# Title: Correlated Gene Modules Uncovered by Single-Cell Transcriptomics with High Detectability and Accuracy

**Authors:** Alec R. Chapman[1,2]*, David F. Lee[1,3]*, Wenting Cai[1,4], Wenping Ma[5,6], Xiang Li[5,6], Wenjie Sun[5,6], X. Sunney Xie[5,6]†

**Affiliations:**

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA.

[2]Current address: Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA

[3]Current address: Color Genomics, Burlingame, CA 94010, USA.

[4]Current address: McKinsey & Company, Boston, MA 02210, USA.

[5]Beijing Advanced Innovation Center for Genomics, Peking University, Beijing, China.

[6]Biomedical Pioneering Innovation Center, Peking University, Beijing, China.

*These authors contributed equally to this work.

†Correspondence to: sunneyxie@pku.edu.cn

**Abstract:**

Single cell transcriptome sequencing has become extremely useful for cell typing. However, such differential expression data has shed little light on regulatory relationships among genes. Here, by examining pairwise correlations between mRNA levels of any two genes under steady-state conditions, we uncovered correlated gene modules (CGMs), clusters of intercorrelated genes that carry out certain biological functions together. We report a novel single-cell RNA-seq method called MALBAC-DT with higher detectability and accuracy, allowing determination of the covariance matrix of the expressed mRNAs for a homogenous cell population. We observed a prevalence of positive correlations between pairs of genes, with higher correlations corresponding to higher likelihoods of protein-protein interactions. Some CGMs, such as the p53 module in a cancer cell line, are cell type specific, while others, such as the protein synthesis CGM, are shared by different cell types. CGMs distinguished direct targets of p53 and exposed different modes of regulation of these genes in different cell types. Our covariance analyses of steady-state fluctuations provides a powerful way to advance our functional understanding of gene-to-gene interactions.

**Main Text:**

Single-cell RNA-seq (scRNA-seq) has greatly expanded our knowledge of gene expression. However, significant advances are still necessary to reach its full potential. Many methods have been developed for single cell amplification (*1-13*), but all suffer from various combinations of poor counting accuracy, low detection sensitivity, or low throughput. While these methods have been successful in cell typing (*5, 6, 14-17*), their ability to shed light on the roles of particular genes has been more limited. To further our understanding of how genes interact to produce

complex cellular behaviors, a technique with high accuracy, sensitivity, and throughput is required. Such an understanding is critical to a wide range of biological problems, for example unraveling the networks of genes controlling cellular differentiation and identifying drug targets, to name only a couple.

To meet these unique technical demands, we designed a novel single-cell mRNA amplification method called Multiple Annealing and Looping Based Amplification Cycles for Digital Transcriptomics (MALBAC-DT) (Figure 1A). Our method improves upon several aspects of our prior work for amplifying DNA and RNA from single cells (*4*, *18*). We improved transcript detection efficiency by optimizing reverse transcription to increase the amount of full length first-strand cDNA produced. First-strand cDNA then is amplified linearly by directly annealing MALBAC random primers along the cDNA strand, followed by exponential amplification by PCR.

Because amplification by MALBAC-DT, in contrast to most single cell amplification methods, does not rely on template switching, we had greater flexibility to choose reverse transcriptases and optimize reaction conditions to maximize first strand cDNA production. Furthermore, with MALBAC-DT it is possible to successfully amplify transcripts that are only partially reverse transcribed, either due to their length or secondary structure. As a result, we detect ~20% more genes from single-cell amounts of RNA compared to Smart-seq2 and obtain a lower percentage of reads corresponding to ERCC synthetic spike-ins which may be shorter or less complex than typical genes (Table S1).

To improve accuracy, we developed a novel unique molecular identifier (UMI) design that can correct previously unrecognized UMI artifacts that occur during amplification and sequencing (Supplementary methods). Although Smart-seq2 does not contain a UMI for absolute quantification of transcripts, we modified the protocol to incorporate the same UMI design to compare with MALBAC-DT and observed approximately twice and many transcripts detected when using MALBAC-DT (Table S2). Finally, our assay incorporates combinatorial cell barcoding to reduce the cost of preparing many single cells. Although we have opted to use UMIs and sequence only the 3' ends of transcripts to improve quantification and reduce costs associated with library preparation and sequencing, we note that it is also possible to perform full length sequencing without UMIs by following standard library preparation protocols.

To demonstrate the ability of our method to generate unique insights into gene function, we amplified and sequenced 768 cells from the U2OS human osteosarcoma cell line, with 738 cells passing quality filters. As expected for a homogenous cell culture, clustering of cells based on gene expression using t-stochastic neighbor embedding (tSNE) (*19*, *20*) did not reveal distinct subpopulations of cells (Figure 1B). However, clustering genes by tSNE did reveal distinct sets of genes that displayed similar patterns of expression across cells (Figure 1C). Clusters of genes that showed similar expression patterns across cells were also revealed by hierarchical clustering (Figure 1D).

To further investigate these clusters of genes, we computed the correlation coefficients for each pair of genes across all cells. Upon hierarchical clustering of the correlation matrix (Figure 2A), we observed 148 correlated gene modules (CGMs), or clusters of 10-200 genes that are highly correlated with each other. Many of these modules consist of genes pertaining to a specific biological function. These include general housekeeping functions—such as cell cycle control and cholesterol (Figure 2B) and protein synthesis (Figure 2C)—as well as functions pertaining

specifically to this cell type such as bone growth and extracellular matrix remodeling (Figure 2D). A full list of CGMs and their associated functional enrichments is provided in Table S3.

It is evident that many CGMs have genes related to specific functions. For example the protein synthesis module (Figure 2B, Table S3 module 45), includes genes responsible for synthesizing tRNAs and amino acids, as well as the machinery required to initiate transcription and translation. The same is true for cholesterol synthesis (Figure 2C, Table S3 module 27). We note gene-to-gene correlation measurements have been widely used, but almost exclusively by means of the perturbative approach (*21-24*), i.e. evaluation of the correlations after introduction of a new experimental condition. This perturbative approach is bound to affect a large number of genes in the cell, usually resulting in large groups of correlated genes. Our method of evaluating correlations of steady state fluctuations of single cells reveals CGMs with a smaller number of intrinsically correlated genes.

While analysis and normalization of cell cycle dependencies is important for cell typing and differential expression analyses (*25*), the CGMs we observed are largely unaffected when expression levels are adjusted to control for cell cycle, with the exception of those CGMs that are directly related to cell cycle activity (Figure S1). Although genes representing different phases of the cell cycle become uncorrelated after normalization, we observe that genes within cell cycle related CGMs remain correlated, as would be expected due to correlations in the stochastic fluctuations that are not removed by normalization.

Detecting CGMs relies on precise measurements of correlations between genes. This requires large numbers of cells, accurate quantification of transcripts, and high detectability. To our knowledge, this is the first study that satisfies all of these criteria. Low cell numbers add sampling noise to the correlation coefficients (Figure 2E), while low detection sensitivity attenuates the correlations (Figure 2F). Simulations demonstrate that neither high cell numbers nor high detection sensitivity alone is sufficient (Figure S2-3), and that the CGMs detected were not the result of spurious correlations due to limited sample size (Figure S4). Previously published data (*8*) generated by 10x Genomics, which has high throughput but low sensitivity is unable to reveal CGMs (Figure S5), further confirming that large cell numbers cannot compensate for low sensitivity. A dataset more recently made available by 10x Genomics is able to detect a small number of modules (40 vs 178 CGMs identified by MALBAC-DT using the same cell line). Of these 40 modules, 53% (21/40) are found to be significantly correlated by MALBAC-DT ($q < .05$, Supplementary methods), while of the 178 CGMs identified by MALBAC-DT, only 9.6% (17/178) are able to be detected as significantly correlated in the 10x Genomics data.

Weighted Gene Correlation Network Analysis (WGCNA) has long been used to identify networks of genes using differential expression data across many cell types and conditions (*26, 27*). While this method of inference is logically distinct from our approach of using correlations within a uniform population, we found that the tools developed for WGCNA could also identify modules in our data. Adding further weight to the biological significance of the CGMs, the modules identified by WGCNA were highly similar to the CGMs we identified (Figure S6, Supplemental methods), although only 19 modules were identified by WGCNA. Modules such as tRNA aminoacylation, mitochondria, translation, and cell cycle are consistently found using both methods. However, modules like glycerolipid metabolism can only be identified with our analyses of the MALBAC-DT data.

Although it is widely understood that related genes will have similar differential expression levels across cell types or in response to perturbations, less consideration has been given to using steady state fluctuations to identify related genes. In general, the transcript levels of two genes under steady state conditions may be correlated if their transcription or degradation rates are correlated. This can arise from a number of biological mechanisms. Overall anabolic and catabolic activity of the cell will affect most genes indiscriminately but is largely removed by normalization. One gene-specific mechanism resulting in correlated transcription rates is coregulation either by a common transcription factor or multiple transcription factors which are themselves coregulated, possibly post-transcriptionally. Other possibilities include correlated changes in epigenetic states such as DNA methylation, histone modifications, or spatial position within the nucleus. Gene-specific mechanisms resulting in correlated degradation rates include common regulatory features in the untranslated regions (UTRs) or regulation by correlated miRNAs.

In light of these mechanisms that potentially result in gene correlations, we expect that genes that are involved in a common function, and hence are coregulated, would exhibit correlations such as those in Figure 2B-D. We examined whether a simple model of coregulation of gene expression is sufficient to produce the magnitude of correlations we observe in the CGMs. Due to the small number of DNA molecules present in a single cell, transcripts and proteins undergo stochastic fluctuations in expression level over time. When a regulatory protein controls the expression of multiple genes, it is natural to expect that fluctuations in the regulator will flow through to its targets, causing the targets to fluctuate in sync with one another. When their expression is measured across many cells, these genes would then be observed to be correlated (Figure 2H). Indeed, we find that conservative assumptions about the regulatory mechanisms underlying transcription and degradation are sufficient to produce correlations of the magnitude that we observe (Supplemental text, Figure S7). In our data, we observe significantly more positively correlated than negatively correlated pairs of genes, indicating that coregulation is a more widespread mechanism of regulation (Figure S8).

More generally, we observed that highly correlated genes were more likely to have been previously identified as being related in databases (*28*) of protein associations, including direct protein-protein interactions (PPI) and inferred functional relations (Figure 2G). Consistent with the observation of function-specific CGMs, this indicates that steady state correlations can indicate functional relationships between genes, and raises the possibility of identifying mechanistically related sets of genes from such measurements.

One CGM we chose for further investigation consists of targets of the key tumor suppressor protein p53 (Figure 3A). This CGM contains 197 genes, most of which have been identified by ChIP-seq studies to contain p53 binding sites. Moreover, all p53 targets identified by a previous ChIP-chip study (*29*) of this cell line are contained in this module.

We performed an shRNA knockdown of p53 followed by MALBAC-DT in order to examine the effect of this module as well as the transcriptome as a whole. Mean p53 transcript levels decreased 15-fold in the knockdown cells compared to the control cells (Figure 3B), and 1337 genes were significantly up- or down-regulated as a result (Figure 3C).

If correlations between genes in our homogenous population of cells reflect coregulatory relationships, then we should expect that these genes would respond similarly to perturbations. Indeed, we find that genes that are strongly correlated in our original dataset tend to be either

both upregulated or both downregulated in response to p53 perturbation, while anticorrelated genes tend to have opposite responses to the perturbation (Figure 3D).

Because correlations reflect regulatory relationships, they have the potential to identify novel genes that act in the same regulatory pathway. Indeed, several genes exhibited a high correlation with p53 expression and were perturbed by p53 knockdown, despite not previously being connected to p53 to our knowledge. As a concrete example, we observed that the deubiquitinase JOSD1 is strongly anticorrelated with p53 activity. Although JOSD1 had not been previously associated with p53, other deubiquitinases are known to modulate p53 activity either directly or through Mdm2. Moreover, a structurally related protein ATXN3 was recently shown to stabilize p53 via deubiquitination (*30*). We therefore hypothesized that JOSD1 might play a role in the p53 pathway. Indeed, JOSD1 was observed to be upregulated upon p53 inactivation, consistent with their anticorrelation in the unperturbed system, and indicating a possible negative feedback loop in which p53 inhibits JOSD1 transcription, while the Josd1 protein stabilizes p53.

The large number of genes that are differentially expressed in response to a perturbation often hinders meaningful analysis of such data, and providing meaningful classifications of these genes for further experiments remains an open challenge. CGMs potentially offer a way to organize these differentially expressed genes into fine-grained modular units. To this end, we looked at the correlations in our original dataset among the 1337 genes which were differentially expressed in response to p53 knockdown (Figure 3E).

These differentially expressed genes clustered into ~10 CGMs, several of which are associated with distinct pathways. One of these CGMs consists almost entirely of genes from the original p53 CGM. Moreover, nearly all of the differentially expressed genes with p53 binding sites are contained in this CGM, indicating that by analyzing correlations we are able to distinguish the direct targets of p53 knockdown from its downstream effects. Strikingly, correlations in our steady-state dataset were able to predict the genes that would be perturbed by p53 knockdown more accurately than ChIP-seq studies. Whereas 49% of genes in the p53 module were downregulated upon knockdown of p53, this was only the case for 33% of genes from a consensus of ChIP-seq studies (*31*), and 9% of genes identified by a ChIP-chip study of the same cell line (*29*). Additionally, genes within CGMs are consistently upregulated or consistently downregulated upon p53 knockdown, in agreement with our model in which correlations among genes arise from coregulation.

Finally, we asked how CGMs compare across cell types. We amplified and sequenced 748 single cells from the HEK293T human embryonic kidney cell line. As expected for dramatically different cell types, several thousand genes were differentially expressed between these two cell lines (Figure 4A-B).

While existing methods are unable to provide meaningful classification of these genes, we find that CGMs provide a natural way to understand the differences between these two cell types. Of the 148 CGMs identified in U2OS, 22 CGMs were also observed to be significantly correlated in HEK293T cells, representing housekeeping machinery shared across vastly different cell types. For many CGMs, the genes are up- or down- regulated as a group (Figure 4C-D). In some of these cases, the CGM is absent in one cell type (Figure 4C), indicating that the module has been switched off. In other cases (Figure 4D), the CGM is present in both cell types, indicating that the genes remain coregulated but at different expression levels.

CGMs can also identify changes in regulatory architecture across cell types even in the absence of differential expression. Although HEK293T is known to lack p53 activity (*32*, *33*), target genes of p53 are not consistently down-regulated in HEK293T compared to U2OS (Figure 4E), possibly indicating their roles in other pathways. Despite the similar expression levels of the component genes, the p53 CGM is absent in HEK293T.

Although in this work we have presented proof of principle analyses on cell lines using CGMs, our results shed light on critical biological processes relevant to many cell types. We expect the analysis of CGMs in a diverse set of cell types and tissues using methods with high sensitivity will produce a wealth of insights not obtainable using differential expression analyses alone.

**References and Notes:**

1. J. Phillips, J. H. Eberwine, Antisense RNA Amplification: A Linear Amplification Method for Analyzing the mRNA Population from Single Living Cells. *Methods* **10**, 283-288 (1996).
2. F. Tang *et al.*, mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Meth* **6**, 377-382 (2009).
3. S. Picelli *et al.*, Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Meth* **10**, 1096-1098 (2013).
4. A. R. Chapman *et al.*, Single Cell Transcriptome Amplification with MALBAC. *PLOS ONE* **10**, e0120889 (2015).
5. Allon M. Klein *et al.*, Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. *Cell* **161**, 1187-1201.
6. Evan Z. Macosko *et al.*, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
7. T. Hashimshony *et al.*, CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biology* **17**, 77 (2016).
8. G. X. Y. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**, 14049 (2017).
9. D. Lovatt *et al.*, Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nature methods* **11**, 190-196 (2014).
10. S. Islam *et al.*, Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research* **21**, 1160-1167 (2011).
11. S. Islam *et al.*, Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. *Nature Protocols* **7**, 813 (2012).
12. S. Islam *et al.*, Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Meth* **11**, 163-166 (2014).
13. M. Hagemann-Jensen *et al.*, Single-cell RNA counting at allele- and isoform-resolution using Smart-seq3. *bioRxiv*, 817924 (2019).
14. A. K. Shalek *et al.*, Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236 (2013).
15. D. Grün *et al.*, Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251 (2015).
16. A. Zeisel *et al.*, Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138 (2015).
17. D. Usoskin *et al.*, Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nature Neuroscience* **18**, 145 (2014).

18. C. Zong, S. Lu, A. R. Chapman, X. S. Xie, Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell. *Science* **338**, 1622 (2012).
19. L. Van Der Maaten, G. Hinton, Visualizing high-dimensional data using t-sne. journal of machine learning research. *J Mach Learn Res* **9**, 26 (2008).
20. L. Van Der Maaten, Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* **15**, 3221-3245 (2014).
21. B. Adamson *et al.*, A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882.e1821 (2016).
22. A. Dixit *et al.*, Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866.e1817 (2016).
23. D. A. Jaitin *et al.*, Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883-1896.e1815 (2016).
24. P. Datlinger *et al.*, Pooled CRISPR screening with single-cell transcriptome readout. *Nat Meth* **14**, 297-301 (2017).
25. A. Scialdone *et al.*, Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54-61 (2015).
26. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
27. B. Zhang, S. Horvath, A General Framework for Weighted Gene Co-expression Network Analysis. *Stat Appl Genet Mol Biol* **4**, (2005).
28. D. Szklarczyk *et al.*, STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-d613 (2019).
29. L. Smeenk *et al.*, Characterization of genome-wide p53-binding sites upon stress response. *Nucleic Acids Res* **36**, 3639-3654 (2008).
30. R. Gao *et al.*, Inactivation of PNKP by Mutant ATXN3 Triggers Apoptosis by Activating the DNA Damage-Response Pathway in SCA3. *PLOS Genetics* **11**, e1004834 (2015).
31. M. Fischer, P. Grossmann, M. Padi, J. A. DeCaprio, Integration of TP53, DREAM, MMB-FOXM1 and RB-E2F target gene analyses identifies cell cycle gene regulatory networks. *Nucleic Acids Research* **44**, 6070-6086 (2016).
32. D. Ahuja, M. T. Sáenz-Robles, J. M. Pipas, SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. *Oncogene* **24**, 7729 (2005).
33. W. T. Steegenga, A. Shvarts, N. Riteco, J. L. Bos, A. G. Jochemsen, Distinct regulation of p53 and p73 activity by adenovirus E1A, E1B, and E4orf6 proteins. *Molecular and cellular biology* **19**, 3885-3894 (1999).
34. A. Santos, L. J. Jensen, R. Wernersson, Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Research* **43**, D1140-D1144 (2014).
35. E. Y. Chen *et al.*, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
36. M. V. Kuleshov *et al.*, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016).

**Acknowledgments:** We thank L. Tan for helpful discussions of analysis methods and S. Mulepati for assistance with knockdown experiments. **Funding:** This work was supported by the Beijing Advanced Innovation Center for Genomics at Peking University, an NIH Director's

Pioneer Award (DP1 CA186693), and two grants from the National Science Foundation of China (21390412 and 21327808) (X.S.X.). **Author contributions:** A.R.C., D.L., and X.S.X. designed the experiments. A.R.C., D.L., and W.M. performed the experiments. A.R.C. and W.C. analyzed the data. A.R.C., D.L., and X.S.X. wrote the manuscript. **Competing interests:** A.R.C., D.L., and X.S.X. are inventors on the patent PCT/US18/34689 filed by President and Fellows of Harvard College. **Data and materials availability:** All sequencing data will be deposited in the NCBI SRA prior to publication.

**Supplementary Materials:**

Materials and Methods

Figures S1-S8

Tables S1-S3

**Fig. 1.** (A) MALBAC-dt protocol and experimental workflow. A homogenous cell population is trypsinized and sorted into individual wells of 96-well plates by flow cytometry. Reverse transcription is carried out using a poly-T primer containing a cell-specific barcode and unique molecular identifier (UMI). First strand cDNA is amplified by random primers using MALBAC thermocycling to ensure linear amplification followed by additional cycles of exponential amplification by PCR. After amplification, samples are pooled together for library preparation and sequencing. (B) Clustering of 738 U2OS cells by t-Stochastic Neighbor Embedding (tSNE). Cells were obtained from a homogenous culture and cDNA was amplified by MALBAC-dt. Consistent with a homogenous culture, no sub-clusters of cells are apparent. (C) Clustering of genes by tSNE. Genes were clustered based on their expression levels across the 738 U2OS cells. Many clusters are present, representing sets of genes that have similar expression patterns. (D) Hierarchical clustering of gene expression data from 738 U2OS cells. As with clustering by tSNE, several sets of genes with similar expression patterns are observed.

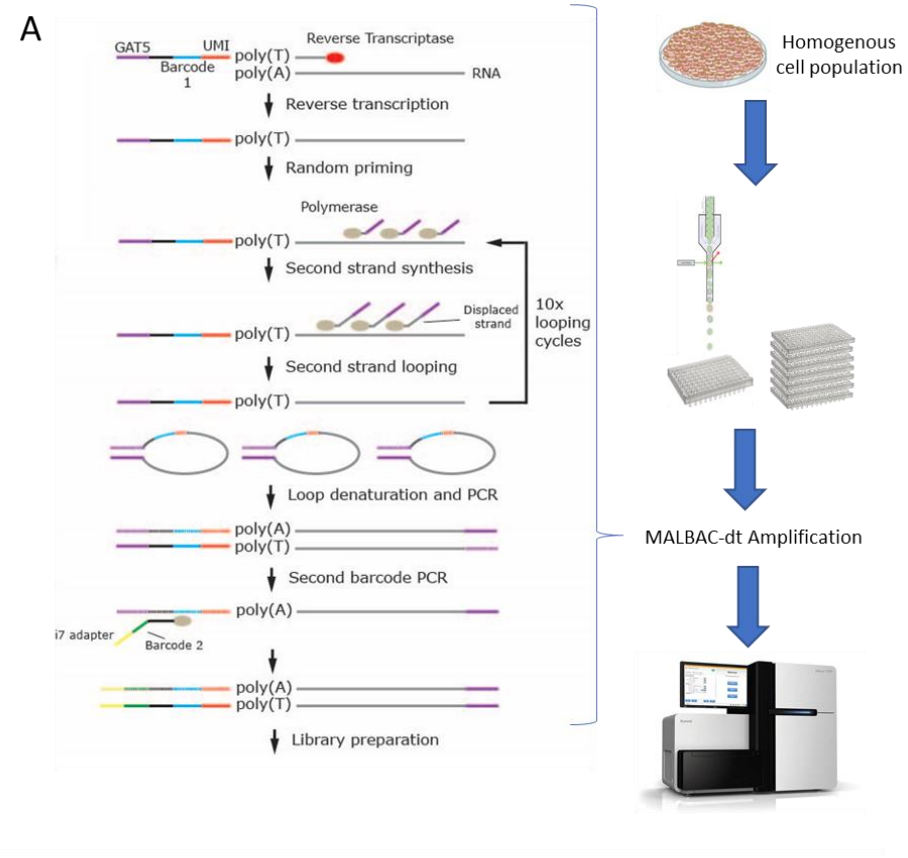
**Fig. 2.** (A) Correlation matrix for ~11,000 genes that were detected in at least 10% of the 738 U2OS cells. Genes are ordered by hierarchical clustering to reveal numerous modules of highly correlated and/or anti-correlated genes. Inset provides an enlarged view to highlight the detail present in many of these clusters. The genes in many of these Correlated Transcriptional Modules (CGMs) are enriched for particular biological function or contain binding sites for specific transcription factors. (B) A CGM related to protein synthesis, with genes responsible for specific processes indicated by arrows. (C and D) CGMs related to sterol synthesis (C) and extracellular matrix remodeling (D). Genes with known functional roles in these processes are labeled in red. (E) Measurement error associated with estimating correlation from a limited number of cells. The Plotted are the distributions of correlations that would be measured for a pair of uncorrelated genes if a given number of cells were sampled. (F) Impact of detection efficiency on correlation measurements. For a pair of genes with a given true correlation, the correlation that would be measured by sampling an unlimited number of cells is plotted as a function of the efficiency of detecting individual transcripts. (G) Genes with high correlations are more likely to be identified as related in protein-protein interaction databases. Gene pairs are binned based on their correlation coefficient. For each bin, the fraction of pairs identified by StringDb is plotted. (H) A schematic model depicting how gene regulatory interactions can result in correlations in a steady state population. Stochastic fluctuations in a transcription factor will result in fluctuations in its target genes, causing them to be correlated. Independently regulated genes, on the other hand, will exhibit no correlation.

**Fig. 3.** (A) CGM related to p53 activity. Genes with significant literature support (*31*) for being targets of p53 are indicated by red arrows, while genes with limited literature support are indicated by black arrows. (B) Distribution of p53 transcript levels in control and shRNA knockdown cells. (C) Mean expression levels of transcripts in p53 knockdown cells vs. control cells. Red points indicate genes which are significantly differentially expressed. (D) Correlation among genes in a homogenous cell population is predictive of their response to perturbation. Genes differentially

expressed in response to p53 knockdown are categorized as strongly anticorrelated (less than -0.2), moderately anticorrelated (between -0.2 and -0.1), weakly anticorrelated (between -0.1 and 0), weakly correlated (between 0 and 0.1), moderately correlated (between 0.1 and 0.2), and strongly correlated (greater than 0.2). For each category, the fraction of gene pairs which are concordantly regulated (both upregulated or both downregulated) or discordantly regulated (one upregulated while the other downregulated) are shown. (E) Hierarchical clustering of genes differentially expressed upon p53 knockdown. Genes cluster into ~10 CGMs. Genes within a CGM have the same directional response to p53 knockdown, consistent with their regulation as a functional unit. Direct p53 targets, indicated by red and black arrows as in (A), are predominantly found in a single CGM, as are the genes originally identified as a CGM related to p53 function. CGMs thus distinguish direct p53 targets from downstream pathways.

**Fig 4.** Shared and cell-type specific CGMs between the U-2 OS and HEK293T cell lines. (A) Mean expression levels of genes in HEK293T vs U2OS. (B) Hierarchical clustering of expression across HEK293T and U2OS cells reveals ~6000 up- and down-regulated genes. (C-F) CGMs organize genes into biologically relevant pathways in a manner that is distinct from and not apparent by differential clustering. In some cases (C) differentially expressed genes correspond to differentially correlated modules. However, CGMs provide a further separation of differentially expressed genes into distinct functional units. In other cases (D), a set of differentially expressed genes can be resolved into a common CGM between multiple cell types. Moreover, differential correlation between cell types can occur without differential expression (E), possibly indicating multiple modes of regulation of the component genes. Finally, correlations can be consistently observed across cell types and organized into distinct CGMs in the absence of differential expression (F).

Figure 1



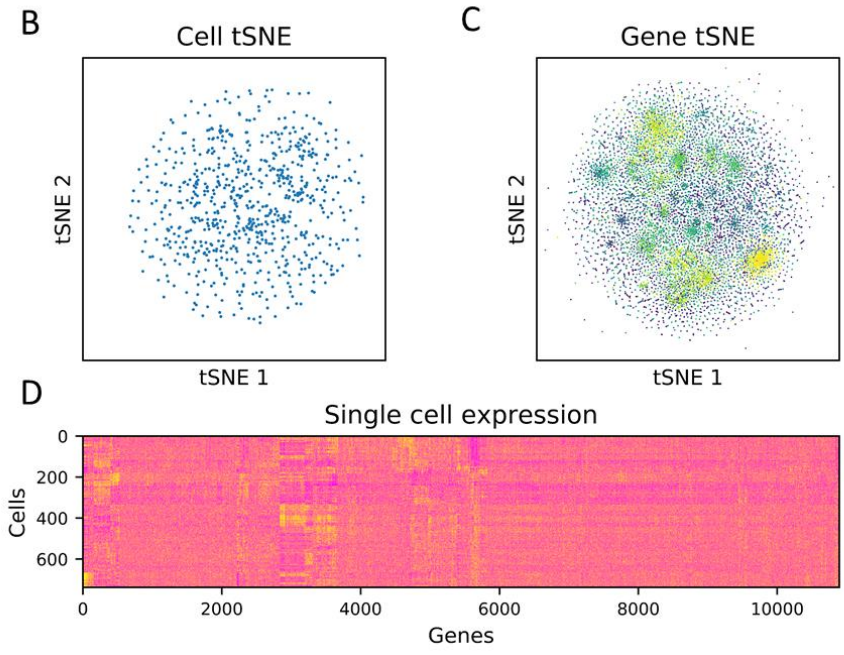



Figure 2

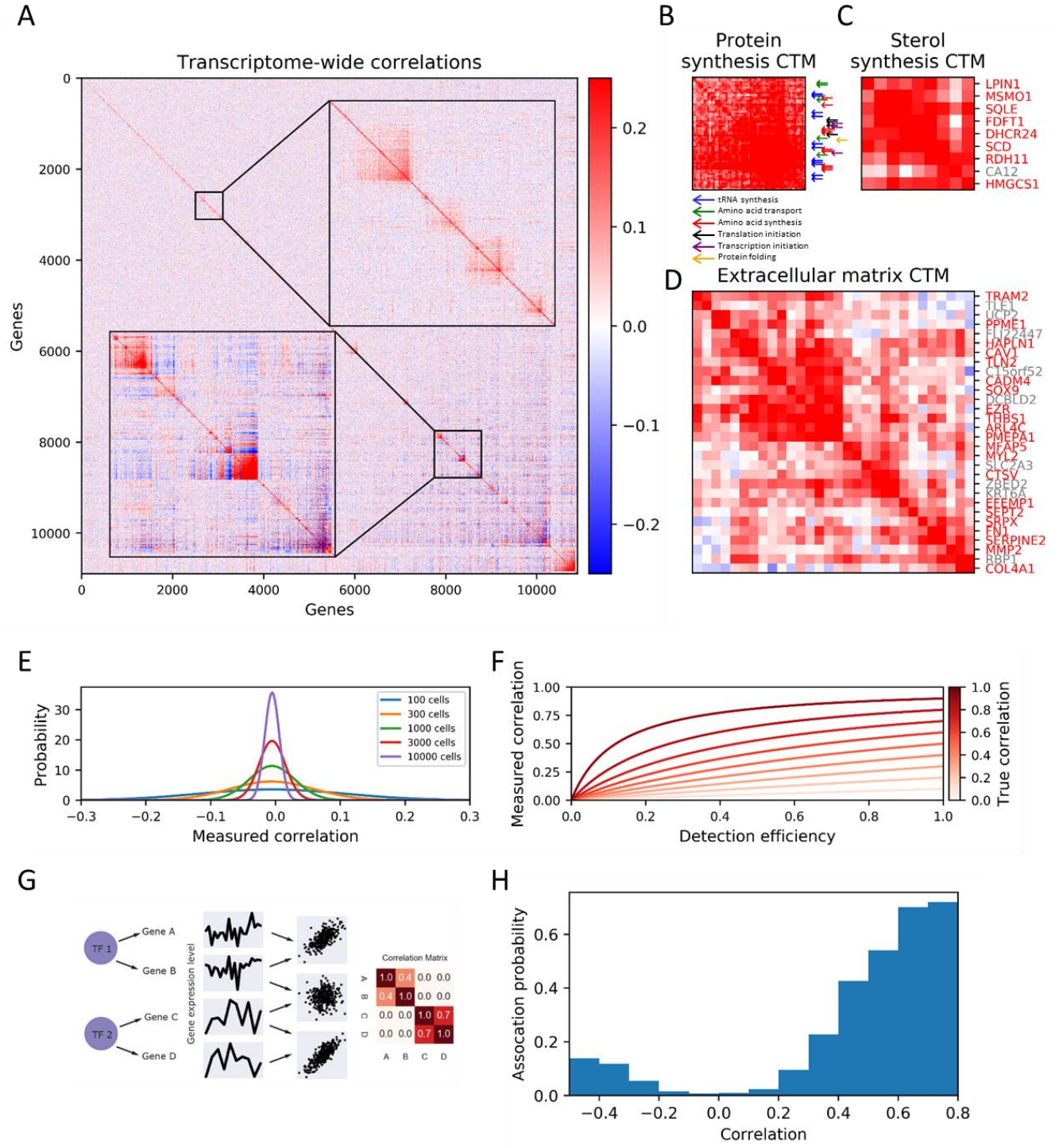


Figure 3

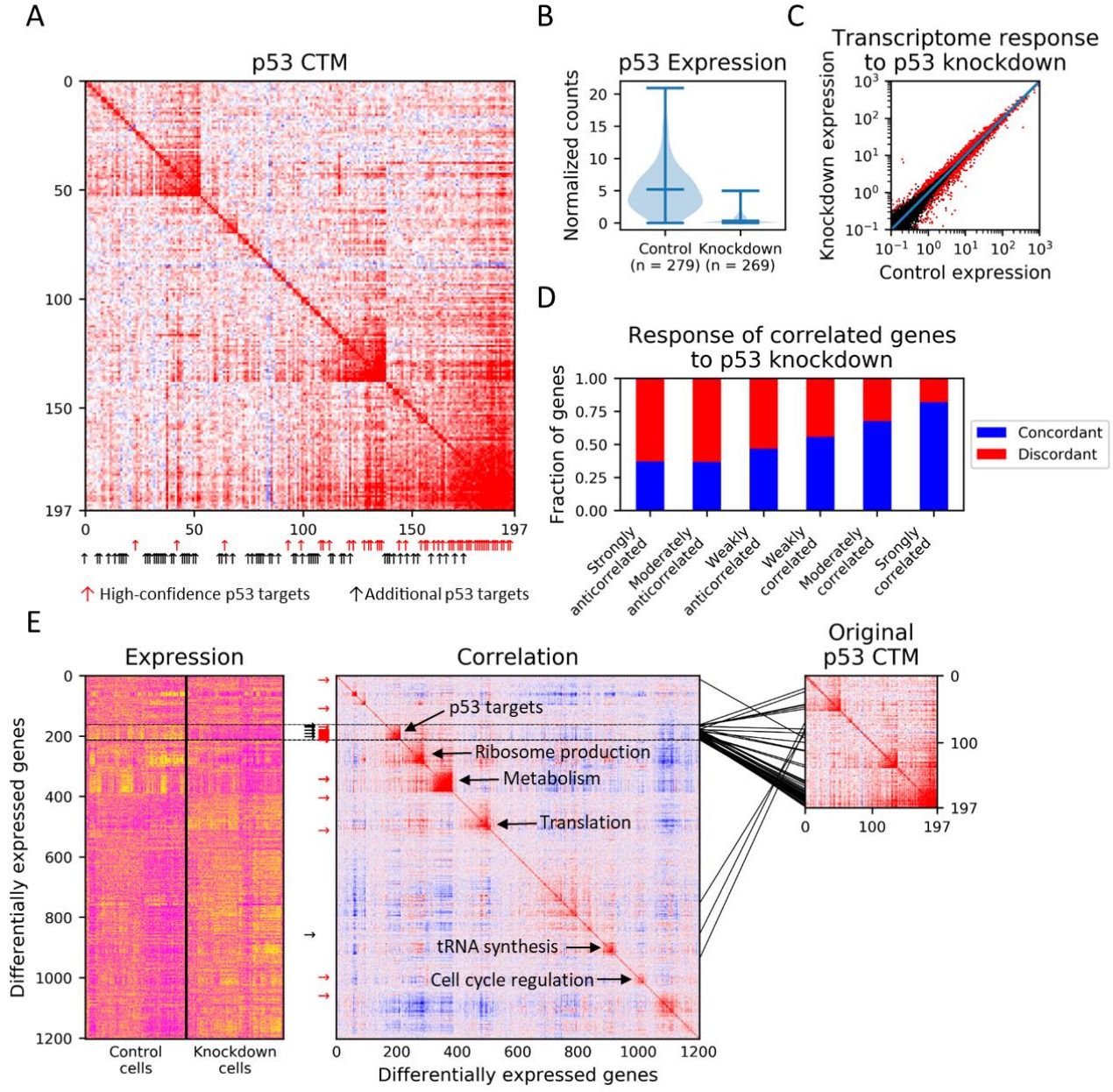
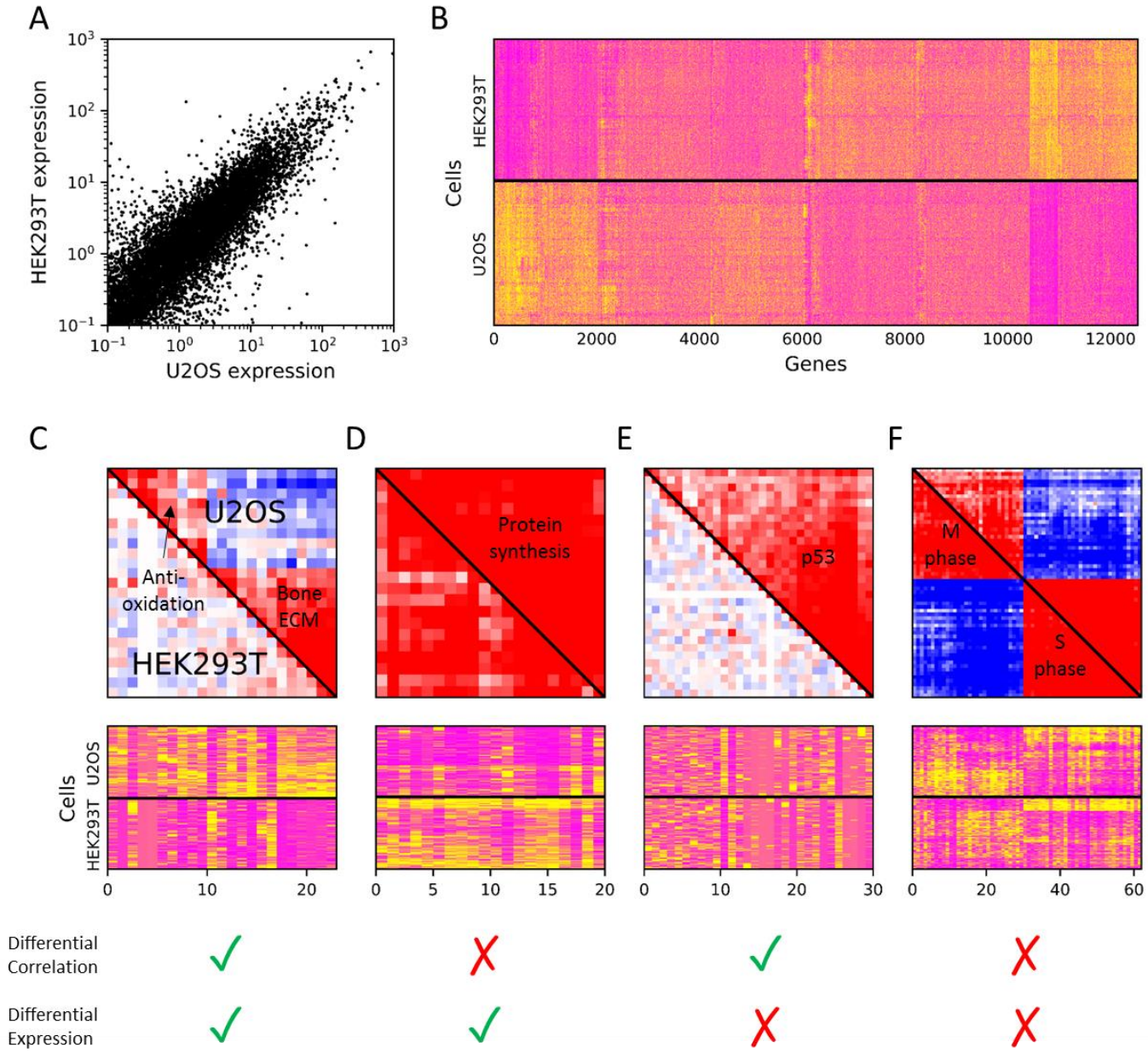


Figure 4



# Supplementary Materials

## Cell culture and handling

U2OS and HEK293T cell lines were obtained from ATCC and cultured at 37°C in RPMI-1640 medium with 10% Fetal Bovine Serum and 1% Penicillin-Streptomycin. To form single cell suspensions for flow sorting, culture medium was removed, cultures were rinsed with Dulbecco's phosphate-buffered saline (D-PBS), and incubated with 1mL of 0.25% trypsin for 5 minutes. Detached cells in D-PBS were pelleted by centrifugation at 300g for 5 minutes and resuspended in D-PBS. Single cell suspensions were kept on ice until flow sorting.

shRNA knockdowns were performed by incubating ~30% confluent U2OS cells with 1ug of either 11653 C3 plasmid for TP53 knockdown cells or with TransIT-LT1 plasmid for control cells. Cells were incubated for 48 hours followed by flow sorting to isolate single cells in lysis buffer.

## MALBAC-DT protocol

Cells were flow sorted into 3uL of lysis buffer consisting of 1uL H2O, 0.6uL 5x SSIV buffer, 0.15uL 10% ICA-630, 0.8uL 5M betaine, 0.05uL SUPERase In, 0.2uL 50uM RT-An primer, and 0.2uL 10mM dNTP mix. Plates are stored at -80°C until ready for amplification. Plates are kept on ice while pipetting and vortexed and briefly centrifuged after all pipetting steps.

To perform reverse transcription, plates are incubated at 72°C for 3 minutes, then 1uL of RT mix is added consisting of 0.264uL H2O, 0.16uL 5x SSIV buffer, 0.2uL 100mM DTT, 0.152uL SUPERase In, 0.024uL 1M MgSO4, and 0.2uL SuperScript IV. Plates are incubated for 10 minutes at 55°C.

Next, excess reverse transcription primers are degraded by exonuclease digestion. 1uL of exonuclease mix is added consisting of 0.1uL ExoI buffer, 0.1uL H2O, 0.6uL ExoI, and 0.2uL 50uM RT-Bn primer. Plates are incubated for 30 minutes at 37°C and then 20 minutes at 80°C.

Amplification is performed by adding 24uL of amplification mix consisting of 18.64uL H2O, 3uL ThermoPol buffer, 0.4uL 10mM dNTP mix, 0.16uL 100mM MgSO4, 0.4uL 50uM GAT-7N, 0.4uL 50uM GAT-COM, and 1uL Deep Vent (exo-). The following thermocycle program is run:

| Step | Temperature | Time |
|---|---|---|
| 1 | 95 | 5:00 |
| 2 | 4 | 0:50 |
| 3 | 10 | 0:50 |
| 4 | 20 | 0:50 |
| 5 | 30 | 0:50 |
| 6 | 40 | 0:45 |
| 7 | 50 | 0:45 |
| 8 | 65 | 4:00 |

| | | |
|---|---|---|
| 9 | 95 | 0:20 |
| 10 | 58 | 0:20 |
| 11 | Goto 2 | 10x |
| 12 | 95 | 1:00 |
| 13 | 95 | 0:20 |
| 14 | 58 | 0:30 |
| 15 | 72 | 3:00 |
| 16 | Goto 13 | 17x |
| 17 | 72 | 5:00 |
| 18 | 4 | 0:00 |

Finally, amplification is completed by adding 0.4uL 50uM Tru2-Gn-RT primers and running an additional 5 cycles of PCR steps 12-15. Amplified plates are stored at -20°C until library preparation.

In early versions of the protocol, a total of 8 RT3-An primers were used per 96 well plate, with one primer corresponding to one row of the plate. During the final amplification step, 12 Tru2-Gn-RT primers were used, with one primer corresponding to one column of the plate. In later versions of the protocol, 96 RT3-An primers were used, with a distinct primer corresponding to each well. In the final step, a single Tru2-Gn-RT primer could then be used. While the former method requires lower upfront costs to synthesize primers, the later method simplifies preparing plates at larger scales and eliminates the possibility of cross-contamination of samples during amplification.

To prepare libraries for sequencing, 1uL from each of the wells are combined and purified using 0.8x Ampure beads. The Nextera library preparation kit is used to add Illumina adapters by tagmentation. During subsequent PCR steps, Ix-Tru2 primers are substituted for Nextera S5XX primers in order to select the 3' ends of transcripts containing cell barcodes and UMIs.

# Sequence processing

Separate fastq files are generated for each cell based on the outer and inner barcode sequences. Barcodes not matching a cell exactly are discarded. Barcodes, adapter sequences, and UMIs are stripped from the reads, and reads are aligned to the human GRCh38.p7 reference using STAR 2.5.2. For each gene, a list of UMIs is obtained for all reads mapping to that gene, excluding regions masked by RepeatMasker. To remove extraneous UMIs resulting from amplification or sequencing errors, UMIs for a particular gene are represented as nodes in a graph, with connections between UMIs differing at no more than 7 bases. Connected components are identified, and the consensus sequence within each component is determined. Consensus sequences matching the $(HBDV)_5$ RT-An pattern and differing from the $(VDBH)_5$ RT-Bn pattern at at least three bases are retained. To avoid potential cross-talk between wells, UMIs observed for the same gene in multiple cells are discarded.

After obtaining UMI counts for all genes and cells, cells for which more than 1% of transcripts are from ERCC spike-ins or contain fewer than 1000 total transcripts are discarded, as are genes

which are observed in fewer than 10% of cells. Counts are normalized relative to the total number of transcripts in each cell prior to computing the correlation matrix.

Hierarchical clustering is performed using the SciPy function scipy.cluster.hierarchy.linkage using method "average," and with a distance metric of $1 - abs(\rho_{ij})$, where $\rho_{ij}$ is the correlation between genes $i$ and $j$. To test the robustness of this clustering algorithm, we randomized the umi counts across all cells for each particular gene and recomputed the correlations between gene pairs, resulting in a distribution of correlation coefficients that would be expected due to limited sample size alone. On top of this background of uncorrelated genes, we set a group of genes to have a stronger correlation and examined whether these genes could then be identified by clustering. We found that for the magnitudes of correlations typically observed in our data, groups of correlated genes could be reliably recovered (Figure S9).

# Cell cycle correction

Pseudo-time inference and cell-cycle correction

Pseudo-time was inferred for each cell by assuming that the expression of cell-cycle genes followed a sinusoidal function along the time trajectory. The actual expression of each cell-cycle gene was further modeled as follows, a normal distribution centered around the level predicted by sinusoidal function, with variance aggregated from both stochastic expression variance and technical noise.

$$y_{g,c} \sim \mathcal{N}(\mu_{g,c},\ v_g^2 + v_{tech}^2)$$

$$\mu_{g,c} = Amp_g\ * \left(\cos\left(\frac{20}{t_c} - T_{peak,g}\right) + 1\right) + AmpShift_g$$

$y_{g,c}$: actual expression of gene g for cell c.

$\mu_{g,c}$: expected expression of g for c from sinusoidal function.

$v_g^2$: gene specific variance from stochastic expression for g.

$v_{tech}^2$: common technical noise.

$Amp_g, AmpShift_g$: amplitude of the sinusoidal function for g.

$T_{peak,g}$: The peak time of g, in the time scale of percentage into the cell-cycle. Retrieved from Cyclebase.org (*34*).

$t_c$: The pseudo-time of cell c.

The transcriptome was fitted against the described model, with a pseudo-time optimized for each cell to maximize the overall likelihood estimation. The MLE process was done using PyTorch.

In order to correct the covariance matrix for cell-cycle effect, cells were then ordered by the assigned pseudo-time, and the expression of each gene was corrected by subtracting the mean of the surrounding rolling window.

# Correlations between coregulated genes

We consider the case of two genes that are regulated by the same transcription factor (Supplementary Figure 1A) with dynamics described by:

$$\frac{\partial[C]}{\partial t} = v_c[B] - d_c[C]$$
$$\frac{\partial[D]}{\partial t} = v_c[D] - d_c[D]$$
$$[B] = f(t)$$

For simplicity we have taken the transcription rate $u$ and degradation rate $v$ to be the same for transcripts C and D. The transcription factor B is allowed to fluctuate in time arbitrarily.

The lifetimes of mRNAs are often short relative to those of proteins. In this case, as B fluctuates, transcripts C and D rapidly adjust and fluctuate independently from one another around the steady state concentration $[C]_{ss} = [D]_{ss} = [B]\, v_c/d_c$, and these fluctuations will follow a Poisson distribution.

Under these assumptions, the covariance between $[C]$ and $[D]$ is:

$$\begin{aligned}\langle \delta c \delta d \rangle &= \sum_b \langle \delta c \delta d | b \rangle p(b) \\ &= \sum_b \langle \delta c | b \rangle \langle \delta d | b \rangle p(b) \\ &= \sum_b \frac{v_c^2}{d_c^2} \delta b^2 p(b) \\ &= \frac{v_c^2}{d_c^2} \sigma_B^2\end{aligned}$$

Following a similar procedure to obtain $\langle \delta c^2 \rangle$ and $\langle \delta d^2 \rangle$, we obtain the correlation coefficient

$$\rho_{CD} = \frac{\langle \delta c \delta d \rangle}{\sqrt{\langle \delta c^2 \rangle \langle \delta d^2 \rangle}} = \frac{\mu_C \mathrm{cv}_B^2}{1 + \mu_C \mathrm{cv}_B^2}$$

where $\mu_C$ is the mean of C, and $\mathrm{cv}_B$ is the coefficient of variation of B.

## Analysis of 10x Genomics datasets

Datasets were downloaded from the website of 10x Genomics. The datasets from Zheng et al (8) consisting of ~2800 HETK293T cells and the newer dataset consisting of a mixture of ~10,000 HEK293T and mouse NIH3T3 cells were used. For the later dataset, only HEK293T cells were used for further analysis. BAM files were downloaded and filtered according to the same mappability criteria used for MALBAC-DT datasets. To determine whether a CGM detected with one method was significantly correlated in the other, genes within a given module were randomly substituted for genes with similar expression levels (among the 50 genes with nearest mean expression), and the average of the absolute value of the correlations in this randomized module were calculated. This randomization was repeated 10,000 times and a Bonferoni-corrected p-value was obtained by comparing the average correlation in the true module to the distribution of average correlations in the randomized modules.

## Comparison to WGCNA

Hierarchical clustering of the correlation matrix is performed using the "hclust" function of R software, with the "average" method and a distance metric of 1-abs(ρ_ij), where ρ_ij is the correlation between genes i and j. Sub-clusters are obtained by cutting the dendrogram using the "cutree" function with parameter h=0.9. Sub-clusters containing more than 10 genes are identified as CGMs. We also compare our module detection method with a widely used gene co-expression analysis package, WGCNA1,2. Modules are identified by the "blockwiseModules" function, with "power = 1, TOMType = "unsigned", minModuleSize = 10" and the other default parameters. Gene set enrichment analysis is performed using the R package "enrichR" with p-value threshold of 1e-5 to associate gene sets in modules with "KEGG_2019_Human", "GO_Biological_Process_2018", "GO_Cellular_Component_2018", "GO_Molecular_Function_2018", and "Reactome_2016" databases. For the U2OS cell line, only 19 modules are identified by WGCNA, and nearly 9,000 genes do not form any module. We use the Dice coefficient as a measure of similarity between modules detected by both methods and found that CGMs and WGCNA modules are highly similar.

## Supplementary Figures

Figure S1. Effect of cell cycle on CGMs. Correlations are shown before (upper right) and after (lower left) adjusting expression data for cell cycle differences across cells. Nearly all CGMs are unaffected by the correction, with individual pairs of genes exhibiting similar correlations before and after adjustment. Genes specifically related to the cell cycle are the exception. Related cell cycle genes generally exhibit weaker correlation after adjustment, and the negative correlations between M and S phase genes are largely eliminated.

Figure S2. Large numbers of cells are needed to observe CGMs. Correlations in the full U2OS dataset (upper right) are compared with correlations calculated using a random subset of 100 cells (lower left). The 100 cell subset is significantly noisier than the full dataset. In many cases, CGMs identified in the full dataset (A) are observed to also be correlated in the 100 cell subset. However, the larger noise in this dataset prevents CGMs from being identified using this reduced dataset alone. When genes are clustered according to their correlation in the 100 cell dataset (B), many spurious clusters are identified. The correlations within such clusters result

solely from measurement error and are not observed in the larger dataset, which has lower measurement error due to the larger number of cells.

Figure S3. High detection efficiency is necessary to detect CGMs. Correlations between genes are shown for the full U2OS dataset (upper right), and a randomly downsampled dataset simulating a 67% lower detection efficiency (lower left). Many correlations are absent or severely attenuated at lower detection efficiency.

Figure S4. CGMs do not result from measurement error. To examine the effect of measurement error in the correlation coefficients on clustering, we randomly permuted the expression counts across all cells for each gene. As a result, there is no expected correlation between genes, and any observed correlation is due to sampling a finite number of cells. In our full U2OS dataset (A), no CGMs are observed after random permutations. Spurious clusters are observed when the permutation is applied to a random subset of 100 cells (B).

Figure S5. Methods with low sensitivity cannot detect CGMs. Correlations are displayed for data generated by MALBAC-DT (upper right) and 10x Genomics (lower right) for the HEK293T cell line. Although 2800 cells were sequenced with 10x compared to 748 cells with MALBAC-DT, correlations and CGMs are only apparent in the MALBAC-DT data.

Figure S6. CGMs are consistent with the modules detected using WGCNA. A) Left panel: hierarchical clustering of gene-gene correlation across U2OS cells. Modules associated with specific functions are highlighted using color bars for both methods. Right panel: heatmap of Dice's coefficient indicates the similarity between the CGMs and the WGCNA modules. Modules associated with specific functions are detected by both methods with shared gene sets over 40% of the average number of genes in the two modules. 84% of the modules detected by WGCNA can be found in CGMs with Dice's coefficient > 0.4. B) Hierarchical clustering of gene-gene correlation across HEK293T cells. The four function-associated modules share gene sets over 40% of the average number of genes in the two modules. 87.5% of the modules detected by WGCNA can be found in CGMs with Dice's coefficient > 0.4.

Figure S7. Model of correlations resulting from shared transcriptional regulation (see also supplementary note). A) A transcription factor B regulates the transcription of genes C and D, with the rate of transcription given by $[B]v_C$. Transcripts C and D are degraded with rate $d_C$. The protein B fluctuates in time, driven by an arbitrarily complex mechanism regulating its production and degradation. Under the assumption that the mRNA lifetimes of C and D are short relative to the timescale of fluctuations of B, the correlation between the two genes is given by $\rho_{CD} = \frac{\mu_C \mathrm{cv}_B^2}{1+\mu_C \mathrm{cv}_B^2}$. B) Heatmap showing the magnitude of correlation coefficients that are obtained for various values of the coefficient of variation of B ($\mathrm{cv}_B$) and mean expression level of C and D ($\mu_C$). C) Correlations under a particular model in which fluctuations in B are driven by a simple model of transcription and degradation at constant rates with no bursting. In this case, under the assumption that the protein B is long-lived relative to its transcript A, the coefficient of variation is given by $cv_B^2 = \frac{1}{\mu_B}(1 + \frac{v_B}{d_A})$.

Figure S8. Distribution of correlation coefficients. A) Histogram of measured correlation coefficients for all pairs of genes. The portion of the distribution corresponding to genes with positive correlation is show in red, and the portion corresponding to negatively correlated genes is shown in blue. The distribution is roughly symmetrical with most gene pairs being

uncorrelated. B) Zoomed in view of panel (A) highlighting the tails of the distribution. Negative correlations are displayed on the positive axis for better contrast with positive correlations, and the frequency is displayed on a log scale. Strongly correlated gene pairs are significantly more prevalent than strongly anticorrelated gene pairs.

Figure S9. Clustering is robust at the magnitudes of correlation observed in our data. A set of 100 genes are set to have a fixed correlation against a background of correlations due entirely to sampling error. Clustering is performed, and the fraction of pairs of genes which are in the same cluster is recorded. The mean fraction is plotted across 10 simulations over different random backgrounds.

Figure S1

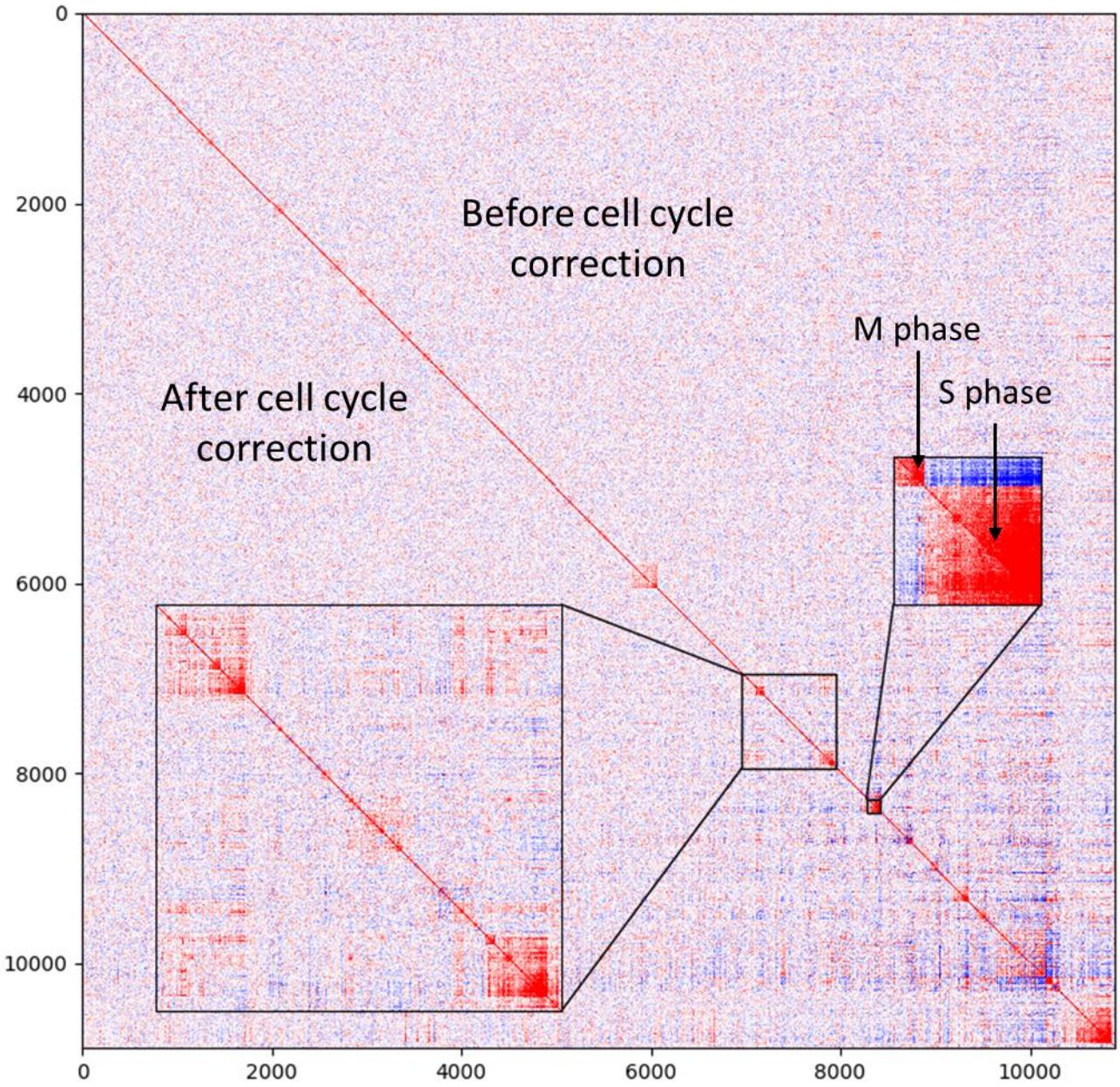


Figure S2

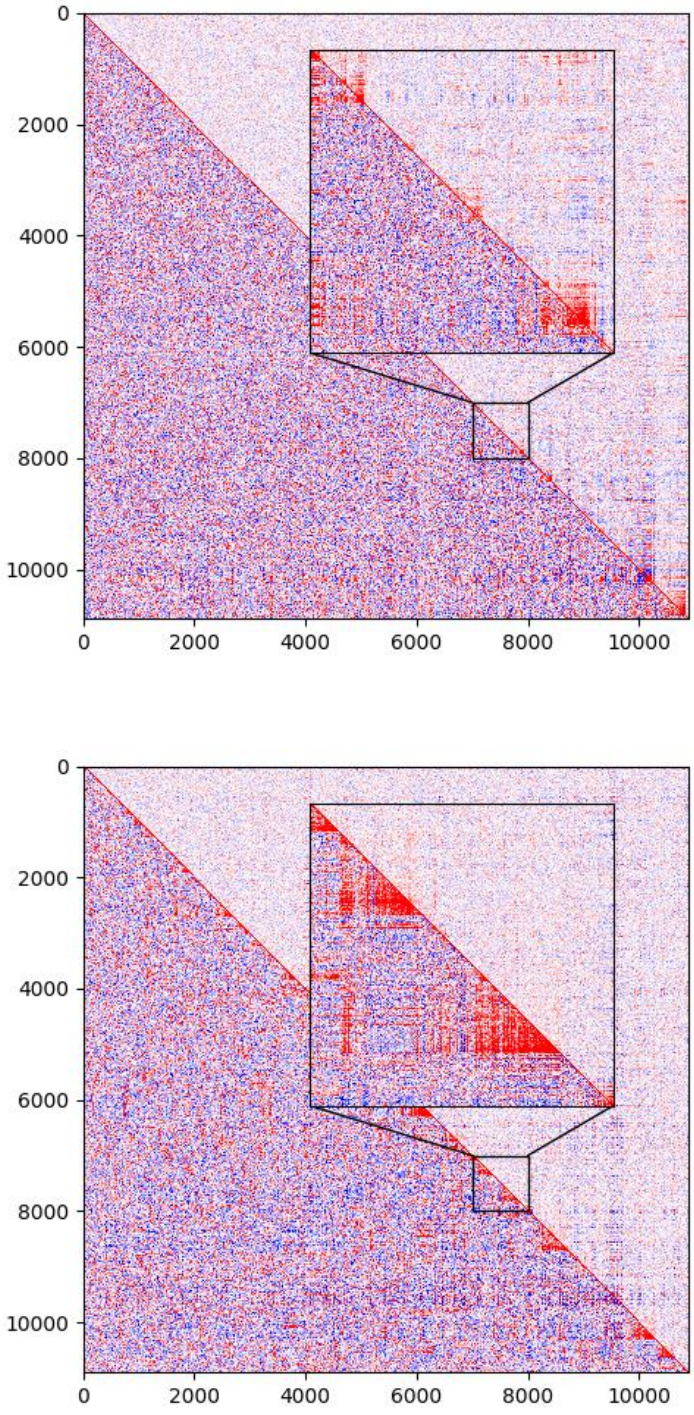


Figure S3

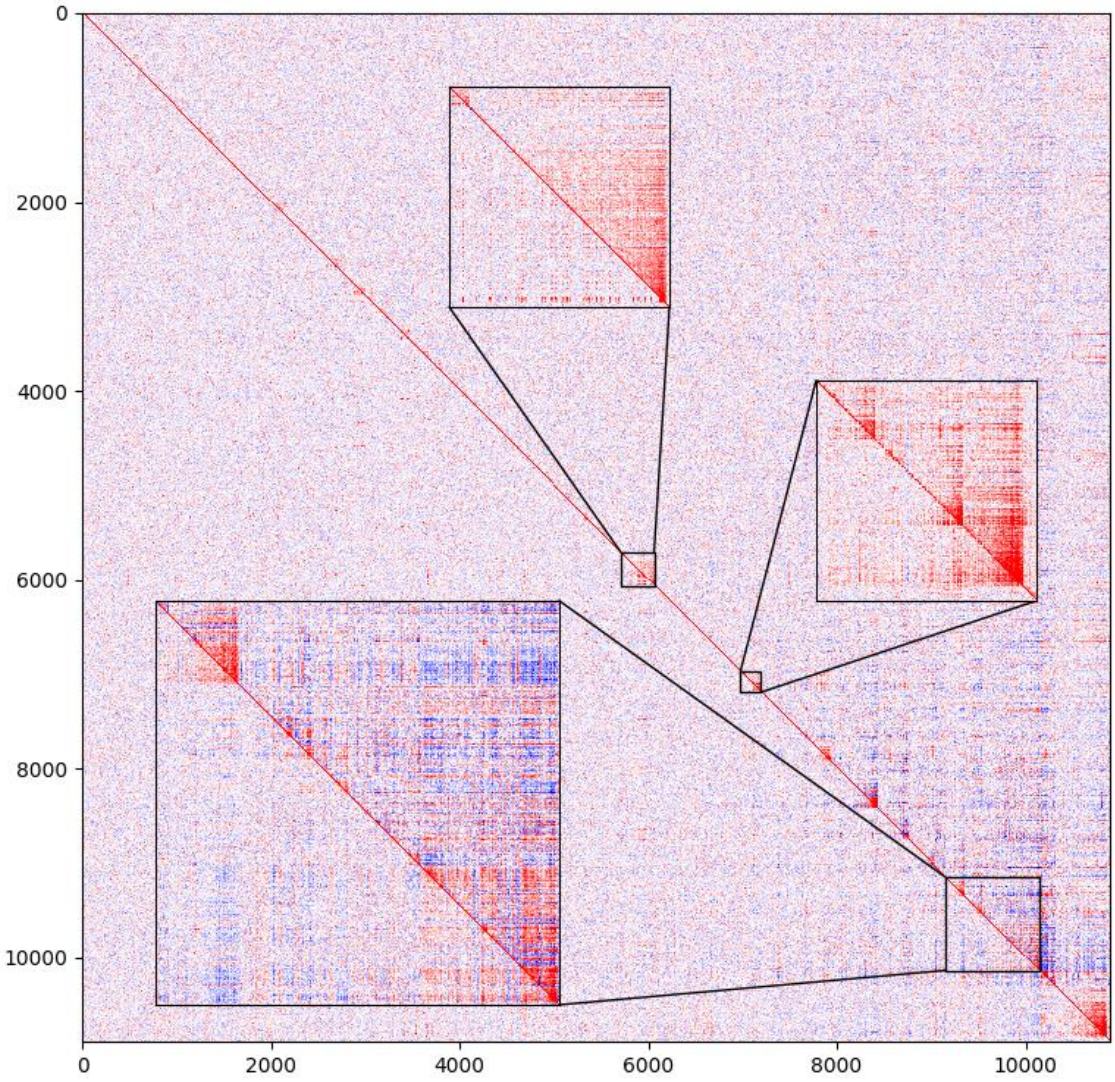


Figure S4

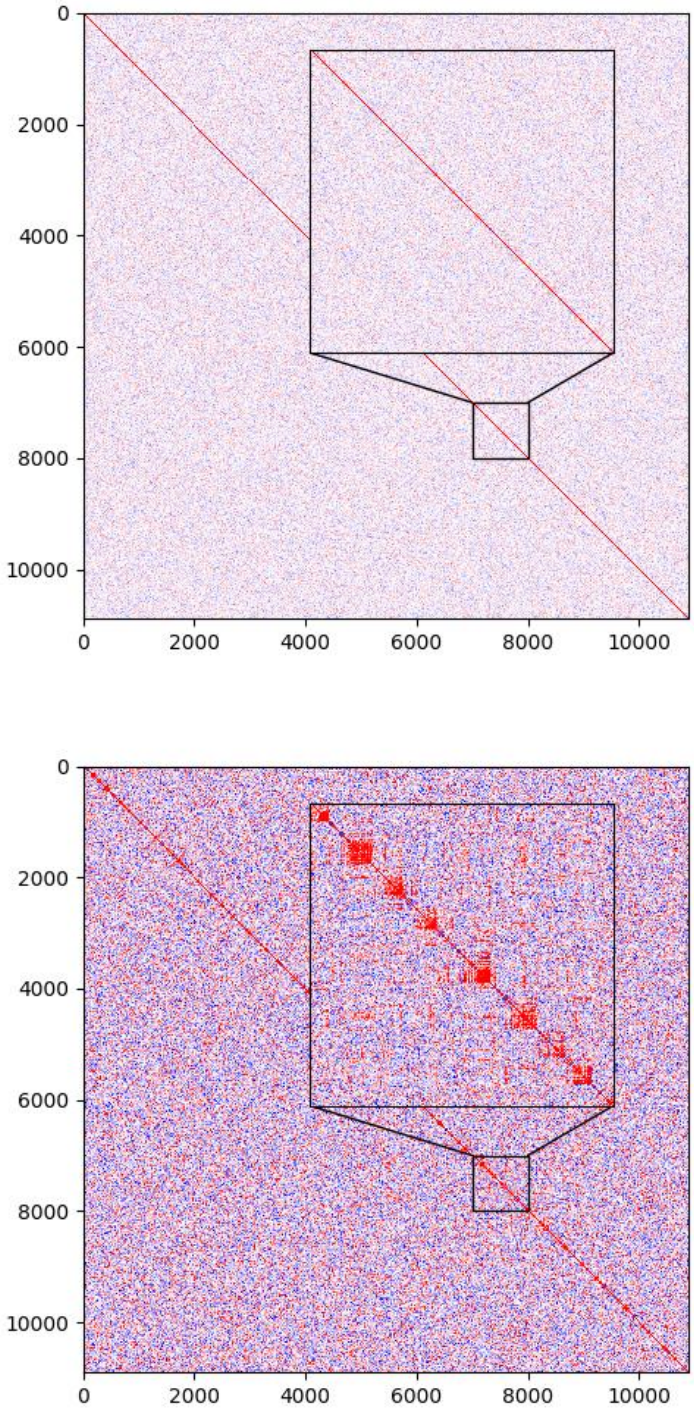


Figure S5

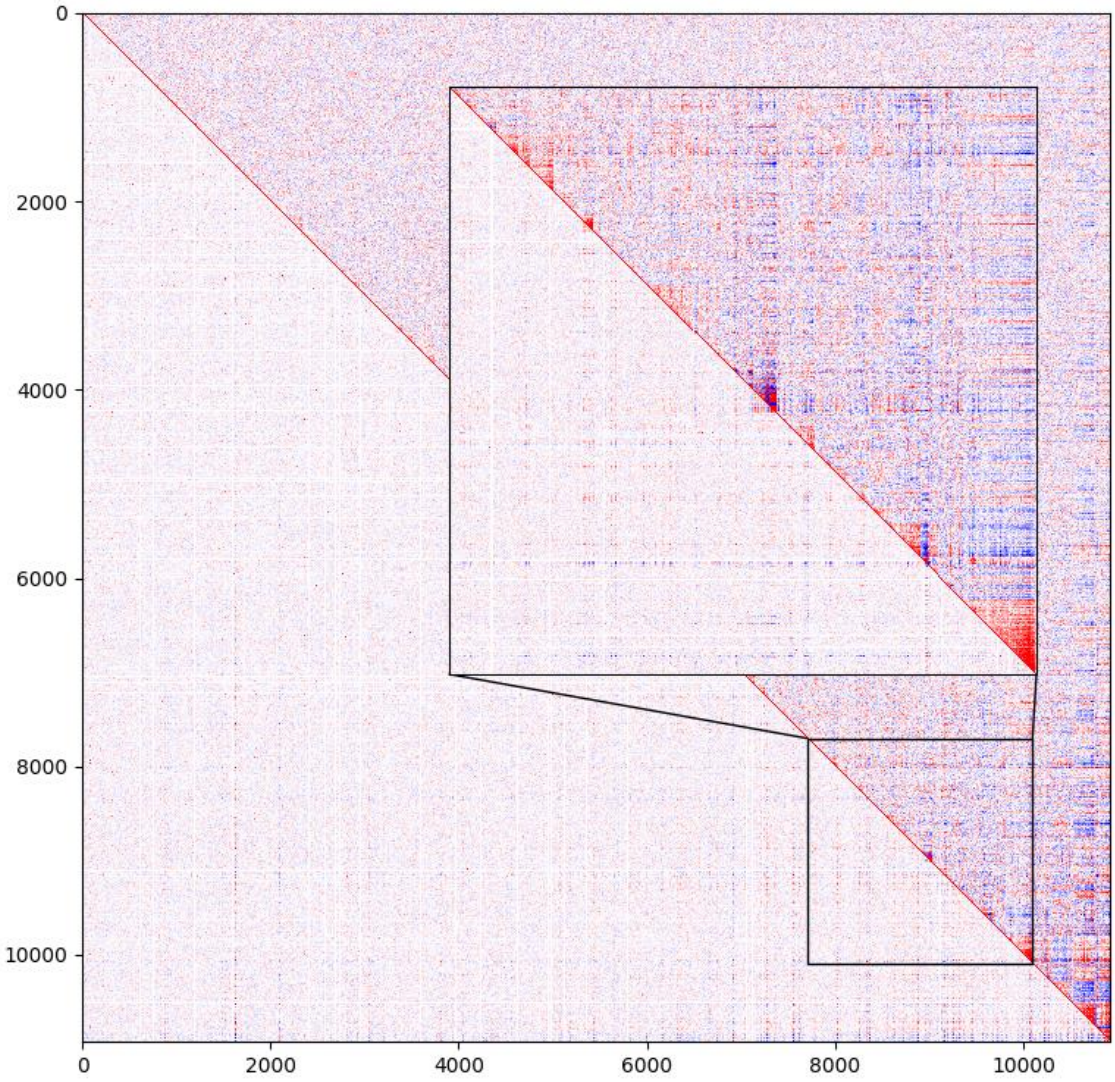


Figure S6

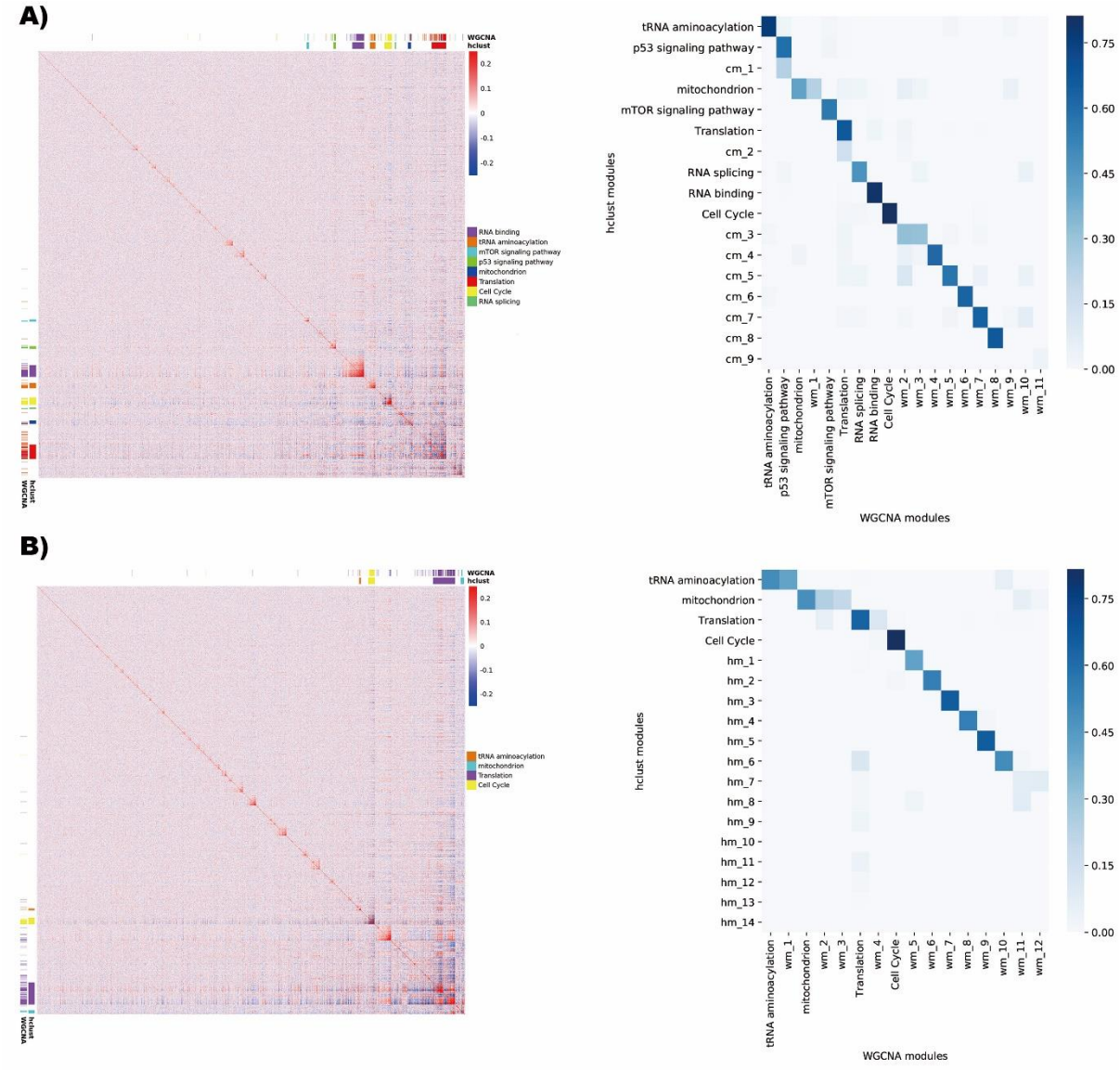


Figure S7

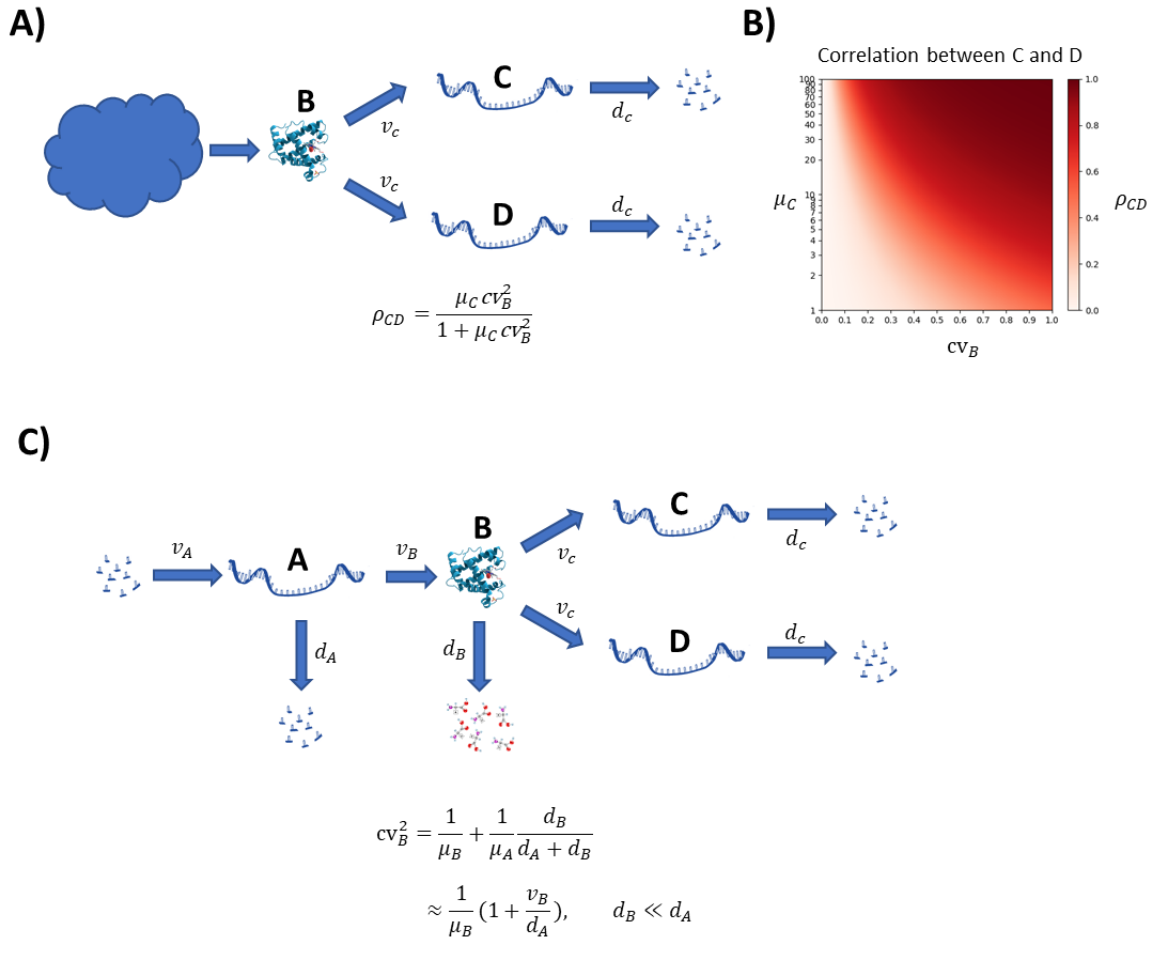


Figure S8

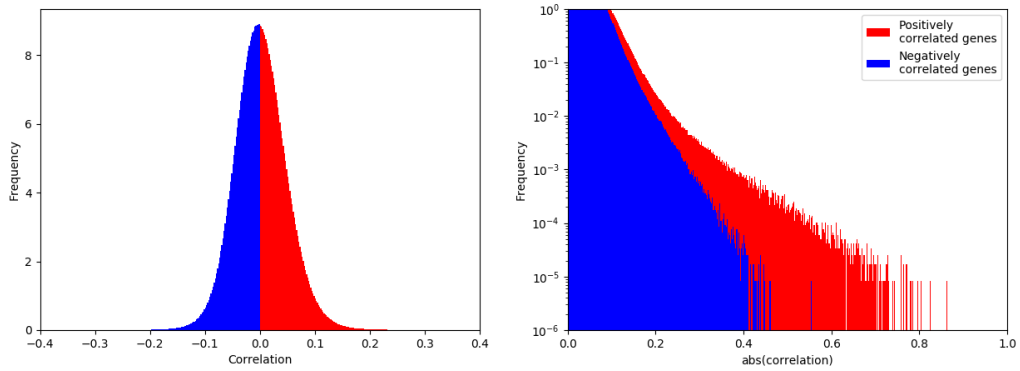
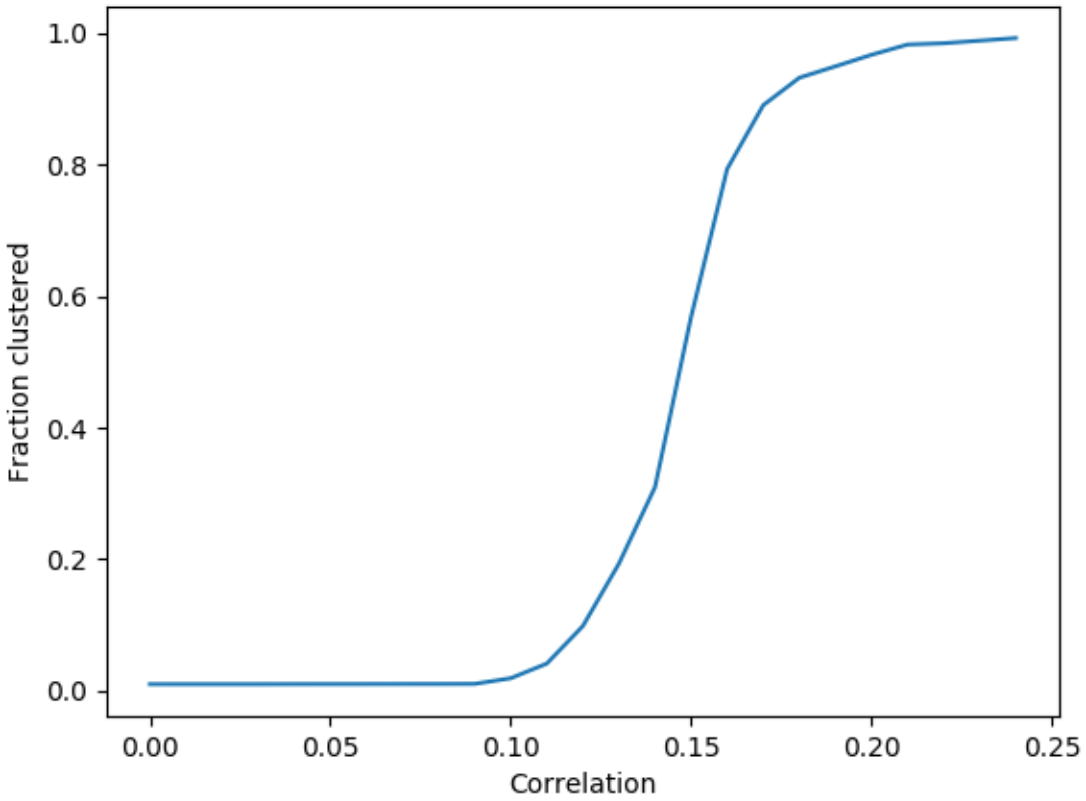


Figure S9



# Supplementary Tables

| Sample | Mapped reads | Exonic fraction | ERCC fraction | Genes detected |
|---|---|---|---|---|
| MALBAC-DT Replicate 1 | 250000 | 89% | 0.2% | 4984 |
| MALBAC-DT Replicate 2 | 250000 | 88% | 0.2% | 5567 |
| Smart-seq2 Replicate 1 | 250000 | 92% | 2.6% | 4326 |
| Smart-seq2 Replicate 2 | 250000 | 93% | 2.8% | 4605 |

**Table S1**. Performance characteristics of MALBAC-DT compared to Smart-seq2. All samples have been downsampled to 250,000 mapped reads for comparison. In order to prevent inclusion of ambiguously mapped reads and inflated gene counts, reads have been stringently filtered to exclude regions flagged by RepeatMasker.

| Sample | Mapped reads | Transcripts detected |
|---|---|---|
| MALBAC-DT Replicate 1 | 380,000 | 48,389 |
| MALBAC-DT Replicate 2 | 380,000 | 47,736 |
| MALBAC-DT Replicate 3 | 380,000 | 47,711 |
| MALBAC-DT Replicate 4 | 380,000 | 49,830 |
| Smart-seq2 with UMIs, Replicate 1 | 380,000 | 24,664 |
| Smart-seq2 with UMIs, Replicate 2 | 380,000 | 20,946 |

**Table S2**. Performance of MALBAC-DT compared to a modified Smart-seq2 protocol containing the same UMI design as for MALBAC-DT. All samples have been downsampled to 380,000 mapped reads for comparison, and the number of transcripts is presented after correcting for amplification and sequencing artifacts.

**Table S3**. Functional enrichments of CGMs identified in U2OS. For each of the 148 CGMs identified in the U2OS dataset, the set of genes comprising the CGM is presented along with the enriched pathways and transcription factors identified by Enrichr (*35*, *36*).

## Table S3

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **1** | 15 | CERK, TMEM109, CNDP2, AK4, SLC35A1, TCF3, HDAC1, ECHDC1, SCP2, EXT2, TMED10, DCTN1, ELOVL5, DDOST, RNF10 | Asparagine N-linked glycosylation_Homo sapiens_R-HSA-446203 (Reactome_2016 q=0.001630) |
| **2** | 436 | EFHD2, MIEF2, PEX26, PIM1, INPPL1, TNKS1BP1, EEF1A2, ZNF224, GEMIN7, JRK, TSC2, GMPPB, EVA1C, CIRBP, LMNB2, YOD1, PPP1R14B, RPS3P7, LMNA, GLRX5, MAML1, AMBRA1, HACD2, C12orf29, IKBKAP, WWP1, TRMT6, DRAP1, PUS1, RP11-95D17.1, SMPD1, PTOV1, TTC38, RSRC2, ATF7, MSRB1, RASAL2, ATF7IP, POP1, RWDD1, SEPT9, PNKD, MGRN1, EPHA1-AS1, VPS37C, SEPN1, COG7, HIRIP3, RRAGB, FAM160B2, SLC25A39, SF3A1, RRP7A, TTLL12, SLC29A1, TBRG4, GCDH, TXN2, OGFOD3, SDC1, CST3, PLCD1, APLP2, CES2, RPL17P25, FAM58A, SLC7A5, DCXR, RRP8, AK2, METTL5, PDLIM7, ADCK2, TMEM256-PLSCR3, VAPB, LDLRAP1, RASL10B, MARCKSL1, PMVK, PTPRS, VASP, RRAS, REL, PSMD7, TBCC, CLPP, NCOA5, SLC39A13, PLEKHG3, LIN7B, CCDC152, DIAPH3, COA6, CSNK1G3, IFFO2, ADGRL1, NUP62, MGME1, HSPA9, MATR3, NLGN2, PDXDC1, FGFR1, C11orf58, TOR4A, CD2BP2, NAT14, ADD1, VPS39, PRKAR1B, FAM234A, COA5, PIGU, MKKS, FBLN1, MTMR14, RANBP10, PCBD1, PMM1, CAD, MRPS16, RRP9, SNX8, EIF3B, ETV4, SMYD5, RRP12, SNRPC, TIMM23, ZNF511, RCC1, COA4, NFIX, WASF2, RAD23A, BRD4, SLC35F2, SRRM2, CAPRIN1, UBR4, EPB41L2, HECTD3, PPP1R7, HDLBP, C12orf43, SMARCA4, PDCD5, RBM3, AP003068.18, ST13, APH1A, SYNGR2, KRT18P29, CTD-2349P21.1, LRR1, KRT18P57, KRT18P18, NDUFB2, DSTN, PSMD8, EDF1, PSMB6, B4GALT2, PQLC1, ADRM1, KIAA2013, TMEM248, HN1, TMEM222, KRT18P31, HAGH, DOCK6, JMJD4, HS6ST1, HK1, RECQL4, ATF5, EVA1A, NME4, SUMO3, DRG2, SGO1, TACC3, ACSF3, FN3KRP, SLC38A10, MKNK2, FAM65A, CTDNEP1, EXOC3, CEP72, NISCH, FMNL3, R3HDM4, SNX17, VPS28, CDK16, KRT18P43, KRT18P37, KRT18P20, IVD, SEPT8, SPCS2, SPG7, MAFG, ZNF574, AGTRAP, CYB5D2, KIF1C, AAMP, | Metabolism_Homo sapiens_R-HSA-1430728 (Reactome_2016 q=0.009233), TNF-alpha NF-kB Signaling Pathway_Mus musculus_WP246 (WikiPathways_2016 q=0.033119), Metabolic pathways_Homo sapiens_hsa01100 (KEGG_2016 q=0.017831), JARID1A_20064375_ChIP-Seq_MESCs_Mouse (ChEA_2016 q=0.000000), MAX_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | RCC1L, NPRL3, POMGNT1, C19orf33, SLC29A2, CBX6, ZFYVE19, MPV17, JMJD8, CIC, PI4KA, MRPL2, APRT, NCS1, SLC10A3, POLD2, FTSJ1, C7orf61, GNPTG, IP6K1, ACTR5, KRT18P28, KRT18P5, ERI3, ASL, YIF1A, IMPA2, P3H1, UTP23, LLGL1, TNNT1, APBB1, KRT18P60, ATP6V0D1, ATPAF2, KCNN4, SPHK1, MRPL54, HSPB1, ALYREF, RPL18, EEF1DP5, KRT18P6, R3HCC1, KEAP1, CHCHD1, ALKBH2, TMEM9, PLD2, DUS2, DDRGK1, PLAUR, MDK, USP11, SLC27A4, ZNF618, UPF1, LYAR, SYNCRIP, PDXK, TSPAN17, MAPK12, ATN1, EPB41L3, MAP7D1, AKIRIN1, NDUFA11, KLHDC3, PYGB, TAX1BP3, BCAT2, ZNF316, RPS15, UBL4A, CINP, PAK4, NAA38, ADCY3, HTT, ZDHHC18, KAT2A, E2F4, UNC119, TADA3, SHISA5, FLOT2, MRPL21, ANAPC11, CAPN2, DAP, HEXA, DLGAP4, MRM3, COMMD4, RANGAP1, CLTB, KRT18P51, ZYX, SMTN, GRB2, CENPT, TK1, YKT6, KRT18P8, EPHA2, TIMM13, PSMG3, EXTL3, NINJ1, PLEKHJ1, MGAT1, NADSYN1, CHID1, CHMP1A, PTTG1IP, FKBP3, GNAI2, AP3D1, PAFAH1B3, TP53I11, CHST10, TNIP2, KMT2B, KRT18P38, KRT18P11, KRT18P52, EEF1DP1, PGS1, PRKD2, RUVBL1, QDPR, RBFA, C19orf48, TPD52L2, PABPC4, PSMA7, KRT18P10, WBSCR22, RP5-1056L3.3, PRPF6, POLR3H, RPS14, TUBB4B, CHMP3, PHPT1, TUBG1, TNIP1, MISP, ROMO1, ZNHIT1, LAMTOR4, TRIP6, MRPS34, POLR2L, FKBP2, ARPC4, BRMS1, SART1, COX5B, UXT, EXOSC5, MED10, RPP25L, AMPD2, C12orf10, CUEDC1, ILF3, NF2, TUBB2A, KIF22, NDUFA13, NECTIN2, PPP2R1A, KRT18P25, ARAF, MAD2L2, CCNY, MCM5, ETFB, SNHG6, CKB, THY1, CPSF1, ACADVL, GTF3C5, TRPC4AP, PLOD3, MYBL2, SNRPN, PPP4C, SMG5, STRA13, NPLOC4, MTCH1, TPM2, ELOF1, SNF8, NUDT1, ACTN1, IPO4, NARF, FAM168B, NUDC, FLII, RPL28, GUK1, PQBP1, FAU, NDUFV1, CIZ1, KRT18P17, KRT18, CLPTM1L, NAA10, GSTP1, UBXN1, ACOT7, RHOC, CCT7, ALDOA, HDAC7, PSMD4, CYC1, PSMC3, TNFRSF12A, INPP5K | |
| **3** | 550 | AP3B1, CEP83, ASH1L, PPIP5K2, AC004893.11, FAM199X, EID1, UBC, DDX5, ATXN2, ZNF609, | GTP hydrolysis and joining of the 60S ribosomal subunit_Homo sapiens_R-HSA- |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | MSI2, EIF2AK2, WWC2, NDUFB6, COX6C, NDUFB9, EIF3E, EIF3H, RPL26, RPL38, PAICS, VDAC1, GLUD1, PGRMC1, POLDIP2, MRPL51, PHB2, PSMF1, FKBP1A, ATP5O, CCT8, SOD1, MTHFD1, EIF2S1, GGCT, ENY2, H2AFV, LSM5, SLIRP, PSMA2, CYCS, HNRNPA2B1, CBX5, ARHGAP35, CHD3, NUCKS1, NUCKS1P1, ERCC6L2, SMC2, SMC1A, CEP152, RNF213, AQR, PBRM1, UTRN, MIB1, RBBP8, ANKRD12, ROCK1, CEP192, SMCHD1, DEK, ESCO2, CDK5RAP2, KIAA0368, HECTD1, WNK1, ERC1, KIF13B, MAN1A2, GNAS, PPP1R9A, PTPRM, UACA, ZC3H7A, KIAA1109, KIF13A, SNAP23, RAI14, PCM1, RMDN3, EIF2AK4, OIP5-AS1, RTF1, ZNF106, ARID2, ZMYM2, ZCCHC11, ARHGAP5, COL12A1, AHNAK, FAT1, MACF1, DST, FBN2, PHLDB2, SPTAN1, IL6ST, TRIO, SACS, KIAA0586, SLC11A2, ATP2B1, POLR2H, SDHA, CHD2, C5orf51, RAB29, UVRAG, C5orf42, RB1CC1, MYCBP2, DCAF5, NR1D2, PHF21A, KIDINS220, SOS1, USP33, HOOK3, MSL2, PIBF1, COL4A3BP, SLU7, DDB1, PLEKHA5, ANKRD28, DCAF1, AC022182.2, GTF3C1, NFIA, DNMT1, SMC5, SMC3, PSME4, ATP1B1, CANT1, MTIF2, ZFR, GPATCH4, CCAR1, CCDC59, SLTM, DIEXF, LTN1, U2SURP, SUPV3L1, HEATR1, UTP20, MRPL24, PPFIA1, STX17, YIPF6, FAF2, NSD1, ATP5F1, USP47, NCBP3, HELZ, ATXN7, KDM5A, CIR1, FARSA, MSL1, SAFB2, MIA3, CEP290, CEP350, TAF3, TYW3, ATP1A1, FXR1, TMED9, MFGE8, CRTAP, LAMP1, TMED3, TMEM219, MTDH, SETX, SF3B5, RNF20, HEXB, RNASEH2A, ROCK2, CCDC88A, ENAH, CDC42BPA, MAP1B, TUSC2, SMARCC2, PCID2, HUWE1, SETP14, BAZ1B, PRKDC, PAPOLA, YBX1, CSDE1, PABPC1, SET, EIF4G2, STAU1, TPM3, HDGF, MPHOSPH10, WDR43, BIRC6, CEBPZ, GNL3L, SASS6, CNTLN, CCDC93, MYO5A, ARIH2, PHF14, SPATS2, GCC2, ZCCHC7, RAD50, EFL1, LSG1, NSRP1, SIN3A, NCOR1, SPOP, ZNF281, ZC3H13, RPL22, TICRR, BPTF, SMG1, CNOT1, ATP5I, RAB5C, SNRNP70, NCAPH2, MED13L, TAOK1, AC016739.2, TOMM7, GLTSCR2, BCL11A, CHD7, UBR5, UBAP2L, PHF3, BAZ1A, SPEN, CHD1, POGK, | 72706 (Reactome_2016 q=0.000000), Cytoplasmic Ribosomal Proteins_Homo sapiens_WP477 (WikiPathways_2016 q=0.000000), Ribosome_Homo sapiens_hsa03010 (KEGG_2016 q=0.000000), MYC_19030024_ChIP-ChIP_MESCs_Mouse (ChEA_2016 q=0.000000), TAF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | PDS5A, CLINT1, NPC2, CHP1, ANKRD17, G3BP2, C17orf89, SFN, S100A10, S100A2, CHD8, SUPT4H1, CUEDC2, RBX1, TMEM43, SRFBP1, MYO1B, CUL3, GIGYF2, TRIP12, ARPC2, LRRFIP1, TBCA, ATP5L, ATP5J, ATP5G3, NDUFC2, RBM25, NAF1, TMEM258, ARF1, DDX6, NDUFA10, CDC5L, IQGAP1, GOLIM4, GALNT1, MTR, MNS1, TOPBP1, KIF21A, PRIM1, MYH10, SMC6, PSIP1, BRCA1, TPR, TOPORS, NEMF, PPP4R3A, NIN, KTN1, PPP4R2, UBTF, ZNF638, DNAJC7, BAZ2A, ACIN1, SBNO1, THOC2, PNISR, ITSN2, AFF4, SLK, ARID4B, PTGES3, NIPBL, RSF1, GOLGB1, FNBP4, MALAT1, CD2AP, AKAP13, DNAJC21, PPA1, FAM208B, ZEB1, KIF5B, MAP4K4, SMARCC1, TLN1, TMEM14A, STMN1, H2AFZ, VOPP1, PSMC5, ICE1, YLPM1, QSER1, TXNDC12, SKA3, RBM28, NIFK, RPF2, UTP14A, NOL8, ESF1, PMM2, DNTTIP2, GTPBP4, PRPF40A, PPIG, PRPF38B, MDN1, TXLNG, SSB, GOLGA4, IWS1, RIF1, CHD4, EIF4G1, STIP1, HSPH1, NOLC1, EIF3A, SRP72, DDX21, SON, PRRC2C, EIF5B, NCL, THRAP3, HNRNPU, GNL2, LTV1, NASP, PNN, CWC22, LEO1, RP11-435F13.2, NDUFAB1, RP11-234A1.1, STOML2, HSBP1, CHRAC1, NDUFB8, TGS1, ZNF24, MRPL37, LAMTOR1, QSOX1, HNRNPM, XRN2, TUBB, PTMA, ACTG1, PTRF, ACTB, BCAP31, TAGLN2, LGALS1, TMSB10, NDUFS6, RPS13, EIF3I, RP11-478C6.4, GPX1, RPL36, PRDX5, EEF2, IRAK1, RPS2, RPS19, RPL29, PKM, COTL1, SNRPD3, CHCHD2, ERH, SUB1, TOMM22, NPM1, RACK1, RPS24, RPL34, RPL34P31, RPS3, TPT1, RPL24, UBA52, RPL10A, RPLP0, RPL32, RP11-69M1.6, RPL35, RPL12, RPL3P4, RPL27, RPS8, RPL31, RPL37A, RPS11, RPS16, ENO1, RPS20, RPL11, RPL8, RPS4X, RPL23, RPL19, NACA, RPL5, RPS6, RPL4, RPS27A, RPL27A, RPLP2, RPS21, RPS12, RPLP1, COX7A2, CD63, NME1, SNRPD2, NDUFS5, COX7C, UQCRQ, PFDN5, ERP29, AHCY, PRDX4, FIP1L1, SMARCA5, HMGN2, ST13P15, PTGFRN, PLAGL2, NSMCE1, SERPINB6, MTCH2, CDK5RAP3, PSME2, QARS, PSME1, BRK1, ARF3, NRDC, MLEC, UBA1, PGD, SND1, MAOA, LAS1L, SRRM1, SRSF11, TCOF1, FUS, RAB34, UQCRC1, | |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | TUFM, CTNNA1, SPARC, AP2M1, SKP1, HADHA, PAGE2, RAB13, STRAP, LDHB, SERF2, RUVBL2, RPS5, AKR1B1, ATP5B, RHOA, CAPNS1, TIMM10, TPI1, ANXA2, PSMD2, MYL6, NEDD8, RTFDC1, C1QBP, EIF6, GPI, PGK1, UBL5, WBP11, PSMB4, PARK7, UQCRH, POMP, DAD1, ATIC, SNRNP200, MAP4, MAGOH, STARD7, LARP1, UTP18, COASY, MRPS7, SF3B2, CCT3, PSMA4, TRMT112, ILF2, BIRC5 | |
| **4** | 10 | UPF2, CEP250, GSE1, MAP3K2, UBXN4, MLLT10, TNRC6B, SEPT11, SCAF11, TOP1 | |
| **5** | 11 | TTC3P1, TMEM255A, SWAP70, SLIT2, RAB11FIP1, PDIA3P1, PDIA6, PDIA3, CALR, NEDD1, GINS1 | Calnexin/calreticulin cycle_Homo sapiens_R-HSA-901042 (Reactome_2016 q=0.000931), Protein processing in endoplasmic reticulum_Homo sapiens_hsa04141 (KEGG_2016 q=0.000651) |
| **6** | 27 | SFT2D1, KIAA0100, PEF1, SAR1A, YRDC, AHCYL1, H3F3C, H3F3B, CCSER2, LCOR, PDCD6, VPS13B, SLC39A10, TTC3, JMY, MXI1, PAFAH1B1, MAD2L1, RPA1, NOC3L, BCCIP, LARS, CWC27, NLN, ABI1, RNF168, REST | ZNF384_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.002830) |
| **7** | 17 | CCND1, PSMB5, EIF2B3, DYNC1I2, ZFYVE9, RBBP9, ACVR1B, BRWD1, TFDP1, SAE1, GNL3, AP1M2, SEPT7, PPP2R5E, MED1, WDR36, RSL1D1 | Cell Cycle, Mitotic_Homo sapiens_R-HSA-69278 (Reactome_2016 q=0.009275), TGF-beta Signaling Pathway_Homo sapiens_WP366 (WikiPathways_2016 q=0.009286), TGF-beta signaling pathway_Homo sapiens_hsa04350 (KEGG_2016 q=0.002329), TCF7L2_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.005646) |
| **8** | 10 | EPS15, CMTR1, NUP107, JAK1, RBM19, ATAD2B, STRN, C17orf53, RP11-157G21.2, MPHOSPH8 | Antiviral mechanism by IFN-stimulated genes_Homo sapiens_R-HSA-1169410 (Reactome_2016 q=0.041208), EGFR1 Signaling Pathway_Mus musculus_WP572 (WikiPathways_2016 q=0.036474) |
| **9** | 12 | MT2A, MT1E, GLRX, RPS4XP8, EHBP1, BNIP3, SOAT1, GSTO1, NRIP3, CD59, SSR3, CD44 | Metallothioneins bind metals_Homo sapiens_R-HSA-5661231 (Reactome_2016 q=0.000380), Zinc homeostasis_Homo sapiens_WP3529 (WikiPathways_2016 q=0.003041), Mineral absorption_Homo |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | | sapiens_hsa04978 (KEGG_2016 q=0.006623) |
| **10** | 12 | GPHN, ZNF711, SIPA1L1, DICER1, KLHL9, EEA1, GAN, POC1A, FLNC, CHD9, DDX46, AHI1 | EGR1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.011357) |
| **11** | 40 | RPL7P13, RPL7P50, SNHG8, UQCRB, SNHG5, GAS5, ZFAS1, RPL37, RPS29, ZCRB1, RPL34P6, RPS20P2, RPLP1P6, RPSAP54, ATP5H, MRPL11, RPL13AP7, TSFM, NUDT5, COPS3, MRPL33, WSB2, NDUFB1, TMEM14B, NDUFA8, CMPK1, COX17, THOC3, ARPC3, SRP9, SLC25A5, PSMA3-AS1, ANXA5, BANF1, NDUFB3, MINOS1, PEBP1, CYB5A, CNN3, CD9 | Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins._Homo sapiens_R-HSA-163200 (Reactome_2016 q=0.000409), Electron Transport Chain_Homo sapiens_WP111 (WikiPathways_2016 q=0.000000), Oxidative phosphorylation_Homo sapiens_hsa00190 (KEGG_2016 q=0.000004), TAF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.029773) |
| **12** | 15 | USP4, FAM13B, COQ9, SETD5, COPB1, TP53BP2, AEBP2, SOCS4, FBXW7, HIPK3, FRYL, SMAP1, TBC1D16, TGOLN2, ARFGEF1 | Histone Modifications_Homo sapiens_WP2369 (WikiPathways_2016 q=0.009858), UBTF_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.004239) |
| **13** | 18 | P4HA1, SSR2, ANAPC16, FURIN, ATP6V0E1, USB1, MTPN, KNOP1, GABARAPL2, SLC5A6, C8orf82, CLNS1A, TIMM50, DENR, DAZAP1, GNS, COX7A2L, SPG21 | |
| **14** | 81 | LA16c-329F2.2, UBA2, XPO1, ZBTB2, TARDBP, TIMM17A, PPP1CC, GAR1, PLEKHB2, DDX3X, NOL4L, MN1, HMGB3, PTTG1, CDKN3, DYNLL1, CDC25B, CDC20, CCNB1, CALM2, CCNB2, PSMC6, RBMX, CBX3, SFPQ, SRSF2, HNRNPDL, SRSF3, GRPEL1, EXOSC3, SSH1, CXorf23, WDR77, SNX12, PHACTR4, MRFAP1, NAA50, CNBP, SRSF5, PPM1G, FAM136A, BOD1L1, RALGAPA2, SH3BP4, GTF2A2, PGAM5, P4HB, LL22NC03-80A10.6, EIF2S3, PSMA1, NELFCD, HNRNPD, HNRNPK, NONO, DUT, MORF4L1, SSBP1, TCEB1, KHDRBS1, AGPS, KDM4A, HNRNPH1, HNRNPA3, CACYBP, HNRNPF, SRSF1, CCT6A, HSPE1, RAN, HNRNPH3, SNRPD1, DHX9, SMS, RP11-20O24.4, TMEM167A, RPL26L1, SNRPE, BTF3, MAGOHB, RBM8A, CKS1B | mRNA Splicing - Major Pathway_Homo sapiens_R-HSA-72163 (Reactome_2016 q=0.000000), mRNA Processing_Homo sapiens_WP411 (WikiPathways_2016 q=0.000000), Spliceosome_Homo sapiens_hsa03040 (KEGG_2016 q=0.000000), FOXM1_23109430_ChIP-Seq_U2OS_Human (ChEA_2016 q=0.000003), TAF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **15** | 28 | KATNAL1, DTNA, MRPS23, YTHDF1, SH2D5, ZNF207, NET1, SNHG1, MRPS27, SNHG16, RPL13A, KRIT1, PTP4A2, PRDX6, INTS10, SNHG17, RSL24D1, PFDN2, SLC39A14, PCNT, RTKN2, CTCF, SETD2, PPP1R12A, TBC1D23, KIF1BP, ARID1A, NEXN | YY1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.041173) |
| **16** | 11 | ANKRD36C, ANKRD36, OTUD4, NBN, BAHD1, NUP93, ZBTB43, BACH1, SPR, CEP135, RPGR | |
| **17** | 13 | AGO2, ZFP91, DOT1L, NF1, PPP1R15A, YWHAZ, IPO7, NFAT5, USP10, XPO4, KPNA3, PUM3, ATP6V1E1 | |
| **18** | 165 | RP11-592N21.1, RFXANK, MEMO1, KLHDC8B, RP11-114H7.1, ASTN2, RP11-133K1.1, RPLP0P6, RPS19P1, CREBL2, UNKL, NKIRAS2, C1orf43, ACAT2, AHSA1, PYCR2, LASP1, SALL2, CMC2, CH17-472G23.2, MAGEA6, MKS1, RP11-296P7.4, RP5-1014D13.2, EIF4H, PLP2, MRPL32, RP11-272G22.3, SH3BGRL3, IFRD2, CDC42, GNG12, YWHAQ, PPP1CB, LUZP1, AXL, BOD1, TTF1, CAPZB, SEC62, RP11-36C20.1, RP11-667M19.2, RP11-730G20.2, TTC5, RNASEH1, DESI2, HSP90AB2P, RPL5P8, YBX1P2, GAPDHP40, MLLT3, PPIH, DBI, ATP6V1G1, SRP14, HMGB1, KCNN3, MORF4L1P1, RPL3, RP11-1036F1.1, RP11-129B9.1, RPL14P3, RPL14P1, CLP1, YBX1P10, YBX1P1, HLA-B, FTH1P16, RP11-257P3.3, FTH1P2, FTH1P20, FTH1P3, FTH1P8, RP11-270C12.3, DHRS7, MAT2A, CCT5P2, RP11-8H2.1, TMEM177, IMP3, RNF11, FUNDC2, IMMT, RPL23A, RPL7L1, TATDN1, SF1, RBM8B, RAC1P2, MIF4GD, NT5DC2, MAGED2, ARFGAP2, RPS11P5, P4HA2, RPUSD3, YWHABP1, PHF10, CTD-2192J16.15, RPSAP12, CDKN1A, FADS1, OST4, SMARCE1, CTSA, RP11-393N4.2, RP11-552O4.2, PLIN3, PXN, S100A16, RAC1, KDELR2, SNRPA1, TCP1, PSMB1, SNRPG, RP11-372E1.1, RPL7P47, RPS17P5, SC22CB-1E7.1, DALRD3, RPL7P9, NME7, POLR2B, B2M, ZWINT, HARS, TRMT112P6, CTD-2256P15.4, PGAM1, CCT4, CCT5, RP1-278E11.3, RP11-778D9.4, EEF1GP5, MRPS18C, UBQLN4, RPS7, RP11-425L10.1, RPL13, EIF4B, UBB, HNRNPAB, HSPA8, FKBP4, RPS3P6, SUV39H1, RPS23P8, HNRNPC, NACA3P, | GTP hydrolysis and joining of the 60S ribosomal subunit_Homo sapiens_R-HSA-72706 (Reactome_2016 q=0.000004), Cytoplasmic Ribosomal Proteins_Homo sapiens_WP477 (WikiPathways_2016 q=0.013495), Ribosome_Homo sapiens_hsa03010 (KEGG_2016 q=0.002718), TAF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000002) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | XRCC6, COX4I1, EIF3F, RCN1, MEA1, RPL10, RPSA, RP11-371A22.1, DHPS, ZNF354C, HMGN1, EIF3M, CCDC73, RPS20P10, MYL6B | |
| **19** | 122 | CFDP1, MTCO1P40, VPS54, CENPN, CUL4A, CDC16, AVEN, RP11-169F17.1, MTND4P35, MT-TG, MT-TL2, MT-TH, MT-TS2, MT-TT, RP11-809N8.5, MT-TC, MT-TY, MT-ATP8, MT-ND4L, MT-CO2, MT-CO3, MT-ATP6, MT-CYB, MTRNR2L3, MTCO1P12, MTRNR2L8, MTRNR2L1, MTRNR2L12, MT-RNR2, MT-CO1, MTATP6P1, MT-TD, MT-TP, MT-ND3, MT-TR, MT-TV, MTND4P12, MTND5P11, MT-RNR1, MTND2P28, MT-ND2, MT-ND6, MT-ND1, MT-ND4, MT-ND5, LRRC42, KDELR1, MFAP1, YTHDF3, NORAD, VAMP3, VMP1, SNX1, MAGEA8, ATG3, UBA5, PANK3, MTMR2, PDHB, PSMD6, METAP1, LCMT1, ECE1, MRPL40, GEMIN6, TSEN2, FHOD1, MRPS25, SS18L2, PHB, FTSJ3, MPZL1, DCAF13, SAMM50, DRG1, RTCB, POLR1E, EIF3D, ADSL, PPIF, SNU13, ZC3H15, CCNH, NAA15, POLR1C, XPO5, KPNB1, PSMD1, HSP90AB1, SRSF7, HSPA4, DNAJA1, BMS1, DKC1, FUBP1, NOP58, NOL9, WDR12, ASUN, GRSF1, UBE2D3, MRPL22, TXNL1, MORF4L2, MAGEA12, ANXA7, NCOR2, LAMTOR5, TMEM126B, FAF1, DDX1, PTDSS1, OLA1, CCDC58, MSANTD3, DARS, ODC1, EBNA1BP2, REXO2, RPL14, SSRP1, RANBP1 | Gene Expression_Homo sapiens_R-HSA-74160 (Reactome_2016 q=0.006011), TNF-alpha NF-kB Signaling Pathway_Mus musculus_WP246 (WikiPathways_2016 q=0.046194), MYC_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |
| **20** | 91 | DNMT3A, PPARA, TNRC18, CEBPB, RPL39, SERPINH1, CNOT2, RTN4, CKAP4, SDC4, ELK1, VWA9, ALKBH5, XRCC5, INTS5, TUBAP2, CCT2, C14orf166, SUMO2, PFN1, VIM, HNRNPA1P48, IMPDH2, SEC61G, PTBP1, NDUFS3, SNRPF, DPM1, MTCYBP18, UBE2I, AATF, IST1, ZNF573, RPL37P6, JDP2, NHSL2, SAMD4B, THOC7, LDHA, CASC3, MT-TE, FTH1P10, FBXO7, NPM1P27, AJUBA, GUCD1, YWHAH, POLDIP3, HMGB1P20, STIP1P3, HMGB1P8, HMGB1P16, RPL30, RP11-832N8.1, NDUFV2, DSP, RPL10P3, GABARAP, RP3-445O10.1, RP11-404F10.2, SQSTM1, RP11-563H6.1, ZNF302, NPM1P40, SRXN1, SH3KBP1, TPT1-AS1, RPN1, SENP7, RPL9, PTMAP5, TXNDC5, HNRNPA1, HLA-DQA1, RP11-87N24.3, UQCRHL, RPL13AP5, AC105399.2, RP3- | SRP-dependent cotranslational protein targeting to membrane_Homo sapiens_R-HSA-1799339 (Reactome_2016 q=0.000160), Cytoplasmic Ribosomal Proteins_Mus musculus_WP163 (WikiPathways_2016 q=0.002008), Ribosome_Homo sapiens_hsa03010 (KEGG_2016 q=0.034152), TAF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.003824) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | 370M22.8, NACAP1, RPLP0P2, RPS23, CTD-3035D6.1, RPS10, PA2G4, ANP32B, EEF1GP1, PTMAP8, RP13-93L13.2, RPL5P4, YBX1P6 | |
| **21** | 12 | PEX5L, GOLGA3, DARS2, NOTCH2, NUP37, PRPSAP1, RBM6, RBM5, TMEM69, MRPL48, HACL1, DCP1A | Peroxisome_Homo sapiens_hsa04146 (KEGG_2016 q=0.008745), GABPA_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.035503) |
| **22** | 18 | GPN1, DPY30, GNAI3, ATP6V1C1, RAP1A, SAP18, SELT, FAM49B, PSMD14, PFN2, IARS2, ACTL6A, TOMM20, TMA7, DHX15, DDX23, CAMK2G, HNRNPH2 | Signaling by Insulin receptor_Homo sapiens_R-HSA-74752 (Reactome_2016 q=0.014537), cAMP signaling pathway_Homo sapiens_hsa04024 (KEGG_2016 q=0.030673) |
| **23** | 11 | DDX31, TIMM21, UBA3, ZNF239, PTRH2, NDN, DENND4A, H2AFY, SNHG15, UTP3, PITPNB | |
| **24** | 22 | C8orf76, ZNF263, MRGBP, RAB22A, TST, TPRG1L, IPO5, RPP40, CNIH4, RPL36AL, WBP2, CDK9, CHTF8, RNPS1, AP1S1, C19orf52, GTF2H5, TGIF2, EPHB4, PRR34-AS1, SEC22B, UQCC2 | Infectious disease_Homo sapiens_R-HSA-5663205 (Reactome_2016 q=0.012219), ZKSCAN1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.017311) |
| **25** | 16 | RIOK2, RINT1, TMA16, UPF3B, TFAM, TCERG1, UBE3A, MRPL18, CHORDC1, DDX18, MRPS9, CWC15, NAE1, CNKSR2, OMA1, CCDC66 | Organelle biogenesis and maintenance_Homo sapiens_R-HSA-1852241 (Reactome_2016 q=0.045668), Spliceosome_Homo sapiens_hsa03040 (KEGG_2016 q=0.023620), TAF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.014870) |
| **26** | 12 | RFPL4A, RFPL4AL1, H1F0, NCOR1P3, NCOR1P1, DENND5B, RGS9, RPL7P48, MTFP1, HNRNPA3P6, DHRS2, ENO3 | |
| **27** | 175 | ERLIN2, PROSC, THOC1, RP11-174O3.3, PLEKHA2, STK39, MAGEC1, SLC5A3, CASC10, CACUL1, INSIG1, SALL1, LPIN1, MSMO1, SQLE, FDFT1, DHCR24, SCD, RDH11, CA12, HMGCS1, JAKMIP2, HIST1H1C, HIST1H2AC, HIST1H2BD, HIST2H2BE, HIST1H2BJ, HIST2H4A, HIST1H4H, HIST1H2BC, MCM8, SLC25A40, TMEM245, CROT, ORAI2, CASP7, PCOLCE2, IL31RA, HSPA12A, ZNF395, CKMT1B, AGO1, FRA10AC1, PDZD8, ZSCAN21, PDE1C, SEMA3A, DDX60L, NUP153, ADNP2, PRPF4B, EXOSC1, TWSG1, RAB31, SPIRE1, FAM210A, RNMT, USP14, YES1, PPP4R1, PTPN2, AFG3L2, SEH1L, VAPA, NAPG, | Cellular responses to stress_Homo sapiens_R-HSA-2262752 (Reactome_2016 q=0.000219), Cholesterol Biosynthesis_Homo sapiens_WP197 (WikiPathways_2016 q=0.000595), Alcoholism_Homo sapiens_hsa05034 (KEGG_2016 q=0.008913), ER_23166858_ChIP-Seq_MCF-7_Human (ChEA_2016 q=0.000000), BCLAF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000077) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | SOD2, CCDC167, VCP, MRPL14, PPIL1, BYSL, CACNA2D1, WHSC1L1, ADAM9, RP11-64K7.1, GTPBP10, ABCE1, DCTD, NOL7, VCL, VAT1L, LINC00958, LINC01029, RNF7, SNW1, SNCA, RIMS2, MLF1, GRK3, SPCS3, POP7, MTERF1, IFT22, PRKRIP1, PLRG1, HP1BP3, ANXA3, PCYT1A, THOC5, ZNF3, ZKSCAN5, MEPCE, SDHAF3, TMEM14C, SRI, NEGR1, NNMT, CDK14, SYPL1, DLD, SPP1, MLLT11, MAP1A, COL6A3, SEPHS2, BZW2, NCBP2, TFRC, RPL35A, TACC1, SERPINE1, MYL12B, MYL12A, TUBB6, ELP3, FZD3, INTS9, CLU, GNAI1, HMBOX1, BAIAP2L1, TM9SF3, SERPINB7, CBR1, GHITM, MAGEA4, WAPL, GLO1, GTSF1, ZNF655, PMPCB, PCLO, CDK6, CCDC25, PSMC2, MCM7, SHFM1, COPS6, DCTN6, PBK, SARAF, LEPROTL1, TXNL4A, PDAP1, FIS1, ARPC1B, BRI3, TXN, ARPC1A, ATP5J2, SLC25A13, SRPK2, DNAJC2, PUS7, AKAP9, LAMB1, ZKSCAN1, TRRAP, CUX1, TRIM56, ANKIB1, KMT2E, SNX30, SETD7, MAGI2-AS3 | |
| **28** | 22 | SUN1, WIPI2, NUDCD3, SIRPA, MFSD1, AZIN1, HPS4, PPP1R2, SVIL-AS1, PLEK2, NPC1, ATP6V1B2, OCIAD2, AEBP1, ELP6, AC018816.3, UQCR10, GRHPR, DKK3, MGST3, CLTA, CSTB | |
| **29** | 14 | ARID1B, PRKCE, SLC26A2, MAP4K5, YTHDC2, SPOCK1, TAX1BP1, SCRN1, PARD3, ARHGAP12, MAX, TRAPPC10, PTGFR, ARHGAP18 | SOX9_24532713_ChIP-Seq_HFSC_Mouse (ChEA_2016 q=0.007865) |
| **30** | 38 | ATRAID, GTF2A1, SMURF2, ALCAM, ERBIN, UBAP2, TNRC6A, LUC7L3, EDEM3, USP34, CEP128, NOP10, NES, JPH4, OPA1, ZC3H14, RRP1B, SKIL, FAM114A2, DLG1, ALPK2, NTM, CCDC80, CABYR, CDH2, RBM18, PSMB7, ARPC5L, SEL1L, PNMA1, STX4, PPP2R3A, BICD1, PDRG1, MAPRE1, PRELID3B, RNF216, EIF2AK1 | ER Quality Control Compartment (ERQC)_Homo sapiens_R-HSA-901032 (Reactome_2016 q=0.040380) |
| **31** | 10 | CIAO1, NME2P1, RP11-234N17.1, OXSR1, MOB1A, STXBP5, UBE2E3, WRNIP1, CAMK2N1, POLE3 | ZBTB33_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.011742) |
| **32** | 11 | FUT10, LGALS3BP, NDUFAF2, PAPSS2, TACC2, SFXN4, TIAL1, CUTC, MCMBP, OXR1, RPP30 | |
| **33** | 17 | MTURN, ZNF181, F2RL2, CDR2L, MERTK, RASSF5, RAB11A, DNAJB11, NUCB2, CRELD2, AKR1C3, SEC11C, LMAN1, MAP2, ABLIM1, PDGFRB, DCLRE1A | NR1H3_23393188_ChIP-Seq_ATHEROSCLEROTIC-FOAM_Human (ChEA_2016 q=0.044362) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **34** | 11 | TOP2B, LZTFL1, MARCH6, TTC37, SUPT7L, GPATCH11, CTNNBL1, DNTTIP1, CBWD2, CHCHD3, SPATA5L1 | |
| **35** | 183 | AUH, GOLPH3L, PPP1R13L, NAIF1, C3orf58, ADAMTS1, GPSM2, KPNA6, AGFG1, HAUS4, NAV2, ITFG2, FAM161A, EPB41L1, CTDSPL, AKIRIN2, OSBP, POC5, DDX10, PAFAH2, NIF3L1, TOMM70, DNMT3B, TTF2, CDKN1B, JADE1, ZNF404, STK17B, NRAV, RASSF1, FAM110A, SLC25A38, POLD3, WDHD1, MSH2, LIMK2, ZNF367, CHEK1, DCLRE1B, USP1, RFC4, RAD51, MCM2, WDR76, MASTL, CDC25A, TIPIN, DSCC1, GINS2, HELLS, CCNE1, PCNA, EXO1, FAM111B, CDC6, MCM6, MCM3, MCM10, CLSPN, ATAD2, DTL, UNG, MSH6, SLBP, SPRTN, ARHGEF39, CCSAP, TMEM138, NEIL3, ERCC6L, CDR2, WSB1, ATL2, TGIF1, ARL4A, KBTBD2, TMEM60, PHF19, ESPL1, JADE2, SMAD3, KATNA1, TRIM59, RHNO1, STIL, PNRC2, CCDC77, CDKN2C, NEURL1B, CDC27, DEPDC1B, ZMYM1, SCLT1, ZNF148, KIAA1524, SMC4, CIT, SPTBN1, LMO7, HYLS1, NCAPH, KDM5B, MCM4, CDC25C, SRGAP2, MZT1, SHISA3, TTK, BUB3, FAM64A, TROAP, PSRC1, SOGA1, VANGL1, AURKB, NCAPD2, PRR11, BRD8, NCAPG, RAD21, SPDL1, CKAP2L, KIF18B, HJURP, CDK1, CDCA3, MIS18BP1, KIF20B, KIF11, CEP55, CNTRL, ECT2, HMMR, HMGB2, NUSAP1, CDCA8, UBE2C, DCAF7, DBF4, NUF2, PIF1, DEPDC1, SHCBP1, GTSE1, RACGAP1, BUB1B, GAS2L3, UBALD2, G2E3, ANLN, KIF4A, CASC5, CKAP5, CKAP2, ASPM, MKI67, SGO2, KIF14, PRC1, KIF18A, KIF23, TOP2A, CENPE, CCNF, CCNA2, NDC80, AURKA, CENPF, TPX2, CDCA2, CENPA, FAM83D, BUB1, PLK1, SPAG5, NEK2, KIF2C, KIF20A, ARL6IP1, CKS2, KPNA2, DLGAP5, KNSTRN | Cell Cycle_Homo sapiens_R-HSA-1640170 (Reactome_2016 q=0.000000), Cell Cycle_Homo sapiens_WP179 (WikiPathways_2016 q=0.000000), Cell cycle_Homo sapiens_hsa04110 (KEGG_2016 q=0.000000), FOXM1_23109430_ChIP-Seq_U2OS_Human (ChEA_2016 q=0.000000), E2F4_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |
| **36** | 23 | LYPD6, CCDC82, ZNF581, MELK, PARP2, TMPO, MCL1, UBE2T, NEDD9, SAV1, CARHSP1, HIST1H1A, HIST1H4C, FANCI, RRM1, FAM111A, SYNE2, DHFR, CDC45, RRM2, TYMS, FEN1, CDCA5 | G1/S-Specific Transcription_Homo sapiens_R-HSA-69205 (Reactome_2016 q=0.000000), Fluoropyrimidine Activity_Homo sapiens_WP1601 (WikiPathways_2016 q=0.000001), Pyrimidine metabolism_Homo sapiens_hsa00240 (KEGG_2016 q=0.002374), AR_21909140_ChIP- |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | | Seq_LNCAP_Human (ChEA_2016 q=0.004688), E2F4_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |
| **37** | 23 | GMNN, ADNP, SP3, SPIN1, TBC1D7, C9orf64, ICE2, ERCC-00009, ARGLU1, ERCC-00108, ERCC-00113, ERCC-00043, ERCC-00111, ERCC-00136, ERCC-00074, ERCC-00046, ERCC-00096, ERCC-00002, ERCC-00130, LRRC58, RMI1, PIGT, ERCC-00145 | |
| **38** | 26 | RER1, TBCB, SLC31A1, CNIH1, SEC61A2, TMEM123, RPL24P8, RPL6, SEP15, PSMA3, ACP1, RRAS2, PPP1R15B, WDYHV1, NDUFAF5, MMADHC, RIPK2, DAXX, DCUN1D5, ALDH18A1, PRDX3, CDC123, VDAC2, UTP11, KARS, MPHOSPH6 | FAS pathway and Stress induction of HSP regulation_Mus musculus_WP571 (WikiPathways_2016 q=0.017617), NRF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.003712) |
| **39** | 12 | MKLN1, NSUN4, HOXB6, JMJD1C, CLIP1, TRIM33, SIRT1, NBEA, FEM1B, CDC37, PHF20L1, IREB2 | BRCA1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.003763) |
| **40** | 16 | SNRPB2, URB1, PPRC1, USP36, HOMER1, DYRK2, AKAP1, TBL1X, COA7, MGA, HK2, MINA, PTX3, PPARGC1B, RP11-259N19.1, SNTB2 | Transcriptional activation of mitochondrial biogenesis_Homo sapiens_R-HSA-2151201 (Reactome_2016 q=0.000178), Mitochondrial Gene Expression_Homo sapiens_WP391 (WikiPathways_2016 q=0.000961), MYC_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000037) |
| **41** | 10 | BLOC1S1, CHMP2A, NEFL, BEX3, CSNK2A1, NOP56, TCEAL8, RABEP1, PPAT, ZNF485 | Lysosome Vesicle Biogenesis_Homo sapiens_R-HSA-432720 (Reactome_2016 q=0.014731), Ribosome biogenesis in eukaryotes_Homo sapiens_hsa03008 (KEGG_2016 q=0.011192) |
| **42** | 52 | SLC44A1, UGP2, CUL4B, KLF7, KLF6, PAPSS1, ZNF827, TUFT1, TNS1, TGFB2, DCBLD1, PRKCA, RHOBTB3, AMIGO2, MOB3B, TIMP3, PLK2, CPA4, DPYSL2, EHD2, SYNPO, TNS3, ZFHX4, MPP5, RHOBTB1, CGN, BMP4, FOXN3, TCF4, BMPR2, COL11A1, BAZ2B, ARHGAP29, RP11-879F14.2, RHOB, REEP1, SCARA3, TGM1, TMOD3, DOCK5, ANXA8L1, FSTL1, SEMA3C, NEK7, CTSC, LAYN, TENM3, ANXA10, PALLD, LRRN1, ADAMTS12, PKP2 | PodNet: protein-protein interactions in the podocyte_Mus musculus_WP2310 (WikiPathways_2016 q=0.012435), ESR1_22446102_ChIP-Seq_UTERUS_Mouse (ChEA_2016 q=0.000947), AR_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.033172) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **43** | 11 | MAPK1IP1L, GPC4, SDC3, TLR4, RORA, C1orf198, LBH, MEIS1, KLHL42, IGF1R, ATP9A | Defective EXT1 causes exostoses 1, TRPS2 and CHDS_Homo sapiens_R-HSA-3656253 (Reactome_2016 q=0.000615), Inflammatory bowel disease (IBD)_Homo sapiens_hsa05321 (KEGG_2016 q=0.024137) |
| **44** | 37 | ING5, IRS1, DNAJC5, PRTFDC1, WAC, LIMCH1, YME1L1, TRAM2, TLE1, UCP2, PPME1, FLJ22447, HAPLN1, CAV1, TLN2, C15orf52, CADM4, SOX9, DCBLD2, EZR, THBS1, ARL4C, PMEPA1, MFAP5, MYL2, SLC2A3, CTSV, ZBED2, KRT6A, EFEMP1, SEPT2, SRPX, FN1, SERPINE2, MMP2, RBP1, COL4A1 | Extracellular matrix organization_Homo sapiens_R-HSA-1474244 (Reactome_2016 q=0.000008), Focal Adhesion-PI3K-Akt-mTOR-signaling pathway_Mus musculus_WP2841 (WikiPathways_2016 q=0.006482), Focal adhesion_Homo sapiens_hsa04510 (KEGG_2016 q=0.000088), TRIM28_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.038650) |
| **45** | 182 | PXDC1, COPG1, FAM177A1, CHIC2, MIR6797, RBM38, COL17A1, ATP6V1D, KHDRBS3, CLDN1, CCDC50, ZNF674, DDIT3, TSC22D3, MXD1, NFIL3, GABARAPL1, YPEL5, NDRG1, MAP1LC3B, FAM24B, FOXF2, RP11-874J12.4, DLGAP1-AS2, IL1RL1, RPL13AP20, CLIP4, GCFC2, VPS41, THEM4, RBSN, PTER, MYZAP, WNT2B, PLXNA2, PID1, PTPN11, AP2B1, NCKAP1, GLT8D2, DNAJC10, CCNB1IP1, N4BP2L2, OGT, IFITM3, SPX, DOK5, PEA15, COMMD7, ETV5, BMP5, ITPR2, MSN, PSAP, GREB1, DUSP11, SLFN5, SESN2, LAMA1, UPP1, MMP3, CREBRF, HRK, GNPDA1, IFI16, GPNMB, THNSL2, BEX2, MCTP1, EPG5, TUBE1, RSPH3, SGPL1, XPOT, RAB3GAP1, CASP4, SLC48A1, CSTA, NLRP1, FAM21C, LARP6, LUCAT1, ZFP69B, TRIM25, NCK2, CCNI, CLIC4, TMBIM6, NMNAT2, GFPT1, ALDH2, VAT1, IDH1, ASS1, GRB10, MOCOS, CLTCL1, SLC38A2, SLC38A1, VEGFA, NCOA7, SEL1L3, OGFRL1, TARS, CARS, MTHFD1L, SLC7A1, ST6GALNAC3, DHRS3, CTH, FUT1, CHAC1, LMO4, IARS, PAPPA2, WARS, HERPUD1, EIF2S2, CEBPG, NUPR1, EIF1, ATF4, PSPH, SHMT2, PHGDH, EIF4EBP1, PCK2, SLC1A5, FKBP9, CCND2, EPRS, ALDH1L2, YARS, BCAT1, GPT2, XBP1, SLC1A4, UHRF1BP1, DDR2, GARS, MARS, SARS, PSAT1, MTHFD2, ASNS, HAX1, SUN3, NARS, RAB39B, AARS, TRIB3, SH2B3, ANKRD11, VLDLR, | Cytosolic tRNA aminoacylation_Homo sapiens_R-HSA-379716 (Reactome_2016 q=0.000000), Trans-sulfuration and one carbon metabolism_Homo sapiens_WP2525 (WikiPathways_2016 q=0.000000), Aminoacyl-tRNA biosynthesis_Homo sapiens_hsa00970 (KEGG_2016 q=0.000000), ATF3_27146783_Chip-Seq_COLON_Human (ChEA_2016 q=0.000000), CEBPB_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | FAM107B, IFRD1, FAM155A, DAB2, FYN, TNC, DYNC1H1, ZPR1, SERPINB8, PIR, MGST1, G6PD, URI1, CTSL, HTATIP2, HKDC1, ADK, GSR, ASPH, GCLM, APCDD1L-AS1, ME1, NQO1, FTL, TRIM16, TRIM16L, SLC7A11, TXNRD1 | |
| **46** | 10 | NRAS, HNRNPA0, KIAA1551, BLMH, CCND3, GDI2, ATP5C1, XPO7, NEK4, KIAA0930 | |
| **47** | 31 | PHF6, FAM127C, STX8, HTATSF1, SERINC1, TERF2IP, PPOX, UFC1, TANGO2, LAMC1, STT3B, FBLN5, RNPEP, CD46, ACOT9, SYNE1, ABCC9, SNX2, NDFIP1, RIOK3, POFUT2, SP100, ITPR1, SP110, AFF1, PHTF1, EDEM1, STK17A, UFD1L, RCAN1, ATP6AP2 | RUNX1_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.042784) |
| **48** | 14 | RAB7A, CDV3, TOMM34, HSPD1, HSP90AB3P, CAPZA1, API5, GAPDHP65, GAPDH, GAPDHP1, HSP90AB6P, CSE1L, PSMD11, PSME3 | Regulation of activated PAK-2p34 by proteasome mediated degradation_Homo sapiens_R-HSA-211733 (Reactome_2016 q=0.004163), Proteasome Degradation_Mus musculus_WP519 (WikiPathways_2016 q=0.005040), Proteasome_Homo sapiens_hsa03050 (KEGG_2016 q=0.007619), TAF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.030891) |
| **49** | 23 | C12orf65, LACTB, MNAT1, COPB2, FNDC3B, AIDA, MRPL35, DVL1, NDUFA12, WWC1, MRTO4, NOP16, DDX27, RAE1, CEP112, GPX8, NBR1, NDUFB5, SNAP29, ATRX, MT-TM, IMP4, ERGIC2 | YY1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.003289) |
| **50** | 10 | AUP1, UQCRFS1, HSPA14, SUPT5H, CCNL1, PSMD13, TMX2, WDR33, GOT1, CRTC3 | Parkin-Ubiquitin Proteasomal System pathway_Homo sapiens_WP2359 (WikiPathways_2016 q=0.008005), YY1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.031846) |
| **51** | 52 | GSK3B, TJP2, NABP1, AREG, CYR61, EDN1, CTGF, VEGFC, FOSL1, ENC1, MYO10, NEDD4L, NRG1, SKAP1, CH17-472G23.4, PDE4DIP, RGS4, MET, DNMBP, NAV3, TGFBR2, KIAA1549L, PLCXD2, ITGA6, ITGA2, ESM1, HMGA2, PAX8-AS1, LINC00911, CPEB4, BIRC2, ADAMTS6, LAMB3, FRMD6, DGKI, ANTXR2, FGF5, KRT15, LAMC2, MYH16, TGM2, MAOB, RP11-78C3.1, NFATC2, | Laminin interactions_Homo sapiens_R-HSA-3000157 (Reactome_2016 q=0.000101), Focal Adhesion_Homo sapiens_WP306 (WikiPathways_2016 q=0.000000), Pathways in cancer_Homo sapiens_hsa05200 (KEGG_2016 q=0.000000), SMAD4_19686287_ChIP-ChIP_HaCaT_Human (ChEA_2016 q=0.000015), SMAD4_CHEA |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | ADAM22, UAP1, UBASH3B, LSM6, CCNJL, CAB39, PHC2, EGFR | (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |
| **52** | 57 | ZNF185, ANXA1, DCLK1, CORO1C, GPR176, DYNLT1, DCTN2, PGM1, SNX7, SLC25A4, C9orf40, SAMD9L, LINC00862, SCN9A, NIPAL3, NCF2, AOX1, KCNQ3, PGM5, SRGN, AC013461.1, MATN2, MPP7, RNF144B, KCND3, STXBP6, TP53, ZHX3, CTNND2, KRT75, ANKRD13A, MIR137HG, DNAH11, CAMK2D, PHACTR2, SPATS2L, ABL2, SH3RF1, HRAT17, ZC4H2, EXT1, INPP4B, LIMA1, WDR1, ARPC5, LPP, ACTR3, CTPS1, TPM1, CALD1, CSRP1, TPM4, RP11-553A10.1, FGF1, MMGT1, PPP1R3B, SLC20A2 | Muscle contraction_Homo sapiens_R-HSA-397014 (Reactome_2016 q=0.000319), PPARD_23208498_ChIP-Seq_MDA-MB-231_Human (ChEA_2016 q=0.001014), PPARD_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.012062) |
| **53** | 18 | PRKD1, RP11-248J18.2, MB21D2, PDLIM5, USP53, C1GALT1, CREB5, HSPB8, ZBTB46, MAFF, DGKD, DUSP1, ANKRD1, CSRNP1, HBEGF, IER2, BHLHE40, KLF10 | Hypertrophy Model_Homo sapiens_WP516 (WikiPathways_2016 q=0.002232), ESR1_21235772_ChIP-Seq_MCF-7_Human (ChEA_2016 q=0.049973), TCF3_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.001043) |
| **54** | 20 | TNFRSF9, INHBA, RELB, TNFAIP3, NFKBIA, NFKB1, CD83, IL32, NFKBIE, BIRC3, MYC, ZFP36L1, TRIB1, ERRFI1, DUSP5, GEM, PTHLH, GPAT3, ZNF697, ADD3 | Activation of NF-kappaB in B cells_Homo sapiens_R-HSA-1169091 (Reactome_2016 q=0.001808), TNF-alpha NF-kB Signaling Pathway_Mus musculus_WP246 (WikiPathways_2016 q=0.000001), Epstein-Barr virus infection_Homo sapiens_hsa05169 (KEGG_2016 q=0.000001), RELA_24523406_ChIP-Seq_FIBROSARCOMA_Human (ChEA_2016 q=0.000000), RELA_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |
| **55** | 22 | RP11-248J18.3, CALM1, WDR20, ZFYVE21, FERMT2, ARID4A, BAG5, YY1, PPP1R13B, SETD3, CDC42BPB, AHNAK2, RP11-144L1.8, HSP90AA2P, HSP90AA1, EIF5, MARK3, SIX4, ZNF174, PCNX4, HIF1A, SNAPC1 | eNOS activation_Homo sapiens_R-HSA-203615 (Reactome_2016 q=0.006049), TCF3_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.013761) |
| **56** | 14 | COPS7A, EARS2, SV2A, COL1A1, TMEM203, M6PR, VDAC3, CNOT7, EMG1, RPUSD4, AP3M2, METTL1, LCLAT1, ARL6IP5 | ECM-receptor interaction_Homo sapiens_hsa04512 (KEGG_2016 q=0.033662), CREB1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.001994) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **57** | 10 | NEIL2, MRPL15, ST3GAL1, ZNF275, RTN4IP1, EXOSC2, SLC7A2, SIX1, INPP5F, HACD3 | SOX17_20123909_ChIP-Seq_XEN_Mouse (ChEA_2016 q=0.028554) |
| **58** | 10 | CMSS1, KMT2C, PRKCQ, ASXL2, ATP11A, RNF4, TDRD3, USF2, STAM, SP1 | Androgen receptor signaling pathway_Homo sapiens_WP138 (WikiPathways_2016 q=0.022989), MYCN_18555785_ChIP-Seq_MESCs_Mouse (ChEA_2016 q=0.037390), ZNF384_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000733) |
| **59** | 11 | S1PR1, TNIK, RFX3, DDX50, OSBPL10, TRIB2, SPIN2B, MCC, MYOF, TAB3, ATP11C | TNF-alpha NF-kB Signaling Pathway_Mus musculus_WP246 (WikiPathways_2016 q=0.035536) |
| **60** | 102 | PDCD4, MEX3B, PTPRU, AZI2, GPR87, PDGFC, CFAP58, RP11-1069G10.2, STOM, CEP76, FTO, GALNT5, F8, AEN, PMAIP1, ERFE, PRKAB2, BAX, KAT2B, TMED4, PPP1R3F, CTA-392C11.1, ZNF785, HOXC13, FAM46A, POLR2A, ISCU, SNHG12, CPE, LRP10, RP11-245D16.4, KLHL17, C3orf67, GADD45A, CYLD, FAM198B, PLK3, RP11-94H18.1, YBX3, ATF3, RNF19B, TP53I3, TSPYL2, EYA2, TRIM5, DGKA, LINC01468, COBLL1, MITF, TNFRSF10B, E2F7, TENM4, APOBEC3C, SLAMF7, WDR66, RP11-107M16.2, RNF182, MRPL49, EI24, SESN1, VAMP8, BBC3, BLOC1S2, KLHL5, AMZ2, NTPCR, RP11-421F16.3, MAST4, CSMD3, CCNG1, ERGIC3, MYLK, SUSD6, FBXO22, CCDC90B, RAD51C, PSTPIP2, ANXA4, FAM210B, PARD6G, FAM212B, PPM1D, FAS, CCDC148, ZMAT3, TP53INP1, TRIM22, TIGAR, CMBL, BTG2, RRM2B, NECTIN4, MDM2, FDXR, CYFIP2, SUGCT, TM7SF3, RPS27L, RP11-115D19.1, TRIAP1, PTP4A1, RP11-363E7.4 | Transcriptional Regulation by TP53_Homo sapiens_R-HSA-3700989 (Reactome_2016 q=0.000000), p53 signaling_Mus musculus_WP2902 (WikiPathways_2016 q=0.000000), p53 signaling pathway_Homo sapiens_hsa04115 (KEGG_2016 q=0.000000), TP53_22127205_ChIP-Seq_IMR90_Human (ChEA_2016 q=0.000000), TP53_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.000000) |
| **61** | 11 | SH3BP5-AS1, RP11-93H12.4, ACYP2, KLHL24, SDAD1P1, GPX3, F8A1, CARD6, ZNF12, AKR1B10, TMEM47 | Metapathway biotransformation_Homo sapiens_WP702 (WikiPathways_2016 q=0.031965) |
| **62** | 35 | NR3C1, USP27X, SRSF10, IRF2BPL, PSMD10, FUCA1, RP11-488P3.1, TEX9, LINC01560, LYRM1, AMZ2P1, BANK1, PRRX1, RP11-76C10.6, RNASEL, KCNK1, LRRC27, ZSWIM7, MICAL2, FHL2, CATSPER1, SPANXD, CREB3, SH3BGR, EDA2R, LINC01021, LURAP1L, NRP1, CDH13, EPHB1, IL20RB, CETN2, PHLDA1, CTSB, TIMP1 | |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **63** | 11 | RPL15, NGLY1, TSHZ1, SNRNP48, FAM101B, FGF2, TMEM147, GAREM2, PYGL, UROD, RALBP1 | |
| **64** | 10 | SNX18, PRKAR2A, SLC1A3, SRPK1, PTPN13, FKBP5, LYN, ZNF93, OSBPL6, LINC01579 | |
| **65** | 17 | OLFML3, ILDR2, PLCE1, SGCB, FEZ1, NYNRIN, RP5-1172A22.1, PDLIM1, SEPW1, SRPX2, KRT17, TUBA1A, MAP3K7CL, FLG-AS1, PTPRR, LRP11, GLP2R | |
| **66** | 10 | ZNF516, APMAP, ZNF322, ZCCHC14, PLEKHA8, RNF114, ARF6, CDCA4, EIF4A2, LSM14B | |
| **67** | 10 | OBFC1, GSTCD, GJC1, PACS1, CENPI, KMT2A, MPHOSPH9, IL7R, ZNF217, CYP24A1 | Metapathway biotransformation_Mus musculus_WP1251 (WikiPathways_2016 q=0.029939) |
| **68** | 11 | IDH3B, VPS16, CCDC51, PTPA, ZSCAN2, S1PR2, SEC11A, UNC45A, MCF2L, GALT, ST3GAL2 | |
| **69** | 12 | NUBP1, SKI, TP73-AS1, LINC01128, CCNL2, AURKAIP1, MRPL20, SSU72, GNB1, MORN1, SDF4, PP7080 | Mitochondrial translation elongation_Homo sapiens_R-HSA-5389840 (Reactome_2016 q=0.023103), KDM5A_27292631_Chip-Seq_BREAST_Human (ChEA_2016 q=0.000283) |
| **70** | 10 | RITA1, STK24, EIF4E2, NDUFAF3, ZCWPW1, DMKN, VGLL3, RAD51AP1, TNFAIP1, HOXB7 | |
| **71** | 10 | NUDT2, FRMD8, FOXM1, HYPK, RRN3P3, GRWD1, SRPRB, C10orf2, IPO9, SAP130 | |
| **72** | 10 | RNF103, BAMBI, HCN4, ASPHD2, MAP3K4, CTD-2371O3.3, HEY1, CPM, FBXO32, MAF | |
| **73** | 23 | RNF149, RP11-408P14.1, RP11-791G15.2, MEGF8, MAP3K5, CKMT1A, PLEKHO1, EFNA1, DUSP10, ARRDC3, RNF13, CHKA, PGAP1, CPEB2, PKD1L1, DPP4, BHLHE41, HMOX1, RND3, CLEC2B, RBBP6, CITED2, ARID5B | TCF4_18268006_ChIP-ChIP_LS174T_Human (ChEA_2016 q=0.028564), SALL4_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.048302) |
| **74** | 160 | ZNF765, ZNF761, GFM2, WDR27, TRAPPC6B, TP53BP1, ATF1, ACBD3, DMTF1, DDX60, ARHGEF12, TMEM184C, USP54, SOS2, VPS13C, LINC00467, TAF1A, RNU6-817P, C15orf57, KLHDC4, RNF185, TDRKH, POTEG, GATA3, IFNLR1, DUSP18, COG5, PHKA1, ATF2, PAM, STAM2, CRELD1, ABHD12, CHEK2, ERMAP, TOP3A, GPR156, GALNT10, ARV1, KCNC3, ASF1A, POLR2J, SLC35D1, ZNF319, SLC2A10, FAR2P1, HERC1, ZNF808, OXA1L, RP11- | ATM Signaling Pathway_Homo sapiens_WP2516 (WikiPathways_2016 q=0.035431), GATA2_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.044922) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | 488L18.4, ZFP1, KIRREL3, FAM49A, STON1, WDFY3, CTNND1, CASP6, FNIP2, TNFAIP8, IFNGR1, ATM, MTERF3, PURG, MAP3K9, RP11-313J2.1, ARMCX4, OSER1, CD276, MYO18A, ANKRD62, ZNF101, CC2D1B, ASB13, ANO5, TRIM65, AASS, DHRS1, CPOX, PARP12, HEATR3, IL17RD, RP1-152L7.5, USP13, TRIM6, LMBR1, NCOA6, HMGB1P10, TET1, VASH2, THRA, FXYD6, UNC119B, RP5-890O3.9, DMRTA1, TUFMP1, HAUS2, POLA1, PDE6D, GPR180, STRIP2, CLMP, RP11-671C19.2, MFAP3L, SALL4, ST5, FAAP24, INTS8, UROS, GSDMB, FAM26E, NEAT1, HM13, CEPT1, TIGD2, RP11-529H20.3, SPPL3, TRMU, RP11-1023L17.1, ASNA1, DNAJC6, MGAT4A, NFRKB, ZNF720, TGFBR3, VPS8, SPAG16, ZNF624, SLC35B4, RAB30, SLC22A23, PLCB3, RPL35P1, GPR155, FARP1, PHLDB1, NREP, FLRT3, BZW1P2, FAM46C, ALDH1L1, WDR74, ZNF266, NAP1L4P1, GCNT2, TMEM170B, WDR60, PPP1R21, EPHA5, VWDE, UBE2D1, CDH11, ZNF449, OBSCN, RP11-10C24.3, CPNE4, SETBP1, RP11-95M15.2, PKIG, R3HDM1, SOCS6 | |
| **75** | 12 | GAB1, BTBD1, ZNF680, MON1B, ACBD5, ROS1, IDH3A, GSAP, RMND1, MINPP1, RPL36AP15, ADAMTSL1 | |
| **76** | 10 | POLR1B, OXNAD1, PIGO, ZNF629, SCYL3, TUBGCP3, SRP72P2, IFIT5, FARSB, NUTF2 | FOXO3_22982991_ChIP-Seq_MACROPHAGES_Mouse (ChEA_2016 q=0.046592), ELF1_ENCODE (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.018095) |
| **77** | 22 | BBS7, SASH1, RP11-113I24.1, NDUFA1, C4orf32, FOXC1, CALN1, CBWD5, R3HDM2, UQCR11, PSMD9, C12orf60, DDA1, PARD3B, MAN2A2, RRP1, PPARGC1A, DOCK9, PPM1H, PSMC3IP, C4orf19, EXOC6 | Huntington's disease_Homo sapiens_hsa05016 (KEGG_2016 q=0.016663) |
| **78** | 11 | ARHGEF11, TBC1D30, ZRANB3, RUNX1, CNKSR3, IFT46, ZNF814, PLEKHM3, DLST, AASDH, KCTD5 | |
| **79** | 12 | USE1, C19orf54, ADCY9, PRCAT47, IPMK, WWP2, KBTBD6, SLCO4A1, KCNJ14, PACSIN2, ZFP36L2, CACNB3 | Calcium Regulation in the Cardiac Cell_Homo sapiens_WP536 (WikiPathways_2016 q=0.010423), Oxytocin signaling pathway_Homo |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | | sapiens_hsa04921 (KEGG_2016 q=0.005050) |
| **80** | 17 | GAPDHP72, CCDC126, GABRA2, BUD31, SNX16, GCH1, RB1, MYADM, GVINP1, AC091633.2, SRPK2P, AKAP10, IGFBP7, ZFP90, TOB1, ZBTB20, SYT1 | |
| **81** | 38 | CAP2, ZNF391, PPIL4, HMGB1P39, UBE2L3, FAM175B, TMEM79, ZMIZ1, TAB2, QKI, GPN2, TNPO1, SYT14, RGS17, ZNF608, RP11-191L9.4, ERMP1, ZKSCAN4, TMEM168, NINL, C18orf32, MARK4, DNAJC16, PHC1, ATXN2L, MINK1, MIR29A, SIPA1L3, PA2G4P4, ZNF287, YWHAZP2, RP11-737O24.3, HMGB1P9, IGF2R, LNP1, LINC00909, RBM4, FBXO46 | Diurnally Regulated Genes with Circadian Orthologs_Homo sapiens_WP410 (WikiPathways_2016 q=0.036275) |
| **82** | 45 | PIGC, ANKRD39, RP11-196G11.5, ABHD10, SLAMF9, TMTC4, BRPF1, RPTOR, RP11-152N13.16, SNAPIN, DNAJC9-AS1, ZNF91, PSMG1, PQLC2, RABEPK, CYB5R4, ARL13B, ZNF324, TPM3P9, EOGT, TMEM178B, TPK1, SERAC1, GRK2, UBAC2, KIAA1644, MLXIP, AKAP6, GPR108, ATRIP, ZNF829, ZNF112, PLA2G4A, ILKAP, DMWD, SLC25A18, RP11-54H7.4, TBC1D10B, DNAJB2, ZNF26, RAB14, EDC3, TRNT1, SRRT, IGHMBP2 | |
| **83** | 42 | PWWP2A, FNIP1, ZNF518A, CLHC1, HILPDA, COQ4, SETD1A, FLNA, SAMD1, MB21D1, COLGALT1, DDAH1, TRIM24, CTD-2366F13.1, EEPD1, AGK, ZNF117, SUMO4, ERCC-00022, RP3-399J4.2, SLC35E2, HMGB1P44, DPM2, TMEM161B, RP11-1281K21.1, PNPLA3, C3orf62, DBF4B, CLPTM1, RABL6, WAC-AS1, ARHGAP10, KCNC4, ZNF254, CTD-3145H4.1, RP11-66N5.2, DANCR, PIP4K2A, MOB1B, PABPC3, LRRFIP1P1, SNX6 | |
| **84** | 17 | ARAP3, HOOK2, RNF145, FAM117A, CRLS1, TMSB15A, BEX1, USP37, RP11-83A24.2, CTD-2515C13.1, HBP1, TDRD7, CCDC138, DNHD1, ERICH1, ZFAND2B, RAB27B | |
| **85** | 16 | TMEM181, NOVA1, FAM57A, SEMA3B, GANC, MTCO3P22, MXRA8, NXN, ZNF616, RP3-394A18.1, TMEM173, DUSP22, ADGRL2, PGM3, AHCYL2, NLRP11 | |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **86** | 20 | SDHB, PLEC, PCGF3, MYDGF, FAM185A, ATMIN, FADS2, ADA, RTKN, ERF, BCAR1, ZNF335, MAP3K6, KRT16, ST13P12, PHF11, ZADH2, ARFGAP3, TRA2A, RPL7P32 | |
| **87** | 11 | S100A6, KHDC1L, SEC22A, NEMP1, ZSCAN20, ZNF19, LZTR1, KIF21B, GCLC, RP11-196G18.22, METTL12 | |
| **88** | 32 | TMEM234, GUSBP1, SIRT7, FAM32A, NDUFV3, ARHGAP44, PDSS2, ENTPD5, SLCO3A1, ASB7, ZNF133, ZNF248, MAPK8IP1, MAP2K7, GJB3, LL0XNC01-237H1.2, NBPF3, TATDN1P1, MAP3K3, RP11-69E11.8, DNASE1, RPL21P4, CHI3L2, CTD-2017F17.2, PARVB, ZNF79, TTC31, SLC2A8, LDLRAD4, bP-2189O9.2, ISG20L2, KIAA0391 | MAPK Cascade_Homo sapiens_WP422 (WikiPathways_2016 q=0.017642) |
| **89** | 10 | C1orf52, ABHD15, ZNF70, VWA5A, FHDC1, CCDC102B, RCL1, METTL21B, SHTN1, RP11-818F20.5 | |
| **90** | 30 | LINC00888, FBXO25, RP11-30J20.1, ZNF318, COQ10A, IFI6, SESN3, ORAI3, HNRNPLL, FECH, WBP1L, MYL9, LTA4H, SEMA4F, HIGD1A, STK16, TMOD2, ABCC2, LINC-PINT, SNTB1, LL22NC03-86G7.1, HPCAL1, ROBO3, RWDD2A, RP1-140K8.5, RP11-356J5.12, MAGI2, STX3, ANGEL1, UBE2V1 | |
| **91** | 10 | DCAF6, PPARG, NAB1, HAUS1, UXS1, ALG8, ANKH, BACH2, SLC24A1, UNC13A | |
| **92** | 39 | RP4-773N10.4, EXOSC4, ABCB7, RP11-2J18.1, TCTN2, SEPSECS, VWA8, PPFIBP2, CHST1, MFSD6, ABCG2, SLC15A4, DZANK1, GNG5, CBX1P2, CCDC102A, ZNF561, SNAP47, MRPL41, LYRM5, TRANK1, LMBRD2, POLH, MYO9B, BCAS3, SWT1, ZNF671, ELF2, SETP4, NHSL1, C9orf91, KDM7A, SMOX, RNASEH2B, CCNG2, TULP3, TMEM267, PXK, N4BP2 | |
| **93** | 12 | C5orf34, C12orf76, SSR1, SCYL2, SAP30, CHST14, FNTA, ZNF605, AUTS2, METTL4, FAM229B, SPPL2A | |
| **94** | 10 | KIAA1715, ZBTB10, TEX30, TDP2, CTDSP2, CHTOP, RABIF, TOP1MT, TCEA1, FAM110B | |
| **95** | 45 | LURAP1, USP30, ELL2, ZNF569, KLF11, FAM160A1, RFX5, SETD4, PTMS, RP11-467J12.4, ZSCAN31, ZNF528, PPP2R2B, SPANXN5, RP11- | |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | 295G20.2, SPCS1, ARMC9, MTND2P21, LETM1, RP4-756H11.5, CASD1, FRAS1, ULK3, C2orf69, LINC00526, DPYD, SEMA3E, ZNF519, IRF1, MMAB, AGPAT4, ELF1, ST3GAL6, GPR137C, RPL32P3, ALG11, RP11-655M14.13, PGM2L1, SPTB, ERCC-00131, LGI2, LINC00511, HNRNPUP1, CFAP46, NHS | |
| **96** | 10 | IKBKE, GXYLT1, IMMP1L, RASGRP3, SYTL5, C1orf226, PPP1R14C, NEK11, KANK1, CDS2 | |
| **97** | 11 | CRYBG3, HAUS6, DOCK1, PABPC4L, HAUS6P1, RNF169, RP11-288C17.1, AC073109.2, PTN, STAG1, DIAPH2 | Regulation of actin cytoskeleton_Homo sapiens_hsa04810 (KEGG_2016 q=0.047052) |
| **98** | 47 | RIMKLA, USP51, ZNF362, METTL18, BDH2, CLGN, IDNK, AC083873.4, RANBP3, CENPH, BIVM, COQ7, FAM204A, DIP2A, PALB2, AGAP6, RILPL2, ZNF790-AS1, PDK1, ZBTB33, ELMOD3, AC019129.1, ZNF543, PPM1J, RP11-278C7.1, KIAA0513, ANKRA2, PCED1A, GNPTAB, ZNF276, EDARADD, ZNF8, CTC-505O3.1, RPL21P75, PSMG4, ABCC4, TSNAX, TOB2P1, RP1-308E4.1, BCAS4, NFATC2IP, MAP3K7, DHTKD1, RBM48, ORC5, BRD7P2, SMCR8 | |
| **99** | 82 | OSBP2, CA13, ACCS, BROX, CRABP2, CMB9-55A18.1, STX7, SGCE, TBC1D15, ZNF701, DUSP12, TCTA, DNPH1, PINX1, XK, FANCB, CDO1, RP11-973H7.4, RPL10AP6, PKD2, NRDE2, INTS6L, RASSF3, SUCLG1, CDC42EP1, MARC2, S1PR3, CTPS2, DOCK11, DTYMK, C16orf74, PCBD2, CHTF18, IP6K2, PPP1R9B, CRYZL1, RPL5P1, KCNG1, IQCC, OPA3, HES1, AADAC, ZNF582, RP11-94I2.4, SYNGR1, NCK1, C16orf13, RYR1, MIS12, C7orf26, EHD1, NUS1, PRDX1, PCMTD2, FXYD5, LGMN, CTD-2286N8.2, CECR5, L3MBTL3, FAHD2B, TUBA3C, MALSU1, SLC4A11, MCAT, MIR34A, DPP9, ITM2B, CNIH3, HECA, RORB, ARHGEF17, MYO3A, PTOV1-AS1, C2orf47, BRD1, PMPCA, SMIM8, C16orf87, LEPR, TMC7, FAM214B, KCTD12 | |
| **100** | 69 | RGP1, OGDHL, TXNRD2, PITRM1, BCL6, TGFB3, SCO1, APP, ADCYAP1R1, RHOU, OSBPL1A, EDA, NAP1L5, RP11-347H15.5, PPP1R12B, RP11-803D5.1, CHAMP1, DCAF12L1, NMU, ANKRD6, SCARNA2, LDB1, PMS2, PMS2CL, NPAS2, | |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | PNMA2, RP11-30L15.4, HOXC10, C1orf112, PBX3, EMC3-AS1, MLK7-AS1, KIF5A, ZNF202, EBPL, FBXO42, IFT88, PAQR7, RP11-545M17.2, GUSBP11, CTB-89H12.4, DLL3, FAM81A, TMEM201, TADA2B, CISH, ALG10B, ABHD18, ZNF641, TKFC, EFEMP2, ADGRE5, VPS4A, GAS7, CBX2, PCYOX1L, VMA21, IFT20, CYB561D2, ZNF521, PLCG2, FGFBP3, ARSG, MRPS22, LTB4R, PER2, MYB, RAPH1, SH3BGRL2 | |
| **101** | 13 | PLK4, ZNF234, EPDR1, EMP2, USP31, FUT11, CHAF1B, ZNF200, USP32, TAF7, AC005154.6, TTC28-AS1, SGSM3 | |
| **102** | 39 | PIK3R3, HMGCR, ZNF461, TSHZ2, PLEKHG1, U2AF1L4, SFXN5, PACSIN3, ARMT1, KIF3C, MBNL2, PAPOLG, NCOA2, TTC9, CDK5R1, RP11-643G16.4, EPHB2, POU2F2, GSTA4, RP11-15A1.8, IFT80, SLC35B3, RP5-1136G13.2, RHEBL1, RP5-1065J22.8, DLG4, CNPY4, CDAN1, TPRN, RPL17, FMNL2, TMEM107, ROR1, IVNS1ABP, TMEM135, HOXB-AS4, PLEKHG2, CLK4, ORAI1 | |
| **103** | 11 | DENND2C, TAOK3, TUBA1B, RP11-983G14.1, ERCC2, ZNF841, DEPDC5, EMB, RP11-128M1.1, DHRS4L2, LITAF | |
| **104** | 11 | TTBK2, RP11-367G18.2, CDHR3, ANKRD52, GATA4, ERBB3, KIF5C, LPAR1, RP13-88F20.1, CNOT6L, BRINP1 | Heart Development_Homo sapiens_WP1591 (WikiPathways_2016 q=0.003520) |
| **105** | 25 | QRFPR, GNG11, RUBCN, RP11-290D2.5, ZCCHC10, ACADSB, COX7B, FAM120AOS, C16orf91, NFYC, LRRC57, UHRF2, ARID3B, RGS16, PPM1L, ABCA5, H3F3AP5, NUDT13, PPP1R3E, FLVCR1-AS1, RP11-252K23.2, RPS20P14, STARD7-AS1, PAN3, FAM127B | Calcium Regulation in the Cardiac Cell_Homo sapiens_WP536 (WikiPathways_2016 q=0.048372) |
| **106** | 14 | NECAB1, HCFC1, AC092835.2, RPL3P7, LRP8, TRERF1, SYCE1, ACSS3, MADD, PORCN, ANKRD42, CRAMP1, DUSP2, TTC19 | |
| **107** | 14 | MARK1, STIM1, FAT4, FRG1DP, TCHH, TRUB1, DTWD2, LRRC1, SPIN4, SORT1, EDIL3, SEC14L1P1, HOXD8, AFP | |
| **108** | 14 | ATG4C, TRPS1, CUL5, DDX19B, RRP36, DCTN5, PCDHGC3, VPS26B, SLC30A5, BRI3BP, IRF8, MAF1, SYT2, ZNF235 | |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **109** | 12 | CDKL3, RP11-817I4.1, BCKDK, MTRNR2L10, C14orf159, MTND1P23, PAXIP1-AS1, RPL23AP82, IFIH1, NIPAL4, GORASP1, DYM | IRF8_22096565_ChIP-ChIP_GC-B_Human (ChEA_2016 q=0.048306) |
| **110** | 12 | MPP4, TRMT10C, LIG1, ARHGAP20, ZNF256, TYW1, ANKRD36B, C6orf120, ZNF2, ZFP62, CETN3, THYN1 | tRNA processing_Homo sapiens_R-HSA-72306 (Reactome_2016 q=0.034695) |
| **111** | 32 | MSL3P1, IFT122, RP11-37B2.1, TBC1D8B, GPR158, DPF2, SNIP1, SNX22, ENPP1, SERGEF, LINC00339, ASNSP1, TOM1, UBL7, ZNF225, RP11-436D23.1, FRS2, C3orf38, CTBP1-AS2, MORN2, ZNF696, ZNF484, ALDH9A1, AP000580.1, OTUD5, SORBS2, WHAMMP3, RP11-755J8.1, LINC01278, ZNF552, MED7, ENPP4 | |
| **112** | 66 | SVBP, BCDIN3D, RP11-262M14.2, MBTD1, FBXO48, SUSD5, LRRC37A16P, SRP72P1, NPM1P39, KCTD16, NGF, TSTD2, C18orf54, GS1-309P15.4, MAP2K3, ID1, BTRC, LRRC8C, TBC1D24, RP11-324I22.4, EGLN3, METTL6, CNN2, CSPP1, NPIPB4, NPHP3, KANSL2, TRAF5, TMEM161B-AS1, ZNF510, PRR13, SLC25A36, ARNT2, GPRC5B, TRMT61B, LYPLAL1, ACOX3, LRCH1, ANK2, PCDH17, PTPRJ, SH3GL1, TMEM187, UBE2E2, DYNC1I2P1, NUFIP2, DNM2, KDM3B, BMP1, MECP2, HPSE, PALD1, DCLK2, RNF126, ZDHHC11, MBLAC2, FAM84B, CH507-9B2.5, KLHDC2, LRRC49, DHX36, RUFY2, FAM184A, DHX38, ZC2HC1A, SH3GL3 | |
| **113** | 13 | NTNG1, ABL1, CHD6, ABI2, TUBGCP2, EPHA4, BMP6, RP11-956J14.2, MTA1, RFFL, C1orf143, DKK2, MKX | RHO GTPases Activate WASPs and WAVEs_Homo sapiens_R-HSA-5663213 (Reactome_2016 q=0.015288), Axon guidance_Homo sapiens_hsa04360 (KEGG_2016 q=0.001161), TP53_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.032671) |
| **114** | 10 | TRIM38, KALRN, DDX58, PTPN9, ANK3, ANKRD30B, PYM1, DNAJC15, ABCB10, LGALS8 | IRF8_22096565_ChIP-ChIP_GC-B_Human (ChEA_2016 q=0.037540), IRF8_CHEA (ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X q=0.042728) |
| **115** | 26 | ZFHX3, SLC9A7, AADAT, HIC2, RAPGEF1, KIZ, SPEF2, ELL, C15orf61, MRPL28, CNNM3, C7orf73, NOP14, ARRDC4, PDSS1, SLC16A9, | |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | PARPBP, C8orf37, PIANP, RLF, RAB12, SF3B4, TRIP10, ZNF766, BBOF1, TIAF1 | |
| **116** | 14 | ZC3H4, RPL7P20, GNB1L, MESDC1, SP6, FAT3, GOLGA8VP, LAMA3, VAV2, SLC9A8, SSNA1, PPP1R36, DBF4P1, TMEM87A | |
| **117** | 86 | FUK, MED27, ATAD3A, L1CAM, CD58, KANSL1L, IFIT3, ANO6, ATP8A1, TRIM41, HECW2, COL27A1, ZNF331, C18orf25, SDPR, CHMP2B, AC097374.2, AC068138.1, FRMPD4, ANKRD20A5P, KIAA1841, ABCA13, STX5, POLR3C, CHCHD6, CYSTM1, PLAG1, SLC35D2, IPO8, SLC41A1, ZNF165, PDP2, MTX3, MURC, HUS1, BECN1, NKX3-1, ZNF597, NENF, KCNMB3, PQLC3, TMEM64, KIAA0895L, FAM110D, ZNF576, KRT8, C17orf62, HSPE1P2, EPS8L1, GPCPD1, RP5-1085F17.3, DLG2, CHMP4A, NAXD, RBM15B, TMEM141, PVR, TMEM209, CTD-2550O8.7, KCNAB1, PCDH9, HIRA, LIPA, ERCC-00003, RP11-120B7.1, FGD1, ZNF860, NAT9, MMAA, DGKK, RP4-575N6.2, IFITM2, CHCHD5, WIZ, Y_RNA, B3GAT3, ASB9, MROH1, SND1-IT1, CACNB2, SUZ12P1, BBS10, RPL13AP2, XXyac-YM21GA2.4, SLC35C2, LIF | |
| **118** | 19 | ARNT, MCCC1, C19orf12, MFSD4B, RP11-305B6.1, ACAN, CTC1, LDLR, ST13P6, C15orf40, NELL2, NAT1, ARHGAP28, CASP9, CTSD, RP11-284F21.10, SMARCE1P1, SECISBP2L, RBM41 | Degradation of the extracellular matrix_Homo sapiens_R-HSA-1474228 (Reactome_2016 q=0.048975) |
| **119** | 10 | PNKP, EGF, RPS6KA2, DNAJC19, NR2C1, RCAN3, C1orf123, TM2D1, C2orf42, RP11-87H9.4 | EGFR1 Signaling Pathway_Mus musculus_WP572 (WikiPathways_2016 q=0.045332) |
| **120** | 11 | DCP2, DHX8, ARHGAP19, TRIM52, RNF6, LYPLA2, HELB, RELA, CMTM3, KBTBD8, ZDHHC7 | Androgen receptor signaling pathway_Homo sapiens_WP138 (WikiPathways_2016 q=0.045116) |
| **121** | 20 | PDE5A, KLF17, CTD-2510F5.4, C1RL, SP2-AS1, NOM1, ERI2, RP11-479G22.8, SOX6, TMEM53, PLEKHH1, MEX3A, SLCO2A1, EXPH5, IGF2BP1, TGFA, SPAG1, ZNF568, PBLD, ZNF850 | |
| **122** | 15 | RBFOX3, FST, DHRS4, SETP2, RPL9P9, MYBL1, FGD4, SATB1, ERCC-00112, PLEKHH2, SCN4B, PFAS, AC005307.5, TBC1D8, ZCCHC2 | |
| **123** | 15 | CCDC28B, DGKE, NUDT12, HIST1H1E, ARL15, AC004381.6, NCKIPSD, RASSF4, ALDH1B1, SPATA20, GPAT4, BRAP, BCAS2, TPCN1, MYO1E | Glycerolipid metabolism_Homo sapiens_hsa00561 (KEGG_2016 q=0.018405) |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **124** | 30 | MAK16, IL1RAPL1, PLA2G7, SCAF4, DPY19L4, NXT2, SIN3B, ZNF718, FAM221A, CDKN2AIPNL, NR2C2AP, EDC4, DNAJC25, ZNF708, RGS7, FAM92A1, OGG1, IL15, ALAD, PRAF2, ARL10, CH17-360D5.3, AC098614.2, TBC1D2B, ZNF595, ARSK, LRRC37A17P, RP11-760D2.7, DNALI1, POLM | |
| **125** | 26 | PDLIM4, ANOS1, P2RY1, EXD2, ZNF525, H6PD, CTC-470C15.1, ZSCAN12P1, SETDB2, VGF, PABPC1L, MDM4, ANGEL2, BLZF1, RPS4XP6, PMFBP1, GOLGA8A, ADAP2, PTCHD4, SAC3D1, SNHG7, MOB3C, RUSC2, KLHL21, OSMR, SUGP2 | |
| **126** | 47 | MSX2, DCP1B, MTRNR2L5, RPL41, RBMS1, METTL25, DDX59, GID4, AC079922.3, CTA-204B4.2, KB-1732A1.1, RASGEF1A, TSGA10, TAF4B, MRPS21, GAK, HSP90AB4P, POLG, TMEM44-AS1, KIAA1328, RP3-522D1.1, RP11-111M22.3, KCTD21-AS1, THAP6, MLLT1, PANK1, PIK3AP1, SLC25A24, CIB2, RPS6KA6, RPL36AP26, TMEM25, HOTAIRM1, KCNK6, RPS4XP16, ERCC-00092, DCLK3, ERO1B, IQCD, YIPF1, RAC1P5, RP11-889L3.1, TMEM44, TCEA2, STX6, ZNF639, ZNF594 | |
| **127** | 25 | BRE, MAP2K5, RP11-284F21.9, ERAL1, ASAH2B, TRIM45, ATP6V1F, WFS1, SLC44A2, AAED1, APOLD1, ZEB1-AS1, CNOT6, FKTN, OLMALINC, ZNF285, FAM134C, FAM43B, C8orf48, ANKHD1-EIF4EBP3, RHBDF2, RP5-1198O20.4, FLAD1, RP11-422P24.10, SIL1 | |
| **128** | 13 | CHST12, ERCC5, RNF44, RPS4XP17, PML, ST6GALNAC2, RPS6KC1, RP11-298I3.4, ADORA2B, ING4, SPIN3, ABRACL, B3GALT5 | Globo Sphingolipid Metabolism_Homo sapiens_WP1424 (WikiPathways_2016 q=0.001610) |
| **129** | 27 | AC007238.1, ERAP1, KPNA5, SPIRE2, FAM167A, ZNF136, MYLK-AS1, NAP1L1P1, ZFYVE27, ZNF790, RP4-612B15.2, PIP4K2C, CTC-459F4.3, PHF5A, RP11-10C24.1, RAB20, MIER3, LIMK1, FKRP, STX18, RPL7P23, UPK3BL, FMN1, SLC25A43, SNRPA, GAS6, SLC7A6 | mRNA Processing_Homo sapiens_WP411 (WikiPathways_2016 q=0.034694) |
| **130** | 14 | ZNF678, NUP160, UBE3C, MRRF, TMED7, XYLT2, TFB2M, MEN1, RP11-379H18.1, FAM122A, AC008850.3, WLS, HOXA11, HOXA13 | Transcriptional misregulation in cancer_Homo sapiens_hsa05202 (KEGG_2016 q=0.033350) |
| **131** | 21 | FBXL2, UBE2D2, C6orf1, C20orf27, NBPF11, BLOC1S4, RALGDS, TIMM23B, DMAP1, | |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| | | LINC00115, RP5-855D21.3, TESK2, GPM6A, WDTC1, PRRT3, PICSAR, ZFAT, THAP10, ZNF182, FKBP7, ZNF587 | |
| **132** | 10 | RP11-551G24.2, PDCD5P2, VPS36, SRRM1P3, CDS1, AOC2, AGGF1, C11orf54, IFT172, VPS50 | |
| **133** | 26 | ARFGEF3, LATS1, NGRN, CTC-444N24.11, ZNF107, TCHP, TNPO2, ATP5EP2, ATP5E, MAP2K2, AKAP5, CA2, TUSC1, RCE1, RP11-84C10.4, ATP5D, GPER1, RNF26, C11orf31, PGBD5, PDE9A, KCNIP3, MAPK3, EMD, SLC45A4, THSD1 | Formation of ATP by chemiosmotic coupling_Homo sapiens_R-HSA-163210 (Reactome_2016 q=0.018223), IL-7 Signaling Pathway_Homo sapiens_WP205 (WikiPathways_2016 q=0.006972), Estrogen signaling pathway_Homo sapiens_hsa04915 (KEGG_2016 q=0.021663) |
| **134** | 28 | C8orf44, FAM76B, FLVCR2, GATA6, AMOTL2, PAIP1, USP6NL, LDOC1L, RABGAP1, SLC38A7, TSTD1, SP2, SNHG3, TRMT13, LINC00116, STAT5A, C19orf81, PPP3CB-AS1, ELFN2, PLCB4, FAXC, UTP4, CTC-351M12.1, FLCN, ICA1L, MORN4, FKBP1B, SPTLC3 | |
| **135** | 13 | ZNF620, TUBA1C, ZNF195, HMGB2P1, RNF144A-AS1, AC079922.2, FSIP1, NPM1P6, ZNF778, CACHD1, KBTBD3, RPL5P12, EIF3G | |
| **136** | 12 | RP11-15A1.3, MACROD2, TMEM154, C6orf203, LINC00941, PROS1, KIAA0753, HMGB1P31, FYCO1, TAF5, ADAMTS5, SPANXN3 | |
| **137** | 27 | MTF2, NR6A1, PHF7, EFNA4, POR, RPL36AP21, ATP5L2, KLC4, CBX7, RPS13P2, CTC-518P12.6, AC009403.2, RPS3AP47, DRAM2, SLC25A51, PCDH18, TUBB2B, TALDO1, RPSAP75, RP11-178H8.7, ADRA2C, PRKAG2, AMFR, TMEM231, AIFM1, RAB9A, LINC00630 | Integration of energy metabolism_Homo sapiens_R-HSA-163685 (Reactome_2016 q=0.033976) |
| **138** | 11 | ZFP41, NFYB, CASC4, CFLAR-AS1, LINC00476, RP11-401O9.3, TTC39C, DPM3, TOLLIP, LRRC40, ZNF619 | |
| **139** | 17 | EFCAB11, ZFAND2A, IKZF4, RPL13AP3, EEF1B2P3, GAS2L1, CLCC1, RGS5, SEPT7P7, PTPN21, USP45, CHUK, RP11-368M16.5, COPG2, RP1-159A19.3, RPL4P4, XKR9 | |
| **140** | 13 | AGFG2, PIGB, TSPAN2, SHISA9, RP11-500C11.3, KSR2, ATP6V0A4, CNN1, HSPBP1, RP11-283I3.6, RP11-51O6.1, ABCA4, HOXC6 | |
| **141** | 10 | PIK3CA, DUBR, ZC3H10, TBX18, NKD1, SH3RF2, BNC2, RP11-54D18.4, CSMD2, RP11-688I9.4 | |

| Cluster ID | Number of Genes | Gene Names | Enrichment (Pathway or transcription factor, database, and false-discovery-corrected q-value) |
|---|---|---|---|
| **142** | 10 | SNX33, RP11-163E9.2, VSIG2, RP11-1094M14.11, FKBP10, MTCO1P22, FAM222B, ZNF333, FCHO2, MYO18B | |
| **143** | 19 | RAD23B, ERCC6, SH3TC2, DACT1, SLIT3, LINC00689, CAPRIN2, ATP6V1H, GSN, AF230666.2, NSUN5P1, TSC22D1, FAM76A, LINC00470, RNF150, DIP2C, ECI1, C1orf50, MSTO1 | Nucleotide excision repair_Homo sapiens_hsa03420 (KEGG_2016 q=0.014417) |
| **144** | 15 | DEPDC4, MIATNB, TMOD1, APOBEC3B, NLRP3, AC002456.2, RAB15, ECE2, B4GALNT1, CORO2B, IL11, SIRT3, SULF2, HACE1, RP11-16K12.1 | |
| **145** | 11 | DFFA, C8orf33, TAMM41, RP1-178F15.4, TMEM55A, ZSWIM1, LRRC71, TRAPPC1, TOX4, ZNF75A, NAA60 | |
| **146** | 14 | ZNF394, SRC, RP11-282K24.3, CDK17, PFDN4, FSD1L, XIRP2, ADAMTS15, TP73, MTMR7, ADGRL4, SYDE2, FOXP1, CREM | Alpha6-Beta4 Integrin Signaling Pathway_Mus musculus_WP488 (WikiPathways_2016 q=0.047427) |
| **147** | 10 | ELOVL1, USP38, ZSCAN25, CTDSPL2, ZNF746, GLUD2, RP6-65G23.3, RP11-674N23.4, FDPS, KIAA0922 | |
| **148** | 11 | AMIGO1, TSPAN19, ABHD13, F3, PIGQ, DCUN1D3, PRDM1, AC010761.8, MAPKBP1, ZNF721, YWHABP2 | |