASSOCIATED CONTENT

Supporting Information

Three supplementary tables and supplementary notes that support this article. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*Qing-Yu He: Phone/Fax: +86-20-85227039. E-mail: tqyhe@jnu.edu.cn.
*Gong Zhang: Phone/Fax: +86-20-85224031. E-mail: zhanggong@jnu.edu.cn.
*Tong Wang: Phone: +86-20-85225960. Fax: +86-20-85222616. E-mail: tongwang@jnu.edu.cn.

Author Contributions

†J.Z., Y.C., and J.G. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Drs. Pengyuan Yang, Fudan University, and Ping Xu, Beijing Proteome Research Center, for their critical consultations. We thank Dr. Gilbert S. Omenn, HUPO and University of Michigan, for his insightful discussion and thoughtful suggestions. This work was supported by the National Basic Research Program "973" of China (2011CB910700) to Q.Y.H., the National Natural and Science Foundation of China (81372135, 81272185, and 81000516 to T.W.; 81322028 and 31300649 to G.Z.), the Institutional Grant of Excellence of Jinan University, China (50625072) to G.Z., the Fundamental Research Funds for the Central Universities of China (21612202 and 21612406) to G.Z. and T.W., respectively, and the Doctoral Fund of Ministry of Education of China (20104401120008) as well as the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry of China to T.W.

■ REFERENCES

(1) Schmutz, J.; Wheeler, J.; Grimwood, J.; Dickson, M.; Yang, J.; Caoile, C.; Bajorek, E.; Black, S.; Chan, Y. M.; Denys, M.; Escobar, J.; Flowers, D.; Fotopulos, D.; Garcia, C.; Gomez, M.; Gonzales, E.; Haydu, L.; Lopez, F.; Ramirez, L.; Retterer, J.; Rodriguez, A.; Rogers, S.; Salazar, A.; Tsai, M.; Myers, R. M. Quality assessment of the human genome sequence. *Nature* **2004**, *429* (6990), 365−8.
(2) Vogel, C.; Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **2012**, *13* (4), 227−32.
(3) 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* (3), 221−3.
(4) 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* (4), 2005−13.
(5) Huhmer, A. F.; Paulus, A.; Martin, L. B.; Millis, K.; Agreste, T.; Saba, J.; Lill, J. R.; Fischer, S. M.; Dracup, W.; Lavery, P. The chromosome-centric human proteome project: a call to action. *J. Proteome Res.* **2013**, *12* (1), 28−32.
(6) Lee, H. J.; Jeong, S. K.; Na, K.; Lee, M. J.; Lee, S. H.; Lim, J. S.; Cha, H. J.; Cho, J. Y.; Kwon, J. Y.; Kim, H.; Song, S. Y.; Yoo, J. S.; Park, Y. M.; Kim, H.; Hancock, W. S.; Paik, Y. K. Comprehensive genome-wide proteomic analysis of human placental tissue for the Chromosome-Centric Human Proteome Project. *J. Proteome Res.* **2013**, *12* (6), 2458−66.
(7) Aebersold, R.; Bader, G. D.; Edwards, A. M.; van Eyk, J. E.; Kussmann, M.; Qin, J.; Omenn, G. S. The biology/disease-driven human proteome project (B/D-HPP): enabling protein research for the life sciences community. *J. Proteome Res.* **2013**, *12* (1), 23−7.
(8) Nagaraj, N.; Wisniewski, J. R.; Geiger, T.; Cox, J.; Kircher, M.; Kelso, J.; Paabo, S.; Mann, M. Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **2011**, *7*, 548.

(9) Wu, S.; Li, N.; Ma, J.; Shen, H.; Jiang, D.; Chang, C.; Zhang, C.; Li, L.; Zhang, H.; Jiang, J.; Xu, Z.; Ping, L.; Chen, T.; Zhang, W.; Zhang, T.; Xing, X.; Yi, T.; Li, Y.; Fan, F.; Li, X.; Zhong, F.; Wang, Q.; Zhang, Y.; Wen, B.; Yan, G.; Lin, L.; Yao, J.; Lin, Z.; Wu, F.; Xie, L.; Yu, H.; Liu, M.; Lu, H.; Mu, H.; Li, D.; Zhu, W.; Zhen, B.; Qian, X.; Qin, J.; Liu, S.; Yang, P.; Zhu, Y.; Xu, P.; He, F. First proteomic exploration of protein-encoding genes on chromosome 1 in human liver, stomach, and colon. *J. Proteome Res.* **2013**, *12* (1), 67−80.

(10) Chen, G.; Gharib, T. G.; Huang, C. C.; Taylor, J. M.; Misek, D. E.; Kardia, S. L.; Giordano, T. J.; Iannettoni, M. D.; Orringer, M. B.; Hanash, S. M.; Beer, D. G. Discordant protein and mRNA expression in lung adenocarcinomas. *Mol. Cell. Proteomics* **2002**, *1* (4), 304−13.

(11) Denoeud, F.; Aury, J. M.; Da Silva, C.; Noel, B.; Rogier, O.; Delledonne, M.; Morgante, M.; Valle, G.; Wincker, P.; Scarpelli, C.; Jaillon, O.; Artiguenave, F. Annotating genomes with massive-scale RNA sequencing. *Genome Biol.* **2008**, *9* (12), R175.

(12) Maier, T.; Guell, M.; Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **2009**, *583* (24), 3966−73.

(13) 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.

(14) Akan, P.; Alexeyenko, A.; Costea, P. I.; Hedberg, L.; Solnestam, B. W.; Lundin, S.; Hallman, J.; Lundberg, E.; Uhlen, M.; Lundeberg, J. Comprehensive analysis of the genome transcriptome and proteome landscapes of three tumor cell lines. *Genome Med.* **2012**, *4* (11), 86.

(15) Wang, T.; Cui, Y.; Jin, J.; Guo, J.; Wang, G.; Yin, X.; He, Q. Y.; Zhang, G. Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. *Nucleic Acids Res.* **2013**, *41* (9), 4743−54.

(16) Woo, S.; Cha, S. W.; Merrihew, G.; He, Y.; Castellana, N.; Guest, C.; Maccoss, M.; Bafna, V., Proteogenomic Database Construction Driven from Large Scale RNA-seq Data. *J. Proteome Res.* **2013**.

(17) Shiromizu, T.; Adachi, J.; Watanabe, S.; Murakami, T.; Kuga, T.; Muraoka, S.; Tomonaga, T. Identification of missing proteins in the neXtProt database and unregistered phosphopeptides in the PhosphoSitePlus database as part of the Chromosome-centric Human Proteome Project. *J. Proteome Res.* **2013**, *12* (6), 2414−21.

(18) Ranganathan, S.; Khan, J. M.; Garg, G.; Baker, M. S. Functional annotation of the human chromosome 7 ″missing″ proteins: a bioinformatics approach. *J. Proteome Res.* **2013**, *12* (6), 2504−10.

(19) Jeong, S. K.; Lee, H. J.; Na, K.; Cho, J. Y.; Lee, M. J.; Kwon, J. Y.; Kim, H.; Park, Y. M.; Yoo, J. S.; Hancock, W. S.; Paik, Y. K. GenomewidePDB, a proteomic database exploring the comprehensive protein parts list and transcriptome landscape in human chromosomes. *J. Proteome Res.* **2013**, *12* (1), 106−11.

(20) Liu, S.; Im, H.; Bairoch, A.; Cristofanilli, M.; Chen, R.; Deutsch, E. W.; Dalton, S.; Fenyo, D.; Fanayan, S.; Gates, C.; Gaudet, P.; Hincapie, M.; Hanash, S.; Kim, H.; Jeong, S. K.; Lundberg, E.; Mias, G.; Menon, R.; Mu, Z.; Nice, E.; Paik, Y. K.; Uhlen, M.; Wells, L.; Wu, S. L.; Yan, F.; Zhang, F.; Zhang, Y.; Snyder, M.; Omenn, G. S.; Beavis, R. C.; Hancock, W. S. A chromosome-centric human proteome project (C-HPP) to characterize the sets of proteins encoded in chromosome 17. *J. Proteome Res.* **2013**, *12* (1), 45−57.

(21) Fagerberg, L.; Oksvold, P.; Skogs, M.; Algenas, C.; Lundberg, E.; Ponten, F.; Sivertsson, A.; Odeberg, J.; Klevebring, D.; Kampf, C.; Asplund, A.; Sjostedt, E.; Al-Khalili Szigyarto, C.; Edqvist, P. H.; Olsson, I.; Rydberg, U.; Hudson, P.; Ottosson Takanen, J.; Berling, H.; Bjorling, L.; Tegel, H.; Rockberg, J.; Nilsson, P.; Navani, S.; Jirstrom, K.; Mulder, J.; Schwenk, J. M.; Zwahlen, M.; Hober, S.; Forsberg, M.; von Feilitzen, K.; Uhlen, M. Contribution of antibody-based protein profiling to the human Chromosome-centric Proteome Project (C-HPP). *J. Proteome Res.* **2013**, *12* (6), 2439−48.

(22) Thireos, G.; Griffin-Shea, R.; Kafatos, F. C. Untranslated mRNA for a chorion protein of Drosophila melanogaster accumulates transiently at the onset of specific gene amplification. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77* (10), 5789−93.

(23) Standart, N.; Hunt, T.; Ruderman, J. V. Differential accumulation of ribonucleotide reductase subunits in clam oocytes: the large subunit is stored as a polypeptide, the small subunit as untranslated mRNA. *J. Cell Biol.* **1986**, *103* (6 Pt 1), 2129−36.

(24) Nielsen, F. C.; Ostergaard, L.; Nielsen, J.; Christiansen, J. Growth-dependent translation of IGF-II mRNA by a rapamycin-sensitive pathway. *Nature* **1995**, *377* (6547), 358−62.

(25) Khan, D.; Sharathchandra, A.; Ponnuswamy, A.; Grover, R.; Das, S. Effect of a natural mutation in the 5′ untranslated region on the translational control of p53 mRNA. *Oncogene* **2013**, *32* (35), 4148−59.

(26) Thompson, A. J.; Abu, M.; Hanger, D. P. Key issues in the acquisition and analysis of qualitative and quantitative mass spectrometry data for peptide-centric proteomic experiments. *Amino Acids* **2012**, *43* (3), 1075−85.

(27) Bell, A. W.; Deutsch, E. W.; Au, C. E.; Kearney, R. E.; Beavis, R.; Sechi, S.; Nilsson, T.; Bergeron, J. J.; Group, H. T. S. W. A HUPO test sample study reveals common problems in mass spectrometry-based proteomics. *Nat. Methods* **2009**, *6* (6), 423−30.

(28) Menon, R.; Zhang, Q.; Zhang, Y.; Fermin, D.; Bardeesy, N.; DePinho, R. A.; Lu, C.; Hanash, S. M.; Omenn, G. S. States, D. J., Identification of novel alternative splice isoforms of circulating proteins in a mouse model of human pancreatic cancer. *Cancer Res.* **2009**, *69* (1), 300−9.

(29) Menon, R.; Omenn, G. S. Proteomic characterization of novel alternative splice variant proteins in human epidermal growth factor receptor 2/neu-induced breast cancers. *Cancer Res.* **2010**, *70* (9), 3440−9.

(30) Menon, R.; Roy, A.; Mukherjee, S.; Belkin, S.; Zhang, Y.; Omenn, G. S. Functional implications of structural predictions for alternative splice proteins expressed in Her2/neu-induced breast cancers. *J. Proteome Res.* **2011**, *10* (12), 5503−11.

(31) Menon, R.; Omenn, G. S. Identification of alternatively spliced transcripts using a proteomic informatics approach. *Methods Mol. Biol.* **2011**, *696*, 319−26.

(32) Barkan, A. Proteins encoded by a complex chloroplast transcription unit are each translated from both monocistronic and polycistronic mRNAs. *EMBO J.* **1988**, *7* (9), 2637−44.

(33) Zhang, G.; Hubalewska, M.; Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **2009**, *16* (3), 274−80.

(34) Garcia-Sanz, J. A.; Mikulits, W.; Livingstone, A.; Lefkovits, I.; Mullner, E. W. Translational control: a general mechanism for gene regulation during T cell activation. *FASEB J* **1998**, *12* (3), 299−306.

(35) Evans, M. S.; Ugrinov, K. G.; Frese, M. A.; Clark, P. L. Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat. Methods* **2005**, *2* (10), 757−62.

(36) Woolhead, C. A.; Johnson, A. E.; Bernstein, H. D. Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol. Cell* **2006**, *22* (5), 587−98.

(37) Ingolia, N. T.; Ghaemmaghami, S.; Newman, J. R.; Weissman, J. S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **2009**, *324* (5924), 218−23.

(38) Zhang, G.; Fedyunin, I.; Kirchner, S.; Xiao, C.; Valleriani, A.; Ignatova, Z. FANSe: an accurate algorithm for quantitative mapping of large scale sequencing reads. *Nucleic Acids Res.* **2012**, *40* (11), e83.

(39) Bloom, J. S.; Khan, Z.; Kruglyak, L.; Singh, M.; Caudy, A. A. Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. *BMC Genomics* **2009**, *10*, 221.

(40) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5* (7), 621−8.

(41) Robinson, M. D.; McCarthy, D. J.; Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2010**, *26* (1), 139−40.

(42) Zhang, G.; Lukoszek, R.; Mueller-Roeber, B.; Ignatova, Z. Different sequence signatures in the upstream regions of plant and animal tRNA genes shape distinct modes of regulation. *Nucleic Acids Res.* **2011**, *39* (8), 3331−9.

(43) Meyer, B.; Papasotiriou, D. G.; Karas, M. 100% protein sequence coverage: a modern form of surrealism in proteomics. *Amino Acids* **2011**, *41* (2), 291−310.
(44) Neuhauser, N.; Nagaraj, N.; McHardy, P.; Zanivan, S.; Scheltema, R.; Cox, J.; Mann, M. High performance computational analysis of large-scale proteome data sets to assess incremental contribution to coverage of the human genome. *J. Proteome Res.* **2013**, *12* (6), 2858−68.
(45) Wisniewski, J. R.; Ostasiewicz, P.; Dus, K.; Zielinska, D. F.; Gnad, F.; Mann, M. Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* **2012**, *8*, 611.
(46) Peng, Z.; Cheng, Y.; Tan, B. C.; Kang, L.; Tian, Z.; Zhu, Y.; Zhang, W.; Liang, Y.; Hu, X.; Tan, X.; Guo, J.; Dong, Z.; Liang, Y.; Bao, L.; Wang, J. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat. Biotechnol.* **2012**, *30* (3), 253−60.
(47) Farrah, T.; Deutsch, E. W.; Hoopmann, M. R.; Hallows, J. L.; Sun, Z.; Huang, C. Y.; Moritz, R. L. The state of the human proteome in 2012 as viewed through PeptideAtlas. *J. Proteome Res.* **2013**, *12* (1), 162−71.
(48) Frese, C. K.; Altelaar, A. F.; van den Toorn, H.; Nolting, D.; Griep-Raming, J.; Heck, A. J.; Mohammed, S. Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **2012**, *84* (22), 9668−73.
(49) Aouacheria, A.; Navratil, V.; Lopez-Perez, R.; Gutierrez, N. C.; Churkin, A.; Barash, D.; Mouchiroud, D.; Gautier, C. In silico whole-genome screening for cancer-related single-nucleotide polymorphisms located in human mRNA untranslated regions. *BMC Genomics* **2007**, *8*, 2.
(50) Nicoloso, M. S.; Sun, H.; Spizzo, R.; Kim, H.; Wickramasinghe, P.; Shimizu, M.; Wojcik, S. E.; Ferdin, J.; Kunej, T.; Xiao, L.; Manoukian, S.; Secreto, G.; Ravagnani, F.; Wang, X.; Radice, P.; Croce, C. M.; Davuluri, R. V.; Calin, G. A. Single-nucleotide polymorphisms inside microRNA target sites influence tumor susceptibility. *Cancer Res.* **2010**, *70* (7), 2789−98.
(51) Saunders, M. A.; Liang, H.; Li, W. H. Human polymorphism at microRNAs and microRNA target sites. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (9), 3300−5.
(52) Menon, R.; Im, H.; Zhang, E. Y.; Wu, S. L.; Chen, R.; Snyder, M.; Hancock, W. S.; Omenn, G. S., Distinct Splice Variants and Pathway Enrichment in the Cell Line Models of Aggressive Human Breast Cancer Subtypes. *J. Proteome Res.* **2013**.
(53) Hardwidge, P. R.; Rodriguez-Escudero, I.; Goode, D.; Donohoe, S.; Eng, J.; Goodlett, D. R.; Aebersold, R.; Finlay, B. B. Proteomic analysis of the intestinal epithelial cell response to enteropathogenic Escherichia coli. *J. Biol. Chem.* **2004**, *279* (19), 20127−36.
(54) Zeiser, J.; Gerhard, R.; Just, I.; Pich, A., Substrate Specificity of Clostridial Glucosylating Toxins and Their Function on Colonocytes Analyzed by Proteomics Techniques. *J. Proteome Res.* **2013**.
(55) Jochim, N.; Gerhard, R.; Just, I.; Pich, A. Impact of clostridial glucosylating toxins on the proteome of colonic cells determined by isotope-coded protein labeling and LC-MALDI. *Proteome Sci* **2011**, *9*, 48.
(56) Zhao, M.; Ren, C.; Yang, H.; Feng, X.; Jiang, X.; Zhu, B.; Zhou, W.; Wang, L.; Zeng, Y.; Yao, K. Transcriptional profiling of human embryonic stem cells and embryoid bodies identifies HESRG, a novel stem cell gene. *Biochem. Biophys. Res. Commun.* **2007**, *362* (4), 916−22.
(57) Wanggou, S.; Jiang, X.; Li, Q.; Zhang, L.; Liu, D.; Li, G.; Feng, X.; Liu, W.; Zhu, B.; Huang, W.; Shi, J.; Yuan, X.; Ren, C. HESRG: a novel biomarker for intracranial germinoma and embryonal carcinoma. *J. Neurooncol.* **2012**, *106* (2), 251−9.
(58) Skirnisdottir, I.; Mayrhofer, M.; Rydaker, M.; Akerud, H.; Isaksson, A. Loss-of-heterozygosity on chromosome 19q in early-stage serous ovarian cancer is associated with recurrent disease. *BMC Cancer* **2012**, *12*, 407.
(59) Grzendowski, M.; Wolter, M.; Riemenschneider, M. J.; Knobbe, C. B.; Schlegel, U.; Meyer, H. E.; Reifenberger, G.; Stuhler, K. Differential proteome analysis of human gliomas stratified for loss of heterozygosity on chromosomal arms 1p and 19q. *Neuro-Oncology* **2010**, *12* (3), 243−56.
(60) Hartmann, C.; Johnk, L.; Kitange, G.; Wu, Y.; Ashworth, L. K.; Jenkins, R. B.; Louis, D. N. Transcript map of the 3.7-Mb D19S112-D19S246 candidate tumor suppressor region on the long arm of chromosome 19. *Cancer Res.* **2002**, *62* (14), 4100−8.
(61) Ostler, K. R.; Yang, Q.; Looney, T. J.; Zhang, L.; Vasanthakumar, A.; Tian, Y.; Kocherginsky, M.; Raimondi, S. L.; DeMaio, J. G.; Salwen, H. R.; Gu, S.; Chlenski, A.; Naranjo, A.; Gill, A.; Peddinti, R.; Lahn, B. T.; Cohn, S. L.; Godley, L. A. Truncated DNMT3B isoform DNMT3B7 suppresses growth, induces differentiation, and alters DNA methylation in human neuroblastoma. *Cancer Res.* **2012**, *72* (18), 4714−23.
(62) Stallings, R. L. Are chromosomal imbalances important in cancer? *Trends Genet.* **2007**, *23* (6), 278−83.
(63) Kleivi, K.; Teixeira, M. R.; Eknaes, M.; Diep, C. B.; Jakobsen, K. S.; Hamelin, R.; Lothe, R. A. Genome signatures of colon carcinoma cell lines. *Cancer Genet. Cytogenet.* **2004**, *155* (2), 119−31.