Published in final edited form as:

J Proteome Res. 2018 December 07; 17(12): 4061–4071. doi:10.1021/acs.jproteome.8b00442.

Chromosome 17 Missing Proteins: Recent Progress and Future Directions as part of the Next-50MP Challenge

Omer Siddiqui^{1,2,+}, Hongjiu Zhang^{2,+}, Yuanfang Guan^{2,3}, and Gilbert S. Omenn^{2,3,4,*}

¹Department of Electronic Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 48109

²Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, 48109

³Department of Internal Medicine, University of Michigan, Ann Arbor, MI, 48109

⁴Department of Human Genetics and School of Public Health, University of Michigan, Ann Arbor MI, 48109

Abstract

The Chromosome-centric Human Proteome Project (C-HPP) announced in September 2016 an initiative to accelerate progress on detection and characterization of neXtProt PE2,3,4 “missing proteins” (MPs) with a mandate to each chromosome team to find about 50 MPs over two years. Here we report major progress toward the next-50 MP Challenge with 43 newly-validated Chr 17 PE1 proteins, of which 25 were based on mass spectrometry, 12 on protein-protein interactions, 3 on a combination of MS and PPI, and 3 with other types of data. Notable among these new PE1 proteins were 5 keratin-associated proteins, a single olfactory receptor, and five additional membrane-embedded proteins. We evaluate the prospects of finding the remaining 105 MPs coded for on Chr 17, focusing on mass spectrometry and protein-protein interaction approaches. We present a list of 35 prioritized MPs with specific approaches that may be used in further MS and PPI experimental studies. Additionally, we demonstrate how *in silico* studies can be used to capture individual peptides from major data repositories, documenting 1MP that appears to be a strong candidate for PE1. We are close to our goal of finding 50 MPs for Chr 17.

Graphical Abstract

*Contact: gomenn@med.umich.edu.

⁺Co-first authors

Supporting Information

The following supporting information is available free of charge at the ACS website, <http://pubs.acs.org>.

Keywords

C-HPP, Chromosome-centric Human Proteome Project; neXt-50 Missing Proteins Challenge; PPI, Protein-Protein Interactions; neXtProt Protein existence evidence levels, PE1, 2, 3, 4; SRMATlas, Selective Reaction Monitoring; GTEx, Genotype-Tissue Expression; Human Protein Atlas; PeptideAtlas

INTRODUCTION

The HPP Human Proteome Project (HPP) is a collaborative effort of teams from around the world to make major progress in understanding the human proteome¹. There are two main goals: first, to obtain definitive evidence of at least one protein product from each protein-coding gene in the human genome and identify and characterize the functions and interactions of the sequence variants, splice variants, and post-translational modifications, completing the protein parts list; second, to integrate proteomics into multi-omics studies of protein interactions, networks, and pathways in health and disease.

The HPP relies on PeptideAtlas² for annual updates of protein identifications based on mass spectrometry (MS), including standardized reanalysis of the raw datasets with the TransProteomic Pipeline^{3,4} and compliance with the HPP Guidelines for Interpretation of MS Data⁵. neXtProt⁶ is the curator for human proteins, drawing upon MS data from PeptideAtlas and incorporating and evaluating evidence from multiple other types of protein studies⁷. neXtProt classifies predicted proteins into five categories. Each protein coding gene is assigned a “Protein Existence” (PE) score based on the data available for review. PE1 signifies validation at the protein level. Notably, neXtProt utilizes high-quality protein-protein interaction (PPI) data as determined by the IntAct Molecular Interaction Database of

yeast 2-hybrid (Y2H) assays and other methods. PE2 signifies that there is sufficient evidence of transcription without protein evidence. PE3 recognizes existence of protein products as validated orthologs in closely related species. PE4 signifies evidence from gene models. We exclude PE5 (uncertain/dubious) genes from HPP analyses⁸.

The goal is to find sufficient evidence to categorize every protein-coding gene as PE1. The proteins coded by PE2–4 genes are referred to as “missing proteins” (MP), the targets of the C-HPP neXt-50 MP Challenge, announced in September 2016, to stimulate each Chromosome team to focus on finding Guidelines-compliant evidence for 50 MPs over a period of two years^{9,10}.

In this paper, we report and analyze the progress on Chromosome 17, showing near completion of the neXt-50 goal for Chr 17. Following the lead of Duek et al.¹⁰, and focusing on both MS and PPI data, we devise strategies for further identifications and propose a list of the missing proteins on Chr 17 that should be the most amenable to detect and validate as PE1.

EVALUATING PROGRESS FOR CHR 17 SINCE THE ANNOUNCEMENT OF THE C-HPP NEXT-50 MISSING PROTEIN CHALLENGE IN SEPTEMBER 2016

The neXtProt release 2016–01 listed 148 missing proteins on Chr 17, with 125 PE2, 17 PE3, and 6 PE4. In 2013, the HPP decided to disregard PE5 entries (dubious/uncertain genes) from the denominator of total predicted proteins (PE1–4)⁸. The most recent neXtProt release 2018–01 has 105 missing proteins on Chr 17, of which 88 are PE2, 13 PE3, and 4 PE4 among the total of 1166 predicted protein entries. Much progress has been made since the announcement of the neXt-50 MP challenge with neXtProt release 2016–01 as the baseline; indeed, 43 missing proteins on Chr 17 have been validated as PE1. The methods by which these 43 entries were upgraded to PE1 reveal strategies that can be used to identify at least some of the remaining 105 missing proteins. Figure 1 shows how these 43 entries were upgraded to PE1.

Generally, MS has been the principal means for validation at PE1⁷. The significant representation of PPI is, in part, due to the difficulties with MS on certain classes of missing proteins. Membrane-embedded proteins are difficult to identify with MS due to difficulty of solubilization of these hydrophobic proteins and the paucity of the basic amino acids Lys and Arg for generation of uniquely-mapping peptides of at least 9 aa in length with trypsin. Various protocols have been suggested to specifically target membrane peptides for observation by MS following Triton-X100 solubilization¹¹ or multi-protease digestion with LysargiNase and GluC¹². Other than olfactory receptors, five membrane-embedded proteins are among the 43 new PE1 proteins from Chr 17: SLC16A11, SLC25A39, TMEM107, TMIGD1, and SMIM5. SLC25A39 was identified by MS, SMIM5 by PPI, and the other three through other types of studies such as disease association analysis (Figure 1B). However, TMEM107 has 2 peptides of 8 aa and 19 aa, corresponding to the 2 peptides proposed by SRMatlas.

In addition to these five membrane-embedded proteins, several families of proteins stand out in the MP list of Chromosome 17. Olfactory receptors (ORs) represent 16 of the 1166 total proteins and 13 of the current 105 missing proteins, as well as OR3A4P at PE5. One olfactory receptor, OR1D4, was missing in neXtProt release 2016–01 but classified as PE1 in neXtProt release 2018–01 through PPI in a yeast-two-hybrid (Y2H) assay (IntAct ID: EBI-11988863, EBI-5663627), so lacking tissue expression. The other Chr 17 PE1 OR is OR1D2, discovered in spermatozoa through chemotaxis studies instead of MS or PPI¹³. Since 2013, ORs have been the most challenging identification targets¹⁴. These proteins contribute to important biological functions such as odor recognition and discrimination, stress response, homeostasis, as well as sexual behavior¹⁵. Based on sequence analysis, they are classified as G-protein-coupled receptors. Such proteins contain seven transmembrane helices and are tightly integrated into plasma membranes, making them hard to solubilize for common proteomic assays. Additionally, OR RNA transcript data generally show low to negligible levels of transcription. Their spatially restricted expression in highly specific groups of cells in inaccessible tissue sites make them even more difficult to detect.

A second family of genes that makes up a significant group of Chr 17 is keratin associated proteins (KRTAP), accounting for 33 of all 1166 protein entries in the chromosome. Of the 43 newly validated missing proteins in the neXtProt release 2016–01, there are five keratin associated proteins, along with one keratin, KRT37. Three of the five KRTAPs were validated exclusively by PPI, one by MS, and one by both PPI and MS. There are 8 more KRTAP predicted proteins among the 105 still-missing proteins from Chr 17. Of the 5 KRTAPs which were detected, as well as 5 of the 8 missing KRTAPs on Chr 17 which have entries in GTEx, all ten have median transcription levels less than 1 TPM in every tissue studied (though skin does have many high individual values). Additionally, there are relatively few uniquely-mapping peptides, due to high sequence homology. The overrepresentation of PPI over MS for the 5 new PE1 KRTAPs suggests that PPI data are a more viable means of detecting and then validating the remaining 8 KRTAPs.

Five C17orf5 on Chr 17 were upgraded to PE1 based on MS evidence since 2016. Vandenbrouck, et al. detected C17orf105 in human spermatozoa¹⁶. Wei, et al. detected C17orf50, C17orf74, and C17orf98 in human testis¹⁷. Hendriks, et al. detected C17orf51 in the study of SUMOylation in HeLa and U2OS cell lines¹⁸. As they do not make up a biological family, there is no clear biological intuition for why MS was used more than PPI, other than MS is more common for PE1 detection in general. In terms of tissue expression, four of these genes were most transcribed in testis, and the fifth, C17orf51, was most transcribed in cerebellum, followed by testis (according to GTEx). Thus the effectiveness of MS has been enhanced by several recent studies searching for MPs specifically in sperm and testis samples^{12,16,19}. Separately, we note that the gene name of the entry in neXtProt for C17orf74 has been changed to SPEM2.

We focus next on the 39 genes with corresponding entries in GTEx²⁰. The average level of RNA transcripts of the missing proteins validated by MS in this list of 39 is about three times that of the missing proteins validated by PPI at 96 vs 33 median TPM in the tissue for which each entry had its highest transcription, according to GTEx. For 12 of the 39 entries, the tissue with highest median transcript level was testis. The second most common tissue

for highest expression was cerebellum, at 3 entries. As will be discussed, testis and cerebellum are also the two highest expressing tissues for the remaining 105 missing proteins.

STRATEGIES FOR DETECTING ADDITIONAL CHROMOSOME 17 MISSING PROTEINS

There are 105 missing proteins (PE2,3,4) on Chr 17 in neXtProt release 2018–01 (Figure 2). Among these missing proteins, certain families and groups of proteins are over-represented due to fundamental difficulties with their detection. These groups can be better understood through biological considerations and an examination of how other members of these groups have been identified recently. Of the 105 missing proteins, 88 are PE2 and should have transcript level data. Transcription patterns can be leveraged to identify tissues to target as well as understand why some missing proteins may be easier or more difficult to detect. The two primary means of validating entries at PE1 are through MS data of two non-nested proteotypic peptides of length at least 9 aa and through high-quality protein protein interaction data. Combining the insights from protein families and transcript data into these workflows can guide further work in finding missing proteins.

Families represented in the Missing Proteins in neXtProt release 2018–01

Certain families of proteins make up significant portions of the 105 current MPs on Chr 17. As noted, 13 of the 105 missing proteins are olfactory receptors. Of the 16 total OR genes, 14 are placed contiguously on the short arm of Chr 17 at 17p13.3 and 17p13.2; the other two are together at 17q22. As noted above, olfactory receptors are difficult to validate due to spatially restricted expression as well as low transcription. One theory explaining the apparent low transcription is that very few cells within certain tissues may express many of the olfactory receptors, presumably enriched in the olfactory epithelium of the forebrain. Thus when a sample is taken for RNA sequencing, the transcripts from the highly specific cells are “diluted”, making it appear as if ORs have low transcription. Searches for OR transcripts and proteins have so far been unproductive. A major effort by Pandey and colleagues in Bangalore, India, has been underway for at least 3 years (personal communication). In Duek et al from Chromosome teams 2 and 14, ORs were placed in a non-detectable low-priority category.

Of the 105 Chr 17 missing proteins, 13 belong to the TBC1D family; 10 of these 13 are located in clusters at 17q12. The TBC1D family is characterized by sharing a common TBC1 domain. Their functions vary, with many acting as GTPase-activating proteins for RAB family proteins according to functional annotations in neXtProt. None of the 43 recently validated Chr 17 missing proteins came from this family, suggesting difficulty with detecting these 13. Much of the difficulty stems from similarity in amino acid sequences of the members of the family, causing many of the peptides detected in MS to not be uniquely mapping. In fact, nine of these entries, TBC1D3B, TBC1D3C, TBC1D3D, TBC1D3E, TBC1D3F, TBC1D3G, TBC1D3H, TBC1D3I and TBC1D3K, lack two proteotypic peptides of length at least 9 aa digested by trypsin (Figure 3). Thus, identification of these three proteins by MS methods is unlikely. Interestingly, there are 44 TBC1D entries across all the

chromosomes in neXtProt. Of the 44, 15 are on Chr 17, and 13 of these 15 are currently missing proteins. Every single one of the remaining 29 TBC1D entries on other chromosomes is validated at PE1. These 29 entries are distributed across 17 different chromosomes with 4 on Chr 4, 3 on Chr 6, and either one or two on the remaining 15 chromosomes. Across the proteome, the TBC1D proteins are a diverse group: their classification by having the TBC1 domain does not correlate much with structure or function. One possible explanation for why the Chr 17 TBC1D entries are so similar in terms of sequence is that they may have arisen over time through gene duplication recently enough for the sequences to not diverge much. Supporting this idea, all nine of the TBC1D genes on Chr 17 which don't have two tryptic proteotypic peptides of length at least 9 aa are clustered at 17q12.

Next, 8 of the currently missing proteins are keratin associated proteins. Chr 17 has 27 keratins, 33 keratin associated proteins, and 1 keratin-like protein. All 61 of these genes are located together at 17q21.2 in a nearly continuous block, interrupted just once by TMEM99. Within the block, the keratin associated proteins are all in one block which is sandwiched by keratins on both sides. As noted, PPI were used to detect four of the five KRTAPs plus one keratin validated since 2016. Additionally as noted, all of the KRTAPs with corresponding entries in GTEx show median transcription levels lower than 1 TPM in every tissue listed. As with the TBC1D group, similar amino acid sequences among members of the KRTAP family lead to fewer proteotypic peptides, causing issues with MS based validation, as 3 of these 8 KRTAPs (KRTAP2-1, KRTAP2-2, KRTAP4-9) do not have two at least two predicted tryptic proteotypic peptides of length at least 9 aa (Figure 3). Looking forward, continued use of PPI may be productive for the KRTAP MPs.

Chromosome 17 has 21 C17orfs, of which 7 are MPs as of release 2018-01. C17orfs are overrepresented in MPs at 1 in 3 being a missing protein, whereas less than 1 in 10 are missing proteins across all Chr 17 entries. There is no clear reason for why this is so, as C17orfs are not a biological family and do not have similar biochemistry.

Of the 105 missing proteins, ten are integral to plasma membranes: 6 from the TMEM family, 1 from SMIM, and 3 from SLC. Due to the issues previously discussed, MS approaches aren't effective for membrane embedded proteins. PPI and non-traditional MS approaches specifically targeting membrane bound proteins are the appropriate options for these ten missing proteins.

Targeted Identification of Missing Proteins by Mass Spectrometry

For identification by MS, two non-overlapping proteotypic peptides of length at least 9 amino acids need to be observed and not be accounted for by sequence variants or isobaric PTMs of abundant PE1 proteins. The majority of MS is done with trypsin digestion; as such, it is important to note that, of the 105 Chr 17 MPs, 16 do not have two tryptic non-overlapping proteotypic peptides of length at least 9 aa: FAM106A, KRTAP2-1, KRTAP2-2, KRTAP4-9, MTRNR2L1, SCGB1C2, SMIM6, TBC1D3B, TBC1D3C, TBC1D3D, TBC1D3E, TBC1D3F, TBC1D3G, TBC1D3H, TBC1D3I and TBC1D3K (Figure 3). The reason for 12 of these genes, namely the keratin associated protein and the TBC1D genes, is that they are part of highly similar families of genes. As such, their peptide sequences are

very similar and lack proteotypic peptides as discussed previously. FAM106A generates 6 tryptic peptides of length at least 9 aa, of which one is proteotypic. The other five overlap with variants of FAM106CP and FAM106A. All of SCGB1C2's tryptic peptides of length at least 9 aa overlap with variants of SCGB1C1. MTRNR2L1 and SMIM6 are relatively short proteins with length 24 and 62 amino acids, respectively, and each generates only one tryptic peptide of length at least 9 aa.

The natural next approach for these 16 genes would be to consider a protease other than trypsin for digestion before the MS. The second most commonly used protease is LysC¹²; using a different protease may in some cases give a proteotypic peptide of length at least 9aa. Still, due to the inherent issue with similar sequences this doesn't happen often. In some cases, such as SCGB1C2, one gene is completely subsumed by another, making it impossible to validate by MS approaches. In general, MS with any protease does not seem fruitful for these 16 due to the lack of unique sequences. Thus other approaches must be considered for this list of 16.

Excluding these 16 leaves 89 genes that could be detected by MS, possibly with traditional protocols using trypsin. Of these 89, 27 have already had one proteotypic peptide of length at least 9 observed, as noted in PeptideAtlas and neXtProt (Figure 3). A convenient approach to identify these 27 MPs is to search already collected MS data, ideally from studies registered in ProteomeXchange²¹ and downloadable from PRIDE²² or other public repository, and find high-quality spectra of a second peptide fitting the criteria, at an appropriate level of confidence. We call this scheme⁷ looking for a “stranded peptide” to pair up with a single peptide in PeptideAtlas. All confirmed peptides in PeptideAtlas are already shared with neXtProt. Table 1 shows the results of our search in PRIDE for second proteotypic peptides for this list of 27 MPs on Chr 17.

These peptides have corresponding entries in SRMATlas. Results of our examination and comparison of the spectra in PRIDE and SRMATlas for the 7 second peptides in Table 1 are given in Figure 4 and Supplementary Figures 1–6. Only one peptide (DLLPSQTASSLCISR, uniquely mapped to protein PIRT/NX_P0C851) has MS/MS spectra in PRIDE/accessed in GPMdb closely matching those of the synthetic peptide in SRMATlas (Figure 4). This peptide comes from cancer cell lines by Klaeger, et al, deposited as PXD005336 in ProteomeXchange²³. The remaining second peptides seem to lack sufficiently high-quality, well-matched spectra in either SRMATlas or MassIVE-KB and are provided in the Supplementary Figures 1–6. This result is one more step toward reaching the goal of 50 MPs achieving neXtProt PE1 status for Chromosome 17 by neXtProt release 2019–01.

For new studies, data regarding transcript expression can guide choice of tissues to target for MP search. For this part of the analysis, we considered only PE2 missing proteins, i.e. missing proteins with validated transcript evidence. We treat a median transcription of less than 1 TPM in a tissue as negligible and round down to 0. Of the 82 PE2 missing proteins with corresponding entries in GTEx, 21 genes had negligible transcription in each of the 53 tissues measured. These 21 genes would be difficult, if not impossible, to find anywhere due to low expression levels. Notably, 9 of these are olfactory receptors and 4 are keratin associated proteins. An encouraging finding is the frequency of testis or cerebellum as the

tissue with the highest median TPM value for missing proteins. Specifically, 24 of the 82 have testis as the tissue with the highest median TPM values. Remarkably, 13 of these 24 were expressed *exclusively* in testis, with median expression levels lower than 1 TPM in every other tissue. The 24 are (with exclusives underlined): ANKFN1, C17orf64, CCDC144A, CDRT1, CDRT15L2, CNTD1, EFCAB13, FAM106A, FBXO39, FBXO47, LRRC3C, MFSD6L, OR3A2, SLC35G3, SLC35G6, SMIM6, SPDYE4, TBC1D26, TBC1D28, TBC1D29, TBC1D3B, TBC1D3G, TMEM95, TMEM99. The special role of testis and sperm tissues in the search for missing proteins is known, and targeted studies have been conducted^{16,19}; further search by these investigators and others should be productive. However, we note that many of the 24 MPs most transcribed in testis are difficult to detect by MS as there are 5 members of TBC1D, 1 olfactory receptor, and multiple other membrane-embedded proteins. Using the data in Figure 3, all 13 testis exclusive MPs have at least two tryptic proteotypic peptides of length 9 aa, and 3 of them have one observed proteotypic peptide of length at least 9 aa already annotated in PeptideAtlas and neXtProt. Cerebellum was the tissue with the highest transcription levels for 8 of the 82 missing proteins: CCDC144B, KCNAB3, MEIS3P2, RNF112, SGK494, SMCR5, TBC1D3F, TIAF1. A complete listing of the top three transcribed tissues for each PE2 MP with an entry in GTEx and a corresponding complete list using Human Protein Atlas²⁴ are given in the Supplementary Tables.

Identification by Protein-Protein Interaction

A major data source that has contributed increasingly to the missing protein identifications in the past two years is PPI data. Of all the 43 Chr 17 proteins that have been identified and promoted to the PE1 level in the past two years by neXtProt, 12 of them are solely based on PPI data, and 3 of them have both PPI and MS evidence (Figure 1, Table 3). Among 12 proteins with just PPI evidence, 11 proteins were upgraded to PE1 based on PPI data from IntAct²⁵, cross-referenced with BioGrid²⁶. For these 11 proteins, all gold interaction evidence listed in neXtProt and the majority of non-gold evidence data were collected from yeast two-hybrid (Y2H) assays, achieving an IntAct score of 16 (two pieces of Y2H experiment evidence). A major source of these data is the Human Interactome Project^{27,28}, conducted by the Center for Cancer Systems Biology at Harvard University, focused on “generating a first reference map of the human protein-protein interactome network”. Their approach is to combine Y2H assay and orthogonal validation by alternative binary assays. These 11 proteins are accepted in UniProt as well as neXtProt (release 2018-01) even though no orthogonal experimental evidence seems to have been available. The 12th, TEX19, was upgraded to PE1 by UniProt based on a published study²⁹ in the human embryonic kidney cell line HEK293, showing its function and its interactions with other proteins during embryonic development (without IntAct score). Of the 3 proteins which were upgraded to PE1 based on both MS and PPI evidence, CLEC10A and KRTAP1-1 had Y2H data (IntAct score = 16), while KCNJ16 was found to form dimers with KCNJ10 in a published study on MDCK II and HEK293T cells³⁰.

The contribution from Y2H assay data provides an alternative approach to MS for identifying the remaining missing proteins. Current MS-based methods rely on extracting protein products from cell lysates and identifying MS spectra. For proteins lacking

solubility, tryptic sites, or uniquely mapping peptides for identification, their identification through traditional MS-based approach is not feasible. Y2H relies on transforming the coding sequence into vectors to be expressed inside yeast cells³¹. The proteins are fused with designed binding and activation domains that enable expression of a reporter gene. Therefore, the interaction of the protein products is confirmed through the activation of the reporter gene. Researchers have developed multiple variations of Y2H for studying proteins of different subcellular localization^{32,33}. As a widely used experimental technique for studying PPI, its sensitivity and specificity have been well calibrated²⁷. In order to ensure the quality of PPI data, the IntAct database adopted a scoring metric to assess the confidence of reported protein-protein interactions (<https://www.ebi.ac.uk/intact/pages/faq/faq.xhtml>). The IntAct database requires two pieces of evidence from orthogonal assay techniques to validate the protein-protein interaction, with at least one piece of evidence showing direct physical contact. Here we discuss specific proteins in Table 2 to see the experimental settings in the context of protein groups in order to apply similar methods to other missing proteins.

Among the 11 proteins identified solely based on PPI, 3 are keratin-associated proteins KRTAP17–1, KRTAP2–4, and KRTAP9–8. The current annotations of these 3 KRTAPs in the neXtProt database contain a large number of interaction data. Among the annotated interactors, other KRTAPs and late cornified envelope proteins are the most commonly seen families: these partners often contain cysteine-rich regions, which allow disulfide bonding. Especially KRTAPs are known for their interactions with each other³⁴. These partner proteins would be good baits in a Y2H assay, effectively capturing target KRTAPs. Most KRTAPs are proteins of about 100 amino acid residues. Their sequences contain long and highly repetitive cysteine-, serine-, glycine-rich regions. We looked into interaction data of 8 currently unidentified KRTAPs (Table 3). KRTAP9–1 and KRTAP9–7 show interactions with MCM2 in AP-MS experiments, as was true for KRTAP9–8. Based on our observation above that KRTAPs themselves can be good baits for other KRTAPs, we propose that KRTAP9 proteins would be good baits for identifying other KRTAP9 proteins. We caution that positive results from AP-MS may imply indirect or non-specific interactions. The other 6 unidentified KRTAPs do not have any interaction experimental data in either IntAct or BioGrid databases. Yet based on observed patterns that KRTAPs tend to form interactions with other KRTAPs, we believe KRTAPs themselves can be good baits for assays.

Another interesting protein upgraded from PE2 to PE1 based on Y2H PPI is the olfactory receptor OR1D4. The protein belongs to the G-protein-coupled receptor family. As noted above, the family is notorious for its tight integration in the membrane, which causes great difficulty in solubilization for common proteomic assays. To make the identification even harder, many olfactory receptors have low transcript expression levels, making them even harder targets in the proteomic analysis. For example, OR1D4 has median expression levels lower than 1 TPM according to GTEx. In the current neXtProt annotation of OR1D4, the protein is upgraded to PE1 based on its interactions with many other transmembrane proteins in Y2H assays. The experiment makes the hydrophobicity of the protein an advantage, allowing effective hydrophobic bait-prey interaction³³. Besides OR1D4, there are only 3 more PE1 ORs in neXtProt release 2018–01. OR2J3 (on Chr 6) is considered to be in olfactory epithelium and associated with sensing grassy smell³⁵. OR1D2 (on Chr 17) was

identified in spermatozoa and is associated with sperm chemotaxis¹³. Neither was found through MS or Y2H. OR2AG1 (on Chr 11) show interactions with MPDZ in Y2H assays. We searched IntAct and BioGrid databases for currently unidentified olfactory receptors (Table 3). Four additional ORs (OR1A2, OR3A1, OR3A2, OR1E2) have evidence from affinity purification experiments, which implies either direct or indirect interactions. Among these, OR1A2 shows interactions with several ER membrane protein complex subunits, which may be good baits for further validation experiments. Similar to OR1D4, these proteins have negligible transcript-level expression across tissues, so identification through PPI can be much easier than identification from tissue specimens.

We observed similar interaction patterns in another membrane protein, SMIM5. SMIM5 was also upgraded last year to PE1 solely based on PPI data; two of its interaction partners are membrane proteins PRRT2 and SH3GLB1. IntAct reported that several currently unidentified TMEM proteins (TMEM92, TMEM95, TMEM99, TMEM220) show interaction with other membrane proteins (Table 3), which can be used as baits in validation experiments as well. In general, membrane-embedded proteins work well as baits for other membrane-embedded proteins in Y2H approaches designed specifically to address the difficulties presented by the membrane^{32,33}. Such an approach with Y2H shows promise for the 4 ORs and 4 TMEMs with annotations present in IntAct and Biogrid, especially due to difficulties with membrane solubilization for MS-based approaches.

A large category of proteins that remain unidentified so far is the TBC1 domain family (TBC1D). Many of these proteins have transcript-level evidence. Because of high similarity in sequences among TBC1D proteins, identification with MS remains challenging. We looked into their PPI data (Table 3). Currently TBC1D3G has multiple pieces of evidence from Y2H assays, among which its interaction with ZRANB1 (as a prey) has been measured twice with positive scores. So far, interaction with ZRANB1 is only reported for TBC1D3B among all other TBC1D proteins. While many identified TBC1D proteins show numerous interaction partners from Y2H assays, their partners are quite heterogeneous. It is reasonable to guess that identifying TBC1D through PPI approaches is feasible, but selecting suitable baits can be difficult.

High throughput Y2H may potentially be used to identify a number of assorted MPs without clear choice of bait or existing PPI annotations. Despite the fruitful results from large-scale Y2H assays, we caution that such evidence is based on artificial production of the proteins that serve as baits instead of natural expression found in particular human specimens. The Y2H method does not solve the challenge of identifying the natural production of the protein inside human tissues.

35 High-priority Chromosome 17 Missing Proteins

Of the 105 MPs on Chr 17, 43 are not amenable to MS; these are the 13 ORs, 13 TBC1Ds, 10 membrane-embedded proteins, 5 KRTAPs, and 2 MPs without at least two tryptic proteotypic peptides of length at least 9 aa not already covered by the other groups (Figure 5). This leaves 62 MPs for MS consideration. As discussed with the transcripts of the 43 entries upgraded to PE1 since release 2016–01, higher transcription increases the probability of MS detection. Of the 62 MPs for MS, 25 have a median transcription of at least 10 TPM

in at least one tissue, according to GTEx. We propose these 25 MPs as the proteins that should be found in MS data with the highest probability. Within these 25, 10 already have one observed proteotypic peptide of length at least 9 aa already annotated in neXtProt; thus, we put these 10 as more prioritized than the other 15. In terms of transcription, 7 of the 25 are most transcribed in testis and 2 of the 25 are most transcribed in cerebellum, according to GTEx.

In the list of 43 MPs not amenable to MS, we look for PPI-based approaches for detection. As was discussed earlier and is listed in Table 3, 4 olfactory receptors (OR3A1, OR1E2, OR3A2, OR1A2) and 4 membrane-embedded proteins (TMEM220, TMEM92, TMEM99, TMEM95) show promise for detection via a similar approach as was used to detect OR1D4. Additionally, as was discussed, two KRTAPs (KRTAP9–1, KRTAP9–7) may be detected with KRTAP9–8 as a bait, similar to how KRTAP9–8 was identified. TBC1Ds and the other two MPs are excluded as current data does not show a clear PPI approach, as discussed. Thus ten MPs form our list of proteins that should be targeted with PPI. We also note, as discussed earlier, high throughput approaches with Y2H may detect an even larger group of MPs.

In sum, of the 105 MPs, 25 warrant search for additional proteotypic peptides by MS guided by high transcript expression in target tissues and 10 seem amenable to characterization by Y2H assays using these proteins as bait (**Figure 5, Table 3**).

CONCLUSION

Between the release 2016–01 and release 2018–01 of neXtProt, 43 previously missing proteins (PE2,3,4) coded by genes on Chr 17 were upgraded to PE1 status. We extensively analyzed the reasons why these MPs were successfully detected and validated. We have applied these lessons to the remaining 105 Chr 17 MPs with recommendations for both MS-based and PPI-based experiments to detect up to 35 of these MPs. By careful search of available peptide data, we propose 7 pairs of stranded peptides with PXD identifiers and spectra available in PRIDE to identify 1 additional MP for promotion to PE1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We gratefully acknowledge timely suggestions from Dr. Lydie Lane at neXtProt and reviewers.

This work is supported by NIH grants P30ES017885 and U24CA210967 (G.S.O.) and NSF 1452656 (Y.G.).

References

- (1). Paik Y-K; Jeong S-K; Omenn GS; Uhlen M; Hanash S; Cho SY; Lee H-J; Na K; Choi E-Y; Yan F; et al. The Chromosome-Centric Human Proteome Project for Cataloging Proteins Encoded in the Genome. *Nat. Biotechnol.* 2012, 30 (3), 221–223. [PubMed: 22398612]
- (2). Desiere F The PeptideAtlas Project. *Nucleic Acids Res.* 2006, 34 (90001), D655–D658. [PubMed: 16381952]

- (3). Deutsch EW; Mendoza L; Shteynberg D; Slagel J; Sun Z; Moritz RL Trans-Proteomic Pipeline, a Standardized Data Processing Pipeline for Large-Scale Reproducible Proteomics Informatics. *Proteomics Clin. Appl.* 2015, 9 (7–8), 745–754. [PubMed: 25631240]
- (4). Deutsch EW; Mendoza L; Shteynberg D; Farrah T; Lam H; Tasman N; Sun Z; Nilsson E; Pratt B; Prazan B; et al. A Guided Tour of the Trans-Proteomic Pipeline. *Proteomics* 2010, 10 (6), 1150–1159. [PubMed: 20101611]
- (5). Deutsch EW; Overall CM; Van Eyk JE; Baker MS; Paik Y-K; Weintraub ST; Lane L; Martens L; Vandenbrouck Y; Kusebauch U; et al. Human Proteome Project Mass Spectrometry Data Interpretation Guidelines 2.1. *J. Proteome Res.* 2016, 15 (11), 3961–3970. [PubMed: 27490519]
- (6). Gaudet P; Michel P-A; Zahn-Zabal M; Cusin I; Duek PD; Evalet O; Gateau A; Gleizes A; Pereira M; Teixeira D; et al. The neXtProt Knowledgebase on Human Proteins: Current Status. *Nucleic Acids Res.* 2015, 43 (Database issue), D764–D770. [PubMed: 25593349]
- (7). Omenn GS; Lane L; Overall CM; Corrales FJ; Schwenk JM; Paik Y-K; Van Eyk JE; Liu S; Snyder M; Baker MS; et al. Progress on Identifying and Characterizing the Human Proteome: 2018 Metrics from the HUPO Human Proteome Project. *J. Proteome Res.* 2018.
- (8). Lane L; Bairoch A; Beavis RC; Deutsch EW; Gaudet P; Lundberg E; Omenn GS Metrics for the Human Proteome Project 2013–2014 and Strategies for Finding Missing Proteins. *J. Proteome Res.* 2014, 13 (1), 15–20. [PubMed: 24364385]
- (9). Paik Y-K; Overall CM; Deutsch EW; Van Eyk JE; Omenn GS Progress and Future Direction of Chromosome-Centric Human Proteome Project. *J. Proteome Res.* 2017, 16 (12), 4253–4258. [PubMed: 29191025]
- (10). Duek P; Bairoch A; Gateau A; Vandenbrouck Y; Lane L Missing Protein Landscape of Human Chromosomes 2 and 14: Progress and Current Status. *J. Proteome Res.* 2016, 15 (11), 3971–3978. [PubMed: 27487287]
- (11). Zhao M; Wei W; Cheng L; Zhang Y; Wu F; He F; Xu P Searching Missing Proteins Based on the Optimization of Membrane Protein Enrichment and Digestion Process. *J. Proteome Res.* 2016, 15 (11), 4020–4029. [PubMed: 27485413]
- (12). Wang Y; Chen Y; Zhang Y; Wei W; Li Y; Zhang T; He F; Gao Y; Xu P Multi-Protease Strategy Identifies Three PE2 Missing Proteins in Human Testis Tissue. *J. Proteome Res.* 2017, 16 (12), 4352–4363. [PubMed: 28959888]
- (13). Spehr M; Gisselmann G; Poplawski A; Riffell JA; Wetzel CH; Zimmer RK; Hatt H Identification of a Testicular Odorant Receptor Mediating Human Sperm Chemotaxis. *Science* 2003, 299 (5615), 2054–2058. [PubMed: 12663925]
- (14). Baker MS; Ahn SB; Mohamedali A; Islam MT; Cantor D; Verhaert PD; Fanayan S; Sharma S; Nice EC; Connor M; et al. Accelerating the Search for the Missing Proteins in the Human Proteome. *Nat. Commun.* 2017, 8, 14271. [PubMed: 28117396]
- (15). Su C-Y; Menzies K; Carlson JR Olfactory Perception: Receptors, Cells, and Circuits. *Cell* 2009, 139 (1), 45–59. [PubMed: 19804753]
- (16). Vandenbrouck Y; Lane L; Carapito C; Duek P; Rondel K; Bruley C; Macron C; Gonzalez de Peredo A; Couté Y; Chaoui K; et al. Looking for Missing Proteins in the Proteome of Human Spermatozoa: An Update. *J. Proteome Res.* 2016, 15 (11), 3998–4019. [PubMed: 27444420]
- (17). Wei W; Luo W; Wu F; Peng X; Zhang Y; Zhang M; Zhao Y; Su N; Qi Y; Chen L; et al. Deep Coverage Proteomics Identifies More Low-Abundance Missing Proteins in Human Testis Tissue with Q-Exactive HF Mass Spectrometer. *J. Proteome Res.* 2016, 15 (11), 3988–3997. [PubMed: 27535590]
- (18). Hendriks IA; Lyon D; Young C; Jensen LJ; Vertegaal ACO; Nielsen ML Site-Specific Mapping of the Human SUMO Proteome Reveals Co-Modification with Phosphorylation. *Nat. Struct. Mol. Biol.* 2017, 24 (3), 325–336. [PubMed: 28112733]
- (19). Carapito C; Duek P; Macron C; Seffals M; Rondel K; Delalande F; Lindskog C; Fréour T; Vandenbrouck Y; Lane L; et al. Validating Missing Proteins in Human Sperm Cells by Targeted Mass-Spectrometry- and Antibody-Based Methods. *J. Proteome Res.* 2017, 16 (12), 4340–4351. [PubMed: 28891297]
- (20). GTEx Consortium. The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* 2013, 45 (6), 580–585. [PubMed: 23715323]

- (21). Vizcaíno JA; Deutsch EW; Wang R; Csordas A; Reisinger F; Ríos D; Dienes JA; Sun Z; Farrah T; Bandeira N; et al. ProteomeXchange Provides Globally Coordinated Proteomics Data Submission and Dissemination. *Nat. Biotechnol.* 2014, 32 (3), 223–226. [PubMed: 24727771]
- (22). Jones P; Côté RG; Martens L; Quinn AF; Taylor CF; Derache W; Hermjakob H; Apweiler R PRIDE: A Public Repository of Protein and Peptide Identifications for the Proteomics Community. *Nucleic Acids Res.* 2006, 34 (Database issue), D659–D663. [PubMed: 16381953]
- (23). Klaeger S; Heinzlmeir S; Wilhelm M; Polzer H; Vick B; Koenig P-A; Reinecke M; Ruprecht B; Petzoldt S; Meng C; et al. The Target Landscape of Clinical Kinase Drugs. *Science* 2017, 358 (6367).
- (24). Pontén F; Jirstrom K; Uhlen M The Human Protein Atlas--a Tool for Pathology. *J. Pathol.* 2008, 216 (4), 387–393. [PubMed: 18853439]
- (25). Hermjakob H IntAct: An Open Source Molecular Interaction Database. *Nucleic Acids Res.* 2004, 32 (90001), 452D–455.
- (26). Chatr-Aryamontri A; Oughtred R; Boucher L; Rust J; Chang C; Kolas NK; O'Donnell L; Oster S; Theesfeld C; Sellam A; et al. The BioGRID Interaction Database: 2017 Update. *Nucleic Acids Res.* 2017, 45 (D1), D369–D379. [PubMed: 27980099]
- (27). Rual J-F; Venkatesan K; Hao T; Hirozane-Kishikawa T; Dricot A; Li N; Berriz GF; Gibbons FD; Dreze M; Ayivi-Guedehoussou N; et al. Towards a Proteome-Scale Map of the Human Protein-Protein Interaction Network. *Nature* 2005, 437 (7062), 1173–1178. [PubMed: 16189514]
- (28). Rolland T; Taan M; Charleatoux B; Pevzner SJ; Zhong Q; Sahni N; Yi S; Lemmens I; Fontanillo C; Mosca R; et al. A Proteome-Scale Map of the Human Interactome Network. *Cell* 2014, 159 (5), 1212–1226. [PubMed: 25416956]
- (29). MacLennan M; García-Cañadas M; Reichmann J; Khazina E; Wagner G; Playfoot CJ; Salvador-Palomeque C; Mann AR; Peressini P; Sanchez L; et al. Mobilization of LINE-1 Retrotransposons Is Restricted by in Mouse Embryonic Stem Cells. *Elife* 2017, 6.
- (30). Tanemoto M; Abe T; Uchida S; Kawahara K Mislocalization of K⁺ Channels Causes the Renal Salt Wasting in EAST/SeSAME Syndrome. *FEBS Lett.* 2014, 588 (6), 899–905. [PubMed: 24561201]
- (31). Mehla J; Caufield JH; Sakhawalkar N; Uetz P A Comparison of Two-Hybrid Approaches for Detecting Protein-Protein Interactions. *Methods Enzymol.* 2017, 586, 333–358. [PubMed: 28137570]
- (32). Brückner A; Polge C; Lentze N; Auerbach D; Schlattner U Yeast Two-Hybrid, a Powerful Tool for Systems Biology. *Int. J. Mol. Sci* 2009, 10 (6), 2763–2788. [PubMed: 19582228]
- (33). Snider J; Kittanakom S; Damjanovic D; Curak J; Wong V; Stagljar I Detecting Interactions with Membrane Proteins Using a Membrane Two-Hybrid Assay in Yeast. *Nat. Protoc.* 2010, 5 (7), 1281–1293. [PubMed: 20595957]
- (34). Khan I; Maldonado E; Vasconcelos V; O'Brien SJ; Johnson WE; Antunes A Mammalian Keratin Associated Proteins (KRTAPs) Subgenomes: Disentangling Hair Diversity and Adaptation to Terrestrial and Aquatic Environments. *BMC Genomics* 2014, 15, 779. [PubMed: 25208914]
- (35). McRae JF; Mainland JD; Jaeger SR; Adipietro KA; Matsunami H; Newcomb RD Genetic Variation in the Odorant Receptor OR2J3 Is Associated with the Ability to Detect the “Grassy” Smelling Odor, Cis-3-Hexen-1-ol. *Chem. Senses* 2012, 37 (7), 585–593. [PubMed: 22714804]

Figure 1: Missing Proteins from Chr 17 Upgraded to PE1 between 2016 and 2018.

43 entries on Chr 17 were upgraded to PE1 since the neXtProt release 2016–01. **A:** The first column has the PE levels (2, 3, 4) of each gene in the release 2016–01. Of the 43, 25 were validated solely by mass-spectrometry (MS), 12 solely by protein-protein interaction (PPI), 3 by MS+PPI, and 3 by other means: TMEM107 by disease mutation, SLC16A11 by biological characterization, and TMIGD1 by post-translational modification (PTM). **B:** Subgroup analysis of the 43 reveals that PPI was more significant for keratin, KRTAPs, membrane-embedded proteins, and a single olfactory receptor. **C:** PPI played a major role in detecting 15 of the 43. The genes with an asterisk were validated by both MS and PPI data. The breakdown of the PPI data reveals that most were from yeast two-hybrid (Y2H) assays.

Figure 2. An overview of public available data about 105 Chr 17 missing proteins.

(A) Inputs for building a strategy to search for the most amenable for detection among the 105 Chr 17 missing proteins. (B) The neXtProt criteria for upgrading a protein to PE1 based on MS and PPI evidence.

Figure 3. A listing of the 105 missing proteins on Chr 17, with a basic annotation of the progress towards their validation with MS-based approaches compliant with HPP Guidelines⁵.

Those MPs with two non-nested tryptic, proteotypic peptides of length 9 aa are shown in light blue. The 16 entries without these two peptides cannot be validated by trypsin based MS. Further, 29 entries already have one observed proteotypic peptide of length at least 9 aa annotated in neXtProt (dark blue). These 29 entries require just one more proteotypic peptide of length at least 9 aa to be observed to be upgraded to PE1. Two dark blue entries lack a second predicted proteotypic peptide (KRTAP4-9, SMIM6).

Figure 4: A comparison of the spectra of synthetic and observed DLLPSQTASSLCISSR peptides.

A: Tandem MS/MS spectrum of an observed peptide from GPMdb (Protein Accession: ENSP00000462046). **B:** MS spectrum of the synthetic peptide from SRMatlas (Protein P0C851). Most of the peaks align almost exactly, suggesting a strong match. **C:** Peaks in the observed peptide spectrum. **D:** Peaks in the synthetic peptide spectrum

Figure 5. A strategy for step-wise detection of 35 of the 105 Chr 17 missing proteins.

Non-prioritized proteins are greyed out. The 10 MPs to search for with MS that already have an observed proteotypic peptide are ABCA10, C17orf64, CD300C, CDRT15L2, CDRT4, CNTD1, FBXO39, HLF, PIRT, and RNF112. The other 15 MPs to search for with MS are ADORA2B, C17orf107, C17orf78, EFCAB13, FGF11, GLP2R, KCNAB3, MFSD6L, MMD, MYADML2, NBR2, PRAC1, RASL10B, RPRML, and SPDYE4.

Table 1:
Seven Chr 17 MPs with a validated first peptide in PeptideAtlas and a second “stranded” proteotypic peptide in PRIDE.

	Missing.Protein	First.Peptide	Second.Peptide	log.e.	PXD	PRIDE
1	CD300C	MTVAGPVGGSLSVQCR	DSPEPSHPGSLFSNVR	−8.0	PXD004034	Yes
2	CSHL1	STFTNNLVYDTSDDYHLLK	NYGLLHCFR	−11.2	PXD002391	Yes
3	EVPLL	VTQECAEYCALYEK	MQASADQVER	−10.5	PXD000109	Yes
4	PIRT	VLEVDEKSPEAK	DLLPSQTASSLCISSR	−10.7	PXD005336	Yes
5	TMEM92	GPLELPSIIPPER	CGLILACPK	−6.9	PXD002121	Yes
6	RNF222	RSRALLLITLIAWAWAAILPWVLLVR	HGMPLGEQDSVLPR	−10.3	PXD003028	Yes
7	SLC16A5	QAVAADALERDLFLEAK	ECPPPPPETPALGCLAACGR	−8.5	PXD005748	Yes

Table 2.
Proteins identified through protein-protein interaction data.

Gene name	neXtProt ID	Interaction data
DLX3	NX_O60479	2 interaction annotations from BioGrid, all from Y2H. 2 interaction annotations from IntAct, all from Y2H. Interaction data with BANP are labelled as gold.
FADS6	NX_Q8N9I5	1 interaction annotation from BioGrid, from Y2H. 18 interaction annotations from IntAct, all from Y2H. Interaction data with ARFIP2 and TTPA are labelled as gold.
FOXN1	NX_O15353	24 interaction annotations from BioGrid, all from AP-MS. 36 interaction annotations from IntAct, all from Y2H. Interaction data with DMRT3 and HOXA1 are labelled as gold.
HEATR9	NX_A2RTY3	1 interaction annotation from IntAct, from Y2H. Interaction data with HEMK1 are labelled as gold.
KRT37	NX_O76014	1 interaction annotation from BioGrid, from AP-MS. 46 interaction annotations from IntAct, 41 from Y2H, 3 from coimmunoprecipitation, 2 from AP-MS. Interaction data with HOXA1, KRT1, KRT4, and PKN1 are labelled as gold.
KRTAP17-1	NX_Q9BYP8	41 interaction annotations from IntAct, all from Y2H. Interaction data with ZNF101 are labelled as gold.
KRTAP2-4	NX_Q9BYR9	10 interaction annotations from BioGrid, all from Y2H. 89 interaction annotations from IntAct, from Y2H. Interaction data with 10 proteins are labelled as gold
KRTAP9-8	NX_Q9BYQ0	1 interaction annotation from BioGrid, all from AP-MS. 80 interaction annotations from IntAct, all from Y2H. Interaction data with 24 proteins are labelled as gold.
LHX1	NX_P48742	5 interaction annotations from BioGrid, from AP-MS and reconstituted complex. 3 interaction annotations from IntAct, all from Y2H. Interaction data with C2CD6 are labelled as gold.
OR1D4	NX_P47884	11 interaction annotations from IntAct, all from Y2H. Interaction data with SGCB are labelled as gold.
SMIM5	NX_Q71RC9	35 interaction annotations from BioGrid, all from AP-MS. 7 interaction annotations from IntAct, all from Y2H. Interaction data with ARFIP1 and SGTA from IntAct labelled as gold.
TEX19	NX_Q8NA77	In UniProt: from PMID 28806172 Characterization in HEK293 cell line. Interaction with UBR2. 2 interaction annotations from IntAct, all from Y2H. No IntAct data are labelled as gold.
CLEC10A *	NX_Q8IUN9	2 interaction annotations from BioGrid, all from AP-MS. 25 interaction annotations from IntAct, all from Y2H. Interaction data with PRAF2 from IntAct labelled as gold.
KCNJ16 *	NX_Q9NPI9	In UniProt: from PMID 24561201 Characterization in MDCK II and HEK293T cells. Forms a heterodimer with KCNJ10. 2 interaction annotations from IntAct, from coimmunoprecipitation, fluorescent imaging, and Y2H. No IntAct data are labelled as gold.
KRTAP1-1 *	NX_Q07627	195 interaction annotations from IntAct, all from Y2H. IntAct data with 17 proteins are labelled as gold.

Proteins have both PPI and MS evidence

AP-MS = Affinity-purification mass spectrometry, Y2H= Yeast two-hybrid

Table 3.
Current available protein-protein interaction data from IntAct and BioGrid for unidentified Keratin-associated proteins, olfactory receptors, and TBC1 domain family proteins.

Gene name	neXtProt ID	Protein-protein interaction data
KRTAP9-1	NX_A8MXZ3	One interaction entry with MCM2, through AP-MS.
KRTAP9-7	NX_A8MTY7	One interaction entry with MCM2, through AP-MS.
OR3A1	NX_P47881	One interaction entry with MPDZ, through peptide array. One interaction entry with APOD, through AP-MS.
OR1E2	NX_P47887	One interaction entry with RAD21, through AP-Western Blot.
OR3A2	NX_P47893	Two interaction entries with ALB, through both AP-MS and anti bait coimmunoprecipitation.
OR1A2	NX_Q9Y585	Interaction entries with five proteins (including 4 ER membrane protein complex subunits, EMC3, EMC4, EMC8, and EMC10), through AP-MS.
TMEM220	NX_Q6QAJ8	Interaction entries with seven proteins, through AP-MS.
TMEM92	NX_Q6UXU6	Interaction entries with RTP2 and DERL3, through Y2H. Interaction entries with Sgo2, Nelfa, MRPL9, and PARD6B, through anti tag coimmunoprecipitation. Interaction entries with 13 proteins, through AP-MS.
TMEM99	NX_Q8N816	One Interaction entry with CDK16, through Y2H assays.
TMEM95	NX_Q3KNT9	Interaction entries with SCAND1, UBE3A, and CIB1, through Y2H assays.
TBC1D3B	NX_A6NDS4	One interaction entry with HSPB1, through reverse ras recruitment system.
TBC1D3H	NX_P0C7X1	One interaction entry with SDCBP, through phage display.
TBC1D3G	NX_Q6DHY5	Interaction entries with eight proteins (including ZRANB1), through Y2H assays.
TBC1D26	NX_Q86UD7	One interaction entry with APP, through reconstituted complex.
TBC1D28	NX_Q2M2D7	Interaction entries with DUPD1, KCNIP1, CAPS, through AP-MS.

AP-MS = Affinity-purification mass spectrometry, Y2H= Yeast two-hybrid