Variation in protein coding genes identifies information flow as a contributor to animal complexity

Jack Dean, Daniela Lopes Cardoso and Colin Sharpe*

Institute of Biological and Biomedical Sciences
School of Biological Science
University of Portsmouth,
Portsmouth, UK
PO16 7YH

* Author for correspondence
colin.sharpe@port.ac.uk

Orcid numbers:
DLC: 0000-0003-2683-1745
CS: 0000-0002-5022-0840

Abstract

Across the metazoans there is a trend towards greater organismal complexity. How complexity is generated, however, is uncertain. Since *C.elegans* and humans have approximately the same number of genes, the explanation will depend on how genes are used, rather than their absolute number. Functional diversity is a measure that quantifies the isoforms, domains and paralogues of a gene. In this paper we determine functional diversity for each protein-coding gene in the human genome and its orthologues across eight commonly used model organisms. From this we derive the Compx list of genes that correlate positively with their increase in cell-type number. We then select genes common to the Compx list and the ExAC list, whose genes show minimal variation across many human genomes, and identify genes with common functions, notably in chromatin structure, RNA splicing, vesicular transport and ubiquitin-mediated protein degradation. Together these functions reveal that information flow within the cell relates to cell-type number, used as a measure of organismal complexity.

Key words: Organismal complexity; information flow; chromatin; splicing; vesicular transport; ubiquitin-mediated protein degradation

Introduction

Some animals are more complex than others. For example, the nematode worm, *C. elegans*, widely used as a model organism, has many fewer cell types than vertebrate models, such as the mouse. Consequently, cell-type number has been widely applied as a quantitative measure of complexity[1 2]. Since the number of genes in the genome does not correlate with complexity[3], it is more likely that patterns of gene expression and the nature of the encoded proteins are important determinants.

Patterns of gene expression are determined by cis-acting regulatory elements[4 5]. Since these frequently act independently, the loss or gain of an element will have little effect on the expression pattern of the gene mediated by other elements. Cis-acting regulatory elements can therefore confer the capacity for adaptive evolution as documented by the different spatial patterns of the Pitx1 expression in sticklebacks that cause a variation in body armour, dependent on the environment [6] and their role in the loss of powered flight in some birds[7]. Although the Encode project [8] has mapped many cis-acting regulatory elements within the human genome, comprehensive data is lacking for most other species. Instead, in this paper we focus on a set of factors that influence protein function, to analyse how changes to these factors in orthologues relate to the complexity of multicellular animals.

Several of these factors have been shown to correlate positively with the increase in cell-type number (CTN) across the metazoan phylogeny. For example, *C. elegans* produces on average just over 1 transcript type per gene (protein coding and non-coding), whilst in humans this rises to around 5 (based on Ensembl[9] assembly statistics) through the wider application of alternative splicing and the use of multiple transcript start sites. The resulting transcript diversity has been positively correlated with animal complexity[2]. In addition, the complement of protein domains and motifs within a protein across a range of species has been shown to correlate with complexity, again as measured by CTN[1 10]. Although the overall number of genes is unlikely to contribute, variation in the number of related genes, the paralogues within a species, could also be important[11]. When a gene is duplicated the resulting daughter genes have the opportunity either to take on new activities (neofunctionalization)[12] or to divide the activities between the two new genes (subfunctionalisation)[13].

Combining data on the number of paralogs, transcript isoforms and protein domains associated with a gene defines a value termed the functional diversity (D_F) of the gene[14]. Correlating D_F with cell-type number across a range of metazoans for over 2000 genes

associated with transcription identified genes involved in chromatin dynamics as candidate contributors to animal complexity[14].

This paper determines the functional diversity for each annotated gene in the human genome and then across an additional eight species commonly used as model organisms for biomedical research (Supplementary Data 1). This approach extends previous analysis to the complete human gene list and its orthologues generating data for more than 120 000 genes to identify a list of genes, the Compx list, whose function may underpin complexity. We then identify a prominent overlap between the Compx list and the ExAC list of human genes that show reduced levels of variation within the human population[15]. It is thought that the ExAC genes are under purifying selection because they encode proteins that play important roles within the cell. Analysis of the overlapping genes found in both lists indicates that many encode proteins that determine information flow within the organism, relating increases in cell-type number to the ability to process information. This defines a wide, systems-based context in which to evaluate the contribution of specific genes and cellular processes to animal complexity.

Methods

Data-mining and calculation of functional diversity

For each human gene, information on the number of paralogues was obtained from the Ensembl database[9]. Similarly, the number of protein coding transcripts and the number of protein domains encoded by each gene, as defined by the Ensembl Prosite Profiles[16], was extracted from the Ensembl database using Biomart and adapted protocols (https://github.com/JackDean1/OrganismComplexity). The pipeline was then used to extract the same information for orthologues of the human genes across eight additional species from the nematode worm, *C. elegans* to the Macaque, *Macaca mulatta*. Since not all human genes have orthologues across all the species, the dataset covers 19 908 identifiable human genes (97.5% of the Ensembl human gene list, unmapped genes were excluded) and 127 355 genes in total. Details of the genomes analysed are provided in Supplementary Data 1.

A value for the functional diversity (D_F) of each gene was then calculated. The protocol is essentially as described[14] and is calculated from:

$$D_F = log_2\mathrm{P} + log_2\mathrm{I} + \sum_{I=1}^{I=n} log_2 M$$

Where P is the number of paralogues, I the number of isoforms and M the number of protein motifs and domains associated with the gene. The complete dataset and calculation of D_F is in Supplementary Data 2.

Correlation analysis

Ortho-sets, which are named after the human gene, were selected that contained the human gene and either eight, seven or six orthologues allowing for the absence, or lack of annotation, of some genes in some species. The D_F values for each gene within the ortho-set were correlated with cell type number (CTN)[1 2 17], taken as a measure of the complexity of an organism, using a Pearson's correlation (r) that takes account of the degrees of freedom defined by the number of orthologues in each ortho-set. A similar approach was taken to identify ortho-sets that showed only minimal change in the value of D_F for each gene of the ortho-set. Lists of ortho-sets from the 'complexity' analysis and the 'minimal change' analysis were compiled with increasing statistical probability in a two-tailed t-test for the Pearson's correlation and the overlap between the lists determined. At $p<0.05$ there was a 10% overlap in the ortho-sets between the two lists, but this reduced to a 0.2% overlap at $p<0.01$. Consequently, the complexity ortho-set list at $p<0.01$, called the Compx list, was used in all analyses.

GO-term and STRING analysis of the gene lists

The Compx list of 571 genes (Supplementary Data 3) was subject to a Gene Ontology (GO) -term over-representation analysis using Amigo2 v2.5.12[18] and the Panther Classification system v14[19 20] to identify terms enriched by more than 2.5 fold and with an adjusted probability of less than 0.05 (Fisher's Exact test with Bonferroni correction for multiple testing). Redundancy between terms within and across the three major GO-terms was identified on a Circos plot (gene lists in Supplementary Data 4). Interactions between the 571 Compx genes were identified using the STRING application[21].

Validation of the Compx – ExAC list overlap

A bootstrap approach was taken to determine whether the overlap between the Compx and ExAC gene lists was greater than might be expected at random. Using simply the numbers of genes in each list (571 for Compx, 3230 for ExAC and 19 908 for the human genome) the overlap between 571 random numbers between 1 and 19 908 and 3 230 random numbers between 1 and 19 908 was determined across 1000 iterations and the mean and standard deviation calculated. The process was then repeated replacing 571 with 362, the number of

genes in the 'no change' analysis. These calculated, expected values were then related to the actual overlaps between the gene lists.

Keywords associated with the function of each of the 159 Compx-ExAC overlap genes were extracted from GeneCards[22] and validated by reference to published information. The gene list was then compiled by major function (Supplementary Data 7).

Results

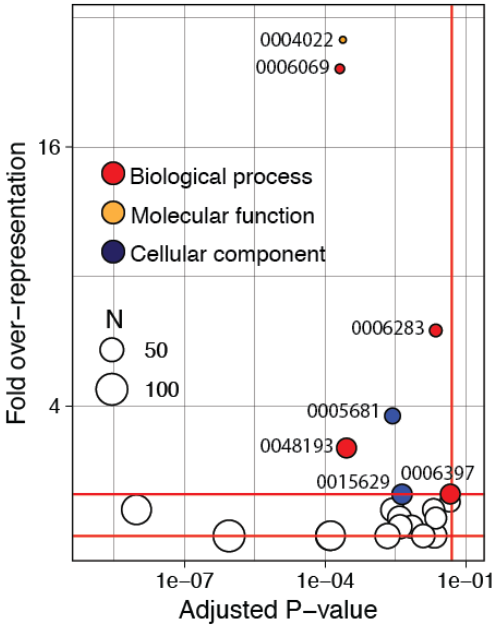
Creating a database of functional diversity for human protein coding genes and their orthologues in eight metazoan species.

This analysis is based on the premise that changes to the biological activity of proteins will be one of a number of factors that contributes to organismal complexity[1 11 14 23]. When comparing orthologous genes, the range of biological activity will depend on three factors; the number of different protein-coding transcripts that can be generated from each gene, the types and numbers of each domain that are encoded in the protein and the number of paralogues present in the genome[14]. Values for each of these factors were extracted from Ensembl, using Biomart, to produce a database for 19 908 human protein coding genes and 127 355 protein coding genes across the nine species considered (Supplementary Data 2) representing a comprehensive coverage of genes from many of the model organisms used in biomedical research. The values were then used to calculate the functional diversity (D_F) of each gene in each species, providing a metric for the capacity of that gene to provide the diverse biological activities that determine complexity. As described in the Methods section, correlating the change in value of D_F with the change in value of the cell type number, taken as a measure of organismal complexity, identified a list of 571 ortho-sets that forms the Compx list

GO-term over-representation analysis identifies a limited number of activities

Whilst it is possible that any one of the 571 ortho-sets may contribute to animal complexity, to identify Compx those with common features, we compared the Compx and human gene lists to identify GO-terms over-represented in the Compx list. The Panther Classification system[19 20] was used to identify GO-terms that are over-represented more than 2.5 fold and with an adjusted p-value of less than 0.05. After the resolution of redundant, nested terms, seven GO-terms met these criteria (Fig. 1 and Supplementary Data 4). These included two associated with the processing of alcohol, two involving mRNA

A



B

C

GO-number	Description	GO-number	Description
0006069	Ethanol oxidation	0006397	mRNA processing
0006283	Transcription coupled nucleotide-excision repair	0004022	Alcohol dehydrogenase activity
		0005681	Spliceosome complex
0048193	Golgi vesicle transport	0015629	Actin cytoskeleton

D

Reactome pathway	N	Fold over-rep	P-value
Ethanol oxidation R-HSA-71384	8	24.3	4.60E-05
Transcription-coupled nucleotide excision repair R-HSA-72163	12	5.6	9.47E-03
mRNA splicing major pathway R-HSA-72163	18	3.7	1.29E-02
Processing of capped intron-containing pre-mRNA R-HSA72203	22	3.4	3.88E-03
DNA repair R-HSA-73894	26	3.2	1.02E-03
Metabolism of RNA R-HSA-8953854	46	2.5	5.96E-05
Membrane trafficking R-HSA-199991	41	2.4	1.65E-03

Figure 1. Over-representation assay for the 571 Compx list.

A. Combined graph for the three primary GO-term fields. 2-fold and 2.5-fold over-representation levels are shown and terms that meet the required criteria are coloured according to the primary GO-term. Terms with >2<2.5-fold over-representation are shown as open circles and are listed in Supplementary Data 4. B. Circos plot showing redundant ortho-sets between the GO terms. Complete lists of ortho-sets are in Supplementary Data 4. C. Identification of the major GO-terms. D. Identification of the major terms within the Panther Reactome for over-represented genes within the Compx list.

processing, including the spliceosome complex, and one each for transcription-coupled nucleotide excision repair, Golgi vesicle transport and the actin cytoskeleton. To further analyse the over-represented genes, the degeneracy between the ortho-sets both within and

across the three primary GO-terms was determined (Fig. 1B, C) to identify a definitive list of 109 ortho-sets (Supplementary Data 4). Combined with a Panther over-representation test for the reactome (Fig. 1D) this identified four main activities: ethanol oxidation, the regulation of splicing, DNA repair and membrane trafficking, specifically of ER-Golgi vesicles.

Identifying interacting subsets using STRING analysis

Identifying interactions between proteins encoded by the Compx list will point to additional functions to which these genes contribute. The bioinformatic STRING program[21], set to identify high confidence, experimentally determined interactions identified 7 activities across 4 networks. (Fig. 2).

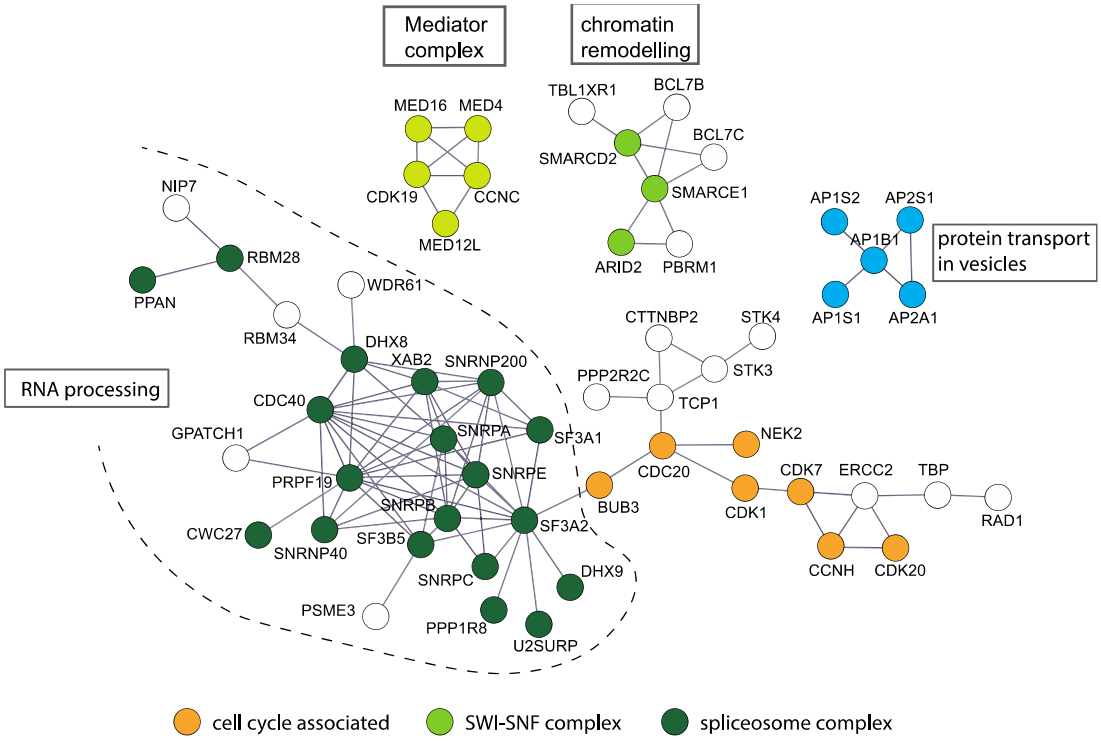


Figure 2: Interaction analysis of the Compx ortho-set

Only experimentally verified (confidence level 0.7) interactions and those networks containing at least one node with three connections are shown. Networks have been coloured by function. The distance between genes and the orientation is random. The analysis identified 7 major functions across 4 networks.

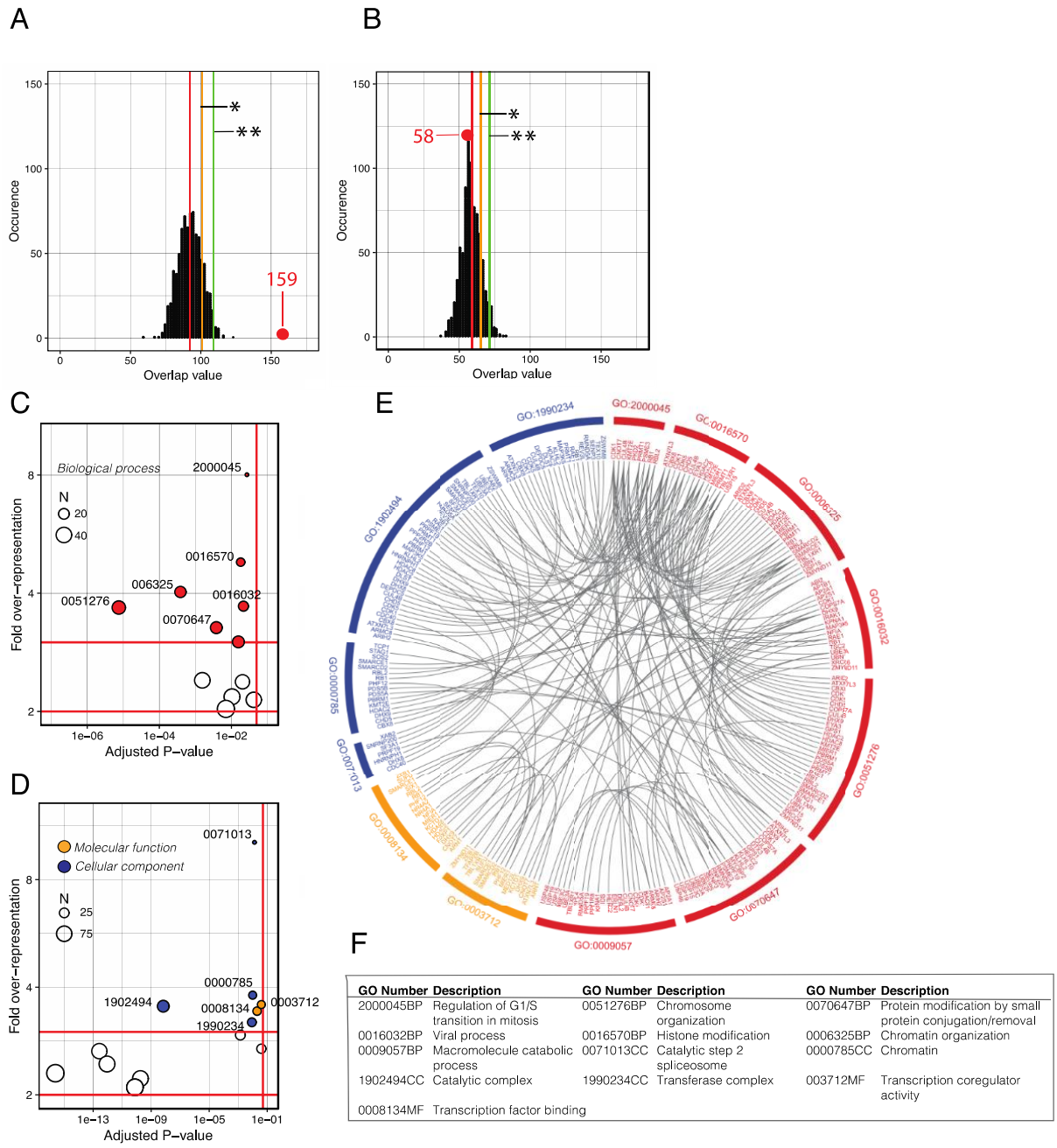
Conserved activities between the two approaches include RNA splicing (the regulation of splicing and snRNP complex) and protein-vesicle transport (associated mainly with the ER-Golgi). In addition, the interactions analysis identified members of the SWI/SNF chromatin

remodelling complex, proteins that impose repressive chromatin conformations, the Mediator complex and proteins that contribute to the cell cycle.

Filtering the Compx list to identify genes with a critical function

The ExAC genes are identified as the human genes that, following sequencing of over 60 000 genomes, show less than expected variation[15]. The suggestion is that the function of genes on the ExAC list cannot be maintained if the gene is subject to the standard rate of mutational change, indicating that their products may play critical roles in the cell[15]. We then hypothesise that genes from the ExAC list may be over-represented in the Compx list. This would then provide a filter to identify genes that are both vital to the cell and contribute to animal complexity.

A direct comparison shows that 159 of the ExAC list of 3230 genes are represented in the 571 gene Compx list (Supplementary Data 5). A bootstrap method of 1000 sampling events shows that the expected overlap between two groups the size of the Compx and ExAC lists, selected at random from a pool the size of the human gene list, is 92. The observed figure of 159 genes is almost 8 standard deviations greater than the expected mean for the random selection (Fig. 3A).



GO Number	Description	GO Number	Description	GO Number	Description
2000045BP	Regulation of G1/S transition in mitosis	0051276BP	Chromosome organization	0070647BP	Protein modification by small protein conjugation/removal
0016032BP	Viral process	0016570BP	Histone modification	0006325BP	Chromatin organization
0009057BP	Macromolecule catabolic process	0071013CC	Catalytic step 2 spliceosome	0000785CC	Chromatin
1902494CC	Catalytic complex	1990234CC	Transferase complex	003712MF	Transcription coregulator activity
0008134MF	Transcription factor binding				

Figure 3: Demonstration of a prominent overlap between the 571 complexity genes set and ExAC list

A. The distribution of 1000 predicted overlap values between samples of size 571 corresponding to the complexity ortho-set list and 3230 corresponding to the ExAC gene list taken from a stock of 19908, corresponding to the number of human genes in the analysis. The actual overlap value 159 is marked by the red dot.

B. The same analysis, repeated using a sample size of 362 corresponding to the minimal-change ortho-set. In each, the mean is marked by a red line, one standard deviation from the mean by the amber line (*) and two standard deviations by the green line (**).

C. GO-term over-representation analysis for the complexity and ExAC overlapping gene list. For the analysis of the general GO-term, Biological Process, the red circles identify terms over-represented by >3x and with adjusted P<0.05.

D. Combined analysis for the general GO-terms Cellular Component (blue circles) and Molecular Function (orange circles) (over-represented by >3x and adjusted P<0.05). The identity of the terms over-represented by >2<3x are listed in Supplementary Data 6.

E. A Circos plot of the individual genes identifies a high degree of redundancy between the over-represented GO-terms and a unique list of 90 genes (Supplementary Data 6).

F. Table of the over-represented GO-terms, the colours refer to the major GO-terms as shown in C and D where red denotes Biological Process, orange is Molecular Function and blue is Cellular Component.

In contrast, the same approach, replacing the Compx list with the 362 minimal-change ortho-set, described above, predicts an overlap with the ExAC list of 58 genes, which is identical to the observed figure of 58 (Fig. 3B). This approach validates the hypothesis of a relationship between genes of the Compx and ExAC lists and produces an overlap-list of 159 genes. Many of the genes resulting from this filter are likely to have both a crucial function in the cell and critically influence animal complexity.

Analysis of the 159 genes that are common to the Compx and ExAC gene lists.

The 159 overlapping genes were subject to GO-term over-representation analysis to highlight commonalities and the hierarchical terms rationalised using Revigo[24]. GO-terms were selected that met the criteria of a greater than 3-fold enrichment, when compared to the standard human genome gene list, and with a probability of less than 0.05 (Fig. 3C and D). This identified terms across the three major GO-term categories, notably associated with chromatin dynamics, transcription coregulation, RNA splicing and the cell cycle. In addition are the terms ‘protein modification by small protein conjugation or removal’ and ‘macromolecule catabolic process’ that refer extensively to protein ubiquitination and related processes including the breakdown of proteins in the proteasome[25] and the elimination of misfolded proteins in the ER-Golgi system by ERAD[26 27] (Fig. 4).

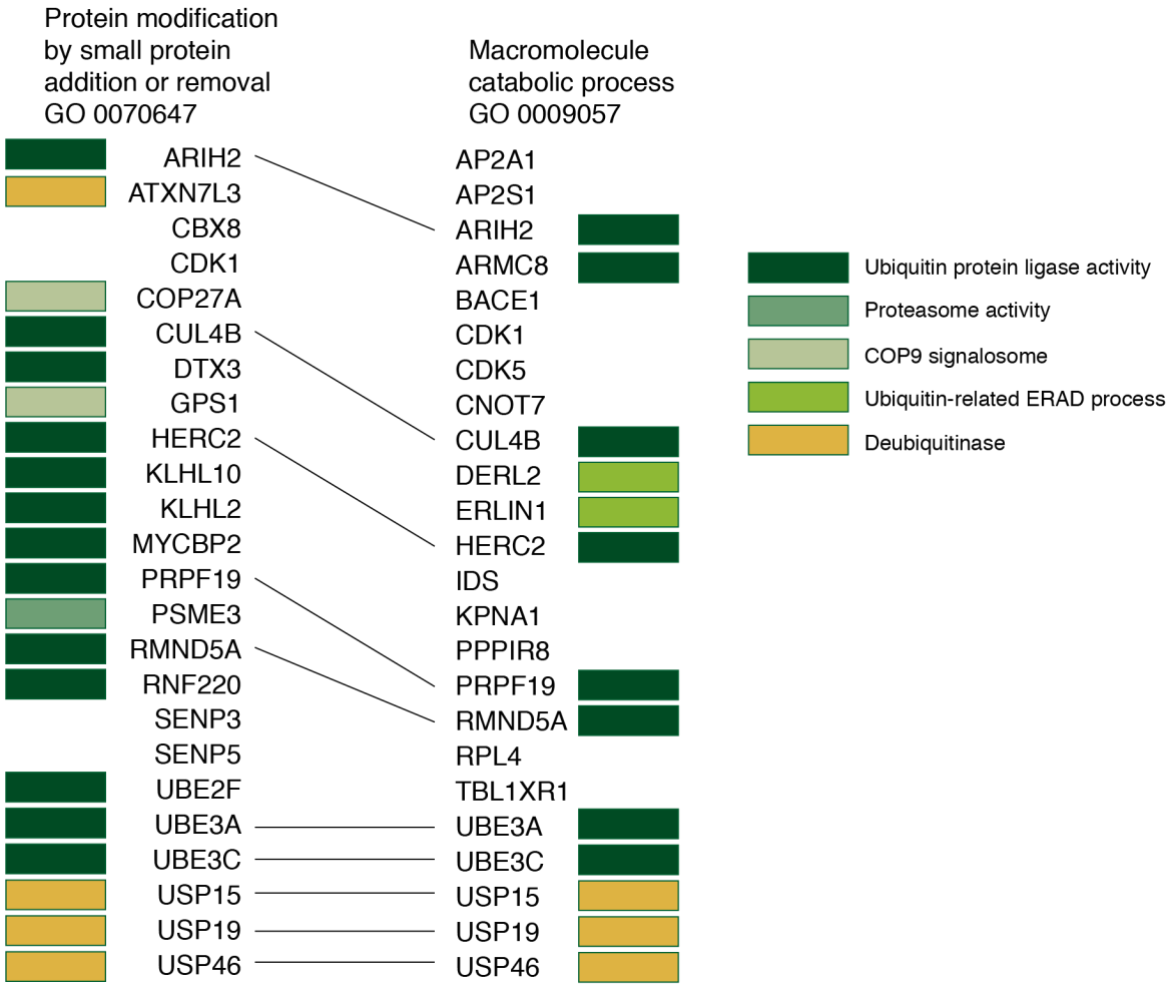


Figure 4. The overlap list contains ortho-sets associated with protein ubiquitination

Isolating the overlap between GO:0070647 protein modification by small protein conjugation or removal and GO:0009057 macromolecule catabolic process, identifies genes involved in the tagging of proteins with ubiquitin and their recognition for breakdown in the proteasome and for the removal of misfolded proteins from the ER-Golgi by the ERAD process.

The extensive redundancy of genes within the GO-terms was plotted using Circos and identifies that 90 of the 159 ortho-sets (61%) are registered in the over-representation GO-term analysis (Supplementary Data 6).

The products of the 159 overlapping genes were analysed for protein interactions using STRING (Fig. 5).

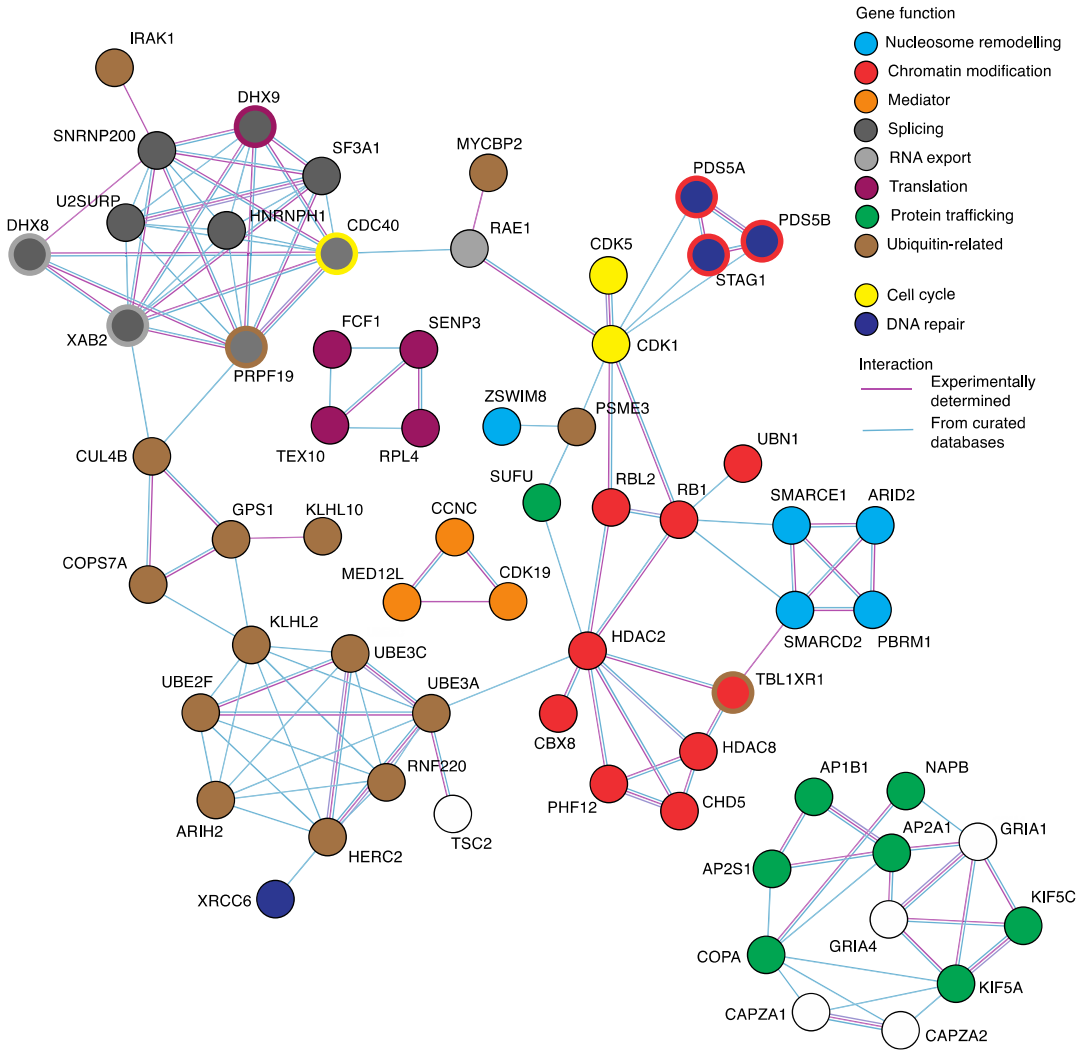


Fig 5. Interaction analysis of the Compx versus ExAC overlap gene list.
Only networks with nodes that contained at least three connections are included. Nodes have been coloured to indicate function. The connectivity of the interacting nodes is more than expected for a random selection of genes with a p=8.0E-06.

STRING analysis identified a range of interacting groups that could be described in terms of the function of the genes, as described in GeneCards and identified eight major and two minor groupings (Fig. 5). Although there may seem to be little in common between the groupings, a closer consideration suggests these are all components that contribute to information flow in the cell, from access to the chromatin to the final removal of the protein. Information flow in the cell can be described as a sequence of seven events (Fig. 6 and see Discussion). In addition, the integrity of the information needs to be maintained by DNA repair mechanisms and information flow integrated with the cell cycle.

To address further the possibility that complexity relates to information flow, the functions of the genes identified in the GO-term analysis that did not appear in the STRING analysis were similarly categorized by reference to GeneCards[22] and published data. In total 81 of the 97 genes could be allocated to events relating to information flow. Of the remaining 61 genes, 25 could be similarly allocated, giving a total of 106 out of 159 genes (67%) that are held in common between the Compx and ExAC gene lists that may play a role in information flow in the cell (Supplementary data 7). These genes were mapped to the seven events (Fig. 6) and those genes that contribute to more than one event indicated. In addition to DNA repair, the integrity of information flow is maintained by the removal of mis-folded proteins in the ER-Golgi system through endoplasmic reticulum-associated degradation (ERAD)[26 27].

These results indicate an important relationship between information flow and a cell-type number, a measure of complexity for multicellular organisms.

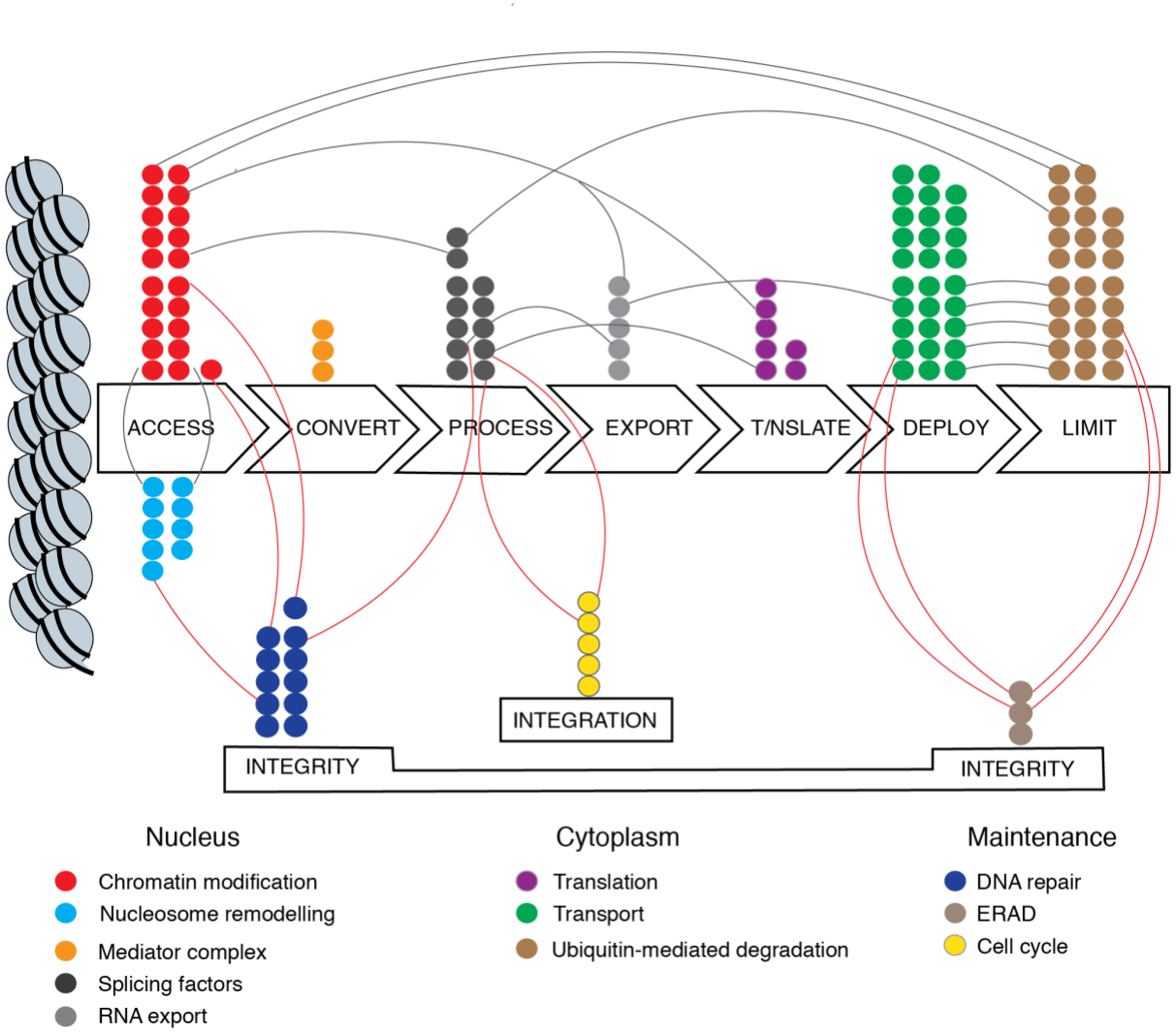


Figure 6 Mapping the Compx-ExAC overlap genes to the elements of information flow in the cell.

Individual genes (coloured circles) are mapped to the seven elements of gene flow and to the two maintenance factors, integrity and integration. Justification for the allocation is given in Supplementary Data 7. The contribution of one gene product to more than one element is marked by a connecting arc.

Discussion

In this paper, we use a simple approach to data-mine three variables that define and quantify the functional diversity (D_F) of proteins derived from each human gene and its orthologues from up to eight other species (termed an ortho-set). The change in D_F for each gene is correlated with the change in number of different cell types in the species, a proxy for organismal complexity, to identify 571 ortho-sets with a positive correlation.

This approach subscribes to a specific model in which incremental changes in functional diversity across an ortho-set relate to an incremental change in complexity. Alternative models, in which, for example, major changes in functional diversity occur at a specific point in the phylogeny, are also likely and can be extracted from our data set (Supplementary Data 2). Whilst individual genes, acting in isolation, are likely to relate complexity in either model, GO-term over-representation and interaction analysis point towards processes and systems and it is these that will contribute to a broader understanding of the basis of organismal complexity.

This paper considers functional diversity, but this will be only one of several factors associated with animal complexity. Amongst others, an increasing diversity in cis-acting regulatory elements (CAREs) across a phylogeny will play a role[4 5 28] and there may also be a contribution to complexity from changes to the complement of regulatory non-coding RNAs, including the miRNAs[29 30] and lncRNAs that have been shown to modulate gene expression[31].

We have not been able to include all the factors that affect protein functional diversity in this analysis. For example, short linear motifs (SLiMs)[23] impact on the function of proteins such as NCoR2 [11] , but these are not currently widely annotated. The functional capabilities of a protein may also be honed by a small number of changes within the sequence that affect neither protein domain number, nor the range of isoforms[32]. Lack of this data may mean that some ortho-sets that contribute to complexity have not been identified.

Other studies have successfully identified the contribution of individual components to complexity, such as protein domains and alternative splicing[1 2]. A previous paper used the approach taken here to assess genes broadly associated with transcription, as listed in the AnimalTFDB 2.0 database of 2087 human genes[14]. The identified ortho-sets that are primarily involved in the dynamic structure and function of chromatin, including nucleosome remodeling, the modulation of chromatin activity and the Mediator function. By extending the analysis to the whole genome we aimed to expand this selection to include events outside the nucleus. GO-term over-representation analysis (Fig. 1A) indicated that the oxidation of alcohol, mRNA processing, notably splicing, aspects of DNA repair and protein transport via vesicles should also be considered. For the oxidation of alcohol, the main component of functional diversity is the reiterative gene duplication followed by sub- and neo-functionalization that have produced a range of enzymes that metabolize ethanol and related alcohols[33 34]. Analysis of interactions amongst the Compx ortho-sets via STRING identified components of the spliceosome and of vesicular transport and in addition highlights

chromatin remodeling and the Mediator complex, consistent with the results from the previous, more limited, survey[14].

In this paper we apply an additional filter by identifying the overlap between the Compx and ExAC genes lists. The ExAC genes have less than expected variation within the human population, suggesting they play important functions that are sensitive to mutation[15]. This might seem to contradict the essence of the Compx list, which selects for increases in functional diversity, but the Compx list reflects changes between orthologues across a wide phylogeny, rather than variation within a single species. There is a greater than expected overlap between the Compx list and the ExAC list that is not seen when the ExAC list is compared to the 'minimal-change' list of genes selected for minimally changing values of functional diversity (Fig. 4). The hypothesis is that the filter will highlight those ortho-sets that both relate to complexity and are important to the function of the cell, allowing the identification of cellular processes that are critical for organismal complexity.

Within the overlap list, GO-term analysis identified sub-terms associated with chromatin dynamics, the spliceosome, protein modification by ubiquitination and the cell cycle. Interaction analysis similarly identified ortho-sets associated with mRNA splicing and chromatin dynamics, notably nucleosome remodeling, ubiquitin-mediated protein turnover and, in addition, ortho-sets that encode proteins involved in the distribution of proteins via vesicles within the ER-Golgi apparatus.

Is there a common theme? Each of these groupings can be directly related to the concept of information flow in the cell. We hypothesise that the contribution of protein functional diversity to complexity is primarily, though not exclusively, at the level of information flow in the cell. Information flow can be explained in terms of seven sequential elements (Fig. 6). First, nucleosome re-modelling and epigenetic marks regulate *access* to chromatin, flipping chromatin between states that either promote or repress gene expression[35 36 37 38 39]. The Mediator complex then promotes interactions between distant enhancer elements and the genes to be expressed[40 41] which in combination with transcription factors and RNA polymerases will *convert* the primary information to the intermediate, RNA. The division of eukaryotic genes into exons and introns requires subsequent *processing* of the primary transcript by the spliceosome[42 43] before *export* of the RNA to the cytosol where it is *translated* to protein. The next step in cellular information flow is one of logistics, *deploying* the product to the right location and in the correct format. Vesicular transport in the cell, associated with the ER-Golgi apparatus, transports proteins to the cell surface and for secretion[15], destinations that will be important as multicellular organisms gain in complexity

and require signaling between cell types. Finally, the specific and regulated addition and removal of ubiquitin[44], coupled with the proteasome, ensures that specific proteins are turned over, to provide a *limitation* to the pathway of information flow[45 46].

We have also identified a number of genes whose function can be ascribed to maintaining the integrity of the information flow. These include DNA repair functions that are important for the stored information and processes such as ERAD that take place in the endoplasmic reticulum to remove defective proteins[27]. Unsurprisingly, information flow needs to be coordinated to the cell cycle and a number of genes that function in both have been identified (Fig. 6 and Supplementary data 7).

We can speculate that the increased range of functions of key proteins at each stage increases the efficiency of information flow through processes of neo- and sub-functionalization of paralogues and isoforms[47]. Whilst this paper identifies a general trend, further analysis of specific genes will be required to determine precisely how these efficiencies are achieved. It is clear, though, that the genes identified in this analysis do not greatly affect the cell-type specific activities of different differentiated cell types, rather they relate to the processing of information to generate those specific activities. Applying the concept of information flow within the cell is important because it will provide the context for a more detailed analysis and understanding of animal complexity.

Acknowledgements

We would like to thank members of the Science Faculty, University of Portsmouth, for critical comments. JD, DLC and CS are grateful for University of Portsmouth Master's support, a PhD bursary with Professor A. Callaghan and research support, respectively.

Contributions

JD collected and analysed the data, DLC analysed the data and CS analysed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The authors declare that all other data supporting the findings of this study are available in the paper, in its supplementary information files (S1 and S3-7 are attached at the end of this document) and S2 is available in the University of Portsmouth repository, PURE, at DOI:

10.17029/3d3b91c7-3ae2-4ef8-9703-852709d04a8f and (S2-7) at https://github.com/colin17/complexity.git .

Code availability

The code used, in R, to extract data from the Ensembl database is freely available at https://github.com/JackDean1/OrganismComplexity.

Supplementary information

S1 Table of the genome data used in this research.

S2 Table of calculations of functional diversity for all genes across the eight species (CSV format).

S3 The Compx list of genes (CSV format)

S4 GO-term details for the Compx gene list (CSV format)

S5 The Compx-ExAC overlap list of genes (CSV format)

S6 GO-term details for the Compx-ExAC overlap gene list (CSV format)

S7 Compx-ExAC overlap gene list, functional details (CSV format)

References

1. Vogel C, Chothia C. Protein family expansions and biological complexity. *PLoS Comput Biol*. 2006. doi:10.1371/journal.pcbi.0020048
2. Chen L, Bush SJ, Tovar-Corona JM, Castillo-Morales A, Urrutia AO. Correcting for differential transcript coverage reveals a strong relationship between alternative splicing and organism complexity. *Mol Biol Evol*. 2014. doi:10.1093/molbev/msu083
3. Baltimore D. Our genome unveiled. *Nature*. 2001. doi:10.1038/35057267
4. Levine M, Tjian R. Transcription regulation and animal diversity. *Nature*. 2003. doi:10.1038/nature01763
5. Wray GA. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet*. 2007. doi:10.1038/nrg2063
6. Chan, Frank Y, Marks, et al. Adaptive Evolution of Pelvic Reduction in Sticklebacks by Recurrent Deletion of a Pitx1 Enhancer. *Science (80-)*. 2009. doi:10.1126/science.1182213
7. Sackton TB, Grayson P, Cloutier A, et al. Convergent regulatory evolution and the origin of flightlessness in palaeognathous birds. *bioRxiv*. 2018. doi:10.1101/262584
8. Ecker JR, Bickmore WA, Barroso I, Pritchard JK, Gilad Y, Segal E. Genomics: ENCODE explained. *Nature*. 2012. doi:10.1038/489052a

9. Frankish A, Vullo A, Zadissa A, et al. Ensembl 2018. *Nucleic Acids Res*. 2017. doi:10.1093/nar/gkx1098
10. Schad E, Tompa P, Hegyi H. The relationship between proteome size, structural disorder and organism complexity. *Genome Biol*. 2011. doi:10.1186/gb-2011-12-12-r120
11. Short S, Peterkin T, Guille M, Patient R, Sharpe C. Short linear motif acquisition, exon formation and alternative splicing determine a pathway to diversity for NCoR-family co-repressors. *Open Biol*. 2015. doi:10.1098/rsob.150063
12. Conant GC, Wolfe KH. Turning a hobby into a job: How duplicated genes find new functions. *Nat Rev Genet*. 2008. doi:10.1038/nrg2482
13. He X, Zhang J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics*. 2005. doi:10.1534/genetics.104.037051
14. Lopes Cardoso D, Sharpe C. Relating protein functional diversity to cell type number identifies genes that determine dynamic aspects of chromatin organisation as potential contributors to organismal complexity. *PLoS One*. 2017. doi:10.1371/journal.pone.0185409
15. Lek M, Karczewski KJ, Minikel E V., et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016. doi:10.1038/nature19057
16. Sigrist CJA, Cerutti L, Hulo N, et al. PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinform*. 2002. doi:10.1093/bib/3.3.265
17. Hedges SB, Blair JE, Venturi ML, Shoe JL. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol Biol*. 2004. doi:10.1186/1471-2148-4-2
18. Blake JA, Christie KR, Dolan ME, et al. Gene ontology consortium: Going forward. *Nucleic Acids Res*. 2015. doi:10.1093/nar/gku1179
19. Mi H, Huang X, Muruganujan A, et al. PANTHER version 11: Expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res*. 2017. doi:10.1093/nar/gkw1138
20. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the panther classification system. *Nat Protoc*. 2013. doi:10.1038/nprot.2013.092
21. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017. doi:10.1093/nar/gkw937
22. Stelzer G, Rosen R, Plaschkes I, et al. GeneCards - the human gene database. The

GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis, Current Protocols in Bioinformatics. doi:https://doi.org/10.1002/cpbi.5

23. Weatheritt RJ, Gibson TJ. Linear motifs: Lost in (pre)translation. *Trends Biochem Sci*. 2012. doi:10.1016/j.tibs.2012.05.001
24. Supek F, Bošnjak M, Škunca N, Šmuc T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011. doi:10.1371/journal.pone.0021800
25. Kwon YT, Ciechanover A. The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy. *Trends Biochem Sci*. 2017. doi:10.1016/j.tibs.2017.09.002
26. Olzmann JA, Kopito RR, Christianson JC. The mammalian endoplasmic reticulum-associated degradation system. *Cold Spring Harb Perspect Biol*. 2013. doi:10.1101/cshperspect.a013185
27. Christianson JC, Ye Y. Cleaning up in the endoplasmic reticulum: Ubiquitin in charge. *Nat Struct Mol Biol*. 2014. doi:10.1038/nsmb.2793
28. Simpson P, Ayyar S. Chapter 3 Evolution of Cis-Regulatory Sequences in Drosophila. *Adv Genet*. 2008. doi:10.1016/S0065-2660(07)00003-X
29. Lee C-T, Risom T, Strauss WM. Evolutionary Conservation of MicroRNA Regulatory Circuits: An Examination of MicroRNA Gene Complexity and Conserved MicroRNA-Target Interactions through Metazoan Phylogeny. *DNA Cell Biol*. 2007. doi:10.1089/dna.2006.0545
30. Mao X, Li L, Cao Y. Evolutionary comparisons of miRNA regulation system in six model organisms. *Genetica*. 2014. doi:10.1007/s10709-014-9758-5
31. Kapusta A, Feschotte C. Volatile evolution of long noncoding RNA repertoires: Mechanisms and biological implications. *Trends Genet*. 2014. doi:10.1016/j.tig.2014.08.004
32. Hardwick LJA, Philpott A. The N terminus of Ascl1 underlies differing proneural activity of mouse and Xenopus Ascl1 proteins. *Wellcome Open Res*. 2018. doi:10.12688/wellcomeopenres.14842.1
33. Gonzàlez-Duarte R, Albalat R. Merging protein, gene and genomic data: The evolution of the MDR-ADH family. *Heredity (Edinb)*. 2005. doi:10.1038/sj.hdy.6800723
34. Thompson CE, Freitas LB, Salzano FM. Molecular evolution and functional divergence of alcohol dehydrogenases in animals, fungi and plants. *Genet Mol Biol*. 2018. doi:10.1590/1678-4685-gmb-2017-0047
35. Struhl K, Segal E. Determinants of nucleosome positioning. *Nat Struct Mol Biol*. 2013. doi:10.1038/nsmb.2506
36. Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol*. 2014. doi:10.1038/nrm3890
37. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature*. 2000.

doi:10.1038/47412

38. Becker PB, Horz W. ATP-dependent nucleosomere modeling. *Annu Rev Biochem*. 2002.

39. Narlikar GJ, Sundaramoorthy R, Owen-Hughes T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*. 2013. doi:10.1016/j.cell.2013.07.011

40. Soutourina J. Transcription regulation by the Mediator complex. *Nat Rev Mol Cell Biol*. 2018. doi:10.1038/nrm.2017.115

41. Allen BL, Taatjes DJ. The Mediator complex: A central integrator of transcription. *Nat Rev Mol Cell Biol*. 2015. doi:10.1038/nrm3951

42. Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol*. 2017. doi:10.1038/nrm.2017.86

43. Herzel L, Ottoz DSM, Alpert T, Neugebauer KM. Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol*. 2017. doi:10.1038/nrm.2017.63

44. Lim, R., Nolte, H., Andrade, J., Zimmermann, B., Shi, C., Doddaballapur, A., Ong, Y.T., Wilhelm, K., Fasse, J.W.D., Ernst, A., Kaulich, M., Husnjak, K., Boettger, T., Guenther, S., Braun, T., Benedito, R., Dikic, I., Potente M. Deubiquitinase USP10 regualtes Notch signaling in the endothelium. *Science (80-)*. 2019;364:188-193. www.sciencemag.org/content/364/6436/188/suppl/DC1.

45. Varshavsky A. The Ubiquitin System, Autophagy, and Regulated Protein Degradation. *Annu Rev Biochem*. 2017. doi:10.1146/annurev-biochem-061516-044859

46. Kleiger G, Mayor T. Perilous journey: A tour of the ubiquitin-proteasome system. *Trends Cell Biol*. 2014. doi:10.1016/j.tcb.2013.12.003

47. Zhang J. Evolution by gene duplication: An update. *Trends Ecol Evol*. 2003. doi:10.1016/S0169-5347(03)00033-8

Supplementary data

Supplementary Data S1, S3-7

S1

Supplementary Table 1

Organism	Classification	Genome version	Gene number[1]	Cell type Number[2]
Homo sapiens	Mammal	GRCh38.p12	20,418	171
Macacca mulata	Mammal	Mmul_8.0.1	21,099	171
Mus musculus	Mammal	GRCm38.p6	22,600	157
Gallus gallus	Bird	GRCg6a	16,878	150
Xenopus tropicalis	Amphibian	JGI 4.2	18,442	121
Takifugu rubripres	Bony fish	FUGU5	20,545	114
Ciona intestinalis	Urochordate	KH	16,671	71
Drosophila melanogaster	Insect	BDGP6	13,931	60
Caenorhabditis elegans	Nematode worm	WBcel235	20,222	29

1. As given on the Ensembl database
2. Cell type number taken from :

Vogel C, Chothia C (2006) Protein family expansions and biological complexity. PLoS Comput Biol 2: e48. pmid:16733546

Hedges SB, Blair JE, Venturi ML, Shoe JL (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol 4: 2. pmid:15005799

S3

```
Compx gene list,,"""Flat"" list",,
ABI2,,ABHD2,,
AC235565.2,,ADORA2A,,
ACOT11,,AGPAT5,,
ACOT7,,AK6,,
ACSS2,,ALAD,,
ADAM15,,ALG1,,
ADAM8,,ALG10,,
ADAMTS3,,ALG10B,,
ADH1A,,ALG8,,
ADH1B,,AMT,,
ADH1C,,ANAPC10,,
ADH4,,ANKZF1,,
ADH5,,ANXA6,,
ADH6,,ARHGDIA,,
ADH7,,ARL5A,,
ADORA3,,ARPC1B,,
ADSS,,ASAH1,,
AGBL4,,ASMT,,
AHCYL1,,ASRGL1,,
AKR1A1,,ATG12,,
ALAS2,,ATG13,,
ALDH1L1,,ATG9B,,
ANKRD50,,ATP13A1,,
AP1B1,,ATP6V0D1,,
AP1S1,,ATP6V1G3,,
AP1S2,,BBS1,,
AP2A1,,BBS2,,
AP2S1,,BCKDHA,,
AP3S2,,BCL6,,
APP,,BTBD3,,
ARAP2,,BYSL,,
ARFGAP1,,C15orf41,,
ARG2,,C1QTNF1,,
ARHGDIB,,C9orf64,,
ARID2,,CACYBP,,
ARIH2,,CAPSL,,
ARL14EP,,CARHSP1,,
ARL15,,CAV2,,
ARMC8,,CCDC124,,
ARNT,,CCDC181,,
ARPC5,,CCDC25,,
ASCL3,,CCNG1,,
ASF1B,,CCS,,
ASNA1,,CCZ1B,,
ATG2A,,CDC16,,
ATP5F1A,,CENPV,,
ATP6V1E1,,CEP78,,
ATXN7L3,,CFAP157,,
BACE1,,CFAP57,,
BANF2,,CHMP7,,
BBS9,,CHRM2,,
BCAP29,,CHSY1,,
BCL7B,,CHSY3,,
BCL7C,,CIRBP,,
BET1,,CLNS1A,,
BET1L,,CLPTM1L,,
```

```
BLM,,CMC1,,
BNC2,,CMC2,,
BPNT1,,CMTR2,,
BSDC1,,CNDP2,,
BUB3,,CNKSR2,,
C1D,,COASY,,
C1QTNF3,,COL4A3BP,,
CA2,,COPB1,,
CA5A,,COX11,,
CALHM3,,CPA1,,
CALHM4,,CPA4,,
CALHM6,,CPLX2,,
CAPN3,,CPSF2,,
CAPZA1,,CPSF7,,
CAPZA2,,CRAMP1,,
CASP2,,CRBN,,
CBLN2,,CSTF1,,
CBX6,,CTCF,,
CBX8,,CTNNB1,,
CCDC151,,CTNNBIP1,,
CCNC,,CTSB,,
CCND1,,DAD1,,
CCNH,,DCAF13,,
CCZ1,,DDB1,,
CCZ1B,,DDX6,,
CD63,,DECR1,,
CDC20,,DGCR8,,
CDC40,,DHRS2,,
CDCA7,,DHX36,,
CDK1,,DHX38,,
CDK11A,,DLX4,,
CDK11B,,DMGDH,,
CDK19,,DMRT2,,
CDK20,,DNAAF3,,
CDK5,,DNAAF5,,
CDK7,,DNAJC18,,
CERS5,,DPH1,,
CHD5,,DPH2,,
CHMP2B,,DPP9,,
CIB2,,DTWD1,,
CIB3,,DUOX1,,
CLASRP,,DUS4L,,
CLIC4,,DYM,,
CLIC5,,DYNC2LI1,,
CLIC6,,EARS2,,
CMTM6,,EDC3,,
CNIH1,,EEF1AKMT2,,
CNIH4,,EEPD1,,
CNOT7,,EFCAB2,,
COBL,,EFHC1,,
COL12A1,,EFHD1,,
COL14A1,,EIF1AD,,
COL6A1,,EIF3B,,
COL6A2,,EIF3C,,
COLQ,,EIF3F,,
COMMD7,,EIF3H,,
COPA,,ELP2,,
COPS7A,,ELP4,,
```

```
COPZ1,,EMC4,,
COQ8B,,EMC7,,
COQ9,,ENKUR,,
CPSF4,,ENO1,,
CRYZL1,,ERCC3,,
CTTNBP2,,EXTL3,,
CUL4B,,FAM185A,,
CWC27,,FAM72A,,
DAP3,,FAM72B,,
DBT,,FAXDC2,,
DEPDC5,,FBXL16,,
DERL2,,FEM1B,,
DHX8,,FGL1,,
DHX9,,FICD,,
DIS3L2,,FMC1,,
DLST,,FNIP1,,
DNTTIP2,,FNTA,,
DOK1,,FRMD6,,
DOK6,,FRS2,,
DOPEY1,,FTCD,,
DPH5,,GAS8,,
DTL,,GATB,,
DTX3,,GATC,,
EEF1B2,,GATM,,
EEF1D,,GCNT3,,
EIF1,,GDF9,,
EIF1B,,GLDC,,
EIF3E,,GOSR1,,
ELFN1,,GOSR2,,
EMC9,,GPALPP1,,
EML4,,GPR89A,,
EMSY,,HAAO,,
ENTPD5,,HADH,,
ERCC2,,HAO2,,
ERGIC1,,HEXB,,
ERLIN1,,HSD17B4,,
ETHE1,,HSPA9,,
EXOC6,,HYDIN,,
EXOC7,,IFT74,,
EXOSC6,,IMMP1L,,
EXT2,,IMP3,,
EYA3,,INTS12,,
FAM120B,,INTS14,,
FAM160A1,,INTS9,,
FAM172A,,IQUB,,
FAM206A,,IRAK4,,
FAM49B,,JOSD1,,
FAM92A,,KATNB1,,
FAM98C,,KCNC4,,
FANCD2,,KDELR2,,
FBLL1,,KDELR3,,
FBXO8,,LACTB,,
FCF1,,LARP7,,
FN1,,LARS,,
FO393400.1,,LCMT1,,
FOXP1,,LGMN,,
FREM1,,LIG4,,
GABARAP,,LONP1,,
```

```
GCAT,,LRRC6,,
GDPD5,,LTN1,,
GFOD1,,MAATS1,,
GFOD2,,MAGEA2,,
GGA3,,MAGEA2B,,
GNA13,,MAGEA4,,
GNG2,,MAPK10,,
GNG3,,MCM4,,
GNL1,,MCTS1,,
GPATCH1,,MED17,,
GPC1,,MED22,,
GPC2,,MED9,,
GPS1,,MFSD8,,
GRIA1,,MIOS,,
GRIA4,,MLEC,,
GRIK1,,MLST8,,
GRIK2,,MOGS,,
GRTP1,,MORN5,,
GSX2,,MPPE1,,
GTF3A,,MRPL11,,
GUCY1B1,,MRPL16,,
H2AFV,,MRPL22,,
HACD2,,MRPL28,,
HADHB,,MRPL43,,
HAND2,,MRPL55,,
HDAC2,,MRPS2,,
HDAC8,,MRPS22,,
HERC2,,MSH6,,
HHIPL1,,MTFP1,,
HM13,,MTRF1,,
HMBS,,MTTP,,
HMGCL,,MVD,,
HNRNPH1,,MYL6B,,
HOOK2,,NAF1,,
HOOK3,,NAT9,,
HOXC4,,NDUFA2,,
HS3ST2,,NDUFAF5,,
HSD17B2,,NDUFAF7,,
ID2,,NDUFB9,,
ID3,,NECAP1,,
IDS,,NELFA,,
IFRD2,,NFS1,,
IFT140,,NHP2,,
IFT20,,NOP16,,
IGSF9,,NOP2,,
IL18RAP,,NOP53,,
ILK,,NUDT2,,
ILKAP,,NUDT4,,
IP6K2,,OSBPL2,,
IPO11,,OTUD4,,
IPP,,PAX6,,
IQGAP1,,PCNX4,,
IRAK1,,PDCD10,,
IRF1,,PDP2,,
IRX4,,PDZD3,,
IRX6,,PES1,,
ITGB4,,PGS1,,
JMJD7,,PHOSPHO1,,
```

```
KCNC2,,PIGL,,
KCND3,,PIGS,,
KCNIP1,,PIN1,,
KCNIP4,,PIP4K2C,,
KCNK15,,PISD,,
KCTD1,,PKMYT1,,
KCTD15,,PLA2G15,,
KCTD3,,PNPO,,
KCTD8,,POLA2,,
KCTD9,,POLH,,
KIF5A,,POLR2D,,
KIF5C,,POLR2G,,
KIFAP3,,PPP2CB,,
KLF6,,PPWD1,,
KLHL10,,PRELID3A,,
KLHL2,,PREPL,,
KMT2E,,PRICKLE1,,
KPNA1,,PRPF8,,
KPNA2,,PSEN1,,
KPNA6,,PSMC2,,
KRI1,,PSMC5,,
KTI12,,PSMD7,,
LANCL1,,PSMG2,,
LIG3,,PTRHD1,,
LLGL2,,RAB10,,
LMAN2L,,RAB27A,,
LMBR1,,RAD51C,,
LMNA,,RAD9A,,
LOXHD1,,RANBP3,,
MAEA,,RASSF8,,
MAFK,,RBM15,,
MAGOH,,RBP5,,
MAGT1,,RCL1,,
MAP3K5,,REEP2,,
MAR7,,RERGL,,
MCCC2,,RFWD3,,
MCRS1,,RHAG,,
MDH1,,RHOH,,
MEAF6,,RNASEH2B,,
MED12L,,RNPS1,,
MED16,,RPL10,,
MED4,,RPLP0,,
METTL15,,RUVBL1,,
METTL2A,,SARAF,,
MIER1,,SARS2,,
MKS1,,SCPEP1,,
MRO,,SCYL1,,
MROH1,,SDF2,,
MROH2A,,SDHD,,
MROH7,,SERP1,,
MROH7-TTC4,,SETD4,,
MROH8,,SFXN2,,
MROH9,,SH3BP4,,
MRPL39,,SH3GL1,,
MRPS18A,,SIK1,,
MSX2,,SLC19A3,,
MTERF3,,SLC25A46,,
MTMR2,,SLC25A48,,
```

```
MTREX,,SLC30A8,,
MXI1,,SLC35F5,,
MYBPC1,,SLC46A1,,
MYBPC3,,SLC52A2,,
MYCBP2,,SLC6A1,,
MYLK,,SLX1A,,
MYOM2,,SLX1B,,
MYOM3,,SMARCB1,,
N6AMT1,,SMARCE1,,
NAPB,,SMN2,,
NAT10,,SNF8,,
NEB,,SNRNP25,,
NEK2,,SNRPF,,
NEK7,,SNUPN,,
NEK8,,SPATA20,,
NETO2,,SPEF2,,
NEUROG2,,SPG7,,
NFIA,,SPRYD7,,
NIP7,,SSR3,,
NIPAL1,,SSU72,,
NIPAL3,,STUB1,,
NIT2,,STX16,,
NKX3-1,,SUOX,,
NME2,,SUPT4H1,,
NME7,,SUPT6H,,
NODAL,,SYS1,,
NR4A1,,TACO1,,
NR4A2,,TARBP1,,
NR4A3,,TATDN1,,
NT5C1B,,TBC1D2,,
NUP43,,TBC1D32,,
OLFM1,,TBCE,,
OLFM3,,TBCE,,
OSBPL1A,,TBCE,,
OSCP1,,TBCE,,
PANK2,,TBL1X,,
PAQR9,,TCTN1,,
PARVA,,TFIP11,,
PARVB,,TLR2,,
PBRM1,,TMEM138,,
PCBD1,,TMEM14B,,
PCBD2,,TMEM169,,
PCMT1,,TMEM41A,,
PDC,,TMEM67,,
PDCL,,TMLHE,,
PDCL3,,TPO,,
PDK3,,TPT1,,
PDS5A,,TRAPPC5,,
PDS5B,,TRAPPC8,,
PDZD11,,TROVE2,,
PDZRN4,,TRPT1,,
PELI3,,TSNAX-DISC1,,
PEPD,,TUBB6,,
PGAP1,,TVP23C,,
PGM1,,TWNK,,
PGM5,,TXNDC15,,
PHF12,,UBA5,,
PHGDH,,UBE3B,,
```

```
PIGO,,UBXN1,,
PIP4K2A,,ULK3,,
PIP4K2B,,UPRT,,
PITPNA,,UTP15,,
PITPNB,,UTP4,,
PKD2,,VPS36,,
PLPPR1,,VSNL1,,
PLXNB2,,WARS,,
PLXNC1,,WASHC4,,
PMS1,,WDR55,,
PNKP,,WDR74,,
POLD1,,WEE1,,
POLR1C,,WRB,,
PORCN,,WWC2,,
PPA2,,YIPF5,,
PPAN,,YPEL5,,
PPAN-P2RY11,,ZC2HC1C,,
PPFIA1,,ZCCHC10,,
PPP1R21,,ZNF830,,
PPP1R8,,,,
PPP2R2B,,,,
PPP2R2C,,,,
PPP2R2D,,,,
PRMT1,,,,
PRPF19,,,,
PSME1,,,,
PSME3,,,,
QTRT2,,,,
RAB1A,,,,
RAB6A,,,,
RAB9B,,,,
RABGGTB,,,,
RAD1,,,,
RAD54L2,,,,
RAE1,,,,
RAP1GDS1,,,,
RASGRP1,,,,
RB1,,,,
RBL2,,,,
RBM12,,,,
RBM28,,,,
RBM34,,,,
RCAN1,,,,
RCAN3,,,,
RETSAT,,,,
REV3L,,,,
RFC5,,,,
RHD,,,,
RMND5A,,,,
RNF115,,,,
RNF126,,,,
RNF157,,,,
RNF220,,,,
RPL22L1,,,,
RPL4,,,,
RPS25,,,,
RPS6KB2,,,,
RRAGC,,,,
```

```
RRM2,,,,
RTRAF,,,,
SALL2,,,,
SAT1,,,,
SATL1,,,,
SBF2,,,,
SCAF4,,,,
SCMH1,,,,
SCRN2,,,,
SCRN3,,,,
SDR42E2,,,,
SEC11A,,,,
SEC11C,,,,
SECISBP2,,,,
SENP3,,,,
SENP5,,,,
SEPT6,,,,
SEPT7,,,,
SESN1,,,,
SESN3,,,,
SETD3,,,,
SF3A1,,,,
SF3A2,,,,
SF3B5,,,,
SFT2D1,,,,
SFT2D2,,,,
SIGIRR,,,,
SIRT6,,,,
SIX1,,,,
SLBP,,,,
SLC10A3,,,,
SLC10A6,,,,
SLC19A1,,,,
SLC25A28,,,,
SLC35A1,,,,
SLC38A7,,,,
SLC7A6,,,,
SLCO4A1,,,,
SMAD3,,,,
SMARCD2,,,,
SMARCE1,,,,
SNRNP200,,,,
SNRNP40,,,,
SNRPA,,,,
SNRPB,,,,
SNRPC,,,,
SNRPE,,,,
SNX1,,,,
SNX10,,,,
SNX12,,,,
SNX3,,,,
SORL1,,,,
SOS2,,,,
SP5,,,,
SPOUT1,,,,
SRP72,,,,
ST3GAL2,,,,
ST7,,,,
```

```
STAG1,,,,
STK3,,,,
STK4,,,,
STMN2,,,,
STMN4,,,,
STXBP1,,,,
STXBP2,,,,
SUFU,,,,
SUMF1,,,,
SUMF2,,,,
TBC1D23,,,,
TBL1XR1,,,,
TBL1Y,,,,
TBP,,,,
TCEA1,,,,
TCF3,,,,
TCP1,,,,
TCP11X2,,,,
TECR,,,,
TEK,,,,
TERT,,,,
TEX10,,,,
TFAP2B,,,,
THOC3,,,,
THSD7A,,,,
TMC7,,,,
TMCC2,,,,
TMED10,,,,
TMEM106B,,,,
TMEM120B,,,,
TMEM131,,,,
TMEM161A,,,,
TMEM161B,,,,
TMEM167B,,,,
TMEM231,,,,
TMEM63B,,,,
TOPBP1,,,,
TPD52,,,,
TPD52L2,,,,
TRAF4,,,,
TRAPPC12,,,,
TRAPPC3L,,,,
TRAPPC6A,,,,
TRIM23,,,,
TRMT112,,,,
TRPC6,,,,
TRPM1,,,,
TRPM2,,,,
TRPM3,,,,
TRPM8,,,,
TSC2,,,,
TSEN2,,,,
TSFM,,,,
TSN,,,,
TSPAN17,,,,
TSPAN3,,,,
TSR1,,,,
TTC1,,,,
```

```
TTC17,,,,
TTC39A,,,,
TTC39C,,,,
TTC7A,,,,
TTC7B,,,,
TTLL5,,,,
TTLL7,,,,
TTLL9,,,,
TUSC3,,,,
TUT7,,,,
TXLNB,,,,
U2SURP,,,,
UBE2A,,,,
UBE2B,,,,
UBE2F,,,,
UBE2L3,,,,
UBE2L5,,,,
UBE3A,,,,
UBE3C,,,,
UBN1,,,,
UGGT2,,,,
UHRF1BP1L,,,,
UNC45A,,,,
UNC5C,,,,
USP15,,,,
USP19,,,,
USP46,,,,
VPS11,,,,
VPS33B,,,,
VPS35L,,,,
VTI1A,,,,
WASHC3,,,,
WDR3,,,,
WDR61,,,,
WDR91,,,,
XAB2,,,,
XRCC6,,,,
ZC2HC1A,,,,
ZC3H14,,,,
ZCCHC24,,,,
ZCCHC4,,,,
ZDHHC17,,,,
ZEB1,,,,
ZFAND2A,,,,
ZFAT,,,,
ZMYND11,,,,
ZNF687,,,,
ZSWIM8,,,,
ZXDA,,,,
ZXDB,,,,
```

S4

,,,,,,Gene summary,,
GO-term Biological Process,N, Enrich,P-value,,,,,
ethanol oxidation (GO:0006069),8,24.3,2.03E-04,,,GO-terms,total,genes
transcription-coupled nucleotide-excision repair
(GO:0006283),12,6.0,2.27E-02,,,5681 6283 6397,2,PRPF19
endoplasmic reticulum to Golgi vesicle-mediated transport
(GO:0006888),21,3.7,8.55E-03,,,,,XAB2
Golgi vesicle transport (GO:0048193),32,3.2,2.86E-04,,,4022 6069,7,ADH5
mRNA processing (GO:0006397),32,2.5,4.66E-02,,,,,ADH4
heart development (GO:0007507),34,2.4,4.53E-02,,,,,ADH1A
mRNA metabolic process (GO:0016071),42,2.3,2.04E-02,,,,,ADH7
cellular component morphogenesis (GO:0032989),49,2.3,2.71E-03,,,,,ADH1C
RNA processing (GO:0006396),51,2.2,3.81E-03,,,,,ADH1B
protein-containing complex subunit organization
(GO:0043933),100,2.0,8.95E-07,,,,,ADH6
chromosome organization (GO:0051276),56,2.0,2.14E-02,,,5681
6397,17,SNRNP200
protein transport (GO:0015031),81,2.0,1.26E-04,,,,,MAGOH
protein-containing complex assembly (GO:0065003),83,2.0,1.31E-
04,,,,,RBM28
,,,,,,,,SNRPE
,,,,,,,,SNRNP40
GO-term Molecular Function,N,Enrich ,P-value,,,,,PPP1R8
alcohol dehydrogenase (NAD) activity (GO:0004022),7,28.4,2.40E-
04,,,,,HNRNPH1
transcription factor binding (GO:0008134),40,2.2,2.27E-02,,,,,CWC27
,,,,,,,,DHX8
,,,,,,,,SF3A1
GO-term Cellular Compartment,N,Enrich,P-value,,,,,CDC40
catalytic step 2 spliceosome (GO:0071013),14,5.9,6.88E-04,,,,,SNRPB
spliceosomal complex (GO:0005681),19,3.8,2.69E-03,,,,,SF3B5
actin cytoskeleton (GO:0015629),34,2.5,4.35E-03,,,,,GPATCH1
catalytic complex (GO:1902494),86,2.3,9.45E-09,,,,,SF3A2
transferase complex (GO:1990234),46,2.1,6.94E-03,,,,,SNRPA
chromosomal part (GO:0044427),51,2.1,3.93E-03,,,,,SNRPC
chromosome (GO:0005694),57,2.0,2.13E-03,,,15629 48193,2,CAPZA2
ribonucleoprotein complex (GO:1990904),48,2.0,1.22E-02,,,,,CAPZA1
,,,,,,15629 6397,1,DHX9
,,,,,,6283 6397,3,CDK7
,,,,,,,,ERCC2
,,,,,,,,CCNH
,,,,,,15629,31,IQGAP1
,,,,,,,,ARPC5
,,,,,,,,PKD2
,,,,,,,,LLGL2
,,,,,,,,CIB2
,,,,,,,,KLHL2
,,,,,,,,ILK
,,,,,,,,TEK
,,,,,,,,PARVA
,,,,,,,,CLIC5
,,,,,,,,CLIC4
,,,,,,,,MYBPC3
,,,,,,,,ADAM8
,,,,,,,,NEB
,,,,,,,,PGM5
,,,,,,,,PARVB

```
,,,,,,,,VPS11
,,,,,,,,MYBPC1
,,,,,,,,TTC17
,,,,,,,,IPP
,,,,,,,,TOPBP1
,,,,,,,,COBL
,,,,,,,,CTTNBP2
,,,,,,,,GABARAP
,,,,,,,,MAEA
,,,,,,,,MYLK
,,,,,,,,MYOM3
,,,,,,,,PGM1
,,,,,,,,TMEM63B
,,,,,,,,SEPT7
,,,,,,,,MYOM2
,,,,,,6069,1,ACSS2
,,,,,,6283,7,RFC5
,,,,,,,,CUL4B
,,,,,,,,GPS1
,,,,,,,,TCEA1
,,,,,,,,LIG3
,,,,,,,,COPS7A
,,,,,,,,POLD1
,,,,,,48193,29,ERGIC1
,,,,,,,,LMAN2L
,,,,,,,,DOPEY1
,,,,,,,,COPA
,,,,,,,,GRIA1
,,,,,,,,PITPNB
,,,,,,,,AP2A1
,,,,,,,,RAB1A
,,,,,,,,KIF5A
,,,,,,,,TRAPPC3L
,,,,,,,,ARFGAP1
,,,,,,,,CNIH1
,,,,,,,,VTI1A
,,,,,,,,CNIH4
,,,,,,,,COPZ1
,,,,,,,,TRIM23
,,,,,,,,GGA3
,,,,,,,,BET1
,,,,,,,,RAB6A
,,,,,,,,PGAP1
,,,,,,,,TMED10
,,,,,,,,BCAP29
,,,,,,,,SNX1
,,,,,,,,TRAPPC12
,,,,,,,,SORL1
,,,,,,,,BET1L
,,,,,,,,EXOC6
,,,,,,,,KIFAP3
,,,,,,,,TRAPPC6A
,,,,,,6397,9,FAM172A
,,,,,,,,SLBP
,,,,,,,,CLASRP
,,,,,,,,THOC3
,,,,,,,,U2SURP
,,,,,,,,APP
```

```
,,,,,,,,CPSF4
,,,,,,,,AHCYL1
,,,,,,,,TSEN2
```

S5

Compx ExAC overlap list

159 genes

ABI2
AHCYL1
ALAS2
ANKRD50
AP1B1
AP2A1
AP2S1
ARID2
ARIH2
ARMC8
ARNT
ATG2A
ATXN7L3
BACE1
BNC2
CAPZA1
CAPZA2
CBX8
CCNC
CCZ1
CDC40
CDK1
CDK19
CDK5
CHD5
CLASRP
CNOT7
COBL
COL12A1
COL6A1
COPA
COPS7A
CUL4B
DEPDC5
DERL2
DHX8
DHX9
DLST
DOPEY1
DTX3
ERLIN1
EYA3
FAM49B
FCF1
FOXP1
GNL1
GPS1
GRIA1
GRIA4
GRIK2
HDAC2
HDAC8
HERC2
HMBS
HNRNPH1

1 HOOK3
2 IDS
3 ILKAP
4 IP6K2
5 IPO11
6 IQGAP1
7 IRAK1
8 KCNIP4
9 KCTD1
10 KCTD3
11 KIF5A
12 KIF5C
13 KLF6
14 KLHL10
15 KLHL2
16 KMT2E
17 KPNA1
18 KPNA6
19 LMNA
20 MAP3K5
21 MDH1
22 MED12L
23 MIER1
24 MYCBP2
25 NAPB
26 NFIA
27 NODAL
28 NR4A1
29 NR4A2
30 OLFM1
31 PBRM1
32 PDS5A
33 PDS5B
34 PGM5
35 PHF12
36 PIP4K2B
37 PITPNA
38 PITPNB
39 PKD2
40 PLXNB2
41 PLXNC1
42 PORCN
43 PPFIA1
44 PPP1R8
45 PPP2R2B
46 PRMT1
47 PRPF19
48 PSME3
49 RAD54L2
50 RAE1
51 RB1
52 RBL2
53 REV3L
54 RMND5A
55 RNF220
56 RPL4
57 RRM2
58 SCAF4

SENP3
SENP5
SESN3
SF3A1
SIX1
SMARCD2
SMARCE1
SNRNP200
SNX1
SOS2
ST3GAL2
ST7
STAG1
STMN2
STXBP1
SUFU
TBL1XR1
TCP1
TEK
TEX10
TFAP2B
THSD7A
TMEM131
TMEM63B
TOPBP1
TSC2
TSN
TTC17
TTC7B
U2SURP
UBE2F
UBE3A
UBE3C
UBN1
UHRF1BP1L
USP15
USP19
USP46
XAB2
XRCC6
ZC3H14
ZDHHC17
ZEB1
ZMYND11
ZNF687
ZSWIM8

S6

,,,,,,#GO term overlap data,,
GO biological process complete,N,Enrich,P-value,,,GO terms,Number,Genes
regulation of G1/S transition of mitotic cell cycle
(GO:2000045),9,8.02,2.69E-02,,,16032 16570 1902494 1990234 2000045
51276 6325 70647 9057,1,CDK1
histone modification (GO:0016570),14,4.8,1.82E-02,,,16032 1902494
1990234 2000045 3712 51276 6325 785 8134,1,RB1
chromatin organization (GO:0006325),22,4.03,3.85E-04,,,16570 1902494
1990234 2000045 51276 6325 70647 9057,1,CUL4B
viral process (GO:0016032),18,3.71,2.16E-02,,,16570 1902494 1990234
3712 51276 6325 70647,1,ATXN7L3
chromosome organization (GO:0051276),30,3.68,7.56E-06,,,16570 1902494
1990234 51276 6325 70647 785,1,CBX8
protein modification by small protein conjugation or removal
(GO:0070647),24,3.27,3.84E-03,,,16570 1902494 1990234 51276 6325 785
8134,1,HDAC2
macromolecule catabolic process (GO:0009057),24,3.01,1.57E-02,,,16570
1902494 1990234 2000045 51276 6325,1,PRMT1
protein localization (GO:0008104),40,2.4,1.56E-03,,,16570 1902494 3712
51276 6325 9057,1,TBL1XR1
protein-containing complex subunit organization
(GO:0043933),34,2.38,2.03E-02,,,16570 1902494 1990234 51276 6325
9057,1,CDK5
macromolecule localization (GO:0033036),42,2.18,1.05E-02,,,1902494 3712
51276 6325 785 8134,1,SMARCE1
nucleic acid metabolic process (GO:0090304),39,2.14,4.19E-02,,,16032
1902494 3712 51276 785 8134,1,DHX9
cellular protein modification process (GO:0006464),49,2.03,7.05E-
03,,,16570 2000045 51276 6325 785,1,KMT2E
,,,,,,16570 51276 6325 70647 9057,1,USP15
,,,,,,16570 3712 51276 6325 8134,1,MIER1
GO molecular function complete,N,Enrich,P-value,,,16570 1902494 51276
6325 8134,1,HDAC8
transcription coregulator activity (GO:0003712),16,3.58,3.98E-
02,,,16570 1902494 51276 6325 785,1,CHD5
transcription factor binding (GO:0008134),18,3.43,1.98E-02,,,2000045
51276 6325 785 8134,1,RBL2
,,,,,,1902494 3712 51276 6325 785,1,SMARCD2
,,,,,,16032 3712 51276 6325,1,ZMYND11
GO cellular component complete,N, Enrich,P-value,,,1902494 51276 6325
785,1,PBRM1
catalytic step 2 spliceosome (GO:0071013),7,10.18,1.32E-02,,,16032
1902494 1990234 51276,1,RAE1
chromatin (GO:0000785),16,3.8,9.83E-03,,,2000045 3712 8134 9057,1,CNOT7
catalytic complex (GO:1902494),39,3.54,6.72E-09,,,16032 1902494 70647
9057,1,UBE3A
transferase complex (GO:1990234),20,3.19,8.67E-03,,,1902494 70647 71013
9057,1,PRPF19
nucleoplasm part (GO:0044451),26,2.94,1.40E-03,,,1902494 1990234 70647
9057,2,RMND5A
chromosome (GO:0005694),22,2.69,3.96E-02,,,,,ARIH2
nucleoplasm (GO:0005654),73,2.65,2.72E-13,,,1902494 3712 785
8134,1,PHF12
nuclear lumen (GO:0031981),79,2.44,9.19E-13,,,16570 51276 6325,1,EYA3
protein-containing complex (GO:0032991),98,2.3,2.53E-16,,,16032 51276
6325,1,UBN1
nuclear part (GO:0044428),79,2.22,1.83E-10,,,16032 51276 70647,1,COPS7A

```
organelle lumen (GO:0043233),87,2.1,7.51E-11,,,1902494 2000045
70647,1,PSME3
intracellular organelle lumen (GO:0070013),87,2.1,7.51E-11,,,16032
1902494 1990234,1,MAP3K5
membrane-enclosed lumen (GO:0031974),87,2.1,7.51E-11,,,1902494 70647
9057,1,UBE3C
,,,,,,1902494 1990234 70647,2,KLHL2
,,,,,,,,SENP3
,,,,,,1902494 1990234 9057,1,ARMC8
,,,,,,51276 6325,1,ARID2
,,,,,,16032 51276,1,XRCC6
,,,,,,51276 70647,1,GPS1
,,,,,,51276 785,3,STAG1
,,,,,,,,PDS5B
,,,,,,,,PDS5A
,,,,,,16032 9057,3,KPNA1
,,,,,,,,AP2A1
,,,,,,,,AP2S1
,,,,,,16032 8134,1,NFIA
,,,,,,70647 9057,3,HERC2
,,,,,,,,USP46
,,,,,,,,USP19
,,,,,,3712 8134,4,ARNT
,,,,,,,,SUFU
,,,,,,,,ZEB1
,,,,,,,,KCTD1
,,,,,,1902494 71013,6,SNRNP200
,,,,,,,,HNRNPH1
,,,,,,,,XAB2
,,,,,,,,DHX8
,,,,,,,,SF3A1
,,,,,,,,CDC40
,,,,,,1902494 1990234,5,REV3L
,,,,,,,,TEX10
,,,,,,,,ZSWIM8
,,,,,,,,DLST
,,,,,,,,DEPDC5
,,,,,,2000045,1,PKD2
,,,,,,16032,4,IRAK1
,,,,,,,,AP1B1
,,,,,,,,TSC2
,,,,,,,,ABI2
,,,,,,70647,6,MYCBP2
,,,,,,,,UBE2F
,,,,,,,,RNF220
,,,,,,,,KLHL10
,,,,,,,,DTX3
,,,,,,,,SENP5
,,,,,,9057,6,BACE1
,,,,,,,,IDS
,,,,,,,,PPP1R8
,,,,,,,,DERL2
,,,,,,,,ERLIN1
,,,,,,,,RPL4
,,,,,,3712,2,TFAP2B
,,,,,,,,MED12L
,,,,,,8134,4,FOXP1
,,,,,,,,SIX1
```

```
,,,,,,,,NR4A2
,,,,,,,,NR4A1
,,,,,,785,2,SOS2
,,,,,,,,TCP1
,,,,,,1902494,1,PPP2R2B
```

S7

```
,,,ELEMENT,,,,,,,,,,,GO-TERM,,,,,,,,,,,
Gene,Remod,Mediator,ChromMod,Splicing,RNAexport,Translation,Transport,U
biq-related,Cell cycle,DNA
repair,STRING,2000045,16570,6325,16032,51276,70647,9057,3712,8134,71013
,785,1902494,1990234,,Gene card functions,Reference DOI:
ARID2,X,,,,,,,,,A,,,X,,X,,,,,,,,,SWI/SNF chromatin remodeling
complex,10.1074/jbc.M111.279968.
CBX8,X,,X,,,,,,,A,,X,X,,X,X,,,,,,X,X,,"Polycomb group (PcG)
multiprotein PRC1-like complex, chromatin remodeling and modification
of histones, monoubiquitination of histone H2A 'Lys-
119',",10.1093/nar/gky1290
PBRM1,X,,,,,,,,,A,,,X,,X,,,,,X,X,,,SWI/SNF chromatin remodeling
complex/numerous DNA repair mechanisms /promoting genome stability
,10.1021/acschembio.7b00541 /10.1016/j.celrep.2014.02.012.
SMARCD2,X,,,,,,,,,A,,,X,,X,,,X,,,X,X,,,SWI/SNF chromatin remodeling
complexes,10.1074/jbc.R111.309302
SMARCE1,X,,,,,,,,,A,,,X,,X,,,X,X,,X,X,,,SWI/SNF chromatin remodeling
complexes,10.1074/jbc.R111.309302
XRCC6,X,,,,,,,,,X,A,,,,X,X,,,,,,,,,Single-stranded DNA-dependent ATP-
dependent helicase / Involved in DNA non-homologous end joining (NHEJ)
/ DNA repair,10.1016/j.mrrev.2014.06.002.
ZSWIM8,X,,,,,,,,,A,,,,,,,,,,,X,X,,Zinc Finger SWIM-Type Containing 8
/ uncertain function,ND
CCNC,,X,,,,,,,,A,,,,,,,,,,,,,string only,Component of the Mediator
complex/The level of mRNAs for this gene peaks in the G1 phase of the
cell cycle,10.3390/biology8010003.
CDK19,,X,,,,,,,,A,,,,,,,,,,,,,string only,one of the components of
the Mediator co-activator complex,10.3389/fcell.2018.00171
MED12L,,X,,,,,,,,A,,,,,,,,X,,,,,,component of the Mediator
complex,10.1038/nrm.2017.115.
CHD5,,,X,,,,,,,,A,,X,X,,X,,,,,X,X,,,Chromatin-remodeling protein /
chromodomain helicase DNA-binding protein / DNA
repair,10.1016/j.mrfmmm.2013.07.008.
HDAC2,,,X,,,,,,,,A,,X,X,,X,,,,X,,X,X,X,,deacetylation of lysine
residues on the N-terminal part of the core histones ,
10.1387/ijdb.150071tz.
HDAC8,,,X,,,,,,,,A,,X,X,,X,,,,X,,,X,,,deacetylation of lysine residues
on the N-terminal part of the core histones / a major 'epigenetic
player' ,10.1016/j.tips.2015.04.013.
RB1,,,X,,,,,,,,A,X,,X,X,X,,,X,X,,X,X,X,,entry into cell division/ acts
as a tumor suppressor/Directly involved in heterochromatin formation
/Recruits and targets histone methyltransferases / recruiting a histone
deacetylase (HDAC) complex,10.1016/j.jaut.2016.05.012
RBL2,,,X,,,,,,,,A,X,,X,,X,,,,X,,X,,,,Key regulator of entry into cell
division/Directly involved in heterochromatin formation /Recruits and
targets histone methyltransferases ,PMID: 12225593
TBL1XR1,,,X,,,,,X,,,A,,X,X,,X,,X,X,,,,X,,,nuclear receptor corepressor
(N-CoR) and histone deacetylase 3 (HDAC 3) complexes / recruitment of
the ubiquitin/19S proteasome complex,PMC4446378
UBN1,,,X,,,,,,,,A,,,X,X,X,,,,,,,,,Acts as a novel regulator of
senescence / HIRA complex/ Adds histone H3.3to nucleosomes -
remodelling replacement,10.1186/s12915-018-0573-9.
PHF12,,,X,,,,,,,,A,,,,,,,,X,X,,X,X,,,recruitment of functional SIN3A
complexes / histone deacetylase (HDAC),10.1128/MCB.21.13.4110-4118.2001
CDC40,,,,X,,,,,X,,A,,,,,,,,,,X,,X,,,Associates with the
spliceosome /second step of pre-mRNA splicing/ control cell cycle
```

```
through splicing of intron-containing pre-mRNAs that encode proteins
important for cell cycle progression.,10.1007/s00294-006-0113-y
DHX8,,,,X,X,,,,,,A,,,,,,,,,,X,,X,,,Involved in pre-mRNA splicing as
component of the spliceosome/ Facilitates nuclear export of spliced
mRNA by releasing the RNA from the spliceosome
,10.1016/j.cell.2016.01.025.
HNRNPH1,,,,X,,,,,,,A,,,,,,,,,,X,,X,,,Mediates pre-mRNA alternative
splicing regulation.,10.1261/rna.036715.112.
PRPF19,,,,X,,,,X,,,A,,,,,,X,X,,,X,,X,,,Ubiquitin-protein ligase / part
of the spliceosome / DNA repair,10.1101/gad.1925010
SF3A1,,,,X,,,,,,,A,,,,,,,,,,X,,X,,,splicing
factor,10.1101/gad.248625.114.
SNRNP200,,,,X,,,,,,,A,,,,,,,,,,X,,X,,,RNA helicase that plays an
essential role in pre-mRNA splicing ,10.1042/BST0381105.
XAB2,,,,X,X,,,,,X,A,,,,,,,,,,X,,X,,,"Involved in pre-mRNA splicing as
component of the spliceosome / transcription-coupled repair /plays
critical role in processes including transcription, transcription-
coupled DNA repair, pre-mRNA splicing, homologous recombination and
mRNA export./mitotic cell cycle regulation",10.1074/jbc.M004936200  /
10.1038/cddis.2016.313.
DHX9,,,,X,,X,,,X,,A,,,,X,X,,,X,X,,X,X,,,"DNA replication,
transcriptional activation, post-transcriptional RNA regulation, mRNA
translation and RNA-mediated gene silencing/Plays a role in DNA
replication at origins of replication and cell cycle progression
/promotes pre-mRNA alternative/  splicing activities/ a role in mRNA
translation",10.18632/oncotarget.8446.
U2SURP,,,,X,,,,,,,A,,,,,,,,,,,,,String only,link the activity of the
spliceosome to the regulation of downstream RNA-binding
proteins,10.1016/j.celrep.2018.09.041.
KIF5C,,,,,X,,X,,,,A,,,,,,,,,,,,,String only,transport of cargo within
the central nervous system. Vesicle associated,10.1007/s00018-009-0253-
6.
RAE1,,,,,X,,,,,,A,,,,X,X,,,,,,,X,X,,involved in RNA export /
nucleocytoplasmic transport,10.1016/j.cell.2005.04.003
FCF1,,,,,,X,,,,,A,,,,,,,,,,,,,String only,pre-rRNA processing and 40S
ribosomal subunit assembly.,10.1016/j.bbrc.2006.05.140
RPL4,,,,,,X,,,,,A,,,,,,,X,,,,,,,,a ribosomal protein that is a
component of the 60S subunit,
SENP3,,,,,,X,,,,,A,,,,,,X,,,,,X,X,,Protease that releases SUMO2 and
SUMO3 monomers from sumoylated substrates, 10.1002/iub.113.
TEX10,,,,,,X,,,,,A,,,,,,,,,,,,X,X,,component of the Five Friends of
Methylated CHTOP (5FMC) complex,10.1074/mcp.M112.017194
AP1B1,,,,,,,X,,,,A,,,,X,,,,,,,,,,at the cytoplasmic face of coated
vesicles located at the Golgi complex/role in protein sorting in the
late-Golgi/trans-Golgi network (TGN) and/or endosomes.
,10.1371/journal.pone.0060866.
AP2A1,,,,,,,X,,,,A,,,,X,,,X,,,,,,,,Adaptor protein complexes function
in protein transport via transport vesicles in different membrane
traffic pathways./nvolved in cargo selection and vesicle formation. /
clathrin-dependent endocytosis /binds polyphosphoinositide-containing
lipids,10.1111/cpr.12522.
AP2S1,,,,,,,X,,,,A,,,,X,,,X,,,,,,,,Adaptor protein complexes function
in protein Transport via Transport vesicles in different membrane
traffic pathways./ involved in cargo selection and vesicle
formation,PMID: 7593184
COPA,,,,,,,X,,,,A,,,,,,,,,,,,,,String only,"Coatomer complex / non-
clathrin-coated vesicular coat proteins /mediate biosynthetic protein
```

```
transport from the ER, via the Golgi up to the trans Golgi
network.",10.1038/ng.3279.
KIF5A,,,,,,,X,,,,A,,,,,,,,,,,,,,String only,Microtubule-dependent motor
required for slow axonal transport of neurofilament proteins/
Microtubule-dependent motor required for slow axonal transport of
neurofilament proteins/ALS,10.1016/j.neuron.2012.10.012.
SUFU,,,,,,,X,,,,A,,,,,,,,X,X,,,,,,Negative regulator in the
hedgehog/smoothened signaling pathway.,
ARIH2,,,,,,,,X,,,A,,,,,,X,X,,,,,X,X,,E3 ubiquitin-protein ligase works
ogether with cullin-5-RING ubiquitin ligase complex activity is
activated upon binding to neddylated form of the ECS
complex,10.1182/blood-2005-04-1450
COPS7A,,,,,,,,X,,,A,,,,X,X,X,,,,,,,,,component of the COP9 signalosome
complex regulates the activity of the ubiquitin conjugation pathway by
mediating the deneddylation of the cullin subunits of SCF-type E3
ligase complexes,PMID: 11861754
CUL4B,,,,,,,,X,,,A,X,X,X,,X,X,X,,,,,X,X,,E3 ubiquitin ligase - may act
as a scaffold protein,10.1016/j.gene.2015.08.064
GPS1,,,,,,,,X,,,A,,,,,X,X,,,,,,,,,component of the COP9 signalosome
complex regulates the activity of the ubiquitin conjugation pathway by
mediating the deneddylation of the cullin subunits of SCF-type E3
ligase complexes,PMID: 11861754
HERC2,,,,,,,,X,,,A,,,,,,X,X,,,,,,,,E3 ubiquitin-protein ligase that
regulates ubiquitin-dependent retention of repair proteins on damaged
chromosomes,10.1158/0008-5472.
KLHL10,,,,,,,,X,,,A,,,,,,X,,,,,,,,,May be a substrate-specific adapter
of a CUL3-based E3 ubiquitin-protein ligase
complex ,10.1095/biolreprod.105.045484
KLHL2,,,,,,,,X,,,A,,,,,,X,,,,,,X,X,,Component of a cullin-RING-based
BCR (BTB-CUL3-RBX1) E3 ubiquitin-protein ligase complex / Actin
binding,10.1016/j.bbrc.2013.06.104.
MYCBP2,,,,,,,,X,,,A,,,,,,X,,,,,,,,,E3 ubiquitin-protein ligase / also
acts as a guanosine exchange factor (GEF) for RAN,10.1038/s41586-018-
0026-1
PSME3,,,,,,,,X,,,A,X,,,,,X,,,,,,X,,,"proteasome regulator/ Facilitates
the MDM2-p53/TP53 interaction which promotes ubiquitination- and MDM2-
dependent proteasomal degradation of p53/TP53, limiting its
accumulation ",10.1007/s10495-015-1149-6.
RNF220,,,,,,,,X,,,A,,,,,,X,,,,,,,,,E3 ubiquitin-protein
ligase,10.1128/MCB.00731-14.
UBE2F,,,,,,,,X,,,A,,,,,,X,,,,,,,,,ubiquitin protein ligase activity /
E2,10.1016/j.molcel.2009.01.011.
UBE3A,,,,,,,,X,,,A,,,,X,,X,X,,,,,X,,,E3 ubiquitin-protein ligase,doi:
10.1146/annurev.biochem.78.081507.101607.
UBE3C,,,,,,,,X,,,A,,,,,,X,X,,,,,X,,,E3 ubiquitin-protein ligase,doi:
10.1146/annurev.biochem.78.081507.101607.
IRAK1,,,,,,,,X,,,A,,,,,,,,,,,,,,,"Serine/threonine-protein kinase
Phosphorylates E3 ubiquitin ligases Pellino proteins/phosphorylates
TIRAP to promote its ubiquitination and subsequent degradation.
Phosphorylates the interferon regulatory factor 7 (IRF7) to induce its
activation and translocation to the nucleus,  When sumoylated,
translocates to the nucleus and phosphorylates
STAT3",10.1016/j.cellsig.2007.08.009
CDK1,,,,,,,,,X,,A,X,X,X,X,X,X,X,,,,,X,X,,key role in the control of the
eukaryotic cell cycle ,10.1007/s00294-018-00926-y
CDK5,,,,,,,,,X,,A,,X,X,,X,,X,,,,,X,X,,essential for neuronal cell cycle
arrest and differentiation / atypical CDK/ neuronal actin cytoskeletal
remodeling,10.1007/s12035-017-0525-3
```

```
PDS5A,,,,,,,,,X,A,,,,,X,,,,,,X,,,,regulator of sister chromatid
cohesion in mitosis / DNA repair,10.1002/bies.201600207
PDS5B,,,,,,,,,X,A,,,,,X,,,,,,X,,,,regulator of sister chromatid
cohesion in mitosis / DNA repair,10.1002/bies.201600207
STAG1,,,,,,,,,X,A,,,,,X,,,,,,X,,,,"a component of cohesin / Cohesin is
a multi-protein complex that tethers sister chromatids during mitosis
and mediates DNA repair,",10.18632/oncotarget.16838.
,,,,,,,,,,,,,,,,,,,,,,,,,,,,,
ATXN7L3,,,X,,,,,X,,,,,X,X,,X,X,,X,,,,X,X,,HAT histone acetylation (HAT)
complex SAGA  / de-ubiquitinase,10.1016/j.molcel.2007.12.011.
EYA3,,,X,,,,,,,X,,,X,X,,X,,,,,,,,,Tyrosine phosphatase /
dephosphorylates 'Tyr-142' of histone H2AX /DNA repair
/apoptosis,10.1038/nature07849.
KMT2E,,,X,,,,,,,,,X,X,X,,X,,,,,,X,,,,Epigeneticwriter/ histone
methyltransferase/ Chromatin interaction is mediated via the binding to
tri-methylated histone H3 at 'Lys-4' (H3K4me3)
,10.1371/journal.pone.0077020.
MIER1,,,X,,,,,,,,,X,X,,X,,,X,X,,,,,,Transcriptional repressor /
recruitment of HDAC1 a histone
deacetylase,10.1371/journal.pone.0169338.
PRMT1,,,X,,,,,,,,,X,X,X,,X,,,,,,,X,X,,Arginine
methyltransferase/monomethylation and asymmetric dimethylation of
histone H4,10.1111/j.1742-4658.2006.05399.x
ZMYND11,,,X,,,,,,,,,,X,X,X,,,X,,,,,,,Chromatin reader that
specifically recognizes and binds histone H3.3 trimethylated at 'Lys-
36' / unconventional transcription co-repressor by modulating RNA
polymerase II,doi: 10.1038/nature13045.
DLST,,,X,,,,,,,,,,,,,,,,,,,,X,X,,In mitochondria/ also lysine
succinylation of histones:,10.1038/nature25003.
PPP2R2B,,,X,X,,,,,,,,,,,,,,,,,,,X,,,phosphatase 2 regulatory subunit B
family / function uncertain,
REV3L,,,X,,,,,,X,,,,,,,,,,,,X,X,,Catalytic subunit of the DNA
polymerase zeta complex/ DNA repair via translesion
synthesis,10.3390/ijms20010100.
CNOT7,,,,,,X,,,,,,X,,,,,,X,X,X,,,,,one of the major cellular mRNA
deadenylases / bulk mRNA degradation,10.1042/BST20120074.
BACE1,,,,,,,X,,,,,,,,,,X,,,,,,,, proteolytic processing of the amyloid
precursor protein,10.2174/156720508783954758
DERL2,,,,,,,X,X,,,,,,,,,X,,,,,,,endoplasmic reticulum-associated
degradation (ERAD),10.1083/jcb.200507057
ERLIN1,,,,,,,X,X,,,,,,,,,X,,,,,,,"mediates degradation of inositol
1,4,5-trisphosphate receptors in the endoplasmic reticulum./he
endoplasmic reticulum-associated degradation
(ERAD)",10.1074/jbc.RA118.004547.
KPNA1,,,,,,,X,,,,,,,X,,,X,,,,,,,involved in nuclear protein
import,10.1083/jcb.201702092.
PPP1R8,,,,,,,X,,,,,,,,,,X,,,,,,,Inhibitor subunit of the major
nuclear protein phosphatase-1 / nvolved in pre-mRNA splicing /
chromatin modification,10.1074/jbc.M805468200. / 10.1074/jbc.M302273200
USP15,,,,,,,X,X,,,,,X,X,,X,X,X,,,,,,,Deubiquitinating enzyme/
Hydrolase that removes conjugated ubiquitin from target proteins /
associates with the COP9 signalosome  / Regulates ER-Golgi vesicle
trafficking ,10.1016/j.cell.2016.05.078.
ABI2,,,,,,,X,,,,,,,,X,,,,,,,,,,Regulator of actin cytoskeleton
dynamics /controls dendritic spine morphogenesis/endocytic trafficking/
WAVE complex,10.3892/ijmm.9.6.591
PKD2,,,,,,,X,,,,,X,,,,,,,,,,,,,"a multi-pass membrane protein that
functions as a calcium permeable cation channel,",10.1113/JP270933.
```

USP19,,,,,,,,X,,,,,,,,,X,X,,,,,,,,Deubiquitinating enzyme/ ER ERAD,https://doi.org/10.3390/ijms17111829

USP46,,,,,,,,X,,,,,,,,,X,X,,,,,,,,Deubiquitinating enzyme/,10.1074/jbc.M808430200

ARNT,,,,,,,,X,,,,,,,,,,,X,X,,,,,,Component of the CTLH E3 ubiquitin-protein ligase complex ,10.1371/journal.pone.0075217

DTX3,,,,,,,,X,,,,,,,,,,X,,,,,,,,E3 ubiquitin ligase ,10.1074/jbc.M301157200

KCTD1,,,,,,,,X,,,,,,,,,,,X,X,,,,,,Transcriptional repression through AP-2 / Interacts with Cullin E3-ubiquitin ligases,10.1042/BCJ20170527

ARMC8,,,,,,,,,,,,,,,,,,X,,,,,X,X,,mitochondrially located enzyme / heme biosynthetic pathway,10.1182/asheducation-2012.1.19.

DEPDC5,,,,,,,,,,,,,,,,,,,,,,,X,X,,component of the GATOR1 complex / an inhibitor of the amino acid-sensing branch of the TORC1 pathway,10.1016/j.nbd.2016.02.010.

FOXP1,,,,,,,,,,,,,,,,,,,,X,,,,,,forkhead box (FOX) transcription factor,10.1101/gad.1637108.

IDS,,,,,,,,,,,,,,,,,,X,,,,,,,, member of the sulfatase family of proteins / lysosomal degradation of heparan sulfate,

MAP3K5,,,,,,,,,,,,,,,X,,,,,,,,X,X,,essential component of the MAP kinase signal transduction pathway,

NFIA,,,,,,,,,,,,,,,X,,,X,,,,,,,,a transcription factor / possible epigenetic role,

NR4A1,,,,,,,,,,,,,,,,,,,,X,,,,,,transcription factor of of the steroid-thyroid hormone-retinoid receptor superfamily,

NR4A2,,,,,,,,,,,,,,,,,,,,X,,,,,,transcription factor of the steroid-thyroid hormone-retinoid receptor superfamily,

RMND5A,,,,,,,,,,,,,,,,,X,X,,,X,,X,X,,Core component of the CTLH E3 ubiquitin-protein ligase complex / relates to cell cycle,10.7554/eLife.35528

SENP5,,,,,,,,,,,,,,,,,X,,,,,,,,,Protease that releases SUMO2 and SUMO3 monomers from sumoylated substrates, 10.1002/iub.113.

SIX1,,,,,,,,,,,,,,,,,,,,X,,,,,,Transcription factor ,

SOS2,,,,,,,,,,,,,,,,,,,,,,X,,,,SOS Ras/Rho Guanine Nucleotide Exchange Factor 2 / interaction of Sos1 with the COP9 signalosome,10.1128/MCB.00049-18. / 10.1038/s41389-018-0111-1.

TCP1,,,,,,,,,,,,,,,,,,,,,,X,,,,As part of the TRiC complex may play a role in the assembly of BBSome,

TFAP2B,,,,,,,,,,,,,,,,,,,X,,,,,,,Sequence-specific DNA-binding protein transcription factor,

TSC2,,,,,,,,,,,,,,,X,,,,,,,,,,String,"Acts as a GTPase-activating protein (GAP) for the small GTPase RHEB, / May also play a role in microtubule-mediated protein transport",

ZEB1,,,,,,,,,,,,,,,,,,,X,X,,,,,,encodes a zinc finger transcription factor /Represses E-cadherin promoter and induces an epithelial-mesenchymal transition (EMT) by recruiting SMARCA4/BRG1,10.1038/onc.2010.102.

,,,,,,,,,,,,,,,,,,,,,,,,,,,

,,,,,,,,,,,,,,,,,,,,,,,,,,,

TOPBP1,X,,,,,,,,,X,,,,,,,,,,,,,,,,,Recruits the SWI/SNF chromatin remodeling complex /initiation of DNA replication and regulation of checkpoint responses./DNA repair,10.1016/j.cell.2005.12.041/10.1016/j.dnarep.2014.06.004.

ZNF687,X,,X,,,,,,,,,,,,,,,,,,,,,,,,May be involved in transcriptional regulation/ interacts with The nucleosome remodeling and deacetylase (NuRD) complex,10.1111/febs.12972

```
LMNA,,,X,,,,,,,,,,,,,,,,,,,,,,"components of the nuclear lamina /
important role in nuclear assembly, chromatin organization, nuclear
membrane and telomere dynamics",10.4161/19491034.2014.990855.
ZC3H14,,,X,,X,X,,,,,,,,,,,,,,,,,,"Involved in poly(A) tail length
control in neuronal cells/may influence mRNA stability, nuclear export,
and translation. / quality control of mRNA splicing and export
",PMID: 27868086
CLASRP,,,,X,,,,,,,,,,,,,,,,,,,,,alternative splicing
regulator/,10.1074/jbc.M206504200
SCAF4,,,,X,,,,,,,,,,,,,,,,,,,,,provide a link between transcription
and pre-mRNA splicing.,10.1074/jbc.M803540200
AHCYL1,,,,,,,X,,,,,,,,,,,,,,,,,,"Multifunctional, but regulates some
receptor levels at the cell surface / involves
IP3",10.1016/j.bbamcr.2014.01.031
ANKRD50,,,,,,,X,,,,,,,,,,,,,,,,,"retromer-mediated sorting/Involved
in the endosome-to-plasma membrane trafficking and recycling of SNX27-
retromer-dependent cargo proteins,",10.1242/jcs.156299.
ATG2A,,,,,,,X,,,,,,,,,,,,,,,,,,binds to ligand-bound aryl hydrocarbon
receptor and aids in the movement of this complex to the nucleus
TRANSPORT,10.1016/j.bcmd.2010.01.005.
CCZ1,,,,,,,X,,,,,,,,,,,,,,,,,,"a guanine exchange factor (GEF) for
RAB7,/required for endosomal maturation and fusion at the
vacuole/lysosome",10.1038/ncomms14034.
DOPEY1,,,,,,,X,,,,,,,,,,,,,,,,,,Golgi to endosome
trafficking,10.1002/glia.22698
HOOK3,,,,,,,X,,,,,,,,,,,,,,,,,,Component of the FTS/Hook/FHIP complex
(FHF complex). The FHF complex may function to promote vesicle
trafficking and/or fusion via the homotypic vesicular protein sorting
complex (the HOPS complex)./ an adapter protein linking the dynein
motor complex to various cargos ,doi: 10.1021/acs.biochem.8b01065.
KCNIP4,,,,,,,X,,,,,,,,,,,,,,,,,,Trafficking of potassium channels from
the ER,10.1074/jbc.M306142200
KPNA6,,,,,,,X,,,,,,,,,,,,,,,,,,involved in nuclear protein
import,10.1083/jcb.201702092.
NAPB,,,,,,,X,,,,,,,,,,,,,,,,,,SNAP beta isoform/vesicular transport
between the endoplasmic reticulum and the Golgi
apparatus,10.1111/cge.12648
PITPNB,,,,,,,X,,,,,,,,,,,,,,,,,,Catalyzes the transfer of PtdIns and
phosphatidylcholine between membranes. / Golgi
transport,10.1016/j.devcel.2018.02.025.
SNX1,,,,,,,X,,,,,,,,,,,,,,,,,,"retrograde transport of cargo proteins
from endosomes to the trans-Golgi network/ member of the sorting nexin
family/Interacts with membranes containing phosphatidylinositol 3-
phosphate (PtdIns(3P)) or phosphatidylinositol 3,5-bisphosphate
(PtdIns(3,5)P2)/",10.1016/j.ejcb.2015.07.002
STXBP1,,,,,,,X,,,,,,,,,,,,,,,,,,a component of the synaptic vesicle
fusion machinery / SNARE complex ,10.1523/JNEUROSCI.0007-16.2016.
UHRF1BP1L,,,,,,,X,,,,,,,,,,,,,,,,,,Syntaxin associated,10.1111/j.1600-
0854
ZDHHC17,,,,,,,X,,,,,,,,,,,,,,,,,,"Palmitoyltransferase specific for a
subset of neuronal proteins, including SNAP25,/ May be involved in the
sorting or targeting of critical proteins involved in the initiating
events of endocytosis at the plasma membrane ",10.1038/nn1702
IPO11,,,,,,,X,,,,,,,,,,,,,,,,,,"Functions in nuclear protein import as
nuclear transport receptor / mediate nucleocytoplasmic transport of
protein and RNA /nuclear import is to protect PTEN from cytosolic
degradation initiated by E3 ubiquitin ligase
components,",10.1083/jcb.201612014
```

```
KCTD3,,,,,,,,X,,,,,,,,,,,,,,,,, Interacts with Cullin E3-ubiquitin
ligases,10.1042/BCJ20170527
ST7,,,,,,,,X,,,,,,,,,,,,,,,,,May act as a tumor suppressor.,
BNC2,,,,,,,,X,,,,,,,,,,,,,,,,transcription factor specific for skin
keratinocytes/ cell cycle,10.1371/journal.pgen.1000744
ILKAP,,,,,,,,X,,,,,,,,,,,,,,,, protein serine/threonine phosphatase
of the PP2C family./ may play a role in regulation of cell cycle
progression / nucleocytoplasmic transport of ILK via nuclear
pore,10.4161/cc.7.14.6241
TSN,,,,,,,,,X,,,,,,,,,,,,,,,"DNA-binding protein that specifically
recognizes consensus sequences at the breakpoint junctions in
chromosomal translocations, / DNA repair",10.1074/jbc.272.17.11402
ALAS2,,,,,,,,,,,,,,,,,,,,,,,,,catalyzes the transfer of sialic acid
from CMP-sialic acid to galactose-containing
substrates,10.1093/glycob/cww098.
CAPZA1,,,,,,,,,,,,,,,,,,,,,,,,String,F-actin-capping protein /regulate
actin cytoskeleton remodelling,doi: 10.1186/s13046-016-0474-0.
CAPZA2,,,,,,,,,,,,,,,,,,,,,,,,String,F-actin capping protein,
COBL,,,,,,,,,,,,,,,,,,,,,,,,reorganization of the actin cytoskeleton
/Regulates neuron morphogenesis and increases branching of axons and
dendrites / actin nucleator Cobl is critical for neuronal
morphogenesis,10.1016/j.devcel.2018.03.007.
COL12A1,,,,,,,,,,,,,,,,,,,,,,,,alpha chain of type XII collagen,
COL6A1,,,,,,,,,,,,,,,,,,,,,,,,alpha chain of type 6 collagen,
FAM49B,,,,,,,,,,,,,,,,,,,,,,,,repressing Rac activity and modulating
cytoskeleton reorganization.,10.1073/pnas.1801340115
GNL1,,,,,,,,,,,,,,,,,,,,,,,,GTPase,
GRIA1,,,,,,,,,,,,,,,,,,,,,,,,String,Ionotropic glutamate receptor. L-
glutamate acts as an excitatory neurotransmitter at many synapses in
the central nervous system,
GRIA4,,,,,,,,,,,,,,,,,,,,,,,,String,Receptor for glutamate that
functions as ligand-gated ion channel in the central nervous system and
plays an important role in excitatory synaptic transmission,
GRIK2,,,,,,,,,,,,,,,,,,,,,,,,Ionotropic glutamate receptor. L-
glutamate acts as an excitatory neurotransmitter at many synapses in
the central nervous system.,
HMBS,,,,,,,,,,,,,,,,,,,,,,,,heme biosynthetic pathway,
IP6K2,,,,,,,,,,,,,,,,,,,,,,,,inositol phosphokinase (IPK)
family,10.1038/ncomms5178
IQGAP1,,,,,,,,,,,,,,,,,,,,,,,,"regulating the dynamics and assembly
of the actin cytoskeleton / a highly conserved cytoplasmic scaffold
protein, is able to curb, compartmentalize, and coordinate multiple
signaling pathways in a variety of cell
types.",10.1016/j.molimm.2015.02.012.
KLF6,,,,,,,,,,,,,,,,,,,,,,,,Kruppel-like family of transcription
factors,10.1038/s41598-018-31101-5.
MDH1,,,,,,,,,,,,,,,,,,,,,,,,"enzyme that catalyzes the NAD/NADH-
dependent, reversible oxidation of malate to oxaloacetate",
NODAL,,,,,,,,,,,,,,,,,,,,,,,,secreted ligand of the TGF-beta,
OLFM1,,,,,,,,,,,,,,,,,,,,,,,,exact function of the encoded protein is
not known /Olfm1 may regulate receptor trafficking from the
intracellular compartments to the synaptic membrane microdomain/
synaptosomes,10.1111/jnc.14231.
PGM5,,,,,,,,,,,,,,,,,,,,,,,,phosphotransferases involved in
interconversion of glucose-1-phosphate and glucose-6-phosphate,
PIP4K2B,,,,,,,,,,,,,,,,,,,,,,,,member of the phosphatidylinositol-5-
phosphate 4-kinase family / a molecular sensor that transduces changes
```

```
in GTP into changes in the levels of the phosphoinositide PtdIns5P
,10.1016/j.tibs.2016.04.003.
PITPNA,,,,,,,,,,,,,,,,,,,,,,,,,,Catalyzes the transfer of PtdIns and
phosphatidylcholine between membranes. / Golgi
transport,10.1016/j.devcel.2018.02.025.
PLXNB2,,,,,,,,,,,,,,,,,,,,,,,,,,transmembrane receptors that
participate in axon guidance,10.1074/jbc.M206005200
PLXNC1,,,,,,,,,,,,,,,,,,,,,,,,,,transmembrane receptors that
participate in axon guidance,10.1074/jbc.M206005200
PORCN,,,,,,,,,,,,,,,,,,,,,,,,,,membrane bound O-acyltransferase (MBOAT)
that is involved in the acylation (palmitoylation) of Wnt proteins /
Post-translational modification/ Vesicle transport of Wnt containing
vesicle from the ER,10.1371/journal.pone.0212711.
PPFIA1,,,,,,,,,,,,,,,,,,,,,,,,,,regulate the disassembly of focal
adhesions/ important for the trafficking of Kif7 and Gli proteins to
the tips of cilia/facilitates exocytosis ,10.1126/scisignal.2005608.
RAD54L2,,,,,,,,,,,,,,,,,,,,,,,,,,DNA helicase /ARIP4 a SNF2 chromatin
remodeling factor (potential),10.1091/mbc.E08-12-1247
RRM2,,,,,,,,,,,,,,,,,,,,,,,,,,Ribonucleotide Reductase Regulatory
Subunit M2 / relates to cell cycle / makes deoxy
nucleotides,10.3892/or.2018.6420
SESN3,,,,,,,,,,,,,,,,,,,,,,,,,,May function as an intracellular leucine
sensor that negatively regulates the TORC1 signaling
pathway,10.1016/j.cmet.2012.08.004.
ST3GAL2,,,,,,,,,,,,,,,,,,,,,,,,,,autophagosome formation/ Atg2-Atg18
complex tethers the PAS to the ER to initiate membrane expansion during
autophagosome formation.,10.1073/pnas.1806727115.
STMN2,,,,,,,,,,,,,,,,,,,,,,,,,,Regulator of microtubule stability /
SCG10 induced APP accumulation in post-Golgi vesicles and on the cell
surface by facilitating its secretory pathway,10.1093/hmg/ddt339.
TEK,,,,,,,,,,,,,,,,,,,,,,,,,,receptor tyrosine kinase,
THSD7A,,,,,,,,,,,,,,,,,,,,,,,,,,Plays a role in actin cytoskeleton
rearrangement.,
TMEM131,,,,,,,,,,,,,,,,,,,,,,,,,,Transmembrane protein / unknown
function,
TMEM63B,,,,,,,,,,,,,,,,,,,,,,,,,,Transmembrane protein / Acts as an
osmosensitive calcium-permeable cation channel.,
TTC17,,,,,,,,,,,,,,,,,,,,,,,,,,role in primary ciliogenesis by
modulating actin polymerization,10.1371/journal.pone.0086476.
TTC7B,,,,,,,,,,,,,,,,,,,,,,,,,,Component of a complex required to
localize phosphatidylinositol 4-kinase (PI4K) to the plasma
membrane,10.1128/MCB.01047-08.
```