Spring 2019

Identification of Candidate Human Satellite II RNA Binding Proteins in a Human Cancer Cell Line

Anthony J. Velleca , '19

Identification of Candidate Human Satellite II RNA Binding Proteins in a Human Cancer Cell Line

Anthony Velleca

A Thesis
Submitted to the Department of Biology
Swarthmore College
April 2019

Introduction

Overview of RNA Binding Proteins

RNA often associates with one or more RNA binding proteins (RBP(s)) to carry out the variety of functions it performs in the cell[1]. The complex that forms is referred to as a ribonucleoprotein particle (RNP)[2]. Movement, permanence and breakdown of all classes of RNA is regulated by RBPs, with examples ranging from miRNA processing to tRNA biogenesis[1]. One canonical role of RNPs is in post-transcriptional gene regulation, in which the translation of mRNA into a functional protein is regulated by one or more RBPs[2]. RNPs are also involved in RNA splicing, as RBPs are a crucial component of the spliceosome, which is responsible for catalyzing the removal of introns and joining of exons that must occur for pre-mRNA to be processed into mRNA[3]. Classically, when RNA was considered in the context of its relationship with RBPs, RNA was thought to be a bystander molecule regulated by proteins, as is the case for mRNA when proteins bind to it and regulate transcription levels[2]. Recent discoveries have established a more RNA-centric view of the function of some RNPs and illuminated the role of RBPs in novel cellular processes[2].

Non-coding RNAs and Expansion of RBP Library

While the relationships between mRNAs, tRNA, miRNAs and the binding proteins these RNAs associate with are more well defined, the relationships, and the function of the RNPs that form between non-coding RNA (ncRNA) and RBPs are just beginning to be investigated[4]. Recent studies have greatly expanded the number of ncRNAs, especially long ncRNAs (lncRNAs), and have revealed a range of previously unknown functions, including gene regulation and paraspeckle formation[4-7]. The functions of many well-studied ncRNAs are known to require the formation of RNPs[5,6,8]. Thus, determining the RNA binding proteins of ncRNAs that have been recently implicated in a cellular process, or do not have a known function, has the potential to provide insight into the full range of ncRNA functions as well as how ncRNAs complex with RBPs to perform these functions[4]. The systematic identification of RNA

binding proteins by several studies has greatly expanded the number of RBPs, elucidated new functions and caused a shift in our understanding of the function of RNA in RNPs[1,2,9,10]. Central to these methods is a crosslinking step that covalently links bound proteins to RNA, making sure that legitimate binding relationships are not lost as RNA-protein complexes are isolated[11]. Baltz et. al. found the mRNA-bound proteome to contain nearly 800 proteins in a human embryonic kidney cell line, a third of which were not previously annotated as RNA binding[10]. Another 15 percent of these proteins were previously not predicted to interact with RNA based on computational methods[10]. They also used next generation sequencing methods to show the protein occupancy on mRNA transcripts[10]. In a separate study, Castello et. al. identified 860 proteins by biochemical and statistical methods that classify as RBPs[9]. The authors emphasized that their list adds more than 300 RBPs to those previously identified[9]. They also noted how this systematic identification revealed new roles for RNA-binding enzymes of intermediary metabolism and RNA binding architectures[9].

Diverse roles for RNA in RNPs

Rather than playing a simply passive role in RNPs, RNA has been shown to modulate protein function. The activation of toll-like receptors (TLRs) involved in the recognition of pathogens by binding with double-stranded RNA, single-stranded RNA and ribosomal RNA is an example of the ways in which RNA serves to directly regulate protein activity[12]. Kinases involved in viral replication pathways have also been found to be regulated by RNA, with dsRNA triggering activation of protein kinase R (PKR) by dimerization and autophosphorylation[13]. Thus, RNA has the potential to serve as a critical molecule for regulating protein function through the formation of RNPs. Another significant finding from the systematic identification of RBPs is that a variety of enzymes involved in critical steps in intermediary metabolism have been identified as RBPs, including most of the enzymes involved in glycolysis[14]. One such example is glyceraldehyde-3-phosphate dehydrogenase (GAPDH), which is a key enzyme in glycolysis that has been shown to interact with the 3'UTR of interferon gamma (IFN-γ)

mRNA and inhibit its translation[15]. For GAPDH, RNA binding and enzymatic activity have been shown to be incompatible, with protein function switching from metabolism to post-transcriptional regulation of gene expression under different cellular conditions[15]. The fact that GAPHD, a critical housekeeping gene, is regulated by an RNA binding relationship indicates how widespread and important RNA-mediated regulation is. Given the number of enzymatic proteins identified as RBPs, there likely exist RNA-dependent regulatory roles for other highly abundant metabolic proteins[2].

RNA also has the ability to serve as a flexible scaffold upon which functional protein complexes can be assembled[2]. One such example is telomerase in yeast. Telomerase is the enzyme responsible for adding repeated DNA sequences to the end of the chromosome, and is made up of an 1157nt RNA subunit, which serves as the template for telomeric DNA synthesis, in addition to several protein subunits[16]. It has been shown that in addition to serving as a template for telomeric DNA synthesis, the RNA subunit is responsible for tethering the protein subunits into a complex[16]. A model has been proposed that classifies RNPs based on the extent to which the RNA component contributes to overall RNP structure[16]. For example, RNPs with specific structures determined primarily by the RNA component itself include the ribosome and ribozyme-protein complexes[16]. The second category comprises RNPs where the specific structure is largely determined by proteins, as is thought to be the case for small nuclear RNPs[16]. The third category, under which the RNA subunit of telomerase falls, have no specific structure for the RNP as a whole and instead the RNA serves as a flexible tethering molecule to localize a number of proteins involved in a cellular process[16]. It is thought that some of the long non-coding RNAs (lncRNAs) that associate with proteins can be classified into this third category[2].

Functions of RNPs that contain lncRNAs

LncRNAs, a heterogenous group of transcripts with diverse functions, are polyadenylated, sometimes spliced and frequently form extensive secondary structure[2]. These transcripts, which have recently gained increased attention from researchers, elicit a range of functions, including acting as

scaffolds, decoys and guides. One common feature is they often associate with proteins to carry out these multitude of functions[17]. In mouse embryonic stem cells (ESCs) knockdown of lncRNAs has been shown to have effects on gene expression comparable to the effects on gene expression when well-known ESC regulatory proteins are knocked down[18]. While the mechanisms for most of these lncRNA-dependent changes in gene expression remain undetermined[7], the proteins that associate with Xist, a lncRNA that is responsible for silencing one X chromosome during development in females, have recently been identified[8]. McHugh et. al. developed an RNA antisense purification (RAP) approach to purify a lncRNA from cells and identify the proteins that interact with the lncRNA by mass spectrometry (RAP-MS)[8]. Ten proteins that specifically associate with Xist RNA were identified by a quantitative mass spectrometry approach[8]. By knocking down each one of these ten proteins using siRNAs and assaying for inability to effectively silence gene expression on the X chromosome following the induction of *Xist* expression, they identified three proteins that are required for Xist-mediated transcriptional repression. One of these proteins, SHARP, is essential for both silencing the inactive X chromosome and excluding RNA polymerase II (Pol II) from the chromosome[8]. Thus, RAP-MS presents an excellent tool to identify the proteins that specifically bind to an RNA of interest and establish a list of proteins with functions to further investigate.

Prevalence of Novel RNA Binding Domains

While it is critical to gain a better understanding of the set of proteins that bind RNAs, understanding the diversity of residues and structures within RBPs that bind RNA is equally important. Studies of the RNA binding domains (RBDs) of RBPs allow for a better understanding of how proteins associate with RNA and how protein and RNA functions might be affected by this complexing. Following the addition of about 300 proteins to the list of proteins with RNA binding capabilities, researchers were interested in the RNA binding domains (RBDs) of these newly identified RBPs, as well as classifying all the RBDs of previously characterized RBPs. Specifically, there was interest in whether RBPs bind RNA by

classical domains that have been well characterized previously, or instead if RBPs tend to bind RNA via unconventional domains. Prior to recent studies that have systematically identified RBDs, the focus of most attention was on the so-called "classical RNA binding domains", which contain well defined RNA binding domains, such as the RNA recognition motif (RRM)[19], K homology (KH) domain, DEAD motif, double stranded RNA-binding motif (DSRM) or a zinc-finger domain[2]. This was primarily due to structural information on RNPs that was obtained by X-ray crystallography, which requires the types of rigid folds found in globular protein domains but not in intrinsically unfolded proteins[2].

Systematic identification of the RNA binding domains of RBPs has shown that many binding relationships do not occur at "classical RNA binding domains" and, instead, that many RBPs rely on disordered regions to bind RNA[2]. Castello et. al. identified RNA binding domains *in vivo* using an approach they termed "RBPmap" and were able to both reidentify classical RBDs as well as identify novel RBDs[20]. Surprisingly, they showed that more than half of RNA-RBP binding sites do not contain a conventional RBD[20]. 1,174 binding sites were identified within 529 HeLa cell RBPs, indicating the presence of multiple binding sites within many of these proteins[20]. The fact that intrinsically disorder regions (regions rich in the amino acids serine, proline, glycine, arginine, lysine and tyrosine[20,21] and natively lack stable three-dimensional structure[20]), make up half of the nearly 1,200 identified binding sites and that for 170 RBPs a disordered RBD is the only detectable RNA binding site, reveals the importance of unstructured protein domains as components of many RNPs[20]. One previously unknown RBP, MeCP2, contains a disordered region, whose RNA binding capability was validated using a binding assay that fused the lysine-rich domain from MeCP2 to eGFP[20]. The authors found RBDs to be well conserved, suggesting important functional roles for these protein domains[20]. Disordered protein regions have also been shown to be used by transcription factors to bind DNA[22] and roles in phase transitions and granule formation through RNA interactions with YGG, a repeat motif common in disordered regions, have been suggested[23]. While these studies are still emerging, the presence and

conservation of distinct motifs within disordered regions across nonhomologous RBPs suggest a range of biological functions[9,20].

RNPs in Disease

A variety of diseases, including neuropathies, muscular atrophies, metabolic disorders and cancer, have been linked to defective RBP expression and function[24]. The list of RBPs implicated in disease has grown following the identification of an increased number of RBPs. Notably, of the over eight hundred RBPs identified by Castello et. al., 86 are associated with human Mendelian disease based on a search of the Online Mendelian Inheritance in Man (OMIM) database[9]. A specific model of RNP formation leading to a disease state is the sequestration of proteins by a class of "decoy" lncRNAs[25]. These decoy lncRNAs do have a function in normal physiology, but when they are expressed from alleles where nucleotide repeat expansion has occurred they can be pathogenic, as sequestration disrupts the normal function of these recruited proteins[25].

One such example is myotonic dystrophy type 1 (DM1)[25], a muscle wasting disease that is characterized by a variety of other disorders, ranging from cataract formation to cerebral atrophy[25]. An individual with DM1 accumulates between 50 and 3000 CTG repeats in the 3'UTR of the dystrophia myotonica protein kinase (DMPK) gene[25]. The critical component of DM1 pathogenesis is expression of RNA from the genomic loci where the repeat expansion has occurred and the formation of nuclear foci that consist of the expanded repeat RNA[25]. Critically, the protein MBNL1 is sequestered by the expanded repeat RNA through binding to the stable hairpin loop that forms in the RNA[25]. The binding of MBNL1 to expanded repeat RNA disrupts the proteins normal function in regulating alternative splicing during development[25]. The resulting misregulation of alternative splicing in development leads to disease pathology[25]. Given that the expression of repetitive DNA sequences and sequestration of critical proteins by the resulting RNA has been shown to disrupt protein function and cause disease, we

are interested in expressed repeats in other pathologies and identification of the proteins that they may bind.

Human Satellite II (HSATII) DNA

Within the ninety eight percent of the human genome that does not code for proteins exists a class of DNA know as satellite DNA, so named because the sequences were first identified from genomic DNA fractions that had slightly different buoyancy densities in $CsSO_4$ than total DNA and thus looked like a satellite, remote from the primary DNA fraction[26]. These DNA fractions were designated Satellite I, II and III[26]. Satellite DNA is primarily located near the centromere and is characterized by tandemly repeating monomer sequences[26]. Categorization of satellite sequences has been refined to the point that there are now thought to be eight classes of human satellite DNA, with their variation residing in the length and sequence of the repeated monomer[26]. For example, alpha-satellite has a monomer length of 171 base pairs, while Human Satellite II (HSATII), the satellite of focus here, repeats every 23 to 26bp[26]. HSATII is defined as repeat arrays of the pentamer ATTCC that are poorly conserved between arrays[26]. In general, these repeats occur as the sequence $(ATTCCATTCG)_2$ followed by either one or two ATG motifs. Fluorescence *in situ* hybridization (FISH) revealed that the bulk of HSATII DNA is located on chromosomes 1, 2, 10 and 16, with several sites residing on other chromosomes[27,28].

HSATII RNA Expression in Cancer and Sequestration Model

While HSATII constitutes two percent of the human genome, it is not expressed at appreciable levels in normal human tissue[29]. Using a next generation digital gene expression (DGE) method to evaluate the transcriptome of primary tumors, Ting *et. al.* determined HSATII RNA to be nearly 150-fold overexpressed in pancreatic ductal adenocarcinomas (PDACs) compared to normal tissues, while other satellites did not have large changes in expression levels between cancerous and normal tissues[29]. Along with identifying increased HSATII expression in PDACs, they also detected robust HSATII expression in lung, kidney, ovarian and prostate cancers, but no expression in a broad range of healthy human

tissues including brain, colon, liver, lung and kidney[29]. HSATII expression is thought to be a result of the epigenetic dysregulation that occurs in cancer[29]. A follow up study found that one result of HSATII expression in cancer cells is an increase in HSATII copy number gain through an unclassified RNA-derived DNA intermediate mechanism[30].

Concurrently, it was found that HSATII RNA forms large nuclear foci in a variety of human cancers and that these sequences can affect the distribution of proteins that regulate chromatin[28]. These foci, called Cancer-Associated Satellite Transcript (CAST) bodies, sequester methyl CpG binding protein 2 (MeCP2) and SIN3A[28]. While MeCP2 is possibly best known for causing the neurodevelopmental disorder known as Rett syndrome (RTT) when it is mutated, this nuclear protein functions in chromatin architecture, regulation of RNA splicing and both repression and activation of transcription[31,32]. As mentioned previously, MeCP2 is a newly characterized RBP that binds RNA through a disordered region[20]. SIN3A is a protein that regulates transcription and is known to complex with MeCP2.[28] By performing RNA immunoprecipitations for MeCP2 and SIN3A, it was found that these two proteins bind HSATII RNA in cancer cells[28]. The accumulation of MeCP2 (and possibly other regulatory proteins) by HSATII RNA may contribute to further epigenetic misregulation if these sequestered proteins are then unable to perform their normal functions. Hall et. al., point out that HSATII RNA foci seem to resemble accumulations of "toxic repeat RNAs" that form in disorders such as myotonic dystrophy type 1, in which sequestration of alternative splicing factors by overexpressed repeat RNA is central to the pathology of the disease[28]. Thus, in order to gain a better understanding of the other proteins that interact with HSATII RNA present in cancer cells, we sought out to identify a comprehensive list of HSATII RNA binding proteins in cancer cell nuclei.

Design of Protocol for Identification of HSATII RBPs

After considering various approaches for systematically identifying RNA-binding proteins for targeted RNA sequences, we decided to adapt the RNA antisense purification with mass spectrometry

(RAP-MS) established by McHugh et. al. to identify HSATII RNA binding proteins[8,33]. One element that distinguishes RAP-MS from other protocols used to identify binding proteins is that crosslinking is done *in vivo*[8], allowing one to identify binding relationships that occur in the cell rather than associations between RNA and protein that may form under *in vitro* conditions, which can be confounding[34,35]. RAP-MS takes a UV-crosslinking and denaturing approach that is used by other methods, including crosslinking and immunoprecipitation[11], to identify only direct RNA-protein interactions, as cross-linking using formaldehyde or other means can lead to the isolation of indirect RNA-protein interactions, even after exposure to denaturing conditions meant to disrupt such interactions[8,36]. While binding relationships between RNA and protein typically rely on non-covalent interactions, exposure to 254nm UV radiation induces the formation of covalent bonds between aromatic ring structures found in several amino acids and all nitrogenous bases[37]. Long biotinylated antisense probes, which form very stable RNA-DNA hybrids, are used instead of antibodies or protein tags because of the ability for this base-pairing interaction to withstand the harsh denaturing conditions needed to effectively purify the RNA[8].

Following cross-linking and hybridization with probes, streptavidin coated magnetic beads are used to isolate RNA-DNA hybrids with bound proteins from a total cellular lysate[33]. The optimization of this protocol for achieving high yields of endogenous RNPs allows for the detection of proteins associated with a given RNA using a mass spectrometer, even when the RNA of interest is likely to make up a small sub-set of total cellular RNA[8]. Proteins are identified from a protein mixture following enzymatic digestion[8]. The formation of fragmented peptides allows for identification using peptide mass fingerprinting, which is done by comparing generated peptides to a database of possible peptide fragments from all known proteins using a database search tool[38]. RNA binding capabilities of identified proteins can then be confirmed by a protein-centric method, such as RNA immunoprecipitation or

Western blot[8]. The end result of RAP-MS is a selective list of binding proteins to a specific RNA of interest whose function can then be further investigated.

A protocol that selectively isolates HSATII RNA and identifies bound proteins was developed by adapting the RAP-MS approach. A probe that selectively captures HSATII RNA and not other satellite sequences or nuclear RNAs was designed and used to isolate proteins bound to HSATII RNA. Bound proteins were identified by mass spectrometry and a system for filtering candidate binding proteins was established and used to generate a candidate list of highest interest candidate HSATII RBPs with functions that are misregulated in cancer. Using this pipeline, it was confirmed that identified proteins were not solely the most abundant proteins present in the cancer cell line used. Identified proteins were also shown to be significantly enriched for RNA binding and nuclear localization. The comprehensive list of proteins that bind HSATII RNA *in vivo* generated here includes many proteins whose functions may be implicated in HSATII expression in cancer.

Materials and Methods

To identify binding proteins of Human Satellite II (HSATII) RNA an RNA antisense purification with mass spectrometry (RAP-MS) protocol developed by McHugh et. al.[33] was adapted. Buffers were made according to concentrations listed in protocol, while updated protocols can be found in the attached supplementary material. Details of this protocol can be found on the Guttman Lab's website (http://guttmanlab.caltech.edu/protocols-RAP-MS-schematic.php).

Probe Design

Transcripts that were sequenced following RNA immunoprecipitation (RIP) with either MeCP2 or SIN3A antibodies were aligned using MacVector nucleic acid alignment, disallowing for gaps. Alignment was scanned for 90nt regions with minimal base pair mismatch. Two such regions were

identified and biotinylated 90nt probes were ordered, with Probe 1 being the exact sequence of a portion of the aligned transcripts and Probe 2 being the complement of a portion of the aligned transcripts.

Cell Culture

A human osteosarcoma cell line (U2OS, ATCC®, HTB-96™) was used to identify HSATII RNA binding proteins because this specific cancer cell line has been found to express HSATII RNA at higher levels than other standard cancer cell lines, such as PC3 (prostate cancer cell line) and Hela (cervical cancer cell line), and have many prominent CAST bodies[28]. U2OS cells were thawed from storage at -80ºC and cultured in 10% FBS media (225mL minimum essential medium (MEM), 25mL FBS (VWR, 89510-186) 2.5mL Penicillin-Streptomycin, 2.5mL L-glutamine). For fluorescence *in-situ* hybridization, cells were fixed on cover slips as described[39]. For cross-linking and harvesting of cells prior to hybridization, cells were split from T75 flasks and grown on 150mm cell culture dishes until 90-100% confluent. If cell growth in 150mm cell culture dishes was uneven, cells were treated with trypsin and redistributed.

Fluorescence *in situ* Hybridization (FISH)

Fluorescence *in situ* hybridization (FISH) was used to determine if Probe 1 and Probe 2 appeared to hybridize to HSATII RNA in cancer cell nuclei. RNA hybridization and detection was performed as previously described[39] using Probe 1, Probe 2, Probe 1 and Probe 2, as well as a 24nt locked nucleic acid (LNA) probe known to robustly detect HSATII RNA[28]. 10 picomoles of each probe was used for hybridization and Dylight® 488 Streptavidin (Vector Laboratories, SA-5488) secondary antibody was used for detection. No difference in hybridization efficiency was observed between 15% and 25% formamide hybridization solutions. Hybridization efficiency of each probe was measured by scoring about 100 nuclei that were hybridized with either Probe 1 or Probe 2. Representative images of each probe were captured with 100X magnification using a Zeiss AxioObserver Z1 equipped with an

Apatome2 and Axiocam 702 CMOS monochrome camera. The more efficient probe was then used for the pulldown portion of the protocol.

In vivo Crosslinking and Cell Harvesting

Crosslinking and harvesting were performed as previously described[33], with the following modification (Supplemental, A). Crosslinking was performed with a BIO-RAD GS Gene Linker UV Chamber. Cell solution was diluted 1:10 and then counted using a BIO-RAD TC20™ Automated Cell Counter, with five plates on average yielding just over 50 million cells. Thus, the crosslinking and harvesting protocol was performed eight times to yield the 400 million cells needed for the next steps in the adapted RAP-MS protocol.

Cell Lysis for Preparation of Nuclear Lysate

Nuclear lysate from 400 million cells was prepared as previously described[33], with the following modification (Supplemental, B). Sonication was performed with a Branson SFX 250 Digital Sonifier®. A nuclear lysate protocol was selected rather than whole cell lysate due to the fact that HSATII RNA has been shown to be present in the nucleus but not in the cytoplasm[28].

Optimizing Amount of Probe used in RAP Protocol

After selecting Probe 1 as the probe we would use in the RAP protocol, the correct probe concentration to use was determined. 30uL of probe per 200 million cells, or 1.5 times more than the amount of probe McHugh et. al. used, was found to be the optimal concentration, as using 20uL of probe per 200 million cells did not yield as much HSATII RNA, while using 40uL of probe surprisingly also yielded less HSATII RNA. Thus, going forward, when scaling the protocol up or down Probe 1 was used at a concentration of 30uL probe per 200 million cells.

Captures Protocol

Captures using a biotinylated HSATII probe were performed as described (Supplemental, C) from 400 million cells. RNA was stored at -80°C prior to RT-qPCR.

Protein Precipitation, SDS-PAGE Gel and Submission for Mass Spectrometry

Protein was precipitated as described (Supplemental, D) and pellet was immediately resuspended in 15uL of 2X Laemmli Buffer (BIO-RAD, 161-0737) with 5% 2-mercaptoethanol (BME), by adding liquid, dislodging pellet with P2 pipette tip, vortexing and pipetting up and down until pellet was no longer visible. This solution was loaded into an SDS-PAGE (BIO-RAD, Mini-Protean® TGX™, 456-1096) gel along with 15uL of ladder (BIO-RAD, Precision Plus Protein™ Dual Color Standards, 1610374) in a separate lane. Gel was run at 300 volts in 1X TBE until dye front was about two thirds of the way down the gel. Gel was stained and destained as described (Supplemental, E). Gel was then stored in water and transported to the Quantitative Proteomics Resource Core at The University of Pennsylvania, where two bands were cut out of gel (one upper and one lower), trypsin digested and submitted to mass spectrometry for protein identification. ThermoFisher Scientific Q-Exactive™ HF-X mass spectrometer was used. Proteome Discoverer (Software Version 2.2, XCALI-97808) was used to identify proteins by peptide mass fingerprinting.

Reverse Transcription – quantitative Polymerase Chain Reaction (-PCR)

Reverse Transcription - quantitative Polymerase Chain Reaction (RT-qPCR) was performed to confirm that protocol was selectively pulling HSATII RNA. HSATII, α-SAT, MALAT1 and β-actin primer sets were used (Table S1). Cycling conditions were as follows: 95°C 3 min initial denaturation followed by 40 cycles of, (95°C 10 sec, 58°C 10 sec, 72°C 20 sec). Relative abundance was calculated by using the formula $2^{(40-C(t)),40}$.

Mass Spectrometry Data Analysis

Upon receiving mass spectrometry data from the Quantitative Proteomics Resource Core at the University of Pennsylvania, the unique known proteins between the upper and lower bands were identified. Out of 405 combined proteins identified in the upper and lower bands (Table S3), 268 unique proteins were identified. Contaminant proteins that appear in the majority of mass spectrometry data

sets, including various keratin proteins, actin and serum albumin were then eliminated from the data set[41]. Ten contaminant proteins were removed from the data set in total, prior to further analysis outlined in Fig 7.

Results

Selection of Protocol to Adapt for Identification of HSATII RNA Binding Proteins

The RNA Antisense Purification with Mass Spectrometry (RAP-MS) approach established by McHugh et. al. to identify Xist RNA binding proteins[8,33] was adapted to identify binding proteins of HSATII RNA. The elements that set this protocol apart from other protocols considered are the use of *in-vivo* UV-crosslinking to form covalent bonds between only directly interacting RNAs and proteins and the capture of the RNA of interest using long biotinylated probes, which allows for the use of harsh denaturing conditions for RNA purification.

Probe Design and Predicted Self-Hybridization

Biotinylated probes were designed with the goal of targeting as broad an array of expressed HSATII sequences as possible. Based on the much smaller size of the HSATII RNA transcripts (250-600bp), compared to Xist RNA (17kb), we reasoned a single 90nt biotinylated probe should be sufficient to capture HSATII RNA transcripts. In order to effectively target expressed HSATII variants, fifteen distinct HSATII RNA transcripts that had been previously identified following RIP with MeCP2 and Sin3A antibody[28] were aligned using MacVector nucleic acid alignment tool, in order to determine regions with minimal base pair mismatch (Fig. 2). From this, we designed two probes that were 90 nucleotides in length (Table S1). Probe 1 represented the exact sequence of a portion of these transcripts, in case HSATII RNA is also expressed from the opposite strand of these genomic loci, while Probe 2 was the complement of the selected 90nt aligned region (Fig. 2). Predicted secondary structure of both probes was analyzed to determine whether self-hybridization might affect the availability of each probe

to hybridize with HSATII RNA in solution. Based on these predicted hybridization patterns, it is unlikely that Probe 1 folds into a secondary structure that would reduce hybridization efficiency, while Probe 2 likely folds into a secondary structure that would negatively affect hybridization efficiency (Fig. 3).

HSATII RNA-specific Probe Validation

After designing the two probes and predicting potential self-hybridization and secondary structure formation, the probes were validated for hybridization by HSATII RNA FISH. RNA fluorescence *in situ* hybridization (FISH) revealed that Probe 1 robustly hybridized with HSATII RNA foci, while Probe 2 did not (Fig. 4), further confirming our prediction of self-hybridization (Fig 3). When Probe 1 was used for hybridization, about one third of U20S nuclei (37 out of 104 or about 35%) had accumulations similar to those observed when a global locked nucleic acid (LNA, Table S1), known to target a broad range of HSATII transcripts[28], was used as a probe. When Probe 2 was used for hybridization, only a small percentage of U2OS nuclei (2 out of 100 or 2%) contained RNA accumulations. When both Probe 1 and Probe 2 were combined and used for hybridization, hybridization with HSATII RNA was not significantly different than hybridization with Probe 1 alone (data not shown). Based on these results Probe 1, but not Probe 2, was used for the capture of HSATII transcripts for RAP-MS (Fig. 1A).

Probe 1 selectively enriches for HSATII RNA

Confirmation that HSATII RAP was selectively pulling down HSATII RNA was analyzed using Reverse Transcription – quantitative Polymerase Chain Reaction (RT-qPCR) to determine the relative levels in the elution (SE) and flow-through samples (SF-T) of HSATII RNA and three control RNAs: α-satellite, which is a pericentromeric satellite repeat sequence with a well-defined 170bp monomer sequence, MALAT1, a 7kb lncRNA expressed from a single genomic locus and β-actin, which is a highly abundant structural protein also expressed from a single genomic locus. RT-qPCR of SF-T and

elution SE using HSATII, α-SAT, MALAT1 and β-actin primer sets (Table S1) confirmed enrichment for HSATII, but not other RNAs, in the elution sample (Fig. 5). The three control transcripts, α-SAT, MALAT1 and β-actin, were reduced in the SE sample compared to SF-T (Fig 5B-D), while amplification of HSATII transcripts was increased in the SE sample compared to SF-T (Fig 5A). This indicated that the level of non-HSATII transcripts was very low in the elution sample and thus that the adapted RAP-MS protocol selectively enriches for and captures HSATII transcripts.

RAP-MS Does Not Identify Only the Most Abundant Proteins in U2OS Cells

After confirming that Probe 1 selectively pulls down HSATII RNA, the entire RAP protocol was performed, including preparation of a protein sample for submission to a mass spectrometry facility (Fig. 6). Upon receiving the results from the mass spectrometry facility, a procedure for arriving at a candidate list of proteins was developed to analyze the mass spectrometry data (Fig. 7).

To confirm we were not simply identifying the most abundant proteins in U2OS cells, as this would indicate that the protocol may not be selective for HSATII RBPs, we compared the collection of the most abundant proteins in U2OS cells[42] to the list of HSATII RNA binding proteins identified by mass spectrometry. It was determined that Probe 1 enriches for more abundant proteins, but not solely highly abundant proteins in U2OS cells (Fig. 8). This is demonstrated by 1) enrichment in the pulldown sample for proteins among the top 50% most expressed proteins, 2) pulldown proteins were not just within the top 10 or 20% of the most expressed proteins, 3) identification of several proteins of low abundance (Fig. 8).

Pulldown Significantly Enriches for RNA-binding Proteins

After determining that RAP-MS does not solely select for the most abundant U2OS proteins, we next wanted to confirm that the identified proteins were enriched for RBPs previously identified by other methods[9]. Capture enriches for RNA-binding proteins, with a larger percentage of proteins identified by mass spectrometry from the pulldown sample being classified mRNA interactome proteins

or candidate RBPs than the percentage of all proteins in HeLa cells that are in the mRNA interactome or are candidate RBPs (Fig. 8, Table S2, 33% of pulldown proteins were found to be in the mRNA interactome or were classified as candidate RBPs[9], while 15% of all HeLa proteins identified by Nagaraj et. al., 2011[43] were found to be in the mRNA interactome or were classified as candidate RBPs[9]).

GO Analysis Confirms RBP Enrichment and Identifies Enriched Proteins Involved in Processes Misregulated in Cancer

Following this confirmation, Gene Ontology (GO) analysis was performed to further confirm that the pulldown data set was enriched for RNA binding proteins, as well as to identify enriched processes within the protein list generated HSATII RAP (Table 1). GO analysis of all 258 proteins identified in pulldown sample indicated that RNA binding and nuclear proteins were significantly enriched in the pulldown sample, with nearly 6-fold more RNA binding proteins and nearly 2-fold more nuclear proteins identified than expected based on abundance (Table 1A). Several processes that are disrupted in cancer were also significantly enriched (Table 1B).

Identification of candidate protein list

We next used this list to generate a curated list of highest interest candidate HSATII RBPs involved in processes that are misregulated in cancer (Table 2). Candidate proteins were identified from the set of identified proteins included under the following Gene Ontology (GO) terms: gene silencing by RNA, chromatin binding, RNA helicase activity and positive regulation of DNA binding (Table 1B). Other identified proteins that were abundant in the sample (Table S3) and had functions that are misregulated in cancer were also added to the candidate list (Table 2).

Figures

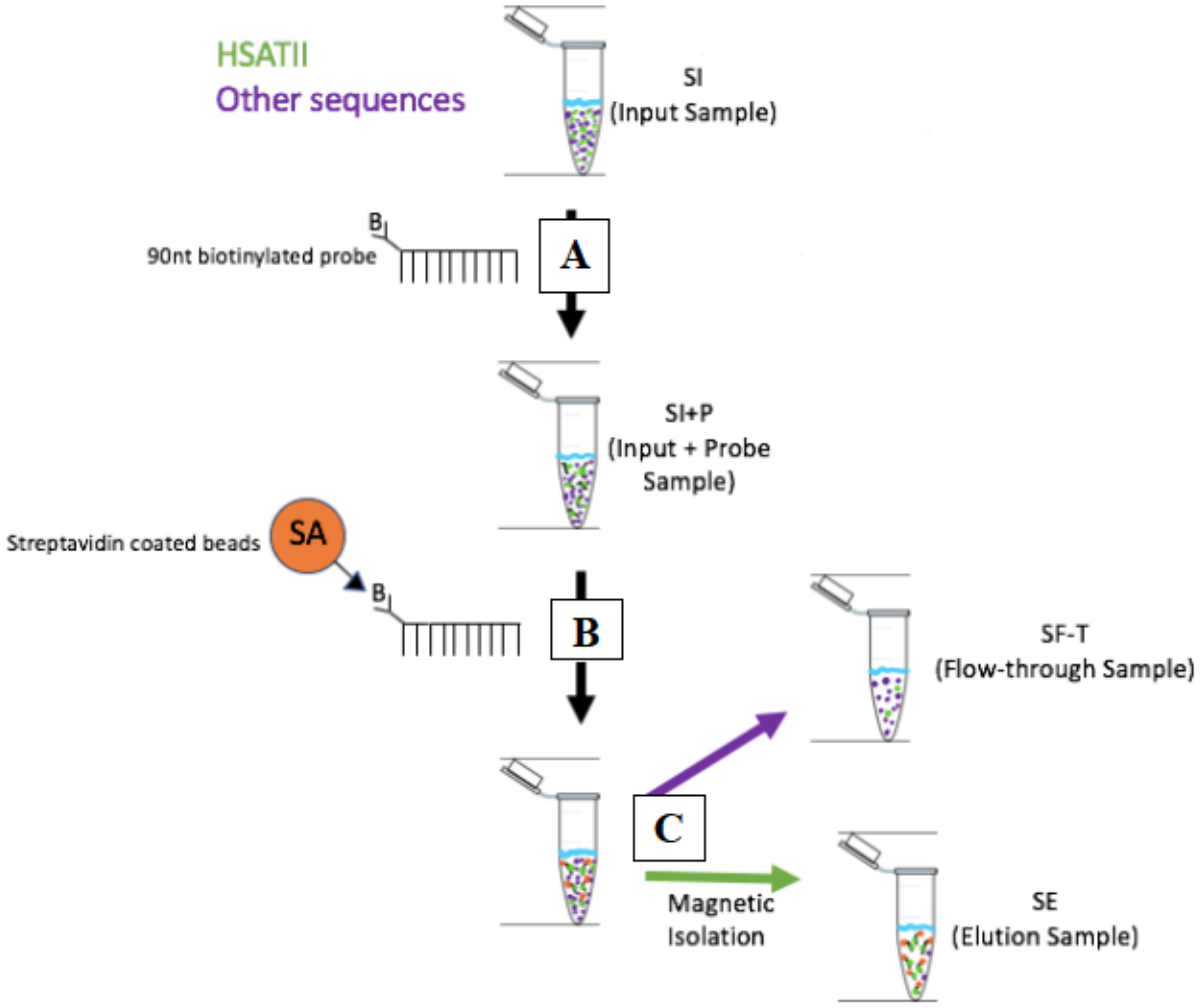
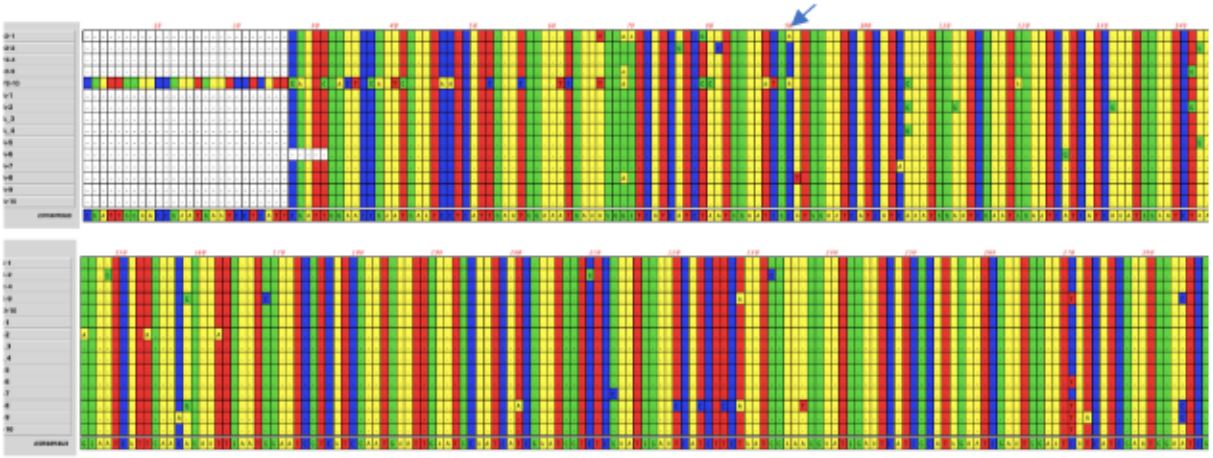


Figure 1. Overview of HSATII RAP-MS protocol starting from nuclear lysate. Input Sample (SI) is composed of nuclear cell lysate after crosslinking. Major steps include hybridization with 90nt biotinylated probe (A), where "B" represents a biotin molecule attached to 5' end of probe, capture of probe-RNP complexes with streptavidin coated magnetic beads (B, beads shown with orange dots), where "SA" represents the streptavidin protein coating the magnetic beads, and isolation of streptavidin-coated beads bound to probe-RNP complexes using magnetic isolation (C). Elution Sample (SE) is expected to be enriched for HSATII RNA (shown with green dots) and contain few other sequences (shown with purple dots), while Flow-through Sample (SF-T) is expected to be enriched for other sequences and contain lower abundance of HSATII RNA.



Fi **(RIPs)**
wi ps
allowed in the alignment, was used. Two 90nt regions with minimal base differences were ordered as Probe 1 and Probe 2. Probe 1 spans aligned bases 180 through 269, while Probe 2 starts at base 90 and ends at base 179. Probe 1 was the exact 90nt sequence indicated, while Probe 2 was the complement of the 90nt sequence shown.

Figure 3. Predicted RNA folding patterns for HSATII RAP probes. Secondary structure prediction generated using RNAbows program (http://rna.williams.edu/rnabows/single.html). Darkness of semicircle indicates the likelihood of hybridization between connected bases occurring. It is unlikely that Probe 1 folds into a secondary structure that would limit hybridization, while Probe 2 likely folds into a secondary structure that limits hybridization, due to the most likely hybridization occurring in the center of the probe sequence and thus not leaving significant bases free to hybridize with HSATII RNA in the lysate.

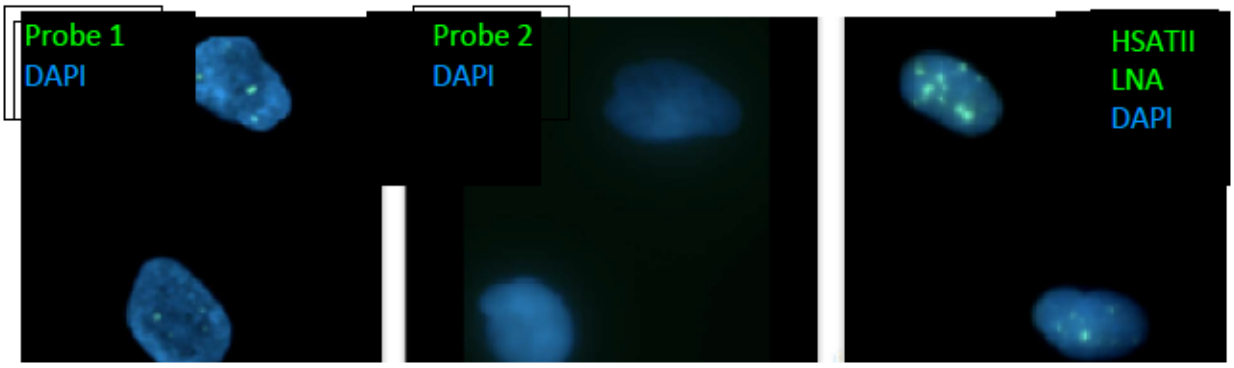


Figure 4. *In situ* validation of HSATII RAP probes. RNA Fluorescence *In Situ* Hybridization (FISH) of Probe 1 (left), Probe 2 (middle) compared to HSATII LNA (right). Biotinylated Probe 1(left) hybridizes to accumulated HSATII RNA foci, as seen by the formation of foci similar to those seen when an LNA known to hybridize to HSATII transcrips was used, in about a third of U2OS nuclei, whereas biotinylated Probe 2 hybridizes in a very low percentage of cells, with the majority showing no HSATII foci.

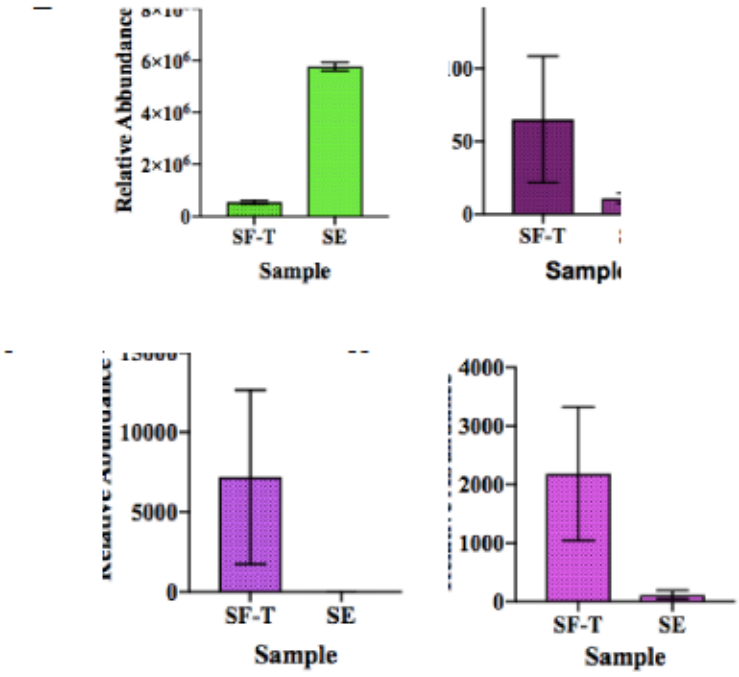


Figure 5. RNA antisense purification protocol selectively enriches for HSATII RNA. Results of RT-qPCR in elution (SE) and flow-through (SF-T) and elution (SE) samples using HSATII (A), **α**-SAT (B), MALAT1 (C) and β-actin (D) primer sets. Relative abundance of three control transcripts, **α**-SAT, MALAT1 and β-actin, was much lower in elution sample than in flow-through sample, whereas the relative abundance of HSATII transcripts was much higher in elution sample than in flow-through sample. Relative abundance was calculated by using the formula $2^{(40-C(t))}$. Means of technical triplicates are graphed, with error bars representing standard deviation.

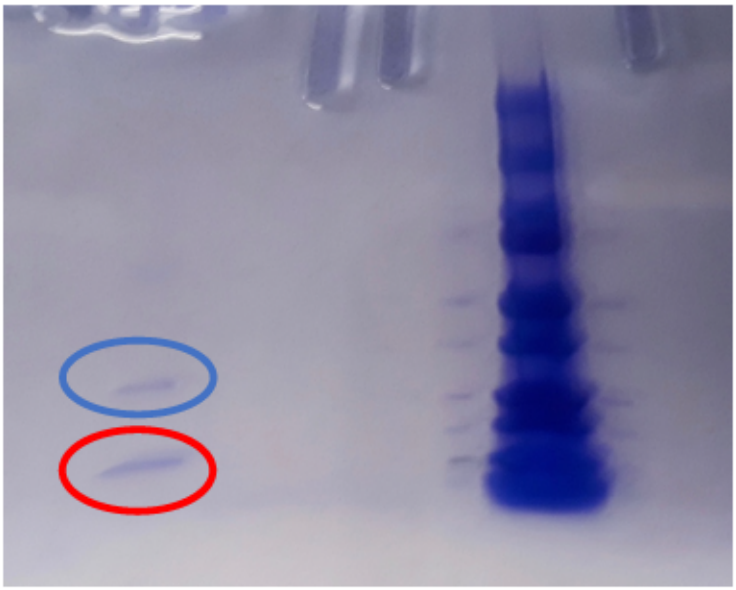


Fig 6. Two protein bands were submitted for mass spectrometry following HSATII RAP. SDS-PAGE gel showing the two bands of protein (red and blue ovals) that were submitted to the Quantitative Proteomics Resource Core at the University of Pennsylvania. 159 proteins were identified from the bottom band (red oval), while 246 proteins were identified from the top band (blue oval). Of these 405 combined identifications in the bottom and top band, 268 proteins were found to be unique.

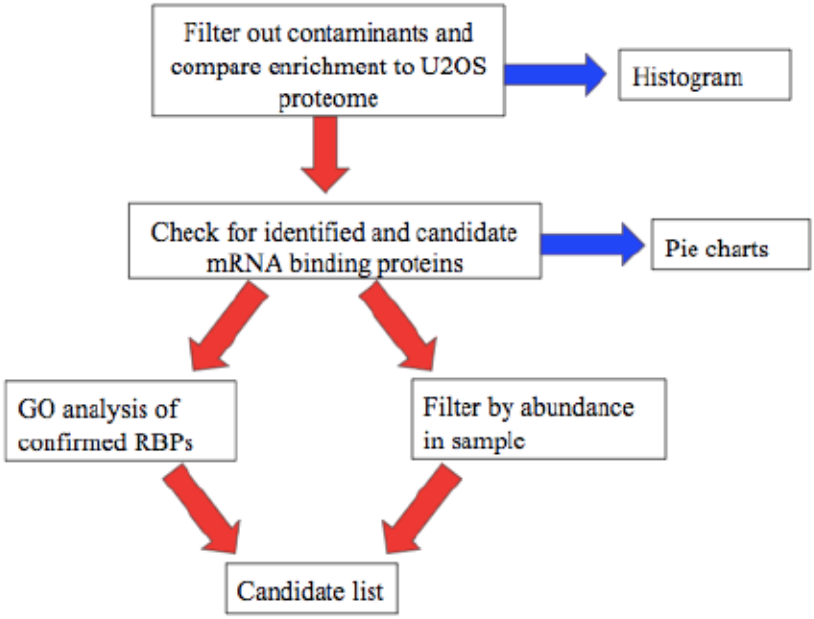


Figure 7. Overview of approach developed to analyze mass spectrometry data. Goals of this pipeline were to 1) filter out common contaminant proteins, 2) compare to the most abundant proteins in U2OS cells and 3) compare to known RBPs in order to generate a candidate list of proteins to investigate further. Results of individual steps within this approach are shown in the following figures.

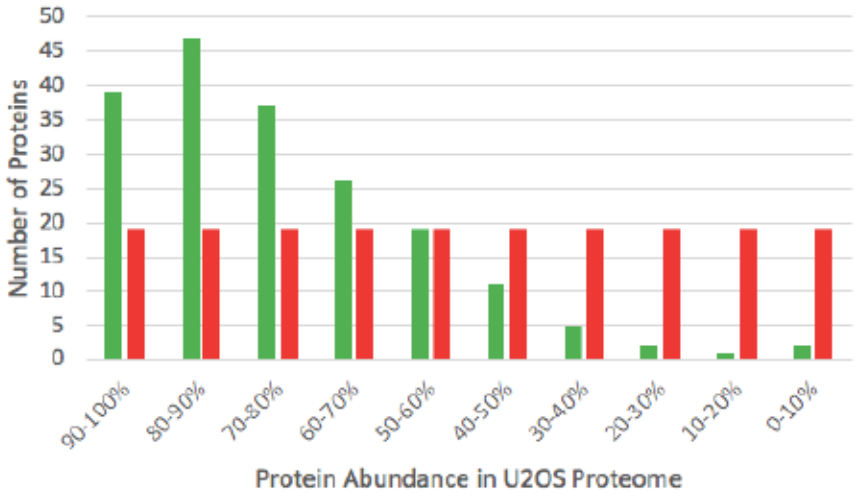


Figure 8. HSATII RAP does not enrich for only most abundant proteins in U2OS cells. Histogram compares the expected distribution (assuming even distribution across ten deciles) of 190 proteins out of 258 total identified proteins present in both the HSATII RAP sample and reference U20S proteome[42] and the actual distribution of this subset of the 258 total non-contaminant identified proteins in pulldown sample. Lack of overlap of all 258 identified proteins with U2OS proteome is likely due to the limitations of whole cell mass spectrometry in identifying less abundant proteins. 190 proteins from HSATII RAP were compared against the top 10%, 10-20%, 20- 30%, etc. of the U2OS proteome to determine the number of proteins from HSATII RAP sample present in each decile.

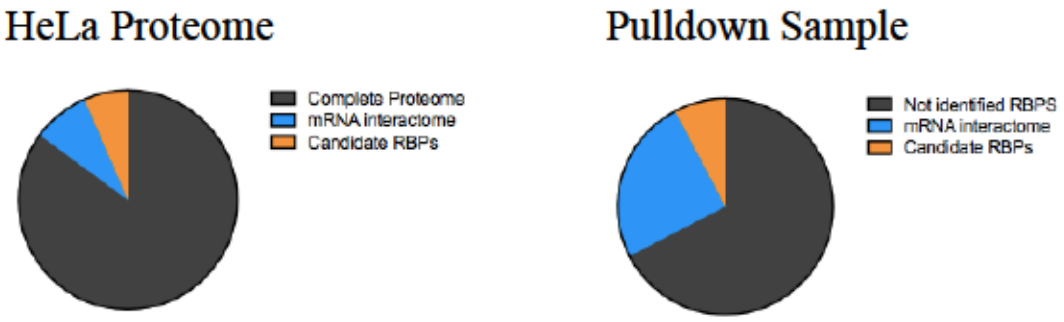


Figure 9. HSATII RAP (pulldown) sample is enriched for RNA binding proteins. The HeLa proteome was used here because a systematic identification of RNA binding proteins has not been done in U2OS cells to our knowledge. The number of mRNA interactome and candidate RPBs in the pulldown sample was determined by comparing the number of these proteins present in the HSATII RAP data set as well as the Castello et. al.[9] data set, which comprises a complete list of RBPs in a HeLa cell line. All of these identified RBPs were included as being in the mRNA interactome or being candidate RBPs within a complete HeLa proteome[43].

Table 1. HSATII RAP-MS proteins are enriched for RNA-binding function, nuclear localization and biological processes involved in cancer. Gene Ontology analysis was done on both all 258 proteins in pulldown sample (A) and just the 84 proteins in pulldown sample categorized as mRNA interactome or candidate RBPs (B). Significant enrichment for RNA-binding and nuclear proteins, with extremely low P-values indicating significance of enrichment, helped confirm that protocol was specific, and that the nuclear lysis protocol was effective. Proteins grouped under significantly enriched biological processes involved in cancer were investigated further to help generate a curated candidate list of highest interest proteins to further explore.

A

Gene Ontology Term	X Enrichment	P-value
RNA Binding	5.8	7.1×10^{-59}
Nucleus	1.97	2.27×10^{-28}

B

Gene Ontology Term	X Enrichment	P-value
Gene Silencing By RNA	20.1	8.44×10^{-4}
Chromatin Binding	4.0	3.34×10^{-4}
RNA Helicase Activity	30.9	2.21×10^{-4}
Positive Regulation of DNA Binding	17.0	1.15×10^{-4}

Table 2. Curated candidate list of proteins with functions involved in processes that are misregulated in cancer. Proteins were identified either by looking through proteins grouped under relevant Gene Ontology terms or by filtering sample proteins by abundance and scanning for relevant functions. Proteins were checked against reference list of confirmed and candidate RNA binding proteins[9] to determine whether they had RNA binding capability. Functions are primarily from the UniProt (https://www.uniprot.org/) page for each protein.

Protein	Identified by	RBP?	Functions
DHX9	GO Term: RNA helicase	mRNA Interactome	ATP dependent RNA helicase with various roles including unwinding DNA:RNA hybrids.
SND1	GO Term: Gene Silencing by RNA	mRNA Interactome	Endonuclease that regulates miRNAs involved in G-to-S phase transition.
SUPT16H	GO Term: Chromatin binding	mRNA Interactome	FACT (facilitates chromatin transactions) complex subunit: general chromatin factor that acts to reorganize nucleosomes.
PARP1	GO Term: Positive Regulation of DNA Binding	mRNA Interactome	Polymerase that plays a key role in DNA repair. Also involved in differentiation, proliferation and tumor transformation
TRIM28	Abundance	No, but other TRIM proteins are	Transcription intermediate factor that coordinates increases in H3K9me decreases in H3K9ac and H3K14ac. Also Deposits HP1.
MCM2	Abundance	No, but MCM3AP is candidate RBP	DNA replication licensing factor that is component of the MCM2-7 complex, which is helicase involved in DNA replication. Required for entry into S phase and in cell division.
MCM4	Abundance	(See above)	Like MCM2, DNA replication licensing factor that is a component of the MCM2-7 complex.
DDB1	Abundance	No	DNA damage binding protein that is required for DNA repair.
HNRNPU	Abundance	mRNA Interactome	Heterogenous nuclear ribonuclear protein involved in several cellular processes including nuclear chromatin organization and mitotic cell progression.
SIN3A	RNA Immunoprecipitation	No	Transcriptional regulator that mediates repression and binds to MeCP2.

Discussion

This work achieved three primary goals. First, a protocol that selectively enriches for HSATII RBPs from a nuclear lysate was established. This included designing a probe that hybridizes efficiently with HSATII RNA (Figs. 2, 4), confirming that the elution sample contained HSATII RNA but not other RNAs (Fig. 5), and running an SDS-PAGE gel (Fig. 6) to be submitted for mass spectrometry analysis in order to identify proteins. Second, a pipeline for analyzing candidate HSATII RBPs was developed (Fig. 7). Known contaminant proteins were filtered out and it was confirmed that identified proteins were not solely the most abundant proteins in cancer cells (Fig. 8) and were enriched for RNA-binding proteins (Fig. 9). Finally, a curated list of candidate HSATII RBPs for further investigation was generated (Table 2) by looking into proteins grouped under enriched processes that are misregulated in cancer (Table 1B), as well as abundant proteins in the sample (Table S3). The achievement of these three goals will be discussed in detail below.

RAP-MS[33] adapted for HSATII RNA proved to be an effective protocol for selectively isolating HSATII RNA and bound proteins. First, a probe intended to hybridize with a range of sequenced HSATII transcripts was designed (Fig. 2) and the probes ability to hybridize with HSATII RNA was validated *in vivo* (Fig. 4). It was next confirmed that the elution sample was enriched for HSATII RNA and not other satellite, non-coding or mRNAs (Fig. 5), which indicated that the proteins present in the sample prior to identification were proteins bound to HSATII RNA and not proteins bound to other RNAs.

The pipeline for analyzing candidate HSATII RBPs developed here was effective in confirming that the protocol was identifying RNA binding proteins and narrowing down the number of proteins of interest, ultimately leading to a curated candidate list (Table 2). Following submission of an SDS-PAGE gel (Fig. 6) for analysis by mass spectrometry, a system for analyzing the 268 identified candidate HSATII RBPs was developed. The goals of this pipeline were to confirm that the protocol was

successful in identifying RNA binding proteins and to generate a curated list of candidate HSATII RBPs (Fig. 7). The first step in this pipeline was filtering out contaminant proteins that appear in the majority of mass spectrometry samples, regardless of assay or cell type[41]. Next, identified proteins were compared against a reference U2OS proteome[42] in order to confirm that identified proteins were not simply the most abundant proteins present in U2OS cells and instead were likely HSATII RBPs (Fig. 8). Then, identified proteins were grouped as either RNA-binding or not based on a previous systematic identification of RBPs in another cell line[9] in order to confirm that the sample was enriched for RBPs (Fig. 9). The fact that not only the most abundant proteins present in U2OS cell were identified and that the identified proteins were enriched for RNA binding confirmed that protocol used was effective in identifying HSATII RBPs.

Gene Ontology (GO) analysis was then performed to further confirm that candidate HSATII RBPs were enriched for RNA-binding capability and nuclear localization (Table 1A), as well as to identify enrichment of proteins involved in processes that are misregulated in cancer (Table 1B). From this set of proteins involved in processes that are misregulated in cancer and enriched for in the sample, proteins were added to a curated candidate list of highest interest HSATII RBPs (Table 2). Along with selecting proteins from distinct processes misregulated in cancer that were enriched for in the pulldown sample, more abundant proteins in the sample (Table S3) with functions that are misregulated in cancer were added to the curated candidate list (Table 2).

The identification of the protein SIN3A in the pulldown sample provided further confirmation that the adapted RAP-MS protocol was effective in identifying HSATII RBPs, as a previous RNA immunoprecipitation (RIP) with SIN3A had indicated that HSATII RNA binds to this protein[28]. SIN3A is a transcriptional repressor responsible for repressing transcription in concert with histone deacetylation[44], as well as repressing genes involved in cell cycle progression[45]. SIN3A is also a known binding partner of MeCP2[28], a protein not identified in the pulldown sample. This suggests that SIN3A,

and not MeCP2, may bind HSATII directly, and that MeCP2 is recruited following SIN3A binding. This is supported by the fact the crosslinking step of the RIP that showed that MeCP2 binds HSATII RNA was performed using formaldehyde[28], which has been shown to induce protein-protein interactions as well as RNA-protein interactions[36]. This implies that a RIP for MeCP2 could have pulled down HSATII RNA even if MeCP2 itself was not directly bound to HSATII RNA, if instead MeCP2 was complexing with SIN3A bound to HSATII RNA. Performing a protein-centric technique, such as crosslinking and immunoprecipitation[11], that uses UV crosslinking rather than formaldehyde crosslinking would be useful in testing the hypothesis stated here that SIN3A, rather than MeCP2, directly binds HSATII RNA.

Additional controls should be performed to confirm that the candidate HSATII RBPs bind HSATII RNA specifically, as it is likely that many of the candidate HSATII RBPs identified have the capacity to bind a broad range of transcripts[2]. This should be done by performing RAP-MS using a probe antisense to U1 RNA, a spliceosome RNA, as well as a probe targeting a ribosomal RNA[8], in order to eliminate binding proteins that are components of highly abundant RNPs from the list of proteins that selectively bind HSATII. The use of control probes could be coupled with a quantitative mass spectrometry approach, such as the use of stable isotope labeling by amino acids (SILAC), to provide sensitive quantification of proteins bound to HSATII and control transcripts in order to determine the proteins that are significantly enriched for HSATII binding[8]. A no-probe control should also be performed to confirm that only proteins bound to HSATII RNA are present in the elution sample and that proteins not bound to HSATII RNA are not being identified. It is also possible that by taking an approach that uses photoreactive nucleoside-enhanced UV-crosslinking[10], crosslinking efficiency could be improved and proteins could be identified from a sample starting with many fewer cells than the 400 million required to identify proteins using the current RAP-MS protocol. This would make performing duplicate samples, as well as including controls, more feasible.

Improved technology and additional background studies would help to improve the accuracy of the pipeline developed here in generating a curated candidate list of HSATII binding proteins. Improved whole-cell mass spectrometry techniques could provide a more complete U2OS proteome than the one used currently to determine abundance of pulldown proteins, as more than a fourth of pulldown proteins were not present in the reference U2OS proteome used[42]. Performing a systematic identification of RBPs in U2OS cells would provide a better reference RBP list than the list currently being compared against to determine enrichment of RBPs, as it is likely that a different set of proteins bind RNAs in U2OS cells than in HeLa cells[2,9,10]. More generally, identifying the RBPs that have the capacity to bind any satellite RNA would be valuable, as it is possible that a large number of proteins have the capacity to bind satellite RNA sequences, and that these may be distinct from mRNA binding proteins identified previously[9,10]. Thus, the large percentage of proteins not classified as RBPs in the pulldown sample, while still a considerably smaller percentage than the percentage of proteins not classified as RBPs among all proteins in HeLa cells, may actually be reduced due to the unknown capability of proteins to bind satellite sequences.

Among the curated candidate list of HSATII RBPs are several proteins involved in processes that are misregulated in cancer. Under the "RNA helicase" Gene Ontology (GO) term, the nuclear protein DHX9 is of interest, as this protein functions to unwind DNA:RNA hybrids[46], which have been shown to form with HSATII RNA in cancer cells[30] and may contribute to HSATII expansion[47]. DNA:RNA hybrids have also been implicated in genome instability through the formation of mutagenic R loop structures[48]. Researchers have suggested testing these mechanisms in cancer models to determine the way in which they contribute to disease progression[48]. Thus, if DHX9 is unable to perform its function of unwinding DNA:RNA hybrids as a result of binding to HSATII RNA in cancer cells, R loop structures may persist, potentially leading to genome instability, increased HSATII expression, and further sequestration of proteins through repeat expansion.

The nuclear protein, SUPT16H, under the "Chromatin binding" GO term, is of interest due to this protein's presence as a subunit of the FACT (facilitates chromatin transactions) complex[49]. The FACT complex functions to reorganize nucleosomes by acting as a chaperone for histone proteins and has the capacity to both increase and decrease the stability of nucleosome structure[49]. If SUPT16H function is affected by binding to HSATII RNA, then this could contribute to the misregulation of DNA accessibility seen in cancer[50]. While TRIM28 is a protein not categorized under any of the four GO terms of interest and has not previously been identified as an RNA binding protein, TRIM28 was added to the curated candidate HSATII RBP list. This protein, which was abundant in the sample, coordinates changes in epigenetic regulation (specifically histone acetylation and methylation)[51], which are processes widely misregulated in cancer cells[52]. The presence of TRIM28 in the pulldown sample suggests that previously unidentified RBPs may bind HSATII RNA and that DNA accessibility and epigenetic regulation may be two processes affected by proteins binding to HSATII RNA in cancer cells.

SUPT16H is also of interest because of the finding that the FACT complex is likely involved in the phosphorylation, and activation of, the tumor suppressor p53 in response to cellular stress[53]. Changes in the activity of p53 have been shown to have large effects on cancer progression, as without an active form of this tumor suppressor, cancer has been shown to progress more rapidly[54]. If the FACT complex is not able to phosphorylate p53 due to the lack of, or presence of a non-functional SUPT16H protein (due to HSATII RNA binding and sequestration), this could have profound effects on a key mechanism necessary to halt cancer progression.

While not grouped under one of the significantly enriched GO terms, HNRNPU was abundant in the sample and has functions of interest. HNRNPU is required for localization of Xist RNA[55] and also plays a critical role in the regulation of chromosome structure[56]. When HNRNPU is depleted Xist RNA is detached from the X chromosome and localization becomes diffuse throughout the nucleoplasm[55]. Thus,

HNRNPU is an example of an RBP required for proper localization of a specific ncRNA and may be required for the localization of other ncRNAs as well. The identification of HNRNPU among the pulldown proteins suggests that this protein may have the capacity to bind a broad range of ncRNAs. In addition to being required for the localization of Xist RNA to the X chromosome, HNRNPU is required for maintaining chromosome folding through interactions with chromatin-associated RNAs (caRNAs)[56]. When HNRNPU is depleted, abnormal chromosome folding, leading to the accumulation of genome damage, occurs[56]. A model has emerged in which HNRNPU and caRNAs form an active net-like structure that is required for organizing chromosome structure and maintaining genome stability[56]. The binding of HNRNPU to Xist and caRNAs to perform divergent functions indicates the important role of RNA in mediating protein localization and function.

Of the proteins in the curated candidate list, it should be determined which of these are significantly enriched for HSATII RNA binding before proceeding with confirmation of binding relationships and functional assays. Prior to investigation of the function of the highest interest candidate HSATII RBPs, binding relationships should be confirmed by additional methods. First, antibodies against proteins of interest should co-hybridized with HSATII in intact nuclei to determine if candidate proteins and HSATII RNA colocalize in the nucleus. If colocalization is confirmed, then RNA immunoprecipitation, using antibodies against candidate proteins, followed by PCR or sequencing should be performed to determine if there is a direct binding relationship. Western blots of the SDS page gel that could be submitted for analysis by mass spectrometry should also be done to confirm the presence of identified proteins. **Pulldown of HSATII RNA with the probe developed here, followed by RNA sequencing, has the potential to improve our understanding of the HSATII transcriptome, as we do not fully understand the range of HSATII RNA transcripts being expressed.** For investigating function further, determining the RNA binding domains of candidate HSATII RBPs with previously identified domains, as well as identifying the RNA binding domains of candidate HSATII RBPs with unidentified domains would be a

good next step for determining how binding with HSATII RNA might impact the function of these proteins. Determining if the localization of the identified proteins changes when bound to HSATII would another worthwhile follow up study. The fact that in cancer cells proteins with diverse functions are binding HSATII RNA, an RNA not present in normal cells, indicates that protein function is likely affected by this binding relationship and warrants further study.

References

1. Gerstberger, S.; Hafner, M.; Tuschl, T., A census of human RNA-binding proteins. *Nature Reviews Genetics* **2014,** *15* (12), 829-845.
2. Beckmann, B. M.; Castello, A.; Medenbach, J., The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. *Pflugers Archiv-European Journal of Physiology* **2016,** *468* (6), 1029-1040.
3. Shi, Y., Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol* **2017,** *18* (11), 655-670.
4. Guttman, M.; Amit, I.; Garber, M.; French, C.; Lin, M. F.; Feldser, D.; Huarte, M.; Zuk, O.; Carey, B. W.; Cassady, J. P.; Cabili, M. N.; Jaenisch, R.; Mikkelsen, T. S.; Jacks, T.; Hacohen, N.; Bernstein, B. E.; Kellis, M.; Regev, A.; Rinn, J. L.; Lander, E. S., Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **2009,** *458* (7235), 223-7.
5. Clemson, C. M.; Hutchinson, J. N.; Sara, S. A.; Ensminger, A. W.; Fox, A. H.; Chess, A.; Lawrence, J. B., An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Molecular Cell* **2009,** *33* (6), 717-726.
6. Rinn, J. L.; Kertesz, M.; Wang, J. K.; Squazzo, S. L.; Xu, X.; Brugmann, S. A.; Goodnough, L. H.; Helms, J. A.; Farnham, P. J.; Segal, E.; Chang, H. Y., Functional demarcation of active and silent chromatin domains in human HOX loci by Noncoding RNAs. *Cell* **2007,** *129* (7), 1311-1323.
7. Rinn, J. L.; Chang, H. Y., Genome Regulation by Long Noncoding RNAs. *Annual Review of Biochemistry, Vol 81* **2012,** *81*, 145-166.
8. McHugh, C. A.; Chen, C. K.; Chow, A.; Surka, C. F.; Tran, C.; McDonel, P.; Pandya-Jones, A.; Blanco, M.; Burghard, C.; Moradian, A.; Sweredoski, M. J.; Shishkin, A. A.; Su, J. L.; Lander, E. S.; Hess, S.; Plath, K.; Guttman, M., The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **2015,** *521* (7551), 232-+.
9. Castello, A.; Fischer, B.; Eichelbaum, K.; Horos, R.; Beckmann, B. M.; Strein, C.; Davey, N. E.; Humphreys, D. T.; Preiss, T.; Steinmetz, L. M.; Krijgsveld, J.; Hentze, M. W., Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **2012,** *149* (6), 1393-1406.
10. Baltz, A. G.; Munschauer, M.; Schwanhausser, B.; Vasile, A.; Murakawa, Y.; Schueler, M.; Youngs, N.; Penfold-Brown, D.; Drew, K.; Milek, M.; Wyler, E.; Bonneau, R.; Selbach, M.; Dieterich, C.; Landthaler, M., The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Molecular Cell* **2012,** *46* (5), 674-690.
11. Ule, J.; Jensen, K.; Mele, A.; Darnell, R. B., CLIP: A method for identifying protein-RNA interaction sites in living cells. *Methods* **2005,** *37* (4), 376-386.
12. Kawai, T.; Akira, S., The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nature Immunology* **2010,** *11* (5), 373-384.
13. Dabo, S.; Meurs, E. F., dsRNA-Dependent Protein Kinase PKR and its Role in Stress, Signaling and HCV Infection. *Viruses-Basel* **2012,** *4* (11), 2598-2635.
14. Beckmann, B. M.; Horos, R.; Fischer, B.; Castello, A.; Eichelbaum, K.; Alleaume, A. M.; Schwarzl, T.; Curk, T.; Foehr, S.; Huber, W.; Krijgsveld, J.; Hentze, M. W., The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. *Nature Communications* **2015,** *6*.

15. Chang, C. H.; Curtis, J. D.; Maggi, L. B.; Faubert, B.; Villarino, A. V.; O'Sullivan, D.; Huang, S. C. C.; van der Windt, G. J. W.; Blagih, J.; Qiu, J.; Weber, J. D.; Pearce, E. J.; Jones, R. G.; Pearce, E. L., Posttranscriptional Control of T Cell Effector Function by Aerobic Glycolysis. *Cell* **2013,** *153* (6), 1239-1251.
16. Zappulla, D. C.; Cech, T. R., RNA as a flexible scaffold for proteins: Yeast telomerase and beyond. *Cold Spring Harbor Symposia on Quantitative Biology* **2006,** *71*, 217-224.
17. Kung, J. T. Y.; Colognori, D.; Lee, J. T., Long Noncoding RNAs: Past, Present, and Future. *Genetics* **2013,** *193* (3), 651-669.
18. Guttman, M.; Donaghey, J.; Carey, B. W.; Garber, M.; Grenier, J. K.; Munson, G.; Young, G.; Lucas, A. B.; Ach, R.; Bruhn, L.; Yang, X. P.; Amit, I.; Meissner, A.; Regev, A.; Rinn, J. L.; Root, D. E.; Lander, E. S., lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **2011,** *477* (7364), 295-U60.
19. Maris, C.; Dominguez, C.; Allain, F. H. T., The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *Febs Journal* **2005,** *272* (9), 2118-2131.
20. Castello, A.; Fischer, B.; Frese, C. K.; Horos, R.; Alleaume, A. M.; Foehr, S.; Curk, T.; Krijgsveld, J.; Hentze, M. W., Comprehensive Identification of RNA-Binding Domains in Human Cells. *Molecular Cell* **2016,** *63* (4), 696-710.
21. Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B., Protein disorder prediction: Implications for structural proteomics. *Structure* **2003,** *11* (11), 1453-1459.
22. Vuzman, D.; Levy, Y., Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Molecular Biosystems* **2012,** *8* (1), 47-57.
23. Zhang, H. Y.; Elbaum-Garfinkle, S.; Langdon, E. M.; Taylor, N.; Occhipinti, P.; Bridges, A. A.; Brangwynne, C. P.; Gladfelter, A. S., RNA Controls PolyQ Protein Phase Transitions. *Molecular Cell* **2015,** *60* (2), 220-230.
24. Cooper, T.; Wan, L.; Dreyfuss, G., RNA and disease. Cell, 2009; Vol. 136, pp 777-793.
25. Morriss, G. R.; Cooper, T. A., Protein sequestration as a normal function of long noncoding RNAs and a pathogenic mechanism of RNAs containing nucleotide repeat expansions. *Human Genetics* **2017,** *136* (9), 1247-1263.
26. Lee, C.; Wevrick, R.; Fisher, R. B.; FergusonSmith, M. A.; Lin, C. C., Human centromeric DNAs. *Human Genetics* **1997,** *100* (3-4), 291-304.
27. Tagarro, I.; Fernandezperalta, A. M.; Gonzalezaguilera, J. J., CHROMOSOMAL LOCALIZATION OF HUMAN SATELLITE-2 AND SATELLITE-3 BY A FISH METHOD USING OLIGONUCLEOTIDES AS PROBES. *Human Genetics* **1994,** *93* (4), 383-388.
28. Hall, L. L.; Byron, M.; Carone, D. M.; Whitfield, T. W.; Pouliot, G. P.; Fischer, A.; Jones, P.; Lawrence, J. B., Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies. *Cell Reports* **2017,** *18* (12), 2943-2956.
29. Ting, D. T.; Lipson, D.; Paul, S.; Brannigan, B. W.; Akhavanfard, S.; Coffman, E. J.; Contino, G.; Deshpande, V.; Iafrate, A. J.; Letovsky, S.; Rivera, M. N.; Bardeesy, N.; Maheswaran, S.; Haber, D. A., Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. *Science* **2011,** *331* (6017), 593-596.
30. Bersani, F.; Lee, E.; Kharchenko, P. V.; Xu, A. W.; Liu, M.; Xega, K.; MacKenzie, O. C.; Brannigan, B. W.; Wittner, B. S.; Jung, H.; Ramaswamy, S.; Park, P. J.; Maheswaran, S.; Ting, D. T.; Haber, D. A., Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proceedings of the National Academy of Sciences of the United States of America* **2015,** *112* (49), 15148-15153.
31. Young, J. I.; Hong, E. P.; Castle, J. C.; Crespo-Barreto, J.; Bowman, A. B.; Rose, M. F.; Kang, D. C.; Richman, R.; Johnson, J. M.; Berget, S.; Zoghbi, H. Y., Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2 (vol 102, pg 17551, 2005). *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (5), 1656-1656.
32. Hite, K. C.; Adams, V. H.; Hansen, J. C., Recent advances in MeCP2 structure and function. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* **2009,** *87* (1), 219-227.
33. McHugh, C. A.; Guttman, M., RAP-MS: A Method to Identify Proteins that Interact Directly with a Specific RNA Molecule in Cells. *Methods Mol Biol* **2018,** *1649*, 473-488.
34. McHugh, C. A.; Russell, P.; Guttman, M., Methods for comprehensive experimental identification of RNA-protein interactions. *Genome Biol* **2014,** *15* (1), 203.
35. Darnell, R. B., HITS-CLIP: panoramic views of protein-RNA regulation in living cells. *Wiley Interdiscip Rev RNA* **2010,** *1* (2), 266-86.
36. Vasilescu, J.; Guo, X.; Kast, J., Identification of protein-protein interactions using in vivo cross-linking and mass spectrometry. *Proteomics* **2004,** *4* (12), 3845-54.
37. Poria, D. K.; Ray, P. S., RNA-protein UV-crosslinking Assay. *Bio Protoc* **2017,** *7* (6).

38. Domon, B.; Aebersold, R., Review - Mass spectrometry and protein analysis. *Science* **2006,** *312* (5771), 212-217.
39. Byron, M.; Hall, L. L.; Lawrence, J. B., A multifaceted FISH approach to study endogenous RNAs and DNAs in native nuclear and cell structures. *Curr Protoc Hum Genet* **2013,** *Chapter 4*, Unit 4.15.
40. Livak, K.; Schmittgen, T., Analysis of Relative Gene Expression Data Using RealTime Quantitative PCR and the 22^(-ddCT) Method. Elsevier: Methods, 2001; Vol. 25.
41. Hodge, K.; Ten Have, S.; Hutton, L.; Lamond, A. I., Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *Journal of Proteomics* **2013,** *88*, 92-103.
42. Beck, M.; Schmidt, A.; Malmstroem, J.; Claassen, M.; Ori, A.; Szymborska, A.; Herzog, F.; Rinner, O.; Ellenberg, J.; Aebersold, R., The quantitative proteome of a human cell line. *Molecular Systems Biology* **2011,** *7*.
43. 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.
44. Yang, X.; Zhang, F.; Kudlow, J. E., Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* **2002,** *110* (1), 69-80.
45. David, G.; Grandinetti, K. B.; Finnerty, P. M.; Simpson, N.; Chu, G. C.; Depinho, R. A., Specific requirement of the chromatin modifier mSin3B in cell cycle exit and cellular differentiation. *Proc Natl Acad Sci U S A* **2008,** *105* (11), 4168-72.
46. Chakraborty, P.; Grosse, F., Human DHX9 helicase preferentially unwinds RNA-containing displacement loops (R-loops) and G-quadruplexes. *DNA Repair (Amst)* **2011,** *10* (6), 654-65.
47. Younger, S. T.; Rinn, J. L., Silent pericentromeric repeats speak out. *Proc Natl Acad Sci U S A* **2015,** *112* (49), 15008-9.
48. Chan, Y. A.; Hieter, P.; Stirling, P. C., Mechanisms of genome instability induced by RNA-processing defects. *Trends in Genetics* **2014,** *30* (6), 245-253.
49. Orphanides, G.; LeRoy, G.; Chang, C. H.; Luse, D. S.; Reinberg, D., FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell* **1998,** *92* (1), 105-16.
50. Corces, M. R.; Granja, J. M.; Shams, S.; Louie, B. H.; Seoane, J. A.; Zhou, W.; Silva, T. C.; Groeneveld, C.; Wong, C. K.; Cho, S. W.; Satpathy, A. T.; Mumbach, M. R.; Hoadley, K. A.; Robertson, A. G.; Sheffield, N. C.; Felau, I.; Castro, M. A. A.; Berman, B. P.; Staudt, L. M.; Zenklusen, J. C.; Laird, P. W.; Curtis, C.; Greenleaf, W. J.; Chang, H. Y.; Network, C. G. A. A., The chromatin accessibility landscape of primary human cancers. *Science* **2018,** *362* (6413).
51. Nielsen, A. L.; Ortiz, J. A.; You, J.; Oulad-Abdelghani, M.; Khechumian, R.; Gansmuller, A.; Chambon, P.; Losson, R., Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *EMBO J* **1999,** *18* (22), 6385-95.
52. Kanwal, R.; Gupta, S., Epigenetic modifications in cancer. *Clin Genet* **2012,** *81* (4), 303-11.
53. Keller, D. M.; Zeng, X.; Wang, Y.; Zhang, Q. H.; Kapoor, M.; Shu, H.; Goodman, R.; Lozano, G.; Zhao, Y.; Lu, H., A DNA damage-induced p53 serine 392 kinase complex contains CK2, hSpt16, and SSRP1. *Mol Cell* **2001,** *7* (2), 283-92.
54. Maclaine, N. J.; Hupp, T. R., The regulation of p53 by phosphorylation: a model for how distinct signals integrate into the p53 pathway. *Aging (Albany NY)* **2009,** *1* (5), 490-502
55. Hasegawa, Y.; Brockdorff, N.; Kawano, S.; Tsutui, K.; Nakagawa, S., The matrix protein hnRNP U is required for chromosomal localization of Xist RNA. *Dev Cell* **2010,** *19* (3), 469-76.
56. Nozawa, R. S.; Boteva, L.; Soares, D. C.; Naughton, C.; Dun, A. R.; Buckle, A.; Ramsahoye, B.; Bruton, P. C.; Saleeb, R. S.; Arnedo, M.; Hill, B.; Duncan, R. R.; Maciver, S. K.; Gilbert, N., SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. *Cell* **2017,** *169* (7), 1214-1227.e18.
57. Hutton, J. R., RENATURATION KINETICS AND THERMAL-STABILITY OF DNA IN AQUEOUS-SOLUTIONS OF FORMAMIDE AND UREA. *Nucleic Acids Research* **1977,** *4* (10), 3537-3555.

Supplementary Material

Supplementary Protocols

A. Cell Harvesting and Crosslinking

1. Grow adherent cells on five 15cm tissue culture plates
2. Remove media from plate and replace with 10mL ice-cold PBS
3. Rock gently for 10 seconds then remove PBS wash
4. Add 10mL ice-cold PBS to plate
5. UV crosslink plate of cells in UV crosslinker at 254nm wavelength with 0.8 J/cm^2. Crosslink with plate cover off. Set our instrument to 800uJ
6. Remove plate of cells from crosslinker and place on ice
7. Scrape cells from plate using cell lifter and transfer 10mL PBS + cells from each plate to a 50mL conical tube (should be about 50mL total volume from five plates)
8. Centrifuge at 1000g for 5 minutes at 4°C to pellet cells
9. Remove supernatant and resuspend cells in 1mL cold PBS, pipetting gently to break up pellet. Transfer resuspended cells in PBS to 15mL conical tube. Pipette up and down several times with 9-inch cotton-plugged Pasteur pipette so that cells are in single-cell suspension. Take 10uL of resuspended cells and add to 90uL PBS in eppendorf tube. Pipette up and down to mix and count cells in 1:10 dilution using cell counter.
10. Centrifuge at 1000g for 5 minutes at 4°C to pellet cells
11. Remove supernatant and flash freeze pellet in liquid nitrogen and store at -80C

B. Cell Lysis: For Preparation of Nuclear Lysate from 400 million cells

- Perform in four batches, using two tubes of 50 million cells per batch. Start next batch once prior batch has started 45 minute incubation.

1. Resuspend 50 million cell pellet (previously stored at -80C) in 1mL of Cell Lysis Buffer (Nuclear I). Transfer to eppendorf tube.
2. Centrifuge at 3,300g for 10 minutes at 4C on microcentrifuge in cold room.
3. Discard supernatant and resuspend cell pellet in 1mL of Cell Lysis Buffer (Nuclear 1 with 0.01% DDM)
4. Incubate for 10 minutes on ice
5. Transfer sample to a dounce tissue homogenizer and use the B (small clearance) pestle 20 times to break cells.
6. Transfer sample back to eppendorf tube

7. Pellet nuclei by centrifugation at 3,300g for 10 minutes at 4C
8. Discard supernatant and resuspend pellet in 580uL of Cell Lysis Buffer (Nuclear II)
9. Incubate for 10 minutes on ice
10. Sonicate with microtip using 5 watts of power (25% duty) for 60 seconds total in pulses of 0.7 seconds on, followed by 3.3 seconds off (should be able to program this into control panel). Sonicate on ice
11. Add 3.75 uL 200X DNase salt solution (1X final concentration) and 165 uL TurboDNase (330U)
12. Incubate for 45 minutes at 37C.
13. Mix lysate with equal volume 2X Hybridization Buffer (750uL)
14. Centrifuge at 16,000g for 10 minutes at 4C
15. Transfer supernatant to fresh tube and flash freeze in liquid nitrogen

C. Captures Protocol for 400 Million Cells

Pre-Clearing Lysate

1. Warm 8 frozen aliquot of lysate (containing 50 million cells) to 37°C using a thermomixer
2. Pool 8 aliquots into one tube

Washing beads

3. Transfer 2.4mL of streptavidin-coated magnetic beads into an eppendorf tube.
4. Seperate on magnetic rack and remove storage buffer from beads
5. Resuspend beads in 2mL of 10mM Tris-HCl pH 7.5
6. Separate on magnetic rack and remove supernatant
7. Repeat washes for a total of 4 washes in Tris and 2 washes in 1X Hybridization Buffer
8. Magnetically separate and remove last wash from beads

Adding Beads to Lysate

9. Resuspend beads in lysate by pipetting gently
10. Incubate for 30 minutes at 37°C with intermittent mixing at 1100 rpm on thermomixer (30 seconds shaking, 30 seconds off)
11. Magnetically separate beads and transfer supernatant to fresh tube. Repeat this step to transfer lysate to fresh tube a second time.
12. Determine lysate volume.
    a. Lysate volume =

13. Remove sample of 1,000,000 cells worth of lysate and transfer to PCR strip tube. This is the RNA input sample. (100,000,000/400,000,000) = 0.0025 = 0.25%. So, remove 0.25% of lysate volume for **RNA input sample (SI)**

    SI volume =

- Total incubation time: 30 minutes

Hybridization, Capture and Protein Elution

1. Denature 63uL probe 1 by heating at 85°C for 3 minutes, then place on ice.
2. Add 60uL probe 1 to lysate
3. Incubate for 2 hours at 67°C with intermittent mixing at 1100 rpm on thermomixer (30 seconds shaking, 30 seconds off)
4. During the 2 hour incubation, prepare streptavidin beads (2.4mL) as previously described (4 washes with 10mM Tris-HCl pH 7.5, 2 washes with 1X Hybridization Buffer)
5. Magnetically separate beads and remove final wash from beads
6. Determine lysate volume
    a. Lysate volume =
7. At the end of the 2 hour incubation, remove sample of 1,000,000 cells worth of lysate and transfer to PCR strip tube. 1,000,000 cells is 0.25% of lysate volume. This is **RNA input + probe sample (SI-P)**
    a. SI-P volume =
8. Resuspend beads in lysate.
9. Incubate for 30 minutes at 67°C with intermittent mixing at 1100 rpm on thermomixer (30 seconds shaking, 30 seconds off)
10. Magnetically separate beads and remove supernatant. Take sample of 1,000,000 cells worth of supernatant and transfer to PCR strip tube. be. This is the **RNA flow-through sample (SF-T)**. Remove equal volume as for RNA input and RNA input + probe sample.
    a. SF-T volume =
11. Wash beads 3 times with at least one bead volume of 1X Hybridization Buffer per wash (2.4mL). Incubate each wash for 5 minutes at 67°C. Leave beads in final 1X Hybridization Buffer wash.
12. Determine volume of beads in 2.4mL 1X Hybridization Buffer.
13. Remove 0.25% of total bead volume and transfer to PCR strip tube. This the **RNA elution sample (SE).**
    a. SE volume =

14. Total incubation time: 2 hours and 48 minutes

Protein Elution Sample Prep

1. Magnetically separate beads and remove supernatant
2. Resuspend beads in 2.0mL of Benzonase Elution Buffer
3. Add 8uL of 1:10 dilution of benzonase non-specific nuclease to tube (1:10 dilution made in Benzonase Elution Buffer previously and stored in freezer).
4. Incubate for 2 hours at 37°C with intermittent mixing at 1100 rpm on thermomixer (30 seconds on, 30 seconds off)
5. Magnetically separate beads and transfer supernatant to a fresh eppendorf tube. Repeat this step for a total of 6 transfers to fresh tubes to remove all traces of streptavidin beads. **This is the protein elution sample**.

RNA Elution and Analysis

1. Take the RNA elution sample from previous Step 11 and separate on magnetic rack.
2. Remove supernatant and resuspend beads in 20uL of NLS buffer.
3. Heat sample for 2 minutes at 95°C (Do in thermocycler)
4. Magnetically separate and transfer supernatant containing eluted RNA to a fresh PCR strip tube.
5. Take the previously collected samples (RNA input, RNA input + probe, RNA flow-through) and dilute each sample to 40uL total volume with NLS Elution Buffer.
6. Add 2uL Proteinase K to SI, SI+P, SF-T and SE samples.
7. Incubate for 1 hour at 52-55°C.
8. RNA samples can be frozen at -20°C for short term storage or -80°C for long term storage.

D. Protein Precipitation

1. Add 10% final concentration of trichloroacetic acid (TCA) to protein elution sample
2. Incubate at 4°C overnight
3. Centrifuge at 16,000g for 30 minutes to pellet protein
4. Remove supernatant and replace with 1mL of cold acetone
5. Centrifuge at 16,000g for 15 minutes
6. Remove supernatant and allow pellet to dry in fume hood
7. Store protein elution sample at -20°C

E. Coomassie Blue Staining Protocol

1. After electrophoresis, incubate gel in a staining container (use tupperware) containing 100mL Coomassie Blue R-250 staining solution (0.1% Coomassie Blue, 40% ethanol, 10% acetic acid). For following steps using microwave, make sure not to overheat the staining solutions.
2. Loosely cover the staining container and heat in a microwave at full power for 1 minute. Do not allow solution to boil.
3. Remove the staining container from the microwave and gently shake the gel for 15 minutes at room temperature on an orbital shaker.
4. Decant the stain and rinse the gel once with DI water.
5. Prepare a destain solution containing 10% ethanol and 7.5% acetic acid. (10mL ethanol, 7.5mL acetic acid, 82.5mL DI water)
6. Place one or two stained gels in a staining container (use same tupperware, just rinse in between) containing the 100mL destain solution.
7. Loosely cover the staining container and heat in microwave at full power for 1 minute
8. Gently shake the gel at room temperature on an orbital shaker overnight. Destain for at least 24 hours. Remove destain solution and store either dry in plastic wrap or in DI water.

Supplementary Figures

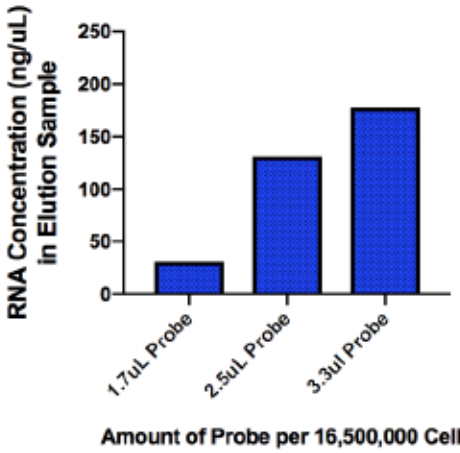


Figure S1. Increasing probe concentration captures more RNA in Elution Sample. Probe amounts represent 1X, 1.5X and 2X of amount of probe used to pulldown Xist RNA[33]. RNA concentration was determined by nanodropping (ThermoFisher Scientific, NanoDrop™ 2000C Spectrophotometer) 1uL of elution sample.

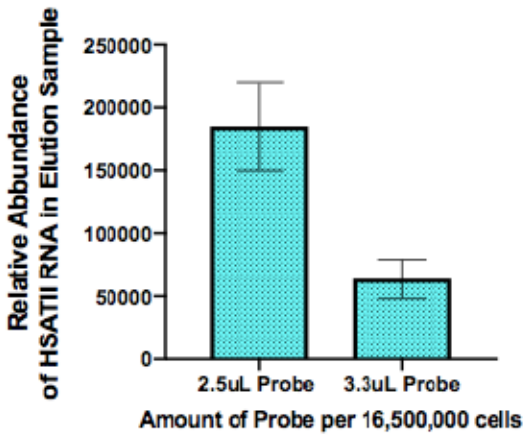


Figure S2. 1.5X probe concentration captures more HSATII RNA than 2X probe concentration. Surprisingly, relative abundance of HSATII RNA in elution sample, as determined by RT-qPCR, was reduced when 2X concentration of probe was used compared to 1.5X concentration. Thus,1.5X probe concentration (60uL probe per 400 million cells) was used for hybridization in RAP-MS protocol. Relative abundance was calculated by using the formula $2^{(40-C(t))}$. Means of technical triplicates are graphed, with error bars representing standard deviation.

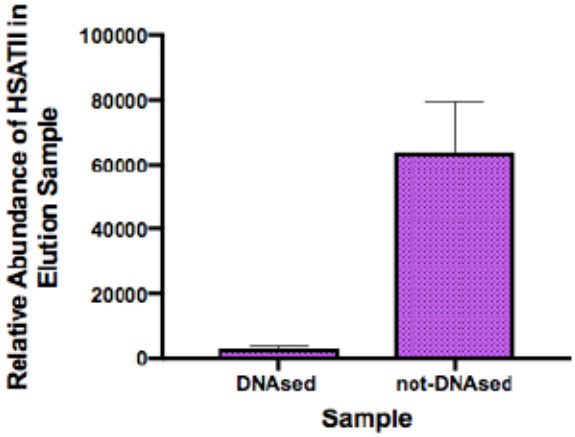


Fig. S3. Elution sample retains a large amount of HSATII DNA. Treating elution sample with DNase prior to analysis by RT-qPCR reduced HSATII amplification considerably, indicating that much of HSATII template present in elution sample was DNA rather than RNA. Much lower RNA:DNA ratio for HSATII in the cell relative to RNA:DNA ratio for Xist, due to high copy number of HSATII DNA and single copy of Xist DNA, likely accounts for extent of DNA contamination. DNase treatment in nuclear lysis protocol, even when amount of DNase enzyme used was increased and length of incubation was increased, was not effective in eliminating a considerable amount of DNA from the elution sample. DNA should not denature at hybridization temperature used in RAP-MS protocol[55] and thus should not be available for hybridization with probe, meaning that it is unlikely that DNA binding proteins, rather than RNA binding proteins, were identified.

Table S1. Sequences for Probe 1, Probe 2 and LNA as well as primers used for RT-qPCR.

Probe	Sequence
Probe 1	GAATGAATTGAATGCAATCATCGAATGGTCTCGAATGGAATCATCT TCTAATGGAAAGGAATGGAATCATCGCATAGAATCGAATGGAAT
Probe 2	GTACCTTAGTAGTAGTTTACCTTAGCTTACCTTAGTAGTAGTTTAC CTTAGATTACCTTAGTAACTTGTCTTAACTTACCTTAGCAGTAG
LNA	ATTCCATTCAGATTCCATTCGATC
MALAT1	**Forward (F):** CAGCAGCAGACAGGATTCCA **Reverse (R):** TCGTTAGCGCTCCTTCCTTC
β-actin	**F:** AGCGAGCATCCCCCAAAGTT **R:** GGGCACGAAGGCTCATCATT
α-SAT	**F:** GAAGCTTAWSTMACAGAGTTKAA **R:** GCTGCAGATCMCMAAGHAGTTTC
HSATII	**F:** ATTCGATTCCATTCGATGATGATTCC **R:** GGAACCGAATGAATCCTCATTGAATG

Table S2. Classification of identified candidate HSATII RNA binding proteins as being mRNA interactome proteins, candidate RBPs or proteins not present in either group. Classification was done by comparing candidate HSATII RBPs, as identified by gene names, to known RNA binding proteins from HeLa cells[9].

mRNA Interactome Proteins	Candidate RBPs	Other Proteins
SSBP1	SF3B3	PKN1
UBA1	GAPDH	MICALL1
HSP90AA1	GTF2I	ELP1
MOV10	CLTC	MENT
EFTUD2	DDX42	MYL6B
UBAP2L	PGK1	MTREX
EIF3A	NUP155	TUBB2A
SAFB2	DSG1	NASP
SEC23IP	ACLY	NOMO2
PNN	KIF11	HBB
DHX30	PGAM1	CORO7
MDH2	PRPF40A	EEF1A2
NCL	CFL1	AP3B1
SUPT16H	TUBB	CPSF1

EEF2	TRIM33	MYO6
HNRNPA2B1	HSPH1	TUBA1B
ADAR	S100A8	PKM
RPS27A	PHB	CASP14
DDX24	U2SURP	ALDOA
SART1	HSPA4	NCAPD2
KTN1		DEFA1
RBM25		DEFA1B
XRN2		SOD2
ATXN2L		MTHFD1L
SND1		MSH6
HDLBP		SMC4
SF3B1		FLII
DDX54		IARS
SUPT5H		PFAS
RBM15B		TBCD
HIST1H4H		IPO5
DIAPH1		ANKFY1
ABCF1		DBN1
YWHAZ		APOL2
LGALS1		USP11
EEF1A1		VIM
HSPA8		HIST2H2BF
HSPD1		RNF20
DHX9		HBA1
TXN		HBA2
NOP2		NDUFA4
TCERG1		CRMP1

ILF3		SLC25A5
DCD		ATP2A2
PRDX1		UBA6
PARP1		MPHOSPH8
HNRNPUL2		SNU13
HNRNPU		AMOT
SAFB		EIF3CL
NAT10		EMC1
BMS1		GOLGA2
NME1		KDM1A
PPIA		IDE
XPO5		CSDE1
FTSJ3		RAB3GAP1
LRPPRC		CACNA1E
USO1		JUNB
RPL7		CSTA
PPIG		HCFC1
DHX57		CKB
HMGB1		CDKL5
ACTN4		SMC3
NPM1		ZFR
MYBBP1A		HK2
		HIST1H4A
		HIST1H4B
		HIST1H4C
		HIST1H4D
		HIST1H4E
		HIST1H4F

		HIST1H4I
		HIST1H4J
		HIST1H4K
		HIST1H4L
		HIST2H4A
		HIST2H4B
		HIST4H4
		TRIM28
		KIF4A
		AGL
		HP
		HSP90B1
		XPO1
		POLR3B
		LRRFIP1
		IARS2
		PRSS1
		RFC1
		TPM3
		HDAC6
		EPHA1
		NUP133
		MCM2
		MCM4
		XPO7
		SMC2
		RBM5
		DIS3

		RRP12
		PRODH
		DCTN1
		PPP1R12A
		GOLIM4
		ERCC6L
		COPA
		CTSD
		SRRT
		LDHA
		ESF1
		SMARCC1
		SMC1A
		WAPL
		PELP1
		MIEF2
		FANCI
		HSP90AB1
		SIN3A
		DAXX
		PLCG1
		GART
		HIST1H2AB
		HIST1H2AE
		GEMIN4
		MAP4
		NEMF
		DDB1

		PDS5A
		RNF40
		CLUH
		HIST1H1D
		AARS
		HIST1H3A
		HIST1H3B
		HIST1H3C
		HIST1H3D
		HIST1H3E
		HIST1H3F
		HIST1H3G
		HIST1H3H
		HIST1H3I
		HIST1H3J
		LARS
		GANAB
		SPAG9
		SEC31A
		EPRS
		TPP2
		APOL1
		USP48
		ATP5ME
		CAPZA2
		ATP1A1
		NNT
		IRS4

		WDR3
		NCKAP1
		ACTA2
		KIF5B
		UPF1
		VARS
		SF3B2
		DLG1
		PSMD1
		TTC37
		TRIM24
		SEC24C
		ATP5F1A
		EPB41L2
		TF
		HTATSF1
		SMARCA5
		RAI14
		TMX1
		MTHFD1
		TUBB4B
		PRSS3
		CCDC187
		ARF3
		VCL
		EIF3B
		GNAS
		ENO1

		S100A10
		MTR
		DSG2
		ATP5F1B
		DDX46
		ERCC5
		USP28
		PARG
		MSH2
		FLG2
		POLD1
		MATR3
		SF3A1
		GEMIN5

Table S3. Identified candidate HSATII RNA binding proteins from upper and lower bands. Proteins, identified by accession numbers and gene names, are sorted by relative abundance. Relative abundance was calculated by dividing absolute abundance values for each protein by the lowest absolute abundance value. The majority of the proteins that do not have a relative abundance value were identified by a single peptide and as a result the mass spectrometer used was not able to provide a value for absolute abundance.

Accession	Gene Symbol	Description	Relative Abundance	Coverage [%]	# Peptides	MW [kDa]
Q8NE71-1	ABCF1	ATP-binding cassette sub-family F member 1 [OS=Homo sapiens]		3	1	95.9
Q8IUM5-4	MYO5	Isoform 1 of Unconventional myosin-VI [OS=Homo sapiens]		4	2	148.6
Q8IUA8	XPO7	exportin-7 [OS=Homo sapiens]		4	2	123.8
P13639	EEF2	Elongation factor 2 [OS=Homo sapiens]		6	3	95.3
P49756-1	RBM25	RNA-binding protein 25 [OS=Homo sapiens]		2	1	100.1
A0AVT1-1	UBA6	Ubiquitin-like modifier-activating enzyme 6 [OS=Homo sapiens]		2	1	117.9
P04179	SOD2	Superoxide dismutase [Mn], mitochondrial [OS=Homo sapiens]		9	1	24.7
P55884	EIF3B	Eukaryotic translation initiation factor 3 subunit B [OS=Homo sapiens]		3	1	92.4
Q9UNX4	WDR3	WD repeat-containing protein 3 [OS=Homo sapiens]		2	1	106
P18206	VCL	Vinculin [OS=Homo sapiens]		2	1	123.7
P19174-1	PLCG1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1 [OS=Homo sapiens]		2	1	148.4
Q86W56	PARG; BPHL	Poly(ADP-ribose) glycohydrolase [OS=Homo sapiens]		2	1	111
Q14152	EIF3A	Eukaryotic translation initiation factor 3 subunit A [OS=Homo sapiens]		1	1	166.5
Q8IX12	CCAR1	Cell division cycle and apoptosis regulator protein 1 [OS=Homo sapiens]		2	1	132.7
O14776-1	TCERG1	Transcription elongation regulator 1 [OS=Homo sapiens]		1	1	123.8
Q12906-1	ILF3	Interleukin enhancer-binding factor 3 [OS=Homo sapiens]		3	1	95.3
O75150	RNF40	E3 ubiquitin-protein ligase BRE1B [OS=Homo sapiens]		2	1	113.6
P35573-1	AGL	glycogen debranching enzyme [OS=Homo sapiens]		1	1	174.7
O00267	SUPT5H	Transcription elongation factor SPT5 [OS=Homo sapiens]		3	1	120.9
Q8NVI1-1	FANCI	Isoform 1 of Fanconi anemia group I protein [OS=Homo sapiens]		2	1	142.5
P05787	KRT8	Keratin, type II cytoskeletal 8 [OS=Homo sapiens]		6	3	53.7
Q16643	DBN1	drebrin [OS=Homo sapiens]		4	1	71.4
Q9H501	ESF1	ESF1 homolog [OS=Homo sapiens]		4	2	98.7
Q7Z5K2-1	WAPAL; WAPL	Wings apart-like protein homolog [OS=Homo sapiens]		2	1	132.9
Q8TEQ6	GEMIN5	gem-associated protein 5 [OS=Homo sapiens]		2	1	168.5
Q9P2R3	ANKFY1	Rabankyrin-5 [OS=Homo sapiens]		3	2	128.3
Q6P158-1	DHX57	putative ATP-dependent RNA helicase DHX57 [OS=Homo sapiens]		2	1	155.5
Q9Y2A7-1	NCKAP1	Nck-associated protein 1 [OS=Homo sapiens]		2	1	128.7
P06576	ATP5B	ATP synthase subunit beta, mitochondrial [OS=Homo sapiens]		4	2	56.5
Q13427	PPIG	peptidyl-prolyl cis-trans isomerase g [OS=Homo sapiens]		2	1	88.6
O00461	GOLIM4	Golgi integral membrane protein 4 [OS=Homo sapiens]		2	1	81.8
Q8IZL8	PELP1	Proline-, glutamic acid- and leucine-rich protein 1 [OS=Homo sapiens]		2	1	119.6
P02787	TF	Serotransferrin [OS=Homo sapiens]		3	2	77
O94979	SEC31A	Protein transport protein Sec31A [OS=Homo sapiens]		2	1	132.9
O60271	SPAG9	C-jun-amino-terminal kinase-interacting protein 4 [OS=Homo sapiens]		2	1	146.1
Q8NW08	POLR3B	DNA-directed RNA polymerase III subunit RPC2 [OS=Homo sapiens]		2	1	127.7
Q96KR1	ZFR	Zinc finger RNA-binding protein [OS=Homo sapiens]		1	1	116.9
Q16512-1	PKN1	serine/threonine-protein kinase N1 [OS=Homo sapiens]		2	1	103.9
P07814	EPRS	Bifunctional glutamate/proline--tRNA ligase [OS=Homo sapiens]		1	1	170.5
O95239-1	KIF4A	Chromosome-associated kinesin KIF4A [OS=Homo sapiens]		1	1	139.8
Q9P0K7-1	RAI14	Ankycorbin [OS=Homo sapiens]		2	1	110
Q99549	MPHOSPH8	M-phase phosphoprotein 8 [OS=Homo sapiens]		4	1	97.1
P28340	POLD1	DNA polymerase delta catalytic subunit [OS=Homo sapiens]		1	1	123.6
Q8WUM0	NUP133	Nuclear pore complex protein Nup133 [OS=Homo sapiens]		3	1	128.9
P81605	DCD	Dermcidin [OS=Homo sapiens]		10	1	11.3
Q9Y2J2	EPB41L3	Band 4.1-like protein 3 [OS=Homo sapiens]		3	1	120.6
Q9H9S0-1	HBS1L	HBS1-like protein [OS=Homo sapiens]		2	1	75.4
Q9Y6Y8	SEC23IP	SEC23-interacting protein [OS=Homo sapiens]		2	1	111
Q08379	GOLGA2	Golgin subfamily A member 2 [OS=Homo sapiens]		2	1	113
Q14692	BMS1	ribosome biogenesis protein BMS1 homolog [OS=Homo sapiens]		1	1	145.7
P51784	USP11	Ubiquitin carboxyl-terminal hydrolase 11 [OS=Homo sapiens]		2	1	109.7
P25705-1	ATP5A1	ATP synthase subunit alpha, mitochondrial [OS=Homo sapiens]		2	1	59.7
Q96ST3	SIN3A	Paired amphipathic helix protein Sin3a [OS=Homo sapiens]		1	1	145.1
P62158	CALM1; CALM2; CALM3	Calmodulin [OS=Homo sapiens]		11	1	16.8
Q8N3F8	MICALL1	MICAL-like protein 1 [OS=Homo sapiens]		2	1	93.4
O95625	ZBTB11	Zinc finger and BTB domain-containing protein 11 [OS=Homo sapiens]		2	1	119.3
Q16555-1	DPYSL2	Dihydropyrimidinase-related protein 2 [OS=Homo sapiens]		3	1	62.3
P04406-1	GAPDH	glyceraldehyde-3-phosphate dehydrogenase [OS=Homo sapiens]		6	1	36
Q5VTR2	RNF20	E3 ubiquitin-protein ligase BRE1A [OS=Homo sapiens]		2	1	113.6
Q16777	HIST2H2AC	Histone H2A type 2-C [OS=Homo sapiens]		15	1	14
O60231	DHX16	Putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16 [OS=Homo sapiens]		1	1	119.2
P05062	ALDOB	fructose-bisphosphate aldolase B [OS=Homo sapiens]		4	1	39.4
P84243	H3F3A; H3F3AP4; H3F3B	histone H3.3 [OS=Homo sapiens]		8	1	15.3
P67936	TPM4	Tropomyosin alpha-4 chain [OS=Homo sapiens]		4	1	28.5
Q15042-1	RAB3GAP1	Rab3 GTPase-activating protein catalytic subunit [OS=Homo sapiens]		2	1	110.5
P23528	CFL1	Cofilin-1 [OS=Homo sapiens]		8	1	18.5
P16402	HIST1H1D	Histone H1.3 [OS=Homo sapiens]		4	1	22.3
P19367-1	HK1	Hexokinase-1 [OS=Homo sapiens]		3	1	102.4
Q08554-1	DSC1	Desmocollin-1 [OS=Homo sapiens]		1	1	99.9
O75155-1	CAND2	Cullin-associated NEDD8-dissociated protein 2 [OS=Homo sapiens]		2	1	135.2
Q03113	GNA12	Guanine nucleotide-binding protein subunit alpha-12 [OS=Homo sapiens]		3	1	44.3
P07339	CTSD	Cathepsin D [OS=Homo sapiens]		2	1	44.5
O43795	MYO1B	Unconventional myosin-Ib [OS=Homo sapiens]		1	1	131.9
Q9HVM6	XPO5	exportin-5 [OS=Homo sapiens]		1	1	136.2
P45880	VDAC2	Voltage-dependent anion-selective channel protein 2 [OS=Homo sapiens]		3	1	31.5
P47755	CAPZA2	F-actin-capping protein subunit alpha-2 [OS=Homo sapiens]		3	1	32.9
Q9UPU5	USP24	Ubiquitin carboxyl-terminal hydrolase 24 [OS=Homo sapiens]		1	1	294.2
P22626	HNRNPA2B1	heterogeneous nuclear ribonucleoproteins A2/B1 [OS=Homo sapiens]		3	1	37.4
P31327-1	CPS1	Carbamoyl-phosphate synthase [ammonia], mitochondrial [OS=Homo sapiens]		1	1	164.8
P62937	PPIA	peptidyl-prolyl cis-trans isomerase A [OS=Homo sapiens]		5	1	18
O15164	TRIM24	Transcription intermediary factor 1-alpha [OS=Homo sapiens]		3	1	116.8
Q9H2J6	VAT1L	Synaptic vesicle membrane protein VAT-1 homolog-like [OS=Homo sapiens]		4	1	45.9
O14791	APOL1	Apolipoprotein L1 [OS=Homo sapiens]	3421.81438	24	9	43.9
P02768-1	ALB	Serum albumin [OS=Homo sapiens]	499.879364	43	22	69.3
P19338	NCL	Nucleolin [OS=Homo sapiens]	327.299032	17	7	76.6
P09874	PARP1	Poly [ADP-ribose] polymerase 1 [OS=Homo sapiens]	297.443104	18	13	113
P22314	UBA1	Ubiquitin-like modifier-activating enzyme 1 [OS=Homo sapiens]	282.210066	21	11	117.8
P35908	KRT2	Keratin, type II cytoskeletal 2 epidermal [OS=Homo sapiens]	229.863137	20	13	65.4
P04264	KRT1	Keratin, type II cytoskeletal 1 [OS=Homo sapiens]	217.59129	20	12	66
P13645	KRT10	Keratin, type I cytoskeletal 10 [OS=Homo sapiens]	190.988796	21	11	58.8
P11586	MTHFD1	C-1-tetrahydrofolate synthase, cytoplasmic [OS=Homo sapiens]	106.494394	12	6	101.5
Q00839	HNRNPU	Heterogeneous nuclear ribonucleoprotein U [OS=Homo sapiens]	103.489968	16	7	90.5
P34932	HSPA4	Heat shock 70 kDa protein 4 [OS=Homo sapiens]	96.585 1536	20	10	94.3
P60709	ACTB	Actin, cytoplasmic 1 [OS=Homo sapiens]	93.850231	24	7	41.7
P41252	IARS	Isoleucine--tRNA ligase, cytoplasmic [OS=Homo sapiens]	84.3871696	8	5	144.4
Q4VCS5	AMOT	Angiomotin [OS=Homo sapiens]	82.8192625	2	2	118
O76039-1	CDKL5	Cyclin-dependent kinase-like 5 [OS=Homo sapiens]	73.7971338	5	3	115.5
O75533-1	SF3B1	splicing factor 3B subunit 1 [OS=Homo sapiens]	70.8997676	8	5	145.7
P49321	NASP	Nuclear autoantigenic sperm protein [OS=Homo sapiens]	66.8874378	26	9	85.3
P53396-1	ACLY	ATP-citrate synthase [OS=Homo sapiens]	66.8791076	13	7	120.8
P35527	KRT9	Keratin, type I cytoskeletal 9 [OS=Homo sapiens]	55.3707749	8	4	62
P42704	LRPPRC	Leucine-rich PPR motif-containing protein, mitochondrial [OS=Homo sapiens]	54.5988997	7	7	157.8
Q13263	TRIM28	Transcription intermediary factor 1-beta [OS=Homo sapiens]	50.5173514	9	4	88.5
Q5QNW6-1	HIST2H2BF	Histone H2B type 2-F [OS=Homo sapiens]	48.5907012	43	5	13.9
Q08211	DHX9	Atp-dependent rna helicase [OS=Homo sapiens]	46.2978369	17	12	140.9
Q15029	EFTUD2	116 kDa U5 small nuclear ribonucleoprotein component [OS=Homo sapiens]	44.928 3152	15	8	109.4
P49736	MCM2	DNA replication licensing factor mcm2 [OS=Homo sapiens]	44.2409677	12	6	101.8
Q15459	SF3A1	splicing factor 3A subunit 1 [OS=Homo sapiens]	41.904925	3	1	88.8
Q16531	DDB1	DNA damage-binding protein 1 [OS=Homo sapiens]	41.3581928	9	5	126.9
P62805	HIST1H4A; HIST1H4F; HIST	histone H4 [OS=Homo sapiens]	39.6347185	41	4	11.4
O14980	XPO1	Exportin-1 [OS=Homo sapiens]	38.800296	8	6	123.3
P68371	TUBB4B	Tubulin beta-4B chain [OS=Homo sapiens]	37.9304991	24	9	49.8
P43243	MATR3	Matrin-3 [OS=Homo sapiens]	36.9295445	10	3	94.6
P49588	AARS	Alanine--tRNA ligase, cytoplasmic [OS=Homo sapiens]	35.9442677	15	7	106.7
Q14697-1	GANAB	Neutral alpha-glucosidase AB [OS=Homo sapiens]	34.1929081	12	6	106.8
Q15424-1	SAFB	Scaffold attachment factor B1 [OS=Homo sapiens]	32.2183291	5	3	102.6
Q92900	UPF1	Regulator of nonsense transcripts 1 [OS=Homo sapiens]	31.2214348	8	5	124.3
O00410	IPO5	Importin-5 [OS=Homo sapiens]	30.6319552	8	7	123.6
Q71U36	TUBA1A	tubulin alpha-1A chain [OS=Homo sapiens]	29.0045417	18	6	50.1
Q13435	SF3B2	Splicing factor 3b subunit 2 [OS=Homo sapiens]	27.8137782	4	2	100.2
P53621-1	COPA	coatomer subunit alpha [OS=Homo sapiens]	27.4087104	6	4	138.3
P22102-1	GART	trifunctional purine biosynthetic protein adenosine-3 [OS=Homo sapiens]	25.6519394	12	6	107.7
P68104	EEF1A1	Elongation factor 1-alpha 1 [OS=Homo sapiens]	23.8077967	10	5	50.1
Q14683	SMC1A	structural maintenance of chromosomes protein 1a [OS=Homo sapiens]	23.4680744	5	4	143.1
Q8IUQ7	SMC3	Structural maintenance of chromosomes protein 3 [OS=Homo sapiens]	23.3268535	2	2	141.5
Q71U14	DDX46	probable ATP-dependent RNA helicase DDX46 [OS=Homo sapiens]	23.2294375	6	5	117.3
Q14203	DCTN1	Dynactin subunit 1 [OS=Homo sapiens]	22.2059684	7	5	141.6

Q9Y5B9	SUPT16H	FACT complex subunit SPT16 [OS=Homo sapiens]	22.09275961	11	6	119.8
P53992	SEC24C	Protein transport protein sec24c [OS=Homo sapiens]	18.8370929	1	1	118.2
Q9P2J5	LARS	Leucine--tRNA ligase, cytoplasmic [OS=Homo sapiens]	17.62906021	8	4	134.4
P13647	KRT5	keratin, type II cytoskeletal 5 [OS=Homo sapiens]	16.79745921	12	6	62.3
Q15393-1	SF3B3	Splicing factor 3B subunit 3 [OS=Homo sapiens]	16.58446491	11	7	135.5
O43707	ACTN4	Alpha-actinin-4 [OS=Homo sapiens]	16.45567461	4	2	104.8
P69905	HBA2; HBA1	Hemoglobin subunit alpha [OS=Homo sapiens]	15.84551451	17	2	15.2
Q99460	PSMD1	26S proteasome non-ATPase regulatory subunit 1 [OS=Homo sapiens]	14.00860911	9	5	105.8
P08238	HSP90AB1	Heat shock protein HSP 90-beta [OS=Homo sapiens]	13.81358471	7	5	83.2
Q8N766-1	EMC1	ER membrane protein complex subunit 1 [OS=Homo sapiens]	13.37111281	7	4	111.7
P42285	SKIV2L2	Superkiller viralicidic activity 2-like 2 [OS=Homo sapiens]	13.08072761	6	3	117.7
O75400	PRPF40A	pre-mRNA-processing factor 40 homolog A [OS=Homo sapiens]	12.1155141	2	1	108.7
O15042	U2SURP	U2 snRNP-associated SURP motif-containing protein [OS=Homo sapiens]	12.06832221	5	2	118.2
O43491-1	EPB41L2	band 4.1-like protein 2 [OS=Homo sapiens]	11.47811791	6	2	112.5
Q13885	TUBB2A	Tubulin beta-2A chain [OS=Homo sapiens]	10.99053941	20	8	49.9
P05141	SLC25A5	ADP/ATP translocase 2 [OS=Homo sapiens]	10.76175651	4	1	32.8
O15067	PFAS	Phosphoribosylformylglycinamidine synthase [OS=Homo sapiens]	10.6097881	11	5	144.6
Q86UP2-1	KTN1	Kinectin [OS=Homo sapiens]	10.39460521	2	1	156.2
P06748	NPM1	Nucleophosmin [OS=Homo sapiens]	9.843170041	12	2	32.6
Q8P2L3-1	DIS3	exosome complex exoribonuclease RRP44 [OS=Homo sapiens]	9.444598651	3	1	108.9
P55265-1	ADAR	Double-stranded RNA-specific adenosine deaminase [OS=Homo sapiens]	9.421410421	4	2	136
P33991	MCM4	DNA replication licensing factor MCM4 [OS=Homo sapiens]	9.191340661	7	4	96.5
Q7KZF4	SND1	staphylococcal nuclease domain-containing protein 1 [OS=Homo sapiens]	9.102344371	13	7	101.9
Q8H0A0	NAT10	RNA cytidine acetyltransferase [OS=Homo sapiens]	8.9551617	2	1	115.7
Q14126	DSG2	Desmoglein-2 [OS=Homo sapiens]	8.944277021	13	6	122.2
P68871	HBB	Hemoglobin subunit beta [OS=Homo sapiens]	8.889545471	17	2	16
Q13MZ4	LRRFIP1	Leucine-rich repeat flightless-interacting protein 1 [OS=Homo sapiens]	8.877211411	2	1	89.2
P35232	PHB	Prohibitin [OS=Homo sapiens]	8.757002041	12	3	29.8
P08670	VIM	Vimentin [OS=Homo sapiens]	8.256311501	5	2	53.6
P14618	PKM	Pyruvate kinase PKM [OS=Homo sapiens]	8.248613171	13	4	57.9
P11142-1	HSPA8	Heat shock cognate 71 kDa protein [OS=Homo sapiens]	8.227029291	5	3	70.9
Q9BQE5	APOL2	apolipoprotein L2 [OS=Homo sapiens]	8.177727461	2	1	37.1
B5ME19	EIF3CL	eukaryotic translation initiation factor 3 subunit C-like protein [OS=Homo sapiens]	8.025662661	7	3	105.4
Q92598	HSPH1	Heat shock protein 105 kDa [OS=Homo sapiens]	7.864599841	5	2	96.8
Q00341-1	HDLBP	Vigilin [OS=Homo sapiens]	7.605615281	8	5	141.4
P16615	ATP2A2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 [OS=Homo sapiens]	7.59465064	2	2	114.7
P63104-1	YWHAZ	14-3-3 protein zeta/delta [OS=Homo sapiens]	7.453804171	14	3	27.7
Q8IY81	FTSJ3	pre-rRNA processing protein FTSJ3 [OS=Homo sapiens]	7.236606831	6	3	96.5
P62987	UBA52	Ubiquitin-60S ribosomal protein L40 [OS=Homo sapiens]	7.120879581	13	1	14.7
O60341	KDM1A	Lysine-specific histone demethylase 1A [OS=Homo sapiens]	7.021443331	2	1	92.8
Q9GZR7	DDX24	ATP-dependent RNA helicase DDX24 [OS=Homo sapiens]	6.871666691	4	2	96.3
P14735	IDE	Insulin-degrading enzyme [OS=Homo sapiens]	6.801507691	3	2	117.9
O75694	NUP155	nuclear pore complex protein nup155 [OS=Homo sapiens]	6.631718961	1	1	155.1
Q9NSE4	IARS2	Isoleucine--tRNA ligase, mitochondrial [OS=Homo sapiens]	6.472848131	1	1	113.7
Q9Y5J3	IKBKAP; ELP1	Elongator complex protein 1 [OS=Homo sapiens]	6.395853241	2	2	150.2
Q5T749	KPRP	Keratinocyte proline-rich protein [OS=Homo sapiens]	6.180366011	2	1	64.1
Q5JTH9-1	RRP12	RRP12-like protein [OS=Homo sapiens]	6.17543371	1	1	143.6
P18669	PGAM1; LOC643576	Phosphoglycerate mutase 1 [OS=Homo sapiens]	6.154256591	4	1	28.8
P02533	KRT14	Keratin, type I cytoskeletal 14 [OS=Homo sapiens]	5.990980291	8	4	51.5
O43719	HTATSF1	HIV Tat-specific factor 1 [OS=Homo sapiens]	5.895895991	4	1	85.8
P52732	KIF11	Kinesin-like protein KIF11 [OS=Homo sapiens]	5.819866261	1	1	119.1
Q86XP3-1	DDX42	ATP-dependent RNA helicase DDX42 [OS=Homo sapiens]	5.542815291	5	3	102.9
O14654	IRS4	Insulin receptor substrate 4 [OS=Homo sapiens]	5.531971031	2	2	133.7
Q7L2E3-1	DHX30	Putative ATP-dependent RNA helicase DHX30 [OS=Homo sapiens]	5.503081151	2	1	133.9
P18124	RPL7	60S ribosomal protein L7 [OS=Homo sapiens]	5.380194131	12	2	29.2
Q9BTW9	TBCD	Tubulin-specific chaperone D [OS=Homo sapiens]	5.284955231	1	1	132.5
P43246-1	MSH2	DNA mismatch repair protein MSH2 [OS=Homo sapiens]	5.178673661	3	2	104.7
P05023	ATP1A1	Sodium/potassium-transporting ATPase subunit alpha-1 [OS=Homo sapiens]	5.163344871	6	3	112.8
O60264	SMARCA5	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 [OS=Ho	5.06987322	2	1	121.8
Q68E01	INTS3	Integrator complex subunit 3 [OS=Homo sapiens]	5.044959991	2	1	118
P04075	ALDOA	fructose-bisphosphate aldolase A [OS=Homo sapiens]	4.940961291	14	3	39.4
P27816-1	MAP4	Microtubule-associated protein 4 [OS=Homo sapiens]	4.910726381	1	1	120.9
Q14MD3	HNRNPUL2	heterogeneous nuclear ribonucleoprotein U-like protein 2 [OS=Homo sapiens]	4.805020431	2	1	85.1
Q13045-1	FLII	protein flightless-1 homolog [OS=Homo sapiens]	4.616377301	1	1	144.7
Q9NTJ3-1	SMC4	Structural maintenance of chromosomes protein 4 [OS=Homo sapiens]	4.50860571	1	1	147.1
Q9BXP5	SRRT	serrate RNA effector molecule homolog [OS=Homo sapiens]	4.504200811	2	2	100.6
Q9H307	PNN	Pinin [OS=Homo sapiens]	4.455486641	2	1	81.6
Q13423	NNT	NAD(P) transhydrogenase, mitochondrial [OS=Homo sapiens]	4.34751841	2	1	113.8
Q14157-1	UBAP2L	Isoform 2 of Ubiquitin-associated protein 2-like [OS=Homo sapiens]	4.313102151	6	3	103.9
P15531	NME1	Nucleoside diphosphate kinase A [OS=Homo sapiens]	4.297925481	19	2	17.1
Q5D862	FLG2	Filaggrin-2 [OS=Homo sapiens]	4.283630281	1	1	247.9
O75153	CLUH; KIAA0664	Clustered mitochondria protein homolog [OS=Homo sapiens]	4.19782	4	2	146.6
Q8WWM7-1	ATXN2L	ataxin-2-like protein [OS=Homo sapiens]	4.103362961	1	1	113.3
O60763-1	USO1	General vesicular transport factor p115 [OS=Homo sapiens]	4.076217721	4	2	107.8
O14974	PPP1R12A	Protein phosphatase 1 regulatory subunit 12A [OS=Homo sapiens]	4.015703581	2	1	115.2
P57737	CORO7	Coronin-7 [OS=Homo sapiens]	4.004329051	2	1	100.5
P48741	HSPA7	Putative heat shock 70 kDa protein 7 [OS=Homo sapiens]	3.94466141	7	2	40.2
Q6PGP7	TTC37	Tetratricopeptide repeat protein 37 [OS=Homo sapiens]	3.907047131	1	1	175.4
Q9H0D6-1	XRN2	5'-3' exoribonuclease 2 [OS=Homo sapiens]	3.888183811	4	2	108.5
O76015	KRT38	Keratin, type I cuticular Ha8 [OS=Homo sapiens]	3.742199581	2	1	50.4
P26640	VARS; VARS2	Valine--tRNA ligase [OS=Homo sapiens]	3.70065961	1	1	140.4
Q99RF7-1	PDS5A	Sister chromatid cohesion protein PDS5 homolog A [OS=Homo sapiens]	3.417923881	4	2	150.7
Q92922	SMARCC1	SWI/SNF complex subunit SMARCC1 [OS=Homo sapiens]	3.367580671	3	1	122.8
Q9UPN9-1	TRIM33	E3 ubiquitin-protein ligase TRIM33 [OS=Homo sapiens]	3.342658471	2	1	122.5
O95347-1	SMC2	structural maintenance of chromosomes protein 2 [OS=Homo sapiens]	3.255560151	3	2	135.6
Q8TDD1	DDX54	ATP-dependent RNA helicase DDX54 [OS=Homo sapiens]	3.088604861	3	1	98.5
Q00610-1	CLTC	Clathrin heavy chain 1 [OS=Homo sapiens]	2.89934221	3	2	191.5
P07864	LDHC	L-lactate dehydrogenase C chain [OS=Homo sapiens]	2.878869191	4	1	36.3
P06733-1	ENO1	alpha-enolase [OS=Homo sapiens]	2.856788731	2	1	47.1
P09429	HMGB1	High mobility group protein B1 [OS=Homo sapiens]	2.745048031	7	1	24.9
P14625	HSP90B1	Endoplasmin [OS=Homo sapiens]	2.68359981	2	1	92.4
Q99707	MTR	methionine synthase [OS=Homo sapiens]	2.667639481	3	2	140.4
P35030	PRSS3	Trypsin-3 [OS=Homo sapiens]	2.556915551	4	1	32.5
Q96QG0	MYBBP1A	Myb-binding protein 1A [OS=Homo sapiens]	2.488000321	3	2	148.8
P35251-1	RFC1	Replication factor C subunit 1 [OS=Homo sapiens]	2.452346851	2	1	128.2
Q5JPE7	NOMO3	NODAL modulator 3 [OS=Homo sapiens]	2.421211341	1	1	139.4
P78347	GTF2I	General transcription factor II-I [OS=Homo sapiens]	2.362897171	3	2	112.3
Q02413	DSG1	Desmoglein-1 [OS=Homo sapiens]	2.09439591	1	1	113.7
P51610-1	HCFC1	Host cell factor 1 [OS=Homo sapiens]	2.081360991	2	2	208.6
O43290	SART1	U4/U6.U5 tri-snRNP-associated protein 1 [OS=Homo sapiens]	2.063076081	3	1	90.2
P08727	KRT19	Keratin, type I cytoskeletal 19 [OS=Homo sapiens]	2.030025261	7	3	44.1
Q9UER7-1	DAXX	Death domain-associated protein 6 [OS=Homo sapiens]	1.997101381	2	1	81.3
O60610-1	DIAPH1	Protein diaphanous homolog 1 [OS=Homo sapiens]	1.956138131	3	2	141.3
O00203-1	AP3B1	AP-3 complex subunit beta-1 [OS=Homo sapiens]	1.940817891	2	1	121.2
Q92888-1	ARHGEF1	Rho guanine nucleotide exchange factor 1 [OS=Homo sapiens]	1.910485411	2	1	102.4
Q10570	CPSF1	Cleavage and polyadenylation specificity factor subunit 1 [OS=Homo sapiens]	1.852106851	1	1	160.8
P40926	MDH2	Malate dehydrogenase, mitochondrial [OS=Homo sapiens]	1.802698021	3	1	35.5
P09382	LGALS1	Galectin-1 [OS=Homo sapiens]	1.797443561	6	1	14.7
P10809	HSPD1	60 kDa heat shock protein, mitochondrial [OS=Homo sapiens]	1.77811071	5	2	61
Q86UV5	USP48	Ubiquitin carboxyl-terminal hydrolase 48 [OS=Homo sapiens]	1.763072641	1	1	119
P10599-1	TXN	thioredoxin [OS=Homo sapiens]	1.689991521	9	1	11.7
Q9H3N1	TMX1	Thioredoxin-related transmembrane protein 1 [OS=Homo sapiens]	1.651129871	4	1	31.8
Q96RU2-2	USP28	Isoform 2 of Ubiquitin carboxyl-terminal hydrolase 28 [OS=Homo sapiens]	1.636352461	2	1	119
P29144	TPP2	Tripeptidyl-peptidase 2 [OS=Homo sapiens]	1.624483431	6	3	138.3
Q12959	DLG1	Disks large homolog 1 [OS=Homo sapiens]	1.442037091	2	1	100.4
Q13242	SRSF9	serine/arginine-rich splicing factor 9 [OS=Homo sapiens]	1.427252981	5	1	25.5
Q6UB35-1	MTHFD1L	Monofunctional C1-tetrahydrofolate synthase, mitochondrial [OS=Homo sapiens]	1.414163621	1	1	105.7
P12277	CKB	Creatine kinase B-type [OS=Homo sapiens]	1.397622881	2	1	42.6
Q9BX66-12	SORBS1	Isoform 12 of Sorbin and SH3 domain-containing protein 1 [OS=Homo sapiens]	1	1	1	143.7

Accession	Gene Symbol	Description	Relative Abundance	Coverage [%]	# Peptides	MW [kDa]
Q00341-1	HDLBP	Vigilin [OS=Homo sapiens]		3	2	141.4
Q7KZF4	SND1	staphylococcal nuclease domain-containing protein 1 [OS=Homo sapiens]		5	2	101.9
O00267	SUPT5H	Transcription elongation factor Spt5 [OS=Homo sapiens]		7	3	120.9
Q13423	NNT	NAD(P) transhydrogenase, mitochondrial [OS=Homo sapiens]		4	2	113.8
P05787	KRT8	Keratin, type II cytoskeletal 8 [OS=Homo sapiens]		6	3	53.7
Q9P2J5	LARS	Leucine--tRNA ligase, cytoplasmic [OS=Homo sapiens]		2	1	134.4
P14625	HSP90B1	Endoplasmin [OS=Homo sapiens]		2	1	92.4
Q9H0A0	NAT10	RNA cytidine acetyltransferase [OS=Homo sapiens]		2	1	115.7
Q5JTH9-1	RRP12	RRP12-like protein [OS=Homo sapiens]		1	1	143.6
P49756-1	RBM25	RNA-binding protein 25 [OS=Homo sapiens]		2	1	100.1
Q9Y2L1-1	DIS3	exosome complex exonuclease RRP44 [OS=Homo sapiens]		4	2	108.9
Q13045-1	FLII	protein flightless-1 homolog [OS=Homo sapiens]		1	1	144.7
P48741	HSPA7	Putative heat shock 70 kDa protein 7 [OS=Homo sapiens]		7	2	40.2
Q16777	HIST2H2AC	Histone H2A type 2-C [OS=Homo sapiens]		15	1	14
Q6UB35-1	MTHFD1L	Monofunctional C1-tetrahydrofolate synthase, mitochondrial [OS=Homo sapiens]		2	1	105.7
Q15042-1	RAB3GAP1	Rab3 GTPase-activating protein catalytic subunit [OS=Homo sapiens]		2	1	110.5
O75694	NUP155	nuclear pore complex protein nup155 [OS=Homo sapiens]		1	1	155.1
O60341	KDM1A	Lysine-specific histone demethylase 1A [OS=Homo sapiens]		2	1	92.8
Q9NTJ3-1	SMC4	Structural maintenance of chromosomes protein 4 [OS=Homo sapiens]		3	2	147.1
P04083	ANXA1	annexin A1 [OS=Homo sapiens]		5	1	38.7
P46087	NOP2	Probable 28S rRNA (cytosine(4447)-C(5))-methyltransferase [OS=Homo sapiens]		2	1	89.2
Q15021	NCAPD2	condensin complex subunit 1 [OS=Homo sapiens]		2	1	157.1
Q8IY81	FTSJ3	pre-rRNA processing protein FTSJ3 [OS=Homo sapiens]		2	1	96.5
Q2NKX8	ERCC6L	Dna excision repair protein ercc-6-like [OS=Homo sapiens]		2	1	141
Q13427	PPIG	peptidyl-prolyl cis-trans isomerase g [OS=Homo sapiens]		2	1	88.6
P16615	ATP2A2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 [OS=Homo sapiens]		2	1	114.7
O95239-1	KIF4A	Chromosome-associated kinesin KIF4A [OS=Homo sapiens]		1	1	139.8
P06576	ATP5B	ATP synthase subunit beta, mitochondrial [OS=Homo sapiens]		2	1	56.5
O14654	IRS4	insulin receptor substrate 4 [OS=Homo sapiens]		1	1	133.7
P60660	MYL6	Myosin light polypeptide 6 [OS=Homo sapiens]		9	1	16.9
Q16512-1	PKN1	serine/threonine-protein kinase N1 [OS=Homo sapiens]		2	1	103.9
P55769	NHP2L1; SNU	NHP2-like protein 1 [OS=Homo sapiens]		9	1	14.2
Q9UER7-1	DAXX	Death domain-associated protein 6 [OS=Homo sapiens]		2	1	81.3
Q9HCE1	MOV10	Putative helicase MOV-10 [OS=Homo sapiens]		3	1	113.6
P12277	CKB	Creatine kinase B-type [OS=Homo sapiens]		3	1	42.6
B5ME19	EIF3CL	eukaryotic translation initiation factor 3 subunit C-like protein [OS=Homo sapiens]		3	1	105.4
P55265-1	ADAR	Double-stranded RNA-specific adenosine deaminase [OS=Homo sapiens]		1	1	136
Q04837	SSBP1	Single-stranded DNA-binding protein, mitochondrial [OS=Homo sapiens]		10	1	17.2
P33176	KIF5B	Kinesin-1 heavy chain [OS=Homo sapiens]		2	1	109.6
P25705-1	ATP5A1	ATP synthase subunit alpha, mitochondrial [OS=Homo sapiens]		2	1	59.7
P22626	HNRNPA2B1	heterogeneous nuclear ribonucleoproteins A2/B1 [OS=Homo sapiens]		3	1	37.4
O75534	CSDE1	cold shock domain-containing protein E1 [OS=Homo sapiens]		3	1	88.8
Q9HAV4	XPO5	exportin-5 [OS=Homo sapiens]		1	1	136.2
P15531	NME1	Nucleoside diphosphate kinase A [OS=Homo sapiens]		8	1	17.1
Q8NDT2	RBM15B	Putative RNA-binding protein 15B [OS=Homo sapiens]		2	1	97.1
O00483	NDUFA4	Cytochrome c oxidase subunit NDUFA4 [OS=Homo sapiens]		12	1	9.4
P52732	KIF11	Kinesin-like protein KIF11 [OS=Homo sapiens]		1	1	119.1
P62861	FAU	40S ribosomal protein S30 [OS=Homo sapiens]		17	1	6.6
P05109	S100A8	Protein S100-A8 [OS=Homo sapiens]		12	1	10.8
Q9GZT3-1	SLIRP	SRA stem-loop-interacting RNA-binding protein, mitochondrial [OS=Homo sapiens]		9	1	12.3
P02768-1	ALB	Serum albumin [OS=Homo sapiens]	388.410571	51	25	69.3
O14791	APOL1	Apolipoprotein L1 [OS=Homo sapiens]	202.377639	23	10	43.9
P04264	KRT1	Keratin, type II cytoskeletal 1 [OS=Homo sapiens]	132.642007	31	16	66
P13645	KRT10	Keratin, type I cytoskeletal 10 [OS=Homo sapiens]	124.721441	28	15	58.8
P09874	PARP1	Poly [ADP-ribose] polymerase 1 [OS=Homo sapiens]	113.650261	12	10	113
P07477	PRSS1	Trypsin-1 [OS=Homo sapiens]	108.936996	8	1	26.5
P35527	KRT9	Keratin, type I cytoskeletal 9 [OS=Homo sapiens]	105.7204	23	9	62
P19338	NCL	Nucleolin [OS=Homo sapiens]	101.580752	17	8	76.6
Q00839	HNRNPU	Heterogeneous nuclear ribonucleoprotein U [OS=Homo sapiens]	83.6634318	15	6	90.5
P62805	HIST1H4A; HI	histone H4 [OS=Homo sapiens]	67.1559719	43	4	11.4
P22314	UBA1	Ubiquitin-like modifier-activating enzyme 1 [OS=Homo sapiens]	65.9461702	19	10	117.8
P35908	KRT2	Keratin, type II cytoskeletal 2 epidermal [OS=Homo sapiens]	64.429637	19	11	65.4
P11586	MTHFD1	C-1-tetrahydrofolate synthase, cytoplasmic [OS=Homo sapiens]	53.4815354	8	4	101.5
A5A3E0	POTEF	POTE ankyrin domain family member F [OS=Homo sapiens]	47.5354782	3	3	121.4
O76039-1	CDKL5	Cyclin-dependent kinase-like 5 [OS=Homo sapiens]	36.1600054	3	2	115.5
Q13885	TUBB2A	Tubulin beta-2A chain [OS=Homo sapiens]	33.8280027	27	10	49.9
Q71U36	TUBA1A	tubulin alpha-1A chain [OS=Homo sapiens]	29.1847414	12	5	50.1
P49321	NASP	Nuclear autoantigenic sperm protein [OS=Homo sapiens]	27.3420823	17	6	85.2
P53396-1	ACLY	ATP-citrate synthase [OS=Homo sapiens]	26.9027985	13	7	120.8
Q08211	DHX9	Atp-dependent rna helicase a [OS=Homo sapiens]	21.7016412	12	8	140.9
P52701	MSH6	DNA mismatch repair protein MSH6 [OS=Homo sapiens]	20.7625162	1	1	152.7
Q15459	SF3A1	splicing factor 3A subunit 1 [OS=Homo sapiens]	20.1481919	3	1	88.8
O43196	MSH5	mutS protein homolog 5 [OS=Homo sapiens]	19.6154253	1	1	92.8
O15067	PFAS	Phosphoribosylformylglycinamidine synthase [OS=Homo sapiens]	19.1597271	9	5	144.6
P13647	KRT5	keratin, type II cytoskeletal 5 [OS=Homo sapiens]	16.6180227	9	5	62.3
P84243	H3F3A; H3F3	histone H3.3 [OS=Homo sapiens]	16.5214544	15	2	15.3
P42704	LRPPRC	Leucine-rich PPR motif-containing protein, mitochondrial [OS=Homo sapiens]	16.1121597	3	2	157.8
P49588	AARS	Alanine--tRNA ligase, cytoplasmic [OS=Homo sapiens]	15.5422068	6	3	106.7
Q13263	TRIM28	Transcription intermediary factor 1-beta [OS=Homo sapiens]	15.4687786	9	4	88.5
Q92598	*HSPH1*	*Heat shock protein 105 kDa [OS=Homo sapiens]*	*14.4481821*	*7*	*3*	*96.8*
P34932	HSPA4	Heat shock 70 kDa protein 4 [OS=Homo sapiens]	13.7863085	10	4	94.3
O14980	XPO1	Exportin-1 [OS=Homo sapiens]	13.7574254	5	3	123.3

P43243	MATR3	Matrin-3 [OS=Homo sapiens]	12.994496	9	3	94.6
P41252	IARS	isoleucine--tRNA ligase, cytoplasmic [OS=Homo sapiens]	12.8959972	7	4	144.4
Q16531	DDB1	DNA damage-binding protein 1 [OS=Homo sapiens]	11.7466015	9	5	126.9
O00410	IPO5	Importin-5 [OS=Homo sapiens]	11.3714115	6	4	123.6
P49736	MCM2	DNA replication licensing factor mcm2 [OS=Homo sapiens]	11.1598395	2	1	101.8
Q15393-1	SF3B3	Splicing factor 3B subunit 3 [OS=Homo sapiens]	10.6026015	8	4	135.5
Q14697-1	GANAB	Neutral alpha-glucosidase AB [OS=Homo sapiens]	10.5712287	6	3	106.8
P22102-1	GART	trifunctional purine biosynthetic protein adenosine-3 [OS=Homo sapiens]	10.2655533	5	3	107.7
Q15029	EFTUD2	116 kDa U5 small nuclear ribonucleoprotein component [OS=Homo sapiens]	10.2630293	10	5	109.4
P02787	TF	Serotransferrin [OS=Homo sapiens]	10.1085524	5	3	77
Q13435	SF3B2	Splicing factor 3b subunit 2 [OS=Homo sapiens]	9.13695586	2	1	100.2
Q92900	UPF1	Regulator of nonsense transcripts 1 [OS=Homo sapiens]	8.04480622	9	5	124.3
P33991	MCM4	DNA replication licensing factor MCM4 [OS=Homo sapiens]	6.94653954	7	3	96.5
P69905	HBA2; HBA1	Hemoglobin subunit alpha [OS=Homo sapiens]	6.89451139	17	2	15.2
Q14683	SMC1A	structural maintenance of chromosomes protein 1a [OS=Homo sapiens]	6.27248882	5	4	143.1
Q16778	HIST2H2BE	Histone H2B type 2-E [OS=Homo sapiens]	5.97123358	6	1	13.9
O43707	ACTN4	Alpha-actinin-4 [OS=Homo sapiens]	5.73016621	4	2	104.8
O75533-1	SF3B1	splicing factor 3B subunit 1 [OS=Homo sapiens]	5.54739528	4	2	145.7
P55884	EIF3B	Eukaryotic translation initiation factor 3 subunit B [OS=Homo sapiens]	5.51285366	3	1	92.4
P68104	EEF1A1	Elongation factor 1-alpha 1 [OS=Homo sapiens]	5.34856084	5	2	50.1
P11142-1	HSPA8	Heat shock cognate 71 kDa protein [OS=Homo sapiens]	5.23391921	6	3	70.9
P02533	KRT14	Keratin, type I cytoskeletal 14 [OS=Homo sapiens]	5.02289563	9	5	51.5
P04179	SOD2	Superoxide dismutase [Mn], mitochondrial [OS=Homo sapiens]	4.84022955	10	2	24.7
Q7L014	DDX46	probable ATP-dependent RNA helicase DDX46 [OS=Homo sapiens]	4.83273255	5	3	117.3
Q9Y5B9	SUPT16H	FACT complex subunit SPT16 [OS=Homo sapiens]	4.41694069	8	5	119.8
Q6EKJ0-1	GTF2IRD2B	General transcription factor II-I repeat domain-containing protein 2B [OS=Homo s	4.31374658	2	1	107.2
P08670	VIM	Vimentin [OS=Homo sapiens]	4.27701519	8	3	53.6
P14618	PKM	Pyruvate kinase PKM [OS=Homo sapiens]	4.26266922	9	3	57.9
P10599-1	TXN	thioredoxin [OS=Homo sapiens]	4.11794684	21	2	11.7
P08238	HSP90AB1	Heat shock protein HSP 90-beta [OS=Homo sapiens]	3.79399472	4	2	83.2
P23528	CFL1	Cofilin-1 [OS=Homo sapiens]	3.65619408	7	1	18.5
P53621-1	COPA	coatomer subunit alpha [OS=Homo sapiens]	3.5922249	5	3	138.3
O43491-1	EPB41L2	band 4.1-like protein 2 [OS=Homo sapiens]	3.57832892	6	2	112.5
P13639	EEF2	Elongation factor 2 [OS=Homo sapiens]	3.33921278	9	3	95.3
P31944	CASP14	Caspase-14 [OS=Homo sapiens]	3.31886932	5	1	27.7
O15042	U2SURP	U2 snRNP-associated SURP motif-containing protein [OS=Homo sapiens]	3.18464583	2	1	118.2
P07437	TUBB	tubulin beta chain [OS=Homo sapiens]	3.17515395	24	9	49.6
P05141	SLC25A5	ADP/ATP translocase 2 [OS=Homo sapiens]	3.14904728	4	1	32.8
P16402	HIST1H1D	Histone H1.3 [OS=Homo sapiens]	3.1109134	5	1	22.3
Q9UQE7	SMC3	Structural maintenance of chromosomes protein 3 [OS=Homo sapiens]	3.02659733	2	1	141.5
P09382	LGALS1	Galectin-1 [OS=Homo sapiens]	3.02494487	7	1	14.7
P07814	EPRS	Bifunctional glutamate/proline--tRNA ligase [OS=Homo sapiens]	2.90296688	1	1	170.5
Q1KMD3	HNRNPUL2	heterogeneous nuclear ribonucleoprotein U-like protein 2 [OS=Homo sapiens]	2.84492755	3	2	85.1
P04406-1	GAPDH	glyceraldehyde-3-phosphate dehydrogenase [OS=Homo sapiens]	2.46017278	4	1	36
O60763-1	USO1	General vesicular transport factor p115 [OS=Homo sapiens]	2.37624236	2	1	107.8
Q99460	PSMD1	26S proteasome non-ATPase regulatory subunit 1 [OS=Homo sapiens]	2.21700238	2	1	105.8
P62937	PPIA	peptidyl-prolyl cis-trans isomerase A [OS=Homo sapiens]	2.12979515	7	2	18
P07864	LDHC	L-lactate dehydrogenase C chain [OS=Homo sapiens]	2.10698869	4	1	36.3
P06748	NPM1	Nucleophosmin [OS=Homo sapiens]	2.10094156	10	1	32.6
P67936	TPM4	Tropomyosin alpha-4 chain [OS=Homo sapiens]	2.08011158	4	1	28.5
Q16555-1	DPYSL2	Dihydropyrimidinase-related protein 2 [OS=Homo sapiens]	1.9992287	5	2	62.3
Q9GZR7	DDX24	ATP-dependent RNA helicase DDX24 [OS=Homo sapiens]	1.97128908	2	1	96.3
Q14203	DCTN1	Dynactin subunit 1 [OS=Homo sapiens]	1.89740884	4	3	141.6
P42285	SKIV2L2	Superkiller viralicidic activity 2-like 2 [OS=Homo sapiens]	1.85533284	1	1	117.7
Q8N766-1	EMC1	ER membrane protein complex subunit 1 [OS=Homo sapiens]	1.83574442	3	2	111.7
Q15424-1	SAFB	Scaffold attachment factor B1 [OS=Homo sapiens]	1.81206808	4	2	102.6
O75400	PRPF40A	pre-mRNA-processing factor 40 homolog A [OS=Homo sapiens]	1.7583324	2	1	108.7
Q9H0D6-1	XRN2	5'-3' exoribonuclease 2 [OS=Homo sapiens]	1.7143534	2	1	108.5
P05023	ATP1A1	Sodium/potassium-transporting ATPase subunit alpha-1 [OS=Homo sapiens]	1.70566678	1	1	112.8
O60282-1	KIF5C	kinesin heavy chain isoform 5C [OS=Homo sapiens]	1.70432435	1	1	109.4
P04075	ALDOA	fructose-bisphosphate aldolase A [OS=Homo sapiens]	1.66971982	4	1	39.4
P60903	S100A10	Protein S100-A10 [OS=Homo sapiens]	1.61261458	14	1	11.2
P23246-1	SFPQ	splicing factor, proline- and glutamine-rich [OS=Homo sapiens]	1.60588631	3	1	76.1
P68871	HBB	Hemoglobin subunit beta [OS=Homo sapiens]	1.58542815	6	1	16
A1A4F0-2	C3orf55; POLI	Isoform 2 of Putative uncharacterized protein POLC2L [OS=Homo sapiens]	1.55184773	7	1	13.6
Q8WUM0	NUP133	Nuclear pore complex protein Nup133 [OS=Homo sapiens]	1.53797188	2	1	128.9
Q5JPE7	NOMO2	NODAL modulator 2 [OS=Homo sapiens]	1.50266495	2	1	139.4
O95163	IKBKAP; ELP1	Elongator complex protein 1 [OS=Homo sapiens]	1.49982401	3	2	150.2
A8MTJ3	GNAT3	Guanine nucleotide-binding protein G(T) subunit alpha-3 [OS=Homo sapiens]	1.49636314	3	1	40.3
Q06830	PRDX1	peroxiredoxin-1 [OS=Homo sapiens]	1.46451637	6	1	22.1
Q32MZ4	LRRFIP1	Leucine-rich repeat flightless-interacting protein 1 [OS=Homo sapiens]	1.41588889	2	1	89.2
P56385	ATP5I	ATP synthase subunit e, mitochondrial [OS=Homo sapiens]	1.3714258	19	2	7.9
Q9NSE4	IARS2	Isoleucine--tRNA ligase, mitochondrial [OS=Homo sapiens]	1.34777685	3	1	113.7
P62158	CALM3; CALM	Calmodulin [OS=Homo sapiens]	1.23312993	22	2	16.8
P01876	IGHA1	immunoglobulin heavy constant alpha 1 [OS=Homo sapiens]	1.153241	3	1	37.6
P39019	RPS19	40S ribosomal protein S19 [OS=Homo sapiens]	1.0087972	6	1	16.1
P07205	PGK2	Phosphoglycerate kinase 2 [OS=Homo sapiens]	1	4	1	44.8