Transcriptome-wide alterations in mRNA translation define breast cancer subtypes

Inci S. Aksoylu[1], Hermano Martins Bellato[2], Shuo Liang[1], Fernanda C. S. Lupinacci[2], Laia Masvidal-Sanz[1], Christian Oertlin[1], Glaucia N.M. Hajj[2, †] and Ola Larsson[1, †, §]

[1]Department of Oncology-Pathology, Science for Life Laboratory, Karolinska Institute, Stockholm 171 77, Sweden; [2] Department of Pathology, A.C. Camargo Cancer Center, São Paulo, Brazil

† Co-corresponding senior authors: Glaucia N.M. Hajj (ghajj@accamargo.org.br) and Ola Larsson (ola.larsson@ki.se)

§ **Lead contact**: Ola Larsson ; Department of Oncology-Pathology, Science for Life Laboratory, Karolinska Institute, Stockholm 171 77 ; Email: ola.larsson@ki.se

24

Abstract

Breast cancer, the most prevalent type of cancer in women, is characterized by multiple underlying molecular mechanisms. Tumor classification, including histological characteristics, receptor expression as well as expression of mRNA signatures, provide estimates of disease progression and guide treatment regimen. However, the role of post-transcriptional mechanisms, including mRNA translation, in establishing tumour proteomes, is unknown. Although there is ample support for transcript-selective alterations in mRNA translation as underlying cancer-associated phenotypes, due to technical challenges, their relevance in patients is unclear. Here, we employ polysome-profiling optimized for small samples to quantify transcriptome-wide patterns of mRNA translation in two cohorts of breast cancer patients. The first cohort revealed wide-spread variation in mRNA translation across patients and suggested six subsets of subtypes of breast cancer, independent of cell-composition, that were also validated in the second patient cohort. Mechanistically, altered translation in breast cancer subtypes is driven by interactions between translation programs downstream of translation initiation factors including eIF4E, eIF4A and DAP5, as well as key signalling pathways including mTOR, the integrated stress response (ISR) and extracellular matrix. Moreover, a plethora of cis-regulatory elements in both coding and non-coding sequences of targeted mRNAs associate with selective regulation of translation in each breast cancer translation-subtype. Therefore, transcript-selective modulation of mRNA translation is key characteristics of breast cancer that may be amendable for tailored treatments.

Introduction

Breast cancer, the most prevalent malignancy in women, affects more than 2 million individuals each year(1) . Treatment of breast cancer is tailored depending on receptor status but also activities of gene expression signatures, including the PAM50 classification(2). The PAM50 classification has contributed to the understanding of breast cancer heterogeneity by identifying five instrinsic subtypes, differing in incidence, disease progression, response to treatment and survival (3–6). Despite ample therapeutic options for breast cancer, a significant proportion of patients either do not respond to the therapies or develop resistance and therefore relapse (7).

mRNA translation, a fundamental component of the gene expression program, allows organisms to, in response to intra- and extracellular stimuli, rapidly adapt their proteome by selectively controlling translation of mRNA transcribed from subsets of genes (8). This dynamic process is tightly regulated e.g. at the initiation step, which is commonly considered rate-limiting, by a repertoire of trans-acting factors, including translation initiation factors, and cis-acting mRNA features such as secondary structures and RNA-elements(9,10). Together, cis- and trans-acting factors underlie transcript-selective regulation of mRNA translation depending on intracellular pathways and environmental alterations. For instance, a significant proportion of mRNA molecules supporting malignant phenotypes possess structured 5’ untranslated regions (5'UTRs) and therefore have a high requirement of the eIF4F-complex for initiation of mRNA translation(11). Initiation is followed by elongation and termination, where elongation is influenced by

features in the coding sequence (CDS) of mRNA molecules such as codon composition that interact with the composition of the cellular tRNA pool (9,10,12–14).

Previous studies show that oncogenic pathways influence various mRNA translation steps to alter the translational efficiency of mRNAs transcribed from specific genes(10,15–20). For instance, the PI3K/AKT/mTOR pathway, which is often activated in malignancies, stimulates formation of the eIF4F translation initiation complex, which then bolsters cap-dependent translation. (21). Further supporting mRNA translation as a node downstream of cancer-associated pathways, transcription of genes encoding components of the eIF4F complex is activated by the MYC oncogene (19). In addition to their influence on translation initiation, oncogenic pathways have also been shown to affect elongation, for example by regulating the activity of elongation factors or tRNA modifications in cancer cells(23–26). Consistently, evidence from mouse and *in vitro* studies suggest a role of dysregulated mRNA translation in cancer (16,27–30).

Here we describe the, to our knowledge, first transcriptome-wide profiling of mRNA translation in human cancer tissues and reveal that transcript-selective alterations in mRNA translation is a feature of breast cancer. We further associate these mRNA translation-changes to cis-acting factors in 5'UTR and CDS as well as key upstream signalling events emanating from intracellular and extracellular queues. These alterations in mRNA translation underlie subtypes of breast cancer distinct from the PAM50 subtypes. Overall, these findings suggest a possibility for treatment of breast cancer via targeting mRNA translation in a precision medicine-fashion.

93

94

95

Methods

Polysome profiling of breast cancer samples

Two cohorts of breast cancer samples (n=165, 135 in Cohort I and 30 in cohort II) were obtained from the A.C. Camargo Cancer Center biobank (Sao Paulo, Brazil) with informed consent obtained from the donors in accordance with ethical guidelines (A.C.Camargo Cancer Center ref.1844/13 for sample collection and EPN ref.2018/1055-31 for data analysis). Polysome fractionation of was performed using the optimized sucrose-gradient method and single-end libraries were prepared following the Smart-Seq2 protocol as outlined (32). Sequencing was performed on the Illumina HiSeq2500 platform, generating 50 base single-end reads from prepared libraries.

Preprocessing and quality control of RNA-sequencing data sets

The quality of single-end RNA sequencing reads was assessed using FastQC v0.11.4. Next, RNA sequencing reads that map to Nextera and Truseq adapter as well as ribosomal RNA sequences were trimmed using BBMap (v.36.69; parameters: k=13, ktrim=n, useshortkmers=t, mink=5, qtrim=t, trimq=10, minlength=25 and rRNA reference sequences obtained from SILVA ribosomal RNA gene database). Post-trimming, the reads were aligned to human reference genome (build hg38) using HISAT2 (v.2.1.0) with default parameters for single-end RNA sequencing. The “featureCounts” function in Rsubread (v.2.6.4; parameters: isPairedEnd = allowMultiOverlap = countMultiMappingReads=FALSE) R/Bioconductor package was used to summarize aligned reads for each sample and gene. For each cohort, samples with fewer than 2.5 million reads and genes with 0 mapped RNA sequencing read in one or more samples

were excluded. Resulting count matrices were TMM-Log2 normalized by "calcNormFactors" and "voom" functions of edgeR (v.3.34.1) and limma (v.3.48.3) R/Bioconductor packages, respectively. Finally, principal component analysis (PCA) was performed using the "pca" function of PCAtools (v.2.4.0; https://github.com/kevinblighe/PCAtools; parameters: removeVar = 0.75 and scale = T) on normalized counts and the outlier samples were removed from downstream analyses. PCA plots were generated using the "biplot" function from the PCAtools package.

PAM50 subtyping of breast cancer samples

Molecular subtypes of breast cancer samples were initially determined by assessing the receptor status. Then, using RNA sequencing gene expression data for genes included in PAM50 panel, we computationally determined the PAM50 subtypes using "molecular.subytping" function from genefu (v.2.24.2) R/Bioconductor package (parameters : sbt.model="pam50", data = normalized counts for either total or polysomal mRNA, annot= annotation for each sample) (33).

Modelling of transcriptome-wide alterations in mRNA translation

Linear regression models were employed to determine the relationship between total cytosolic and polysome-associated mRNA levels. Potential influence from confounding factors, including cell composition of breast cancer samples or clinical covariates, were adjusted for. Cell composition of breast cancer samples were determined using CIBERSORTx in "absolute mode" where cell types included in the built-in LM22 signature

matrix and single cell signatures derived from human cell breast cancers (34,35). Univariate and multivariate regression models were then used to identify the cell types and clinical covariates that significantly associate with alterations in polysome-associated mRNA levels. Per gene and per patient residuals and slopes were collected from a final multivariate linear regression model that is corrected for library sizes, cell types and clinical covariates as follows:

$$Ex_{Polysomal} = \gamma^{1\ldots j}_{Lib.¿oly.} + \gamma^{1\ldots j}_{Lib.¿otal} + \gamma^{1\ldots j}_{Clin.Cov.} + \gamma^{1\ldots j}_{Cell\,Abd.} + Ex_{Total} + \varepsilon$$

Where $Ex_{Polysomal}$ polysome associated mRNA levels for a gene across all samples that are corrected for polysome-associated mRNA library size ($\gamma_{Lib.Size.Poly.}$), total cytosolic mRNA library size ($\gamma_{Lib.Size.Total}$), clinical covariates ($\gamma_{Clin.Cov.}$) and cell types ($\gamma_{Cell\,Abd.}$). Ex_{Total} is the linear dependence of polysome-associated mRNA levels to total mRNA levels and ε is the error. Following this, using samples from the previous model we obtain residuals for a second cohort of patients using the following model per gene and per patient in which each patient is added to the model one at a time in each iteration:

$$\varepsilon_{Coh1} = \gamma^{1\ldots k}_{Lib.¿oly.} + \gamma^{1\ldots k}_{Lib.¿otal} + \gamma^{1\ldots k}_{Clin.Cov.} + \gamma^{1\ldots k}_{Cell\,Abd.} + Ex_{Total} + \varepsilon$$

Sample classification using per-gene and across-patient translational efficiencies

To classify patients based on translation, residuals obtained from the multivatiate linear regression models were used as input in the cola (v.2.10.0) R/Bioconductor package (36) . Initial analysis using the “run_all_consensus_partition_methods” function with kmax=10 revealed that k=6 with ATC and mclust combinations yielded the most optimal clustering solution as evidenced by mean silhouette, concordance, Jaccard index and 1-PAC values, plotted by “select_partition_number” function (**Fig.S2**). Following this, clustering was

performed using the “consensus_partition” function (set.seed(1233) and parameters: top_value_metho“= ""TC",partition_meth”d="mclust,max_k = 10 ,partition_repeat = 100). A heatmap of the clustering was plotted using the “get_signatures” function (parameters: k=6, silhouette=0.75, group_diff = log2(1.5), fdr_cutoff=0.01) of cola (v.2.19.0) R/Bioconductor package. Enrichment analysis of resulting gene signatures from this analysis was performed using “enrichGO” function (parameters: orgDb= org.Hs.eg.db, keyType: “ENSEMBL”, ont=”BP”, pAdjustMethod=”BH”, pvalueCutoff=0.01, qvalueCutoff=0.01) and results were plotted using the “barplot” function (parameters: showCategory=5, color=”p.adjust”) from clusterProfiler (v.4.0.5.) R/Bioconductor package for top 5 terms for each gene signature cluster(37).

Measuring transcriptome-wide translational efficiencies of human triple-negative breast cancer cell lines

To quantify mRNA translation at a transcriptome-wide level in human triple-negative breast cancer cell lines we performed polysome-profiling in 6 cell lines starved for 16 hours. Isolated mRNA was quantified using Affymetrix GeneChip Human Transcriptome Array 2.0 (HTA 2.0). Following this, normalized counts were obtained using the “rma” function from affy (v.1.70.0) R/Bioconductor package and log2-transformed expression estimates were obtained using the “exprs” function from Biobase (v.2.52.0) R/Bioconductor package(38,39). Following this, using the model of the first cohort, we obtain residuals for cell lines using the following model per gene and per replicate in which each sample is added to the model separately in each iteration:

$$\varepsilon_{Coh1} = \gamma^{\square}_{Lib.¿oly.} + \gamma^{\square}_{Lib.¿otal} + Ex_{Total} + \varepsilon$$

Analysis of gene signatures in breast cancer samples and human breast cancer cell lines

To assess the link between observed alterations in mRNA translation and upstream regulators of mRNA translation, empirical cumulative distribution functions were employed. This approach provides information of whether a gene set, e.g. those that were previously identified as translationally regulated by an upstream mechanism, were regulated in a patient-selective fashion. The activity of the gene set was determined by the shift of residuals for each regulated gene set relative to background genes (i.e. all genes quantified not part of the gene set) at the 50th quantile (16,17,25,27,40–44) and the significance of the shift was determined using the Wilcoxon rank sum test. Activities in patient groups were asssesed by calculating mean per-gene residuals across patients within a subtype and then the shift and statistics relative to the background. Results were visualized using the "Heatmap" function of ComplexHeatmap (v.2.8.0) R/Bioconductor package(45).

Comparison of mRNA translation between patients and cancer cell lines

To compare translation in subtypes of breast cancer to cancer cell lines, the predict_classess function (set.seed(123) and Parameters: nperm=100, dist_method="correlation") from cola (v.2.10.0) R/Bioconductor package was used on an input matrix that was prepared by calculating per-gene mean residual of 3 replicates from each triple-negative breast cancer cell line, scaling of the matrix was performed by "scale" function (Parameters: center = F). The results were plotted by the same function.

Identification of mRNA features that associate with mRNA translation patterns in subtypes of breast cancer

For assessment of the association of 5'UTR and CDS features with mRNA translation patterns, a three-step process implemented in the "featureIntegration" function from the Anota2seqUtils package was applied on per-gene mean residuals calculated across patients within a subtype. For each group of patients, two sets of genes composing of 500 genes with most activated translation and 500 genes with the most suppressed translation were tested for 5'UTR features (GC content, length, G-quadruplexes identified by pqs finder(46), presence of uORFs with strong Kozak context,presence of RBP-motifs annotated in ATtRACT(47) database and folding energy calculated by Mfold algorithm(48)) and CDS features (codon, dicodon and amino-acid usage) separately using corresponding functions from anota2seqUtils R/Bioconductor package. Following this, 5'UTR and CDS features that are identified as significantly associating with alterations of translational efficiency were assessed in the same model. In addition to the network plots generated by featureIntegration function, an "EffScore" was calculated:

1. $X = -\log_{10}(\text{p-value}) \times 1/r$ [p-value = p-value of correlation and r=correlation coefficient]
2. $X_{Absolute} = \log_{10}(\text{absolute}(X))$
3. EffScore = (sign of X) x $X_{Absolute}$
4. $\text{EffScore}_{Adjusted}$ = EffScore x Variance Explained [Variance Explained = variance explained by the feature of interest]

Heatmap of scaled $EffScore_{Adjusted}$ values for features that were significant (FDR<XXX) in at least two patient subtypes were plotted using the “Heatmap” function in the ComplexHeatmap R/Bioconductor package. Similarly, the activities of signatures were assessed and the heatmap of $EffScore_{Adjusted}$ values were plotted in the same way.

Results

Quantification of mRNA translation in breast cancer patients

Hitherto, mRNA translation has been a largely overlooked molecular characteristic of breast cancer patients, and no studies have been conducted to understand how mRNA translation is regulated at a transcriptome-wide scale. Although conventional proteomics approaches serve as an invaluable way to measure the proteome, after adjustments to variation in total mRNA levels, such methods do not distinguish alterations in the proteome arising via modulation of mRNA translation or protein stability. Consequently, in order to elucidate alterations in protein synthesis due to changes in mRNA translation, direct measurement of mRNA translation is required. To study mRNA translation in breast cancer patients, we used primary non-treated tumors from two patient cohorts (Cohort 1 includes all molecular subtypes of breast cancer while Cohort 2 is limited to triple-negative breast cancer) and performed polysome-profiling tailored for small samples, generating high quality RNA sequencing datasets for downstream analyses (**Fig. 1A-B, Fig. S1A-B**) (32). Polysome-profiling quantifies gene expression at two levels: mRNA associated with ribosomes (the methods enriches for association with >3 ribosomes) and the cytosolic (hereafter denoted “total”) mRNA pool. Subsequently, we used a machine-learning approach to classify patients from Cohort 1 into PAM50 subtypes (33). This revealed that PAM50-classification of patients using total or polysome-associated mRNA was similar (**Fig. 1C**) and thereby confirms the limited post-transcriptional regulation of PAM50-genes observed in a study of breast cancer proteomes (49).

Abundant variation in mRNA translation across breast cancer patients

To understand whether altered mRNA translation is a feature of breast cancer, we used a linear regression-approach to calculate per-gene translation for each patient. The approach calculates residuals as a measure of translation by comparing polysome-associated mRNA to total mRNA levels in a model that is also adjusted for a set of clinical parameters as well as cellular composition (including a repertoire of immune cells and CAFs) of breast cancer samples (**Fig. 2A**). The distribution of slopes and the coefficient of determination (proportion of variation in polysome-associated mRNA levels explained by total mRNA levels) collected from per-gene analysis show that thousands of genes undergo post-transcriptional regulation that affects their translation (**Fig. 2B**). Therefore, alterations in mRNA translation appears to be an overlooked gene-expression mechanism in breast cancer.

Transcriptome-wide alterations in mRNA translation define breast cancer subtypes

We next determined if there are shared patterns of translation across breast cancer patients, which would suggest classes of breast cancer with rewired mRNA translation. Analysis of translation residuals using the cola R/Bioconductor package and assessment of clustering quality metrics such as 1-PAC, mean silhouette, concordance and Jaccard index (**Fig. S2A-D**) suggested six classes (**Fig. 3A**). Importantly, these appeared to be distinct from PAM50 subtypes and clinical parameters that were adjusted for (e.g. clinical stage; **Fig 3A**). The analysis indicated four major gene sets as underlying classification

into the six subclasses (**Fig. 3A**). To understand which cellular functions might be affected by variation in translation across the subclasses, we assesses functions enriched among proteins encoded by mRNAs included in the 4 gene sets (**Fig. 3A**) using gene ontology analysis. This analysis identified many highly enriched functions including antigen processing, mRNA metabolism, histone modification and regulation of translation. (**Fig. 3B**). Therefore, translation modulated across the six translation subclassess appears to target core cellular functions.

To validate these putative subclasses of breast cancer defined by mRNA translation, we used the second cohort of triple-negative breast cancer patients and calculated per-gene and per-patient translational efficiencies (see materials and methods). Next, we predicted classes of newly added samples. This assigned individual patients from this new patient cohort to all identified putative classes of breast cancer, albeit some samples appeared to show some level of similarity to more than one subclass (**Fig. S3A**). As the translation subclasses could be detected in a separate dataset, we regard these as subtypes defined by translation. In summary, these studies uncovered shared patterns of translation across breast cancer patients that could be separated into six subtypes distinct from PAM50 subtypes.

Upstream regulators of mRNA translation associate with translation in breast cancer subtypes

Next we sought to identify mechanisms mediating variation in translation across breast cancer subtypes by first linking this heterogeneity to the activity of a upstream

regulators (e.g. mTOR singaling pathway, activity of subunits of eIF4F complex, cellular stress, DAP5-dependent translation, U34-modifications and ECM) of mRNA translation using a computational approach (16,17,25,27,40–44). To this end we assessed whether gene sets (“signatures”), whose translation is modulated when the activity of these regulators is altered, are regulated in each patient. Per-patient activity of each signature was measured as a shift in residuals relative to the background. This approach revealed large variation in activity of upstream regulators of mRNA translation across patients (**Fig. 4A**). A similar analysis for patients summarized within translation subtypes identified in figure 2 showed that each subtype had a distinct combination in activity of upstream regulators (**Fig. 4B**). For example, subtype 3 shows suppressed translation of e.g. mTOR and high-eIF4E signatures while translation of DAP5 and eIF2a signatures is activated whereas the opposite pattern is observed for subtype 2 (**Fig. 4B**). Taken together, the heterogeneity of mRNA translation across patients overlaps with effects on mRNA translation from key upstream regulators.

In vitro models of triple negative breast cancer recapitulate translation subtypes

Although we sought to assess translation patterns in cancer cells by adjusting for variation in composition of e.g. immune cells and CAFs, it is possible that the observed patterns of translation in breast cancer subtypes originate from non-cancer cells. To assess if identified patterns of translation can arise in breast cancer cells, we sought to identify cancer cell lines mirroring the translation patterns observed in patient samples. To this end, we performed polysome-profiling in 6 triple-negative breast cell lines and derived residuals for each cell line as a measure of translational efficiency. Using the

same approach applied for detection of overlaps between the two patient cohorts, we identified triple-negative human breast cancer cell lines that recapitulate three translation subtypes both in terms of overall profile of mRNA translation as well as the activity of upstream regulators of mRNA translation (**Fig.5A-B**). This supports that observed patterns of translation can arise in cancer cells, also *in vitro*, and such *in vitro* models may therefore be used to study mechanisms underlying translation patterns observed in human breast cancer.

Cis- and trans-acting factors associate with altered mRNA translation in breast cancer subtypes

We next sought to gain a deeper mechanistic understanding of dysregulated translation in translation subtypes of breast cacner. Therefore, we identified and modelled cis- and trans-acting factors that regulate mRNA translation as factors underlying altered translation in breast cancer. This not only helped us to identify complex networks of cis-acting features (**Fig. S4A-F**) but also revelaed that these features associate the with translation heterogeneity observed across translation subtypes. (**Fig. 6A-B**). Strikingly, almost all translation subtypes had an indispensable effect exerted by the codon composition of the CDS. For example, genes with increased translational efficiency in subtype 3 were found to have an enrichment of AAT codons in their CDS. Conversely, in subtype 2, we obserwed that genes with decreased translational efficiency have an enrichment of AAT content. These observations support that codon usage may drive the heterogeneity in translational landscape of breast cancer. We also assessed the contribution of each trans-acting factor to the observed alterations of mRNA translation

by performing step-wise linear regression analysis. In contrast to the previous analysis (Fig.4A-B), this approach allows an assessment of the relative contribution of the selected repertoire of trans-acting factors. As a result, we characterized and quantified how each trans-acting factor associates with mRNA translation in each translation subtype (**Fig. 6C and Fig. S5A-F**). This revealed that at least 40% of the variance in mRNA translation can be explained by cis-acting features (referred to as mRNA features) alone (**Fig. 6D**). In contrast, signatures downstream of trans-acting features only explained a maximum of 45% of the variation in translation in translation subtypes (**Figure 6D**), suggesting that additional pathways are active. Overall, we unveiled complex networks of cis- and trans-acting factors and their influence on translation alterations in breast cancer.

Discussion

Numerous studies have characterized the role of mRNA translation in cancer model systems. In this study, we reveal that there are significant changes in mRNA translation in breast cancer patients. Furthermore, we observed that these alterations contribute significantly to the heterogeneity in breast cancer. This provides a novel molecular characteristics not previously considered in efforts to understand the heterogeneity of breast cancer. However, it remains to be discovered whether the identified translation subtypes of breast cancer have prognostic value in predicting survival and disease progression, an aspect that could not be addressed in this study due to the limited number of patients and follow-up time available.

Furthermore, we linked variation in translation within breast cancer subtypes to activities of known trans-acting modulators of mRNA translation. Our findings suggest a complex interplay of these modulators in shaping the translational landscape of breast cancer. Understanding the clinical implications of these patterns could open new avenues of therapeutic strategies targeting identified modulators, which subsequently may provide personalized treatment options. Although we identified triple-negative human breast cancer cell lines that overlap with some of our translational subtypes, we believe that including cancer cell lines from various molecular subtypes could help identify additional human breast cancer cell lines that reflect the complex translational landscape of those subtypes not represented in our study, potentially expanding the scope of different treatment regimens that could be tested *in vitro*.

We identified cis-acting factors including codon composition and 5'UTR features as strongly associating with alterations in mRNA translation in translation subtypes of breast cancer. These findings provide leads towards a deeper understanding of how alterations in mRNA translation arise. In particular, codon composition appears to be a largely overlooked facet that requires further studies to understand how various steps of mRNA translation are affected and orchestrated to drive cancer-associated alterations in mRNA translation.

Overall, our study has revealed a translational landscape of breast cancer arising by an interplay bewteen cis- and trans-acting factors. Together with the identification of translation subtypes of breast cancer and how each cis- and trans-acting factor is associated with altered mRNA translation within these subtypes, our findings emphasize the potential of mRNA translation as a therapeutic target. In addition, it is also important to highlight the need for further studies to fully understand the clinical implications of our results, particularly in predicting disease progression and treatment outcomes.

433

Figure Captions

Figure 1. Quantification of mRNA translation in breast cancer samples. A. Experimental procedure to study mRNA translation in breast cancer samples using the optimized-sucrose gradient technique for polysome profiling **B.** Clinical characteristics of Cohort I and II breast cancer patients. **C.** PAM50 subtypes determined by a machine-learning based approach using total (first row) and polysome-associated mRNA (second row).

Figure 2. Multiple-regression analysis reveals alterations in translational efficiencies across breast cancer patents. A. Computational workflow to derive per-patient and per-gene measurement of translation**. B.** Densities of slopes (top) and r^2 (bottom) from per-gene models of translation across patients in cohort I. The number of genes showing more extreme values than those indicated with dotted lines is shown.

Figure 3. Translation landscape of breast cancer. A. A heatmap using per-gene and per-patient translation residuals showing 6 translation subtypes of breast cancer. Shown are also PAM50 subtypes as well as a selection of clinical parameters. **B.** Gene ontology analysis using GO-BP terms for each gene set identified in 3A. The 5 most significant functions are shown.

Figure 4. Translation subtypes of breast cancer show distinct activities of trans-acting factors A. Heatmap of per-patient q50 shifts as a measure of patient-specific activities of trans-acting factors. Samples are ordered by translation subtype and

clustered within subtype. **B.** Heatmap of q50 shifts as a measure of activity of each trans-acting factors in translation subtypes.

Figure 5. Breast cancer cell lines recapitulate translation subtypes of breast cancer. **A.** Heatmap of correlations between triple-negative human breast cancer cell lines and translation subtypes (left). Shown is also a table (right) with the p-value of the correlation and the subtype each cell line is assigned to. **B.** Heatmap of q50 shifts as a measure of activities of trans-acting factors comparing patients to cell lines identified in A. Only translation subtypes with identified cell lines are included and the positions of the cell lines are indicated by dotted lines.

Figure 6. Cis and trans-acting factors distinctly influence mRNA translation in breast cancer A. Heatmap of EffScore (see Methods for calculation method) for cis-acting factors across translation subtypes that are siginificantly associated with altered translation a single translation subtype. **B.** Similar as 6A, but showing cis-acting factors that overlap between patient subtypes. **C.** Heatmap of EffScore for significant trans-acting factors across translation subtypes . **D.** Table showing the percentage of variation in mRNA translation explained by mRNA features (i.e. cis-acting factors) and signatures (i.e. trans-acting factors).

480

481

482

483

Additional Files

Supplementary Figure 1. Data set characteristics. A. Barplot of library sizes (i.e. sequencing depth) with a red line indicating threshold of 2.5 million reads. **B.** PCA biplot of samples generated by PCAtools biplot function, colored by library type.

Supplementary Figure 2. Identification of stable patient clusters. A. Plot showing k (number of clusters) vs 1-PAC (score of consensus clustering) with optimal k plotted as red. **B.** Plot showing k (number of clusters) vs mean silhouette with optimal k plotted as red. **C.** Plot showing k (number of clusters) vs concordance with optimal k plotted as red. **D.** Plot showing k (number of clusters) vs 1-PAC (score of consensus stability) with optimal k plotted as red. All plots were generated by "select_partition_number" function of cola R/Bioconductor package.

Supplementary Figure 3. Translation subtypes of breast cancer are captured in triple-negative breast cancer cohort. A. Heatmap of correlations between triple-negative human breast cancer samples and translation subtypes.

Supplementary Figure 4. Networks of cis-acting features explain translation pattern in translation subtypes of breast cancer. A-F . Networks of cis-acting features identified to correlate with alterations in translational efficiencies in subtype 1-6.

Supplementary Figure 5. Networks of trans-acting features explain translation pattern in translation subtypes of breast cancer. A-F . Networks of trans-acting

features identified to correlate with alterations in translational efficiencies in subtype 1-6 .

Supplementary Table 1. Data sets used as "signatures" to assess how upstream regulators of mRNA translation associate with altered mRNA translation.

Declarations

The authors declare that they have no competing interests

530

531

532

References

1. Harbeck N, Penault-Llorca F, Cortes J, et al. Breast cancer. *Nat Rev Dis Primer*. 2019;5(1):1-31. doi:10.1038/s41572-019-0111-2

2. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-752. doi:10.1038/35021093

3. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869-10874. doi:10.1073/pnas.191367098

4. Parker JS, Mullins M, Cheang MCU, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *J Clin Oncol*. 2009;27(8):1160-1167. doi:10.1200/JCO.2008.18.1370

5. Ohnstad HO, Borgen E, Falk RS, et al. Prognostic value of PAM50 and risk of recurrence score in patients with early-stage breast cancer with long-term follow-up. *Breast Cancer Res*. 2017;19(1):120. doi:10.1186/s13058-017-0911-9

6. Prat A, Ellis MJ, Perou CM. Practical implications of gene-expression-based assays for breast oncologists. *Nat Rev Clin Oncol*. 2012;9(1):48-57. doi:10.1038/nrclinonc.2011.178

7. Luque-Bolivar A, Pérez-Mora E, Villegas VE, Rondón-Lagos M. Resistance and Overcoming Resistance in Breast Cancer. *Breast Cancer Targets Ther*. 2020;Volume 12:211-229. doi:10.2147/BCTT.S270799

8. Hershey JWB, Sonenberg N, Mathews MB. Principles of Translational Control. *Cold Spring Harb Perspect Biol*. 2019;11(9):a032607. doi:10.1101/cshperspect.a032607

9. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science*. 2016;352(6292):1413-1416. doi:10.1126/science.aad9868

10. Roux PP, Topisirovic I. Signaling Pathways Involved in the Regulation of mRNA Translation. *Mol Cell Biol*. 2018;38(12):e00070-18. doi:10.1128/MCB.00070-18

11. Pelletier J, Graff J, Ruggero D, Sonenberg N. TARGETING THE eIF4F TRANSLATION INITIATION COMPLEX: A CRITICAL NEXUS FOR CANCER DEVELOPMENT. *Cancer Res*. 2015;75(2):250-263. doi:10.1158/0008-5472.CAN-14-2789

12. Hanson G, Coller J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol*. 2018;19(1):20-30. doi:10.1038/nrm.2017.91

13. Schuller AP, Green R. Roadblocks and resolutions in eukaryotic translation. *Nat Rev Mol Cell Biol*. 2018;19(8):526-541. doi:10.1038/s41580-018-0011-4

14. Novoa EM, Pouplana LR de. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet*. 2012;28(11):574-581. doi:10.1016/j.tig.2012.07.006

15. Liu GY, Sabatini DM. mTOR at the nexus of nutrition, growth, ageing and disease. *Nat Rev Mol Cell Biol*. 2020;21(4):183-203. doi:10.1038/s41580-019-0199-y

16. Larsson O, Li S, Issaenko OA, et al. Eukaryotic Translation Initiation Factor 4E–Induced Progression of Primary Human Mammary Epithelial Cells along the

573 Cancer Pathway Is Associated with Targeted Translational Deregulation of
574 Oncogenic Drivers and Inhibitors. *Cancer Res*. 2007;67(14):6814-6824.
575 doi:10.1158/0008-5472.CAN-07-0752

576 17. Furic L, Rong L, Larsson O, et al. eIF4E phosphorylation promotes tumorigenesis
577 and is associated with prostate cancer progression. *Proc Natl Acad Sci*.
578 2010;107(32):14134-14139. doi:10.1073/pnas.1005320107

579 18. Rajasekhar VK, Viale A, Socci ND, Wiedmann M, Hu X, Holland EC. Oncogenic Ras
580 and Akt signaling contribute to glioblastoma formation by differential recruitment
581 of existing mRNAs to polysomes. *Mol Cell*. 2003;12(4):889-901. doi:10.1016/s1097-
582 2765(03)00395-2

583 19. Lin CJ, Cencic R, Mills JR, Robert F, Pelletier J. c-Myc and eIF4F Are Components of a
584 Feedforward Loop that Links Transcription and Translation. *Cancer Res*.
585 2008;68(13):5326-5334. doi:10.1158/0008-5472.CAN-07-5876

586 20. Robichaud N, del Rincon SV, Huor B, et al. Phosphorylation of eIF4E promotes EMT
587 and metastasis via translational control of SNAIL and MMP-3. *Oncogene*.
588 2015;34(16):2032-2042. doi:10.1038/onc.2014.146

589 21. Rubio A, Garland GD, Sfakianos A, Harvey RF, Willis AE. Aberrant protein synthesis
590 and cancer development: The role of canonical eukaryotic initiation, elongation
591 and termination factors in tumorigenesis. *Semin Cancer Biol*. 2022;86:151-165.
592 doi:10.1016/j.semcancer.2022.04.006

22. Alain T, Morita M, Fonseca BD, et al. eIF4E/4E-BP Ratio Predicts the Efficacy of mTOR Targeted Therapies. *Cancer Res*. 2012;72(24):6468-6476. doi:10.1158/0008-5472.CAN-12-2395

23. Faller WJ, Jackson TJ, Knight JRP, et al. mTORC1-mediated translational elongation limits intestinal tumour initiation and growth. *Nature*. 2015;517(7535):497-500. doi:10.1038/nature13896

24. Santos M, Fidalgo A, Varanda AS, Oliveira C, Santos MAS. tRNA Deregulation and Its Consequences in Cancer. *Trends Mol Med*. 2019;25(10):853-865. doi:10.1016/j.molmed.2019.05.011

25. Lorent J, Kusnadi EP, van Hoef V, et al. Translational offsetting as a mode of estrogen receptor α-dependent regulation of gene expression. *EMBO J*. 2019;38(23):e101323. doi:10.15252/embj.2018101323

26. Wang Y, Tao EW, Tan J, Gao QY, Chen YX, Fang JY. tRNA modifications: insights into their role in human cancers. *Trends Cell Biol*. 2023;33(12):1035-1048. doi:10.1016/j.tcb.2023.04.002

27. de la Parra C, Ernlund A, Alard A, Ruggles K, Ueberheide B, Schneider RJ. A widespread alternate form of cap-dependent mRNA translation initiation. *Nat Commun*. 2018;9(1):3068. doi:10.1038/s41467-018-05539-0

28. Jewer M, Lee L, Leibovitch M, et al. Translational control of breast cancer plasticity. *Nat Commun*. 2020;11(1):2498. doi:10.1038/s41467-020-16352-z

29. Lee LJ, Papadopoli D, Jewer M, et al. Cancer Plasticity: The Role of mRNA Translation. *Trends Cancer*. 2021;7(2):134-145. doi:10.1016/j.trecan.2020.09.005

30. Nasr Z, Robert F, Porco JA, Muller WJ, Pelletier J. eIF4F suppression in breast cancer affects maintenance and progression. *Oncogene*. 2013;32(7):861-871. doi:10.1038/onc.2012.105

31. Bhat M, Robichaud N, Hulea L, Sonenberg N, Pelletier J, Topisirovic I. Targeting the translation machinery in cancer. *Nat Rev Drug Discov*. 2015;14(4):261-278. doi:10.1038/nrd4505

32. Liang S, Bellato HM, Lorent J, et al. Polysome-profiling in small tissue samples. *Nucleic Acids Res*. 2018;46(1):e3. doi:10.1093/nar/gkx940

33. Gendoo DMA, Ratanasirigulchai N, Schröder MS, et al. Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. *Bioinformatics*. 2016;32(7):1097-1099. doi:10.1093/bioinformatics/btv693

34. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37(7):773-782. doi:10.1038/s41587-019-0114-2

35. Wu SZ, Al-Eryani G, Roden DL, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet*. 2021;53(9):1334-1347. doi:10.1038/s41588-021-00911-1

632 36. Gu Z, Schlesner M, Hübschmann D. *cola* : an R/Bioconductor package for consensus
633 partitioning through a general framework. *Nucleic Acids Res*. 2021;49(3):e15-e15.
634 doi:10.1093/nar/gkaa1146

635 37. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for
636 interpreting omics data. *The Innovation*. 2021;2(3):100141.
637 doi:10.1016/j.xinn.2021.100141

638 38. Huber W, Carey VJ, Gentleman R, et al. Orchestrating high-throughput genomic
639 analysis with Bioconductor. *Nat Methods*. 2015;12(2):115-121.
640 doi:10.1038/nmeth.3252

641 39. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy—analysis of *Affymetrix GeneChip* data
642 at the probe level. *Bioinformatics*. 2004;20(3):307-315.
643 doi:10.1093/bioinformatics/btg405

644 40. Guan BJ, van Hoef V, Jobava R, et al. A Unique ISR Program Determines Cellular
645 Responses to Chronic Stress. *Mol Cell*. 2017;68(5):885-900.e6.
646 doi:10.1016/j.molcel.2017.11.007

647 41. Modelska A, Turro E, Russell R, et al. The malignant phenotype in breast cancer is
648 driven by eIF4A1-mediated changes in the translational landscape. *Cell Death Dis*.
649 2015;6(1):e1603. doi:10.1038/cddis.2014.542

650 42. Ramírez-Valle F, Braunstein S, Zavadil J, Formenti SC, Schneider RJ. eIF4GI links
651 nutrient sensing by mTOR to cell proliferation and inhibition of autophagy. *J Cell*
652 *Biol*. 2008;181(2):293-307. doi:10.1083/jcb.200710215

43. Parker MW, Rossi D, Peterson M, et al. Fibrotic extracellular matrix activates a profibrotic positive feedback loop. *J Clin Invest*. 2014;124(4):1622-1635. doi:10.1172/JCI71386

44. Gandin V, Masvidal L, Cargnello M, et al. mTORC1 and CK2 coordinate ternary and eIF4F complex assembly. *Nat Commun*. 2016;7:11127. doi:10.1038/ncomms11127

45. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32(18):2847-2849. doi:10.1093/bioinformatics/btw313

46. Hon J, Martínek T, Zendulka J, Lexa M. pqsfinder: an exhaustive and imperfection-tolerant search tool for potential quadruplex-forming sequences in R. *Bioinforma Oxf Engl*. 2017;33(21):3373-3379. doi:10.1093/bioinformatics/btx413

47. Giudice G, Sánchez-Cabo F, Torroja C, Lara-Pezzi E. ATtRACT—a database of RNA-binding proteins and associated motifs. *Database J Biol Databases Curation*. 2016;2016:baw035. doi:10.1093/database/baw035

48. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003;31(13):3406-3415. doi:10.1093/nar/gkg595

49. Johansson HJ, Socciarelli F, Vacanti NM, et al. Breast cancer quantitative proteome and proteogenomic landscape. *Nat Commun*. 2019;10(1):1600. doi:10.1038/s41467-019-09018-y

675

676

677

678

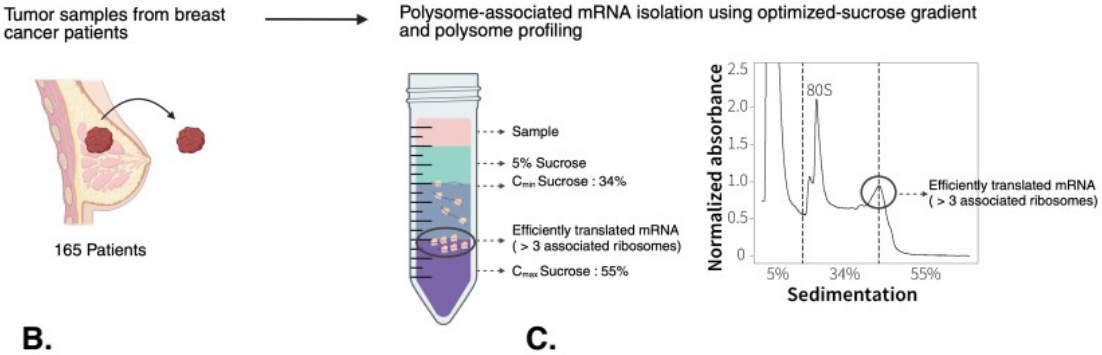
679

680

681

Figure 1.

A.

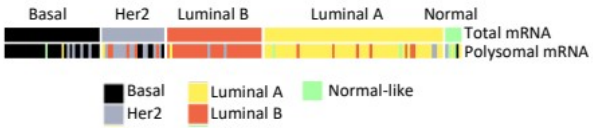


B.

Variable	Overall, N = 165[†]	Cohort I, N = 135[†]	Cohort II, N = 30[†]
Age at Diagnosis	55 (35,89)	55 (35,89)	56 (35,81)
Molecular Subtype			
HER 2 +	10 (6.1%)	4 (3.0%)	6 (20%)
Luminal A	65 (39%)	64 (47%)	1 (3.3%)
Luminal B	51 (31%)	51 (38%)	0 (0%)
Triple Negative	39 (24%)	16 (12%)	23 (77%)
Family History of Cancer	106 (73%)	88 (75%)	18 (67%)
Unknown	20	17	3
Clinical Stage			
I	50 (30%)	45 (33%)	5 (17%)
II	86 (52%)	67 (50%)	19 (63%)
III	27 (16%)	22 (16%)	5 (17%)
IV	2 (1.2%)	1 (0.7%)	1 (3.3%)
Histological Grade			
1	23 (14%)	18 (14%)	5 (17%)
2	71 (45%)	58 (45%)	13 (43%)
3	64 (40%)	53 (41%)	11 (37%)
4	1 (0.6%)	0 (0%)	1 (3.3%)
Unknown	6	6	0
p53			
-	87 (58%)	80 (65%)	7 (25%)
+	49 (32%)	29 (24%)	20 (71%)
+*	15 (9.9%)	14 (11%)	1 (3.6%)
Unknown	14	12	2
KI-67 (%)	25 (0,95)	18 (0,90)	53 (0,95)
Unknown	47	40	7

[†] Mean (Minimum,Maximum); n (%)
+ = weak signal*

C.



682

683

Figure 2.

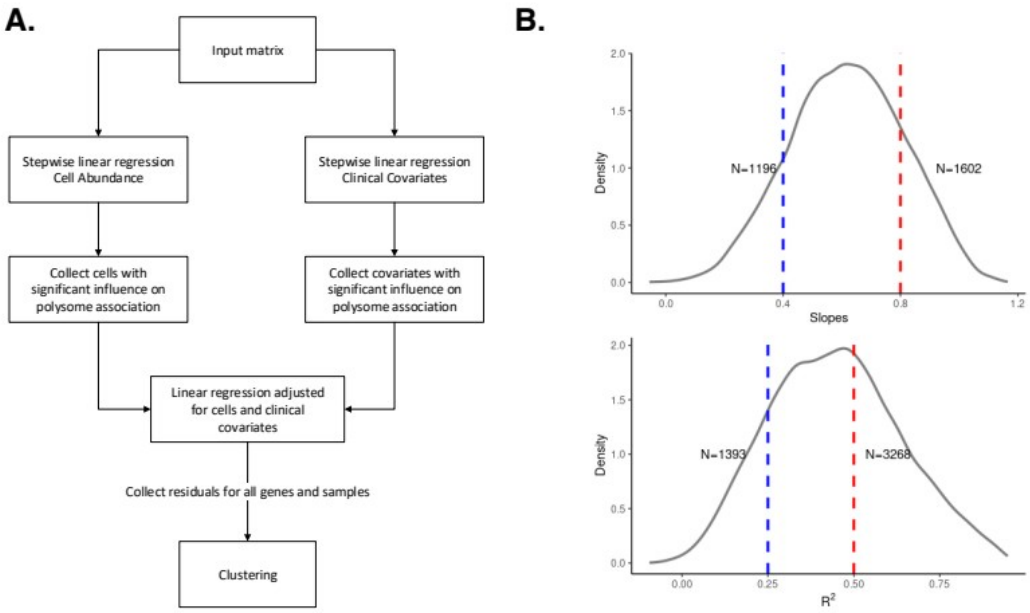
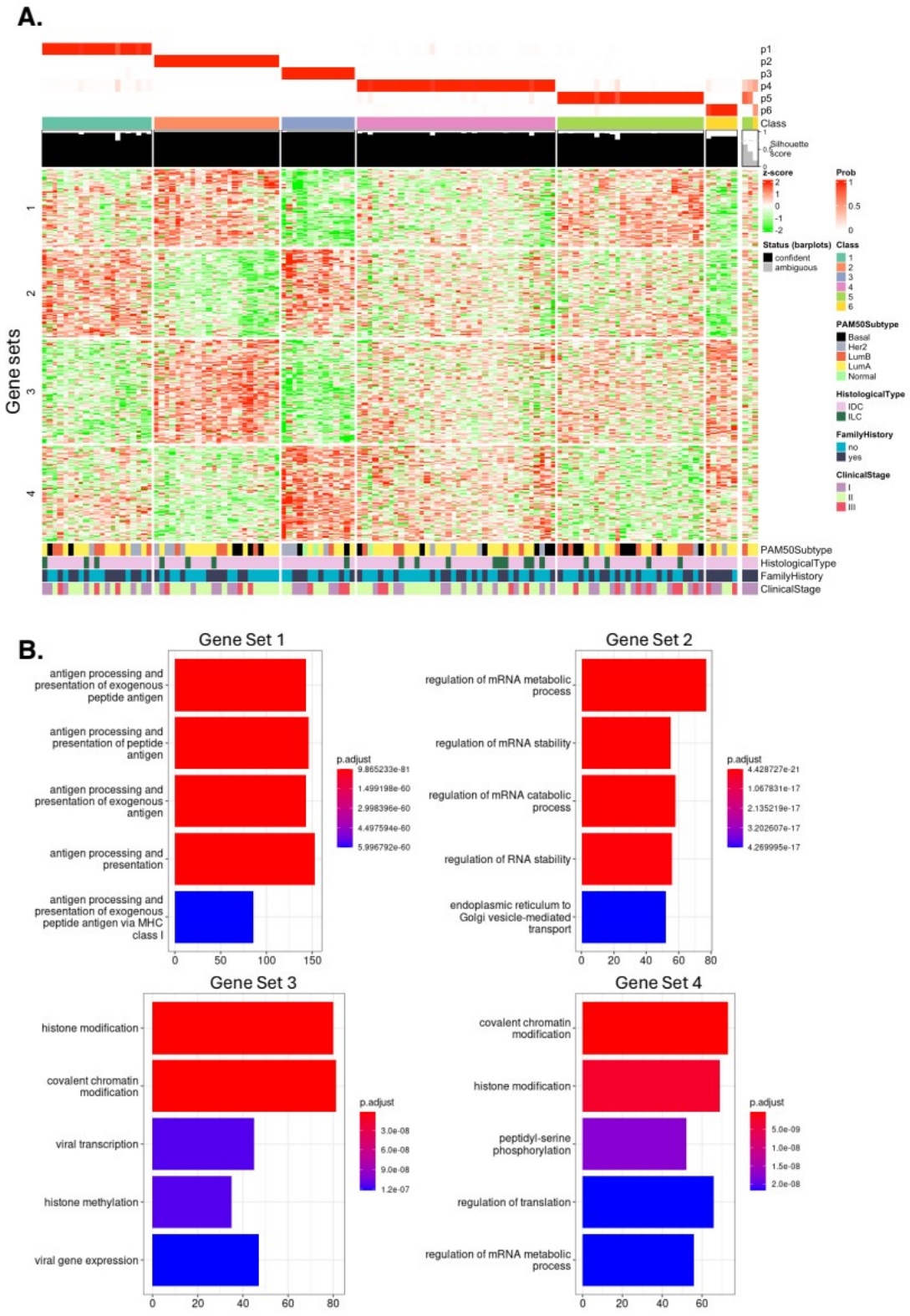


Figure 2. Multiple-regression analysis reveals alterations in translational efficiencies across breast cancer patents. A. Computational workflow to derive per-patient and per-gene measurement of translation. **B.** Densities of slopes (top) and r^2 (bottom) from per-gene models of translation across patients in cohort I.

684

685

Figure 3.



686

687

Figure 4.

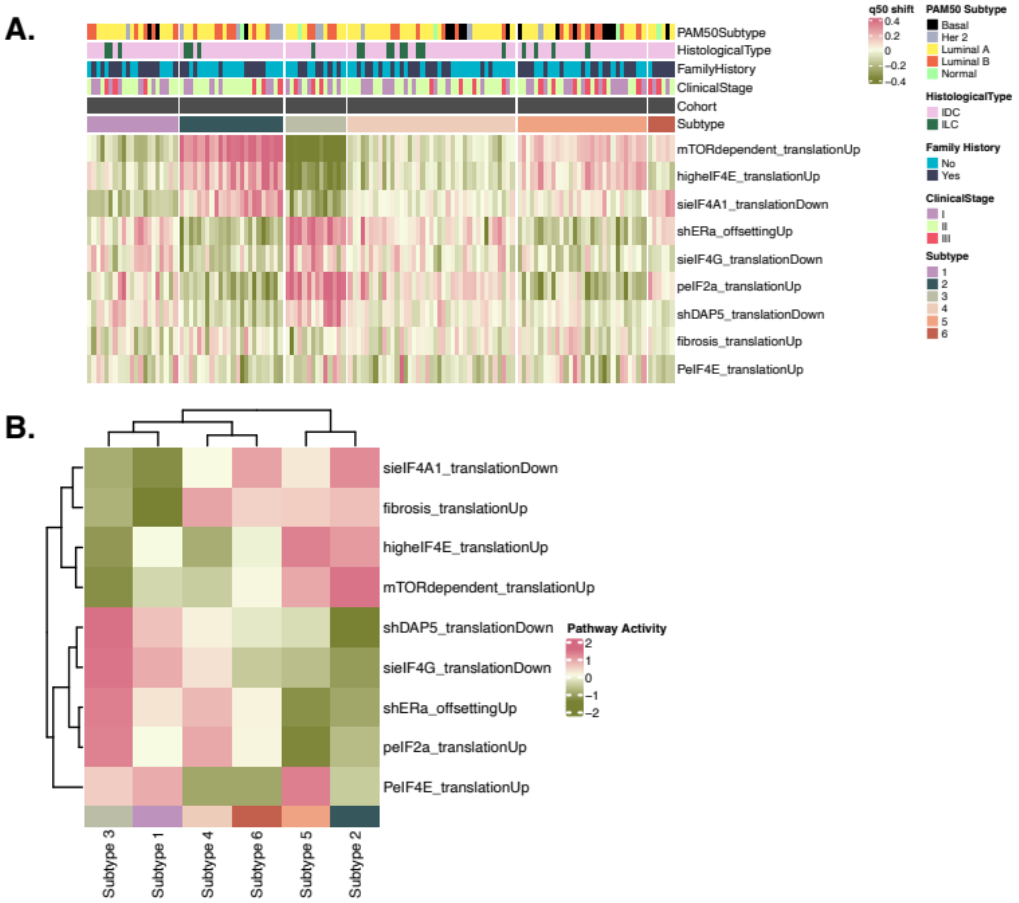
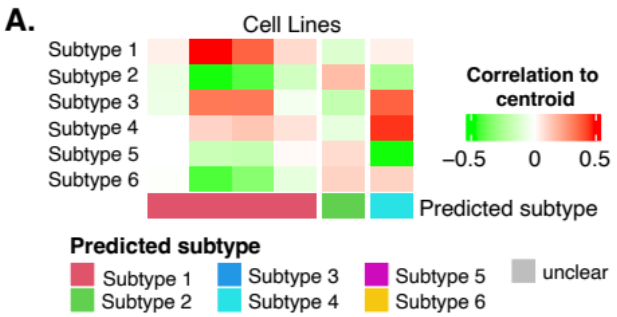
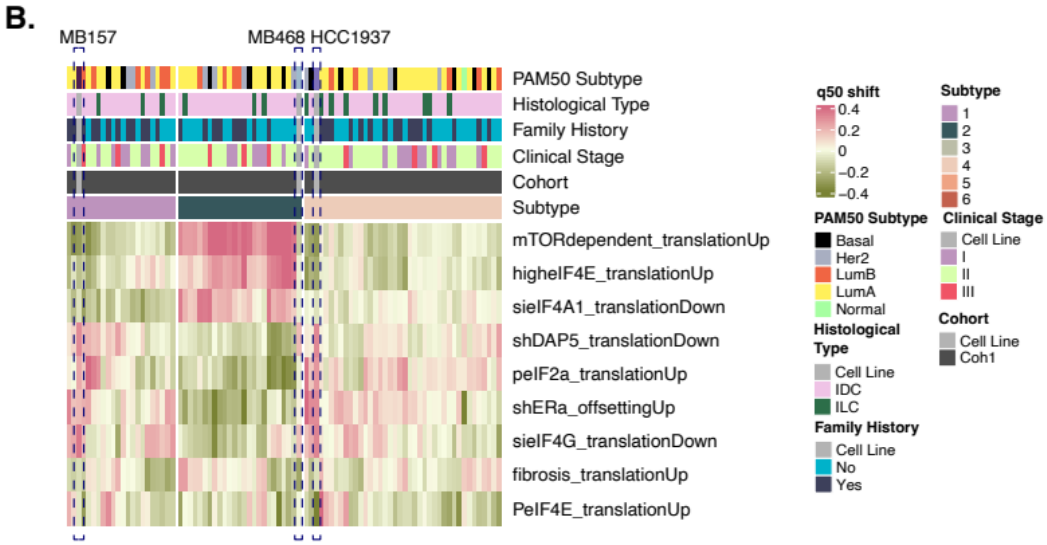

A.
PAM50Subtype
HistologicalType
FamilyHistory
ClinicalStage
Cohort
Subtype
mTORdependent_translationUp
higheIF4E_translationUp
sieIF4A1_translationDown
shERa_offsettingUp
sieIF4G_translationDown
peIF2a_translationUp
shDAP5_translationDown
fibrosis_translationUp
PeIF4E_translationUp
q50 shift
0.4
0.2
0
−0.2
−0.4
PAM50 Subtype
Basal
Her 2
Luminal A
Luminal B
Normal
HistologicalType
IDC
ILC
Family History
No
Yes
ClinicalStage
I
II
III
Subtype
1
2
3
4
5
6
B.
sieIF4A1_translationDown
fibrosis_translationUp
higheIF4E_translationUp
mTORdependent_translationUp
shDAP5_translationDown
sieIF4G_translationDown
shERa_offsettingUp
peIF2a_translationUp
PeIF4E_translationUp
Pathway Activity
2
1
0
−1
−2
Subtype 3
Subtype 1
Subtype 4
Subtype 6
Subtype 5
Subtype 2

Figure 5.



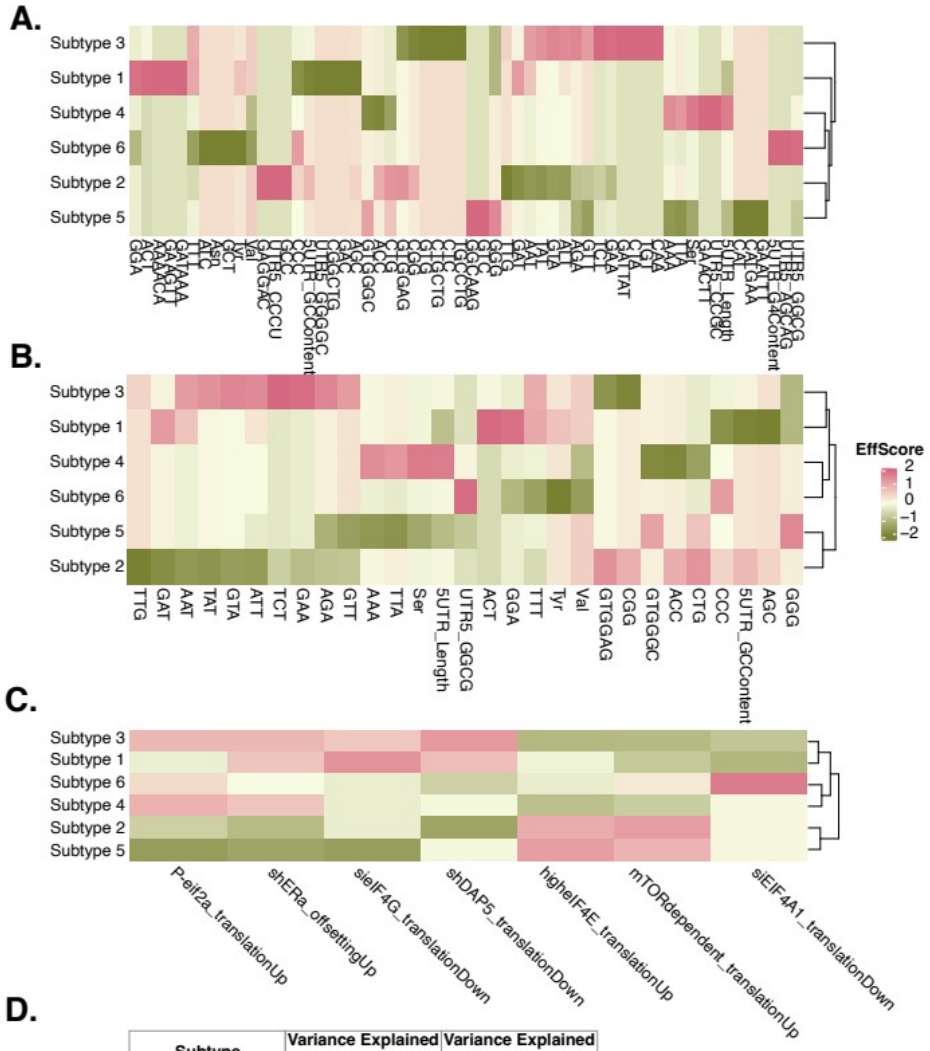
Cell Line	Class	P-value
HCC1937	4	2.14e-20
MB157	1	2.30e-18
MB468	2	3.45e-02



690

691

Figure 6.



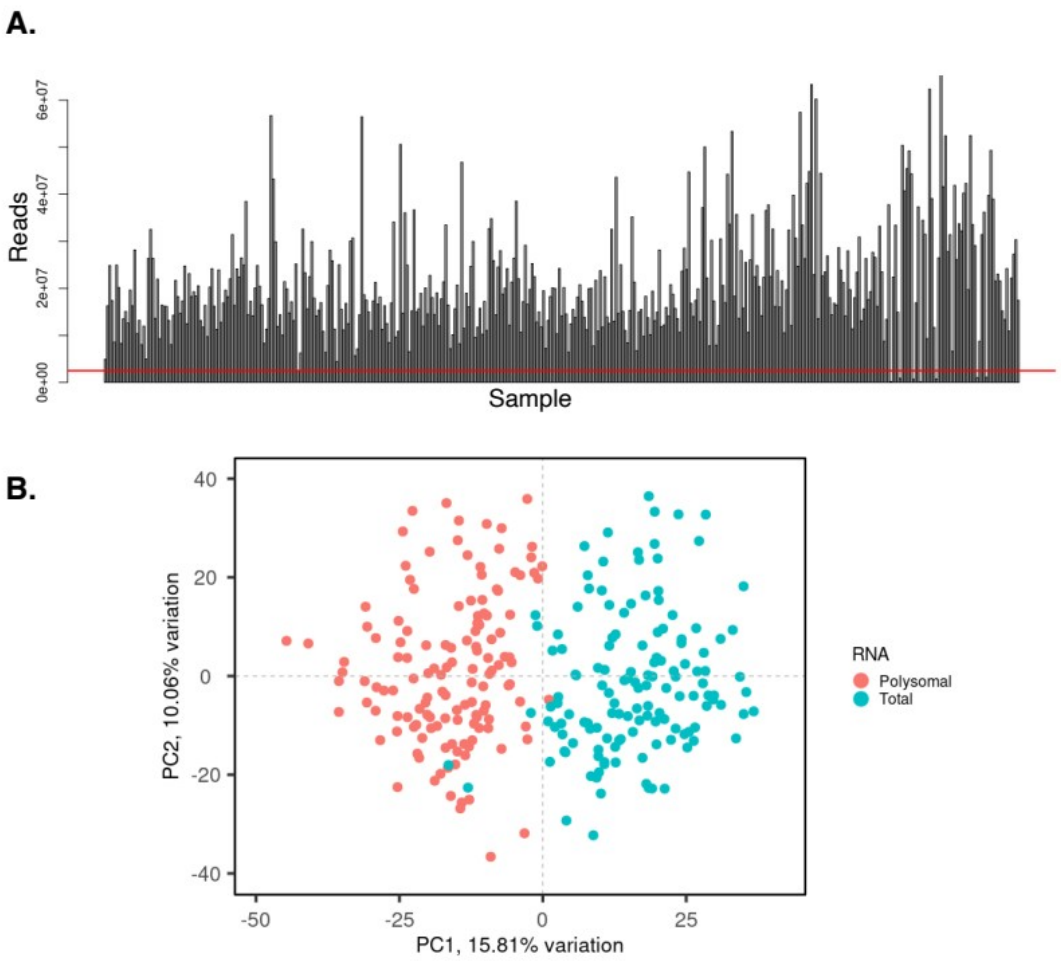
D.

Subtype	Variance Explained mRNA features (%)	Variance Explained Signature (%)
Subtype 1	77.81	15.21
Subtype 2	88.1	38.2
Subtype 3	91.16	45.81
Subtype 4	52.72	28.3
Subtype 5	76.69	38.53
Subtype 6	44.14	6.07

692

693

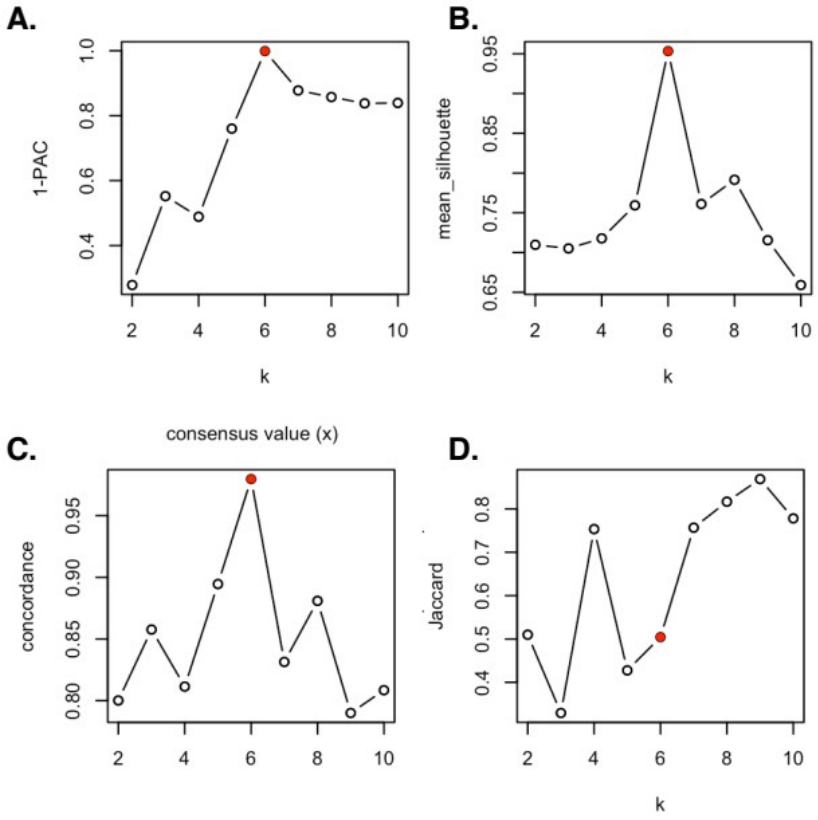
Figure S1.



694

695

Figure S2.

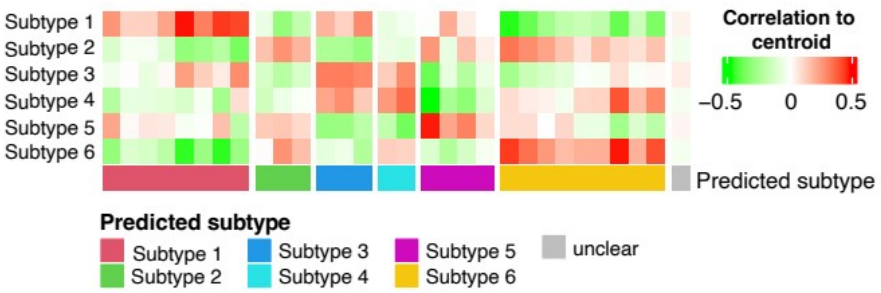


696

697

Figure S3.

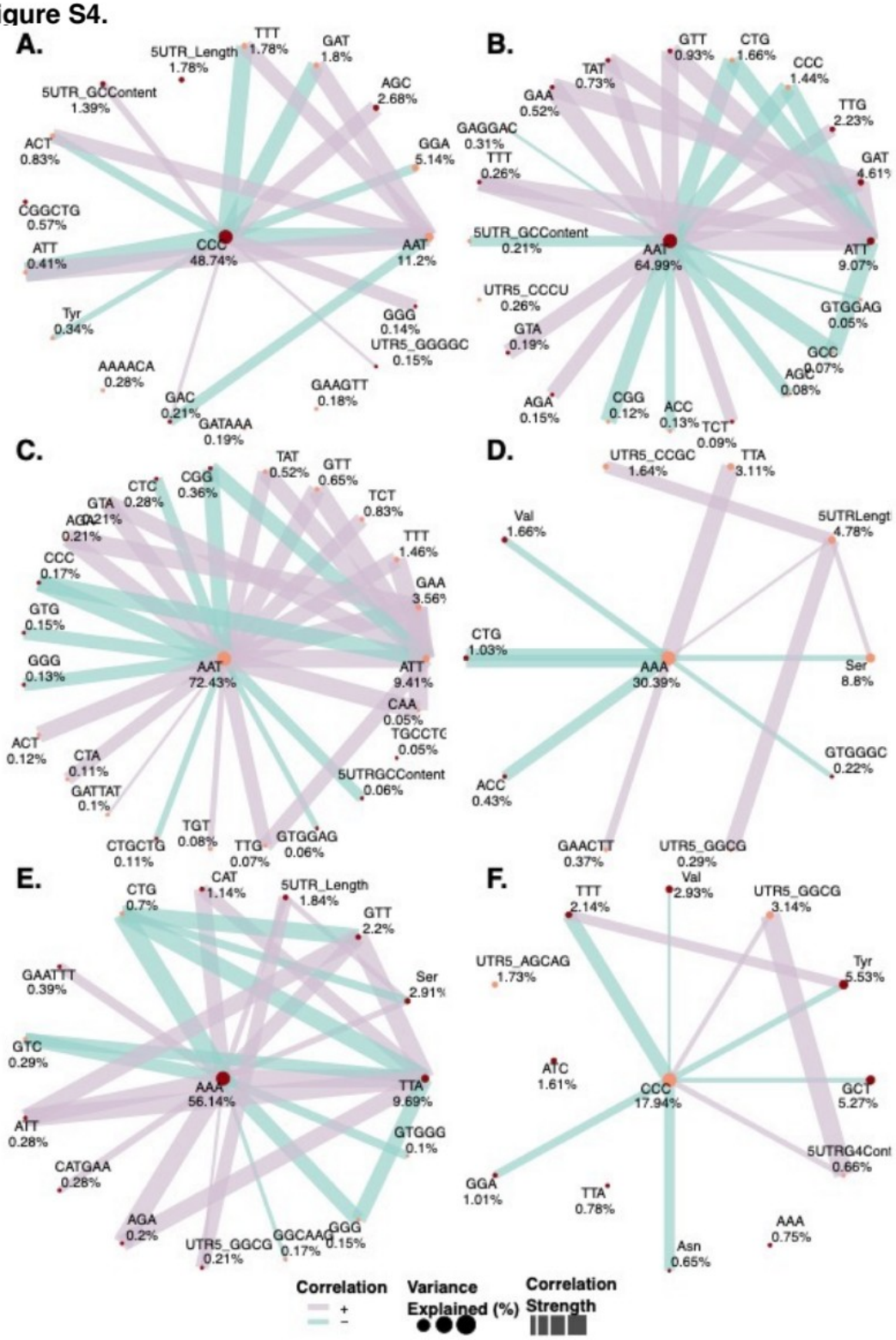
A.



698

699

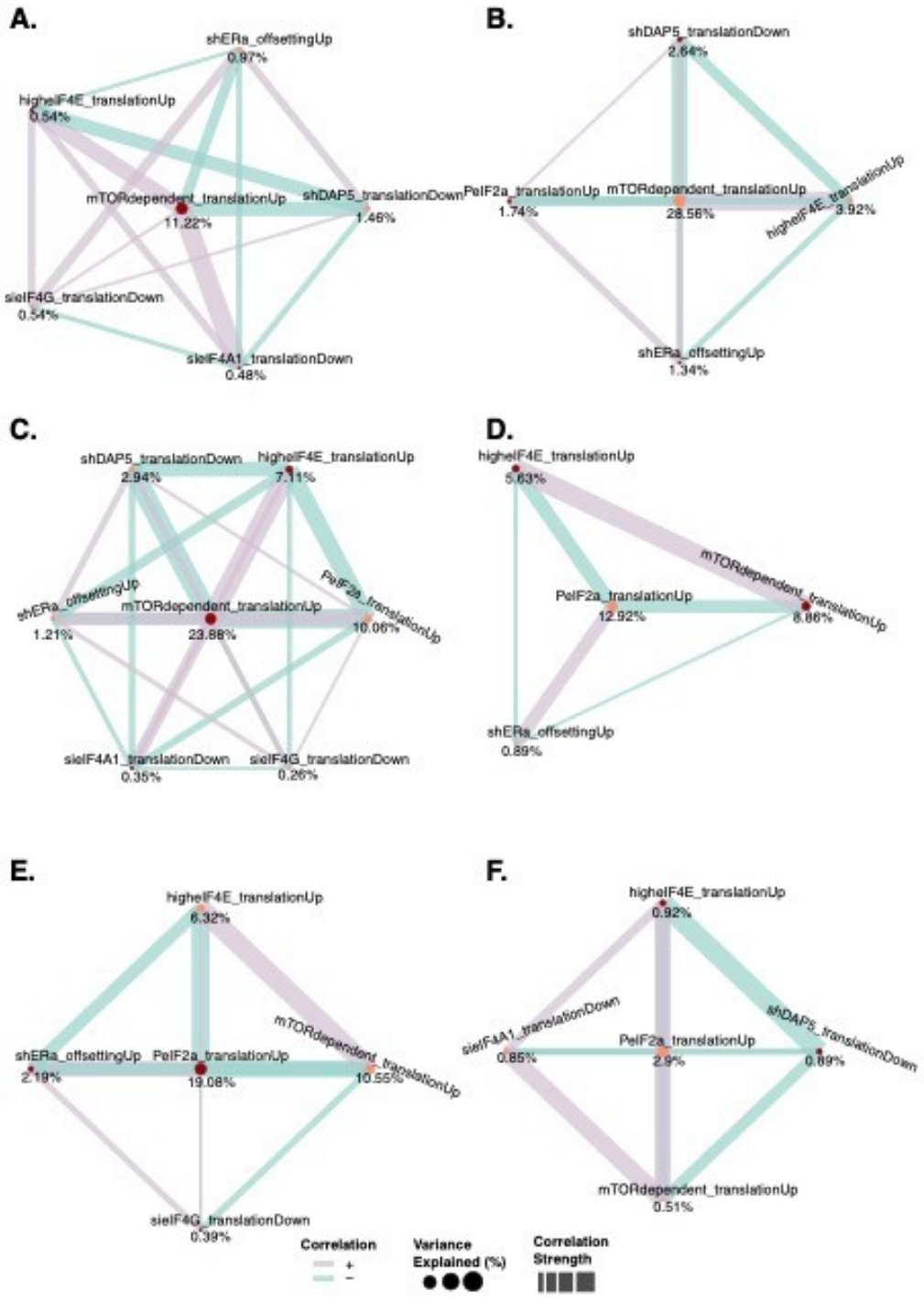
Figure S4.



700

701

Figure S5.



702