TranSNPs: A class of functional SNPs affecting mRNA translation potential revealed by fraction-based allelic imbalance

Highlights
- SNPs can alter mRNA translation potential
- Polysomal imbalance of heterozygous SNPs can mark allele-specific protein expression
- UTR SNPs associated with allele-specific translation demonstrate prognostic value

Samuel Valentini, Caterina Marchioretti, Alessandra Bisio, Annalisa Rossi, Sara Zaccara, Alessandro Romanel, Alberto Inga

alessandro.romanel@unitn.it (A.R.)
alberto.inga@unitn.it (A.I.)

Valentini et al., iScience 24, 103531 December 17, 2021 © 2021 The Author(s).
https://doi.org/10.1016/j.isci.2021.103531

iScience Article
TranSNPs: A class of functional SNPs affecting mRNA translation potential revealed by fraction-based allelic imbalance
TranSNPs: A class of functional SNPs affecting mRNA translation potential revealed by fraction-based allelic imbalance

Samuel Valentini,1,4 Caterina Marchioretti,1,2,4 Alessandra Bisio,1,4 Annalisa Rossi,1 Sara Zaccara,1,3 Alessandro Romanel,1,5,* and Alberto Inga1,5,6,*

SUMMARY
Few studies have explored the association between SNPs and alterations in mRNA translation potential. We developed an approach to identify SNPs that can mark allele-specific protein expression levels and could represent sources of inter-individual variation in disease risk. Using MCF7 cells under different treatments, we performed polysomal profiling followed by RNA sequencing of total or polysome-associated mRNA fractions and designed a computational approach to identify SNPs showing a significant change in the allelic balance between total and polysomal mRNA fractions. We identified 147 SNPs, 39 of which located in UTRs. Allele-specific differences at the translation level were confirmed in transfected MCF7 cells by reporter assays. Exploiting breast cancer data from TCGA we identified UTR SNPs demonstrating distinct prognosis features and altering binding sites of RNA-binding proteins. Our approach produced a catalog of tranSNPs, a class of functional SNPs associated with allele-specific translation and potentially endowed with prognostic value for disease risk.

INTRODUCTION
Single-nucleotide polymorphisms (SNPs) represent one of the largest classes of genetic variations. They underlie and are responsible for inter-individual variations in complex-disease phenotypes, including cancer risk or aggressiveness. Wide attention has been given to SNPs that can lead to allele-specific changes in gene expression, for instance, by modulating the affinity of transcription factor-binding sites in promoter or enhancer regions directly, or indirectly via influencing epigenetic regulation. Included in these examples are functionally relevant SNPs affecting p53 function as a transcription factor or p53 protein expression by altering target binding sites (Bond et al., 2006; Menendez et al., 2007, 2019; Tomso et al., 2005), or association-driven studies that have candidated SNPs as modulators of major cancer drivers, such as AR (Bu et al., 2016; Romanel et al., 2017), ER (Bailey et al., 2016; Dunning et al., 2016), or cMYC (Pomerantz et al., 2009; Takatsuno et al., 2013).

A fraction of SNPs identified in the human population is located within coding regions or UTRs. In this case, both mechanism- and association-driven studies have pursued functional SNPs that can modify aspects of post-transcriptional gene regulation, for example, by altering microRNA-binding sites (Landi et al., 2012).

Specific tools have been implemented to mine the available wealth of RNA-seq-based gene expression data and identify and pursue instances of allele-specific gene expression (Mayba et al., 2014; Pandey et al., 2013; Przytycki and Singh, 2020; Romanel, 2019; Romanel et al., 2015; Rozowsky et al., 2011; Skelly et al., 2011; Wei et al., 2012). The same cannot be said for annotating candidate SNPs driving or being associated with alterations in mRNA translation potential. Recently, we showed that allelic imbalance restricted to polysome-bound mRNAs can be exploited to investigate the functional significance of 5’UTR single-nucleotide variants (SNVs) at the CDKN2A gene (Andreotti et al., 2016), since specific SNVs affected CDKN2A translation. Those results led us to hypothesize that a comparative analysis of allelic-imbalance from total and polysomal mRNAs extracted and sequenced starting from the same cell sample, which is independently genotyped for heterozygous SNPs, could overcome the intrinsic noise in SNP calling and coverage from RNA-seq data. If so, a catalog of coding and UTRs SNPs associated and potentially causative of alterations in mRNA translation potential could be obtained. Here we describe such an approach...
starting from a commonly used breast cancer cell line and two disease-relevant treatments. About 4% of the heterozygous SNPs that could be followed showed a significant level of imbalance within polysomal RNAs. Nearly 25% of these are located in UTR sequences, and some appear to stratify patients with cancer for distinct clinical outcomes.

Thus, our approach led us to distinguish both constitutive and treatment-dependent SNPs that are associated with or can cause changes in mRNA translation efficiency. We propose that our method can lead to identify a class of functional SNPs that we call tranSNPs, endowed with prognostic value for cancer or other diseases.

RESULTS

Using polysome profiling to identify SNPs exhibiting allelic imbalance

In order to develop an approach to map SNPs associated with allelic imbalance within translating ribosomes, we took advantage of a well-characterized cancer cell line with available genotyping data. We thus chose MCF7 cells, p53 wild-type breast adenocarcinoma-derived cells we had previously used for polysomal profiling studies (Zaccara et al., 2014). After determining the set of heterozygous SNPs using DNA data, we exploited RNA-seq reads to compare relative allelic representation between the transcripts isolated in the cytoplasm because of their association with polysomes or lack thereof. Besides a mock condition, we included two treatments, 1 μM doxorubicin or 10 μM Nutlin, that we found can reduce global translation, engage p53 responses, and lead to a significant uncoupling between transcriptome and translome changes. Indeed, many transcripts were significantly modulated by the treatments only at the level of polysomal RNA fractions (Rizzotto et al., 2020; Zaccara et al., 2014). We reasoned that these treatments could uncover specific allelic imbalance events due to changes in translation specificity mediated, for instance, by the action of RNA-binding proteins (Zaccara et al., 2014; Rizzotto et al., 2020). Polysomal profiling was performed using cytoplasmic lysates fractionated by a linear 10%-50% sucrose gradient, as previously described (Rizzotto et al., 2020; Rossi et al., 2019; Zaccara et al., 2014). Cytoplasmic mRNA associated with light fractions, with ribosome subunits, or with the 80S monosome, assumed to be not actively translating, were pooled together and sequenced separately from mRNAs associated with two or more ribosomes. Total RNA was also collected and sequenced. The RNA sequencing of two biological replicates for each treatment and fraction produced on average 32 million unique mapped reads. Differential gene expression analysis and pathway enrichment analysis confirmed that the two treatments activated a canonical p53 response (Figure S1).

Identifying instances of allelic imbalance in polysomal RNA

To identify SNPs exhibiting allelic imbalance in one or more RNA fractions, we first retrieved from public databases an exhaustive list of SNPs that are heterozygous in the MCF7 cell line. We integrated SNP-array and exome-based genotype calls and overall identified 11,544 heterozygous SNPs of which 1,802 were in 3' UTRs and 729 were in 5' UTRs (Table S1). Then, we determined the allelic fractions of these SNPs in our RNA-seq data using a pileup approach based on high-quality reads and bases and requiring a minimum local depth of coverage of 10X. We found an average of 4,100 of the expected 11,544 SNPs per sample with confirmed coverage signal and, overall, we found 3,974 SNPs analyzable in at least one condition.

Our approach focuses on the comparison of SNP allelic counts across biological replicates of the two RNA fractions that were sequenced for each of the three treatments. We first determined which is the distribution of SNP allelic fractions across all experiments and measured to what extent they are variable across biological replicates. Of interest, the variability of SNP allelic fractions was limited across biological replicates (Figure S2) with a mean replicate’s difference of about 7% that was consistent among the different RNA fractions and treatments (Table S1). Then, we determined the allelic fractions of these SNPs in our RNA-seq data using a pileup approach based on high-quality reads and bases and requiring a minimum local depth of coverage of 10X. We found an average of 4,100 of the expected 11,544 SNPs per sample with confirmed coverage signal and, overall, we found 3,974 SNPs analyzable in at least one condition.

We used this approach to identify allelic imbalances between polysomal and total RNA fractions. Specifically, considering SNPs allelic fraction (AF) intervals calculated by summing and subtracting condition-specific AF
variabilities from SNPs AFs, we searched for SNPs presenting polysomal and total RNA non-overlapping AF intervals (Figure 1A). The analysis led to 162 imbalance instances, involving 147 SNPs that were identified in both biological replicates preserving the imbalance direction. This conservative list of 147 SNPs represents the first catalog of tranSNPs: polymorphisms exhibiting allelic imbalance specifically in polysome-bound mRNA fractions (Tables S3 and S4). The list is almost equally divided between constitutive, doxorubicin-dependent, and Nutlin-dependent polysomal allelic imbalances (Figure 1B). Only three tranSNPs were common to all conditions, and only eight to the two different p53-activating treatments. Despite the limited observed intersection among our conservative calls, the comparison of condition-associated tranSNPs allelic fractions across all conditions showed not only an expected clear shift in the condition-associated tranSNPs AF distribution but also a fraction of SNPs with comparable imbalance across two or three conditions (Figure 1C). The condition-associated lists of genes harboring the tranSNPs showed mild enrichment for specific biological processes or molecular functions, including regulation of spindle organization, mitotic cytokinesis, protein kinase regulator activity, and catalytic activity (Figure S3).

Characteristics of tranSNPs are shown in Table 1 and are compared with those in the larger set of 11,544 initial SNPs and those in the set of 3,974 analyzable SNPs. Of the 147 tranSNPs we identified, 39 are located in UTRs and represent top candidates for a direct role on the observed allelic imbalance.
**Table 1. Summary table of MCF7 heterozygous SNPs**

|                          | MCF7 heterozygous SNPs | Analyzable SNPs | tranSNPs |
|--------------------------|------------------------|-----------------|----------|
| Number                   | 11,544                 | 3,974           | 147      |
| Average Global MAF       | 0.26                   | 0.25            | 0.25     |
| Average EUR MAF          | 0.28                   | 0.28            | 0.27     |
| Overlapping genes        | 5,493                  | 2,532           | 139      |
| Intronic                 | 4,724                  | 675             | 26       |
| Coding                   | 6,161                  | 3,007           | 108      |
| UTR                      | 2,553                  | 1,182           | 39       |
| LD blocks                | 8,656                  | 3,202           | 143      |

Summary table presenting the characteristics of tranSNPs together with same features measured from the set of the SNPs heterozygous in the MCF7 cell line initially considered, and the set of SNPs analyzable from the RNA-seq data by our approach. Average Global and EUR MAF (Minor Allele Frequency) were retrieved from 1,000 Genomes Project data. Details on how the LD blocks were defined are provided in the text.

**BRI3BP 3’ UTR rs1055472 and CDKN1A 5’ UTR rs2395655 alleles are functionally distinct**

We chose two UTR SNPs showing polysomal allelic imbalance for validation and opted for a gene reporter assay to evaluate the functional impact of the two pairs of tranSNPs. We selected two Nutlin-dependent tranSNPs and two genes harboring them whose functions are related to p53.

*BRI3BP*, also known as human cervical cancer oncogene 1 (HCCR-1), might act as negative modulator of p53 (Cho et al., 2007; Ha et al., 2008). The two rs1055472 alleles of the BRI3BP 373nt 3’ UTR were cloned downstream of the Firefly cDNA. MCF7 cells were transiently transfected with each of the two alleles, and the activity of the reporter relative to the Renilla control luciferase was measured in untreated cells or in cells treated with Nutlin. In both cases we observed that the alternative allele led to a relative increase in the reporter gene activity (Figure 2A), suggesting an overall differential translation efficiency. Although the SNP was computationally classified as polysomally imbalanced only after Nutlin treatment (Tables S3 and S4), computational and experimental data combined suggest that the two alleles are functionally distinct, impacting on gene expression.

*CDKN1A* (p21) is an important cyclin-dependent kinase inhibitor and one of the major direct p53 transcriptional targets mediating cell-cycle arrest (El-Deiry, 2016; Kreis et al., 2019). The gene is highly regulated at the transcriptional and post-transcriptional levels (Rossi et al., 2019; Wang et al., 2000). *CDKN1A* transcripts were highly induced by both doxorubicin and Nutlin treatment, and the coverage of rs2395655 alleles in the untreated condition was very low (<10X). We evaluated the *CDKN1A* 5’ UTR rs2395655 alleles cloned upstream of the Firefly cDNA. We observed an overall bimodal distribution in the reporter gene activity, whose amplitude is reduced when the alternative allele is present (Figure 2B, left panel). The treatment with Nutlin further mitigated this bimodal distribution with an overall reporter gene activity that is strongly reduced when the alternative allele is present (Figure 2B, right panel). Since the SNP was concordantly classified as polysomally imbalanced after Nutlin treatment, in this case we highlight a scenario with functionally distinct alleles where differential translation efficiency may impact protein expression only under specific conditions.

**TranSNPs can have prognostic significance in breast cancer TCGA data**

Considering that rs1055472 and rs2395655 are tranSNPs with functionally distinct alleles and the two genes harboring them are related to p53, we next investigated their potential clinical impact. We focused on breast cancer for consistency with the cell line model used and exploited the richness of data available in TCGA (Koboldt et al., 2012). Exploring TCGA survival data (Liu et al., 2018) and using Kaplan-Meier curves, we observed that patients harboring the rs1055472 alternative allele show a statistically significant increase in progression-free interval time (p value = 0.042, Figure 3A). Of note, *BRI3BP* and *TP53* transcript levels were not correlated (Figure S4A) and variant rs1055472 was not associated with patients’ *TP53* somatic aberration status (Figure S4B) or with the utilization of DNA-damaging agents (Figure S4C). Furthermore, the analysis repeated on *TP53* aberrant or *TP53* wild-type patients only (Figure S4D) showed similar trends. Overall, this indicates that rs1055472 signal is not dependent on the
Motivated by this result, we then explored more systematically whether tranSNPs could be associated with distinct clinical variables in patients with cancer. To this end, our tranSNPs list was extended including all common SNPs in strong linkage disequilibrium (LD) ($r^2 > 0.8$) with them, obtaining 3,003 SNPs distributed across 120 LD blocks with genotype imputable from TCGA data. The extended SNP list was then used to interrogate clinical data in the breast cancer cohort from TCGA (Liu et al., 2018), using Overall Survival (OS), Disease and Progression Free Intervals (DFI, PFI), and Disease Specific Survival (DSS) endpoints. Kaplan-Meier curves were built stratifying patients for the presence of the alternative allele (AA versus AB + BB) or the presence of the homozygous genotype for the alternative allele (AA + AB versus BB). Cox proportional hazards regression models were built to perform the analysis. To limit false-positive results, for each considered outcome, association results demonstrating a p value < 0.05 were aggregated at the level of LD blocks and >5% of SNPs reproducing the association signal in a block were required to nominate the block as associated. On average, we found 10.7% (min 8.3%, max 14.2%) associated blocks across all outcomes with an average fraction of SNPs associated in a block equal to 56.8% and an average number of SNPs per block of 23.4 (Table S5). Across the associated blocks, we identified 33 SNPs in the UTR sequences of 17 genes demonstrating significant prognostic effects (Figures 3B-3D, and Table S6). Both protective and risk alternative SNP alleles were found.

**UTR tranSNPs could alter RNA-binding protein target sites**

To search for a mechanism that could underlie polysomal imbalance and the observed differential translation, we examined the potential impact on RNA-binding proteins (RBPs)-binding sites for the 33 UTR tranSNPs resulting from the survival analysis. These polymorphisms might indeed directly cause the observed allelic imbalance of SNPs located in UTRs by, for example, impacting on UTR structure or binding sites for RBPs or also microRNAs. RNA binding consensus motifs were retrieved from the RBPDB database (Cook et al., 2011). TESS software (Schug, 2008) was used to compute motifs scores starting from 60-bp sequences with the SNP reference or alternative allele in the center. Overall, we identified 61 putative motif disruptions involving 27 UTR variants and RBPs associated with translation control and mRNA stability (Table S7), such as ELAVL1/HuR, FUS, YBX1, and PABPC1 (Behm-Ansmant et al., 2007; Mazan-Mamczarz et al., 2003; Peixeiro et al., 2012, p. 1; Tanguay and Gallie, 1996; Yasuda et al., 2013).
UTR-specific imbalance in ATF6 transcript

Among the UTR SNPs showing potential prognostic significance, a peculiar allelic imbalance pattern was observed for polymorphisms in the ATF6 transcript. Although only one ATF6 SNP was present in our initial tranSNPs list, an overall trend of imbalance in polysomal-associated RNAs was observed and restricted to SNPs located in its relatively long 3' UTR (Figure 4A). This imbalance trend was consistent across most ATF6 3' UTR SNPs that are heterozygous in MCF7, and the overall imbalance distribution was strongly statistically significant (p value < 1 \times 10^{-6}) when compared with imbalance distributions obtained from 1,000 sets of random sequential heterozygous SNPs of the same cardinality (Figure 4B). On the contrary, the ATF6 heterozygous coding SNPs did not show any significant imbalance. Furthermore, alternative alleles were confirmed to be all in phase using 1,000 Genomes Project phased genotype data, ATF6 transcript level was not associated with clinical outcomes in both univariate and multivariate models, and no differential ATF6 transcript level was observed for patients carrying the clinically distinct haplotype structure.

A gene reporter assay was used to evaluate the functional impact of two different sub-regions of the long ATF6 3' UTR, corresponding to chr1:161930870-161931403 and chr1:161931723-161932422 (hg19) genomic coordinates (Figure 4C). For the downstream region, we observed a consistent statistically significant increase in the reporter gene activity for the allele harboring the alternative SNPs bases (Figure 4C), both in untreated and Nutlin-treated cells. For the upstream region, we observed a similar trend of increased reporter gene activity (although not reaching statistical significance) for the allele harboring the alternative SNPs bases in the untreated cells but not in the Nutlin-treated cells.

Computational predictions suggest that the allele harboring the alternative SNPs bases loses two PABPC1-binding sites that have strong scores when considering the allele harboring the reference SNPs bases.
Table S7), suggesting an allele-dependent distinct processing of the 3’ UTR. Indeed, using RNA immunoprecipitation (RIP) assays, we demonstrate that the \(\text{ATF6}\) transcript, and specifically the 3’ UTR region, is a PABPC1 target in MCF7 cells (Figure 4D).

**DISCUSSION**

The relative abundance of an mRNA in polysomal versus subpolysomal fraction is frequently used as a proxy for protein synthesis and can capture changes both in global and specific translation efficiency, driven by structural or sequence elements in the context of cell stress responses. It is common knowledge that cancer cells experience chronic stress as well as acute stress responses that converge on translation controls, unfolded protein responses, oxidative stress, and proteasome functions. Recent studies, including some from our group, are placing wild-type p53 in an important position to respond and influence these pathways and also are indicating that the gain of function properties of several mutant p53...
proteins may converge on the dysregulation of controls on translation quality, protein folding, and proteasome functionality. Hence, in this study, we chose a p53 wild-type cancer cell line well characterized in terms of p53-induced responses (Andrysik et al., 2017), including post-transcriptional and translational changes (Zaccara et al., 2014). We used both a commonly used chemotherapeutic drug that activates the DNA damage response, and a selective MDM2 inhibitor that results in acute p53 activation without apparent genotoxic response (Rizzato et al., 2020; Tovar et al., 2006). Doses and treatment time points were selected based on previous characterization of both p53 activation pathway and cell outcome (Lion et al., 2013; Nassiri et al., 2019). Furthermore, for both treatments, we have previously shown a global as well as a specific impact on translation (Zaccara et al., 2014).

Since MCF7 cells have been genotyped both by SNP arrays and exome-sequencing as part of the cell-line encyclopedia and NCI-60 studies, we could leverage RNA-seq data of total and polysome-bound mRNAs to focus on allele counts for all transcribed SNPs in the heterozygous state to reveal an imbalance in one fraction over the other. We hypothesized that this approach, although restrictive, could overcome limitations deriving from the quality and quantity of sequence reads that can be attributed to each allele of a SNP pair. Computational approaches that use allele counts of heterozygous SNPs to identify imbalance events have been implemented and widely used to investigate allele-specific expression patterns both in healthy and disease tissues (Pandey et al., 2013; Romanel, 2019; Romanel et al., 2015; Rozowsky et al., 2011; Skelly et al., 2011; Wei et al., 2012). However, only few approaches (Mayba et al., 2014; Przytycki and Singh, 2020) have been developed to compare allelic imbalances across matched samples, and all of the available methods have been specifically designed in the context of cancer (e.g., to compare cancer samples with their matched normal counterparts), hence limiting hence their broad applicability.

Using polysomal profiling and RNA relative quantification from either pooled or individual fractions typically show relatively small differences, that can, however, underlie an important effect at the level of protein amounts. The approach we developed to identify tranSNPs is focused on computing and comparing the relative change of allele counts across mRNA fractions. To account for technical noise, the approach embeds in the calculation the amount of allele counts variability observed in a given condition exploiting data from biological replicates. Of note, such variability was not affected by the cellular treatment employed and was consistent across different RNA fractions. We identified 147 tranSNPs, representing nearly 4% of the overall analyzed SNPs, a fraction that, although conservative, can be considered comparable with the recent fraction of about 2% of allele-specific expressed genes computed across tumor versus normal matched samples (Przytycki and Singh, 2020). Indeed, in Przytycki and Singh (2020), differential allele-specific expression (ASE) of genes was computed by aggregating at gene level heterozygous SNPs data and no statistically significant difference between genes with differential ASE and those without with respect to the number of heterozygous SNPs across the length of the genes was found. We checked the intersection of our tranSNP list with the current list of GWAS variants from the GWAS catalog (Buniello et al., 2019) that are related to cancer (N = 2,657), although it is possible that tranSNPs might not be directly related to the risk of developing a tumor but be related to a distinct clinical outcome. Only an intersection between the LD block defined by our tranSNP rs760077 and association signal with gastric cancer (variant rs760077 and LD variants rs140081212, rs4072037) and breast cancer (LD variants rs2075570, rs2974935) was identified. We further explored the intersection between the tranSNPs and a set of GWAS-implicated cancer risk SNPs identified in 41 genes of the p53 response pathway, of which those mapping at KITLG, CDKN2A, and TEX9 genes were reported to potentially impact drug sensitivity (Stracquadanio et al., 2016; Zhang et al., 2021). Given that MCF7 cells express wild-type p53 and that the doxorubicin or Nutlin treatments elicit a p53-dependent response, we were motivated to examine the potential overlap with our tranSNP list. However, only 13 of those 41 GWAS-implicated cancer risk SNPs could be evaluated in our MCF7 RNA-seq data. They map at six genes (CERS5, ISYNA1, CFLAR, SESN1, AKAP9, CYP51A1) that were not reported to impact drug sensitivity (Stracquadanio et al., 2016; Zhang et al., 2021). Finally, we considered the results of a very recent study where over 12,000 3’UTR variants, including more than 2,000 disease-associated SNPs from the GWAS catalog, were cloned within 100-nt-long fragments and tested in a massively parallel reporter assay measuring relative RNA expression (Griesemer et al., 2021). Surprisingly, about 25% of the tested GWAS SNPs resulted in steady-state changes in reporter transcripts’ abundance. Among the GWAS SNPs included in that screen, 131 are heterozygous in MCF7 and could be analyzed by our method. We found no intersection with our tranSNP list, consistent with our approach being focused on allelic differences that are detected among polysomal-associated mRNAs but not total mRNAs.
A subset of the identified UTR tranSNPs showed apparent prognostic value. Some of these are located in genes that have been already established to play a role in cancer in general and in breast cancer or in relation to p53 functions, in particular. These include BR3BP, also known as human cervical cancer oncogene 1 (HCCR-1), that might act as a negative modulator of p53 (Cho et al., 2007; Ha et al., 2008); the ATF6 gene, an important modulator of endoplasmic reticulum stress that can promote cell survival (Guan et al., 2017; Sicari et al., 2019); and the protein lysine methyltransferase SETD3 gene, which was recently reported as a prognostic marker in patients with breast cancer (Hassan et al., 2020) and also shown to promote apoptosis in response to DNA damage, in part through the modulation of p53 (Abaev-Schneiderman et al., 2019).

Other genes implicated by our results are GFM1, a mitochondrial translation elongation factor that has been linked to oxidative stress and cancer cell survival (Norberg et al., 2017), and PLEC, an important member of a large family of scaffolding proteins that can modulate various aspects of cell function. The plectin gene is frequently found co-amplified with MYC in BRCA1-deficient breast cancer (Annunziato et al., 2019). Sequence variants predicted to be deleterious have also been identified in primary breast cancers (Horvath et al., 2013). Physical interaction of plectin with BRCA2 has also been reported (Niwa et al., 2009). In all these cases, the tranSNPs we uncovered may be modifiers of protein expression, influencing mRNA translation efficiency, with consequences on cancer phenotypes.

To begin exploring a more precise mechanism of action, motifs analysis based on eCLIP or RIP-seq data was exploited to identify UTR tranSNPs that may alter or lead to gain or loss of RNA-binding sites. Based on these predictions and focusing on a peculiar imbalance in ATF6 transcript that we observed only across 3’ UTR SNPs, we performed RIP assays and confirmed that ATF6 is a target of PABPC1, a well-known RNA-binding protein. This, combined with predicted putative loss of two PABPC1-binding sites across one of the two ATF6 alleles, suggests an allele-dependent processing of the 3’ UTR.

Overall, our approach led us to identify a class of SNPs that are associated with changes in mRNA translation efficiency, potentially acting at the post-transcriptional level, and that show prognostic value for cancer, hence implicating the potential identification or stratification of patients with cancer based on genetic markers that are helpful in the prediction of prognosis in regard to death, progression, or recurrence.

Limitations of the study

This study considers a single cell line and two biological replicates for each considered treatment/condition. Survival analysis is focused on breast cancer data. Our catalog of tranSNPs is hence limited to the data we analyzed.

STAR+METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines and culture conditions
- METHOD DETAILS
  - SNP status in MCF7 cells
  - Differential allele-specific expression analysis
  - Survival analysis using TCGA data
  - Identification of SNPs disrupting putative RBP consensus motifs
  - Polysomal profiling and RNA-sequencing
  - Cloning 5’UTR or 3’UTR allelic variants for gene reporter assays
  - Dual-luciferase assays
  - RIP assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103531.
ACKNOWLEDGMENTS

We thank the LaBSSAH-CIBIO Next Generation Sequencing Facility of the University of Trento. The research leading to these results has received funding from “Fondazione AIRC per la Ricerca sul Cancro” under MFAG 2017 - ID. 20621 project - P.I. A.R. and under IG18985 - P.I. A.I..

AUTHOR CONTRIBUTIONS

C.M. performed the differential allele-specific expression analyses from RNA-seq data. S.V. performed survival analysis and RBP motifs analysis. A. Rossi performed the RIP assays. A.B. performed and supervised cloning and gene reporter assays. S.Z. contributed to data analysis and manuscript writing. A. Romanel designed and supervised all the computational approaches and analyses and co-wrote the first draft of the manuscript. A.I. designed the overall study, supervised the validation experiments, and co-wrote the first draft of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 8, 2021
Revised: October 27, 2021
Accepted: November 23, 2021
Published: December 17, 2021

REFERENCES

Abaev-Schneiderman, E., Admoni-Elisha, L., and Levy, D. (2019). SETD3 is a positive regulator of DNA-damage-induced apoptosis. Cell Death Dis. 10. https://doi.org/10.1038/s41419-019-1328-4.

Andreotti, V., Bisio, A., Bressac-de Paillerets, B., Harland, M., Cabaret, O., Newton-Bishop, J., Pastorino, L., Bruno, W., Bertorelli, R., De Sanctis, V., et al. (2016). The CDKN2A/p16INK4a 5’UTR sequence and translational regulation: impact of novel variants predisposing to melanoma. Pigment Cell Melanoma Res. 29, 210–221. https://doi.org/10.1111/pcm.12446.

Andryszk, Z., Galbraith, M.D., Guarnieri, A.L., Zaccara, S., Sullivan, K.D., Pandey, A., MacBeth, M., Inga, A., and Espinosa, J.M. (2017). Identification of a core TP53 transcriptional program with highly distributed tumor suppressive activity. Genome Res. 27, 1645–1657. https://doi.org/10.1101/gr.220533.117.

Annunziato, S., de Ruiter, J.R., Henneman, L., Brambillasca, C.S., Lutz, C., Vaillant, F., Ferrante, F., Drenth, A.P., van der Burg, E., Siteur, B., et al. (2019). Comparative oncogene genomics identifies combinations of driver genes and drug targets in BRCA1-mutated breast cancer. Nat. Commun. 10, 397. https://doi.org/10.1038/s41467-019-08301-2.

Bailey, S.D., Desai, K., Kron, K.J., Mazrooei, P., Sinnott-Armstrong, N.A., Treloar, A.E., Dowar, M., Thu, K.L., Cescon, D.W., Silvester, J., et al. (2016). Noncoding somatic and inherited single-nucleotide variants converge to promote ESR1 expression in breast cancer. Nat. Genet. 48, 1260–1266. https://doi.org/10.1038/ng.3650.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C., Lehár, J., Kryukov, G., Sonkin, D., et al. (2012). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483, 603–607. https://doi.org/10.1038/nature10103.

Behm-Ansmant, I., Gatfield, D., Rehwinkel, J., Hilgers, V., and baumalde, E. (2007). A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. EMBO J. 26, 1591–1601. https://doi.org/10.1038/sj.emboj.7601588.

Bond, G.L., Hirsheld, K.M., Kirchhoff, T., Alexe, G., Bond, E.E., Robins, H., Bartel, F., Taubert, H., Wuerl, P., Hait, W., et al. (2006). MDM2 SNP309 accelerates tumor formation in a gender-specific and hormone-dependent manner. Cancer Res. 66, 5104–5110.

Bu, H., Narisu, N., Schlick, B., Rainer, J., Manke, T., Schäfer, G., Pasqualini, L., Chines, P., Schweiger, M.R., Fuchsberger, C., and Lückner, H. (2016). Putative prostate cancer risk SNP in an androgen receptor-binding site of the melanoephilin gene illustrates enrichment of risk SNPs in androgen receptor target sites. Hum. Mutat. 37, 52–64. https://doi.org/10.1002/humu.22909.

Buniello, A., MacArthur, J.A.L., Ceredo, M., Harris, L.W., Hayhurst, J., Mongalone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 47, D1005–D1012. https://doi.org/10.1093/nar/gky1120.

Cho, G.W., Shin, S.M., Kim, H.K., Ha, S.A., Kim, S., Yoon, J.H., Hur, S.Y., Kim, T.E., and Kim, J.W. (2007). HCCR-1, a novel oncogene, encodes a mitochondrial outer membrane protein and suppresses the UVC-induced apoptosis. BMC Cell Biol. 8. https://doi.org/10.1186/1471-2121-8-50.

Cook, K.B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T.R. (2011). RBPD: a database of RNA-binding specificities. Nucleic Acids Res. 39, D301–D308. https://doi.org/10.1093/nar/gkq1069.

Dassi, E., Greco, V., Sidarovich, V., Zuccotti, P., Arseni, N., Scaruffi, P., Tonini, G.P., and Quattrone, A. (2019). Translational compensation of genomic instability in neuroblastoma. Sci. Rep. 9, 14364. https://doi.org/10.1038/s41598-019-43644-x.

Delaneau, O., Howie, B., Cox, A.J., Zagury, J.-F., and Marchini, J. (2013). Haplotype estimation using sequencing reads. Am. J. Hum. Genet. 93, 687–696. https://doi.org/10.1016/j.ajhg.2013.09.002.

Dunning, A.M., Michailidou, K., Kuchenbaecker, K.B., Thompson, D., French, J.D., Beesley, J., Healey, C.S., Kar, S., Pooley, K.A., Lopez-Knowles, E., et al. (2016). Breast cancer risk variants at 24q5 display different phenotype associations and regulate ESR1/RMND1 and CCDC170. Nat. Genet. 48, 374–386. https://doi.org/10.1038/ng.3521.

El-Deiry, W.S. (2016). p21(WAF1) mediates cell-cycle inhibition, relevant to cancer suppression and therapy. Cancer Res. https://doi.org/10.1158/0008-5472.CAN-16-2055.

Griesemer, D., Xue, J.R., Reilly, S.K., Ulisch, J.C., Kukreja, K., Davis, J.R., Kanai, M., Yang, D.K., Butts, J.C., and Guey, M.H. (2021). Genome-wide functional screen of 3’UTR variants uncovers causal variants for human disease and evolution. Cell 184, 5247–5260.e19. https://doi.org/10.1016/j.cell.2021.08.025.

Guar, B.J., van Hoef, V., Jobava, R., Elroy-Stein, O., Valasek, L.S., Cargnello, M., Gao, X.H., Krokowski, D., Merrick, W.C., Kimball, S.R., et al. (2017). A unique ISR program determines cellular responses to chronic stress. Mol. Cell 68, 885–900.e6. https://doi.org/10.1016/j.molcel.2017.11.007.
Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An L1 interacting with HCCR-1 induces tumorigenesis. Cell 173, 400–416. https://doi.org/10.1016/j.cell.2018.02.052.

Moynihan, C.J., Evans, J.H., Lee, A.V., et al. (2019). HCCRBP-1 directly inhibits the function of p53 in colorectal cancer. iScience 11, 103531, December 17, 2021.

Pandey, R.V., Franssen, S.U., Futschik, A., and Schlo"{o}terer, C. (2013). Allelic imbalance in colorectal cancer. iScience 1, 103531, December 17, 2021.

Schug, J. (2008). Using TESS to predict transcription factor binding sites in DNA sequence. Curr. Protocols Bioinform. https://doi.org/10.1002/0471250953.bi0206s21.

Shi, C.C., Ye, Z., Zhang, Q., et al. (2019). Mutant p53 improves cancer cells’ survival and resistance to therapeutic drugs. Cancer Res. 79, 2590–2600. https://doi.org/10.1158/0008-5472.CAN-18-3092.

Skelly, D.A., Johansson, M., Madeoy, J., et al. (2019). MBASED: allele-specific expression uncovers breast cancer genes dysregulated by cis noncoding mutations. Cell 173, 400–416. https://doi.org/10.1016/j.cell.2018.02.052.
and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. Genome Res. 21, 1728–1737. https://doi.org/10.1101/gr.119784.110.

Stracquadanio, G., Wang, X., Wallace, M.D., Grawenda, A.M., Zhang, P., Hewitt, J., Zeron-Medina, J., Castro-Giner, F., Tomlinson, I.P., Goding, C.R., et al. (2016). The importance of p53 pathway genetics in inherited and somatic cancer genomes. Nat. Rev. Cancer 16, 251–265. https://doi.org/10.1038/nrc.2016.15.

Takatsuno, Y., Mimori, K., Yamamoto, K., Sato, T., Niida, A., Inoue, H., Imoto, S., Kawanou, S., Yamaguchi, R., Toh, H., et al. (2013). The rs6983267 SNP is associated with MYC transcription efficiency, which promotes progression and worsens prognosis of colorectal cancer. Ann. Surg. Oncol. 20, 1395–1402. https://doi.org/10.1245/s10434-012-2657-z.

Tanguay, R.L., and Gallie, D.R. (1996). Translational efficiency is regulated by the length of the 3' untranslated region. Mol. Cell Biol. 16, 146–156. https://doi.org/10.1128/mcb.16.1.146.

Therneau, T.M., and Grambsch, P.M. (2000). Modeling Survival Data: Extending the Cox Model, Statistics for Biology and Health (Springer New York). https://doi.org/10.1007/978-1-4757-3294-8.

Tomso, D.J., Inga, A., Menendez, D., Pittman, G.S., Campbell, M.R., Storici, F., Bell, D.A., and Resnick, M.A. (2005). Functionally distinct polymorphic sequences in the human genome that are targets for p53 transactivation. Proc. Natl. Acad. Sci. U S A 102, 6431–6436.

Tovar, C., Rosinski, J., Filipovic, Z., Higgins, B., Kolinsky, K., Hilton, H., Zhao, X., Vu, B.T., Qing, W., Packman, K., et al. (2006). Small-molecule MDM2 antagonists reveal aberrant p53 signaling in cancer: implications for therapy. Proc. Natl. Acad. Sci. U S A 103, 1888–1893. https://doi.org/10.1073/pnas.0507493103.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Finn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31, 46–53. https://doi.org/10.1038/nbt.2450.

Uroda, T., Anastasakou, E., Rossi, A., Teulon, J.-M., Pellequer, J.-L., Annibale, P., Passey, O., Inga, A., Chilón, I., and Marcia, M. (2019). Conserved pseudoknots in IncRNA MEG3 are essential for stimulation of the p53 pathway. Mol. Cell 75, 982–995.e9. https://doi.org/10.1016/j.molcel.2019.07.025.

Wei, Y., Li, X., Wang, Q., and Ji, H. (2012). IASeq: integrative analysis of allele-specificity of protein-DNA interactions in multiple ChIP-seq datasets. BMC Genomics. 13, 681. https://doi.org/10.1186/1471-2164-13-681.

Yasuda, K., Zhang, H., Loiselle, D., Haystead, T., Macara, I.G., and Mili, S. (2013). The RNA-binding protein Fus directs translation of localized mRNAs in APC-RNP granules. J. Cell Biol. 203, 737–746. https://doi.org/10.1083/jcb.201306058.

Yates, A., Beal, K., Keenan, S., McLaren, W., Pignatelli, M., Ritchie, G.R., Ruffier, M., Taylor, K., Vullo, A., and Flick, P. (2015). The Ensembl REST API: Ensembl data for any language. Bioinformatics 31, 143–145. https://doi.org/10.1093/bioinformatics/btu613.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfile: an R package for comparing biological themes among gene clusters. OMICS 16, 284–287. https://doi.org/10.1089/omi.2011.0118.

Zaccara, S., Tebaldi, T., Pederiva, C., Cirielli, Y., Biso, A., and Inga, A. (2014). p53-directed translational control can shape and expand the universe of p53 target genes. Cell Death Differ. 21, 1522–1534. https://doi.org/10.1038/cdd.2014.79.

Zhang, P., Kitchen-Smith, I., Xiong, L., Stracquadanio, G., Brown, K., Richter, P.H., Wallace, M.D., Bond, E., Sahgal, N., Moore, S., et al. (2021). Germline and somatic genetic variants in the p53 pathway interact to affect cancer risk, progression, and drug response. Cancer Res. 81, 1667–1680. https://doi.org/10.1158/0008-5472.CAN-20-0177.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. 10, 1523. https://doi.org/10.1038/s41467-019-09234-6.
## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT RESOURCE | SOURCE | IDENTIFIER |
|------------------|--------|------------|
| Antibodies       |        |            |
| Anti-PABP        | Abcam  | Cat# bb6125 [10E10] |
| Normal mouse IgG | Merck Millipore | Cat# 12-371 |
| Bacterial and virus strains |        |            |
| E. coli strain   | laboratory stock | DH5alpha |
| Chemicals, peptides, and recombinant proteins |        |            |
| TRIzol          | Thermo Fisher Scientific | Cat# 15596026 |
| Nutlin          | Cayman chemicals | Cat#10004372 |
| Doxorubicin      | Sigma Aldrich | Cat#25316-40-9 |
| Xba I           | New England Biolabs | Cat# R0145S |
| Nco I           | New England Biolabs | Cat# R3193S |
| T4 DNA ligase    | New England Biolabs | Cat# M0205S |
| Protein A Magnetic Beads | Thermo Fisher Scientific | Cat# 88845 |
| Critical commercial assays |        |            |
| TruSeq RNA Library preparation kit v2 | Illumina | Cat#RS-122-2001 |
| Agilent RNA 6000 Nano kit | Agilent | Cat#5067-1511 |
| RevertAid RT Kit | Thermo Fisher Scientific | Cat# K1691 |
| Dual Luciferase Reporter assay system | Promega | Cat# E1910 |
| Deposited data |        |            |
| Sequencing of total and polysomal RNA from normal and treated MCF7 cells | This paper | Bioproject: PRJNA693005 |
| TCGA SNP 6.0 array calls | Genomic Data Commons | http://portal.gdc.cancer.gov/legacy-archive |
| TCGA clinical information | Liu et al. (2018) | https://doi.org/10.1016/j.cell.2018.02.052 (Table S1) |
| MCF7 heterozygous SNPs (array data) | Barretina et al., 2021 | GEO GSM888366 |
| MCF7 heterozygous SNPs (wes data) | Reinhold et al. (2012) | https://doi.org/10.1158/0008-5472.CAN-12-1370 |
| RBPDB | http://rbpdb.cbr.utoronto.ca/ | https://doi.org/10.1093/nar/gkq1069 |
| Experimental models: Cell lines |        |            |
| MCF7 | ICLC | N/A |
| Oligonucleotides: Cloning primers |        |            |
| BRI3BP_F | Metabion | aacttagAGGTCAGCGCGCCGGCGGGGTCACAC |
| BRI3BP_R | Metabion | aacttagAACTTGAACCAATTGCCTATT |
| ATF6_852F | Metabion | aacttagATCATCAGCTAGCTGGGGAGT |
| ATF6_852R | Metabion | aacttagGCAACCCCCAAAAAGGTAC |
| ATF6_848F | Metabion | aacttagGCAGACAGCTGTAAAGG |
| ATF6_848R | Metabion | aacttagGCTGGAAAGCCAAAAGC |
| p21_UTR_ref_F | Metabion | CATGGGGCGGCTGGCAAAAGAAATCCCTGGGTTA |
| p21_UTR_ref_R | Metabion | CATGGGGCGGCTGGCAAAAGAAATCCCTGGGTTA |

(Continued on next page)
RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Alberto Inga (alberto.inga@unitn.it).

Materials availability
This study did not generate new materials.

Data and code availability
- RNA-seq BAM files have been deposited at BioProject under the accession number PRJNA693005.
- Any additional information required to reproduce this work is available from the Lead Contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines and culture conditions
MCF7 cells were cultured in standard RPMI (Lonza) supplemented with 10% FBS (brand), 100 units/ml penicillin, 100 mg/mL streptomycin antibiotic mix, and 2 mM glutamine. Cells were periodically tested to ascertain the absence of mycoplasma contamination. The authenticity of the parental cells was confirmed by a commercial genotypic service (BMR genomics, Padova, Italy). Doxorubicin was purchased from Sigma. Nutlin was purchased from Cayman chemicals.

| REAGENT RESOURCE | SOURCE | IDENTIFIER |
|-------------------|--------|------------|
| p21_UTR_ALT_F     | Metabion | CATGGGGTGCGTATTGGCTTTGGGCTGCTGTTTCCAGCTGCTGCAACACAGGAGGTTTTTCTCTGTCAGGCACCC |
| p21_UTR_ALT_R     | Metabion | CATGGGGCGCTGAACAGAAAGACCCCTGTGCTGTCAGACAGCAGGAGGAGAGGACCCAAAGACAAATAAGGACCC |

Oligonucleotides: PCR and sequencing primers

| Oligonucleotide | Source | Identifier |
|-----------------|--------|------------|
| GAPDH_F         | Metabion | TCCAAATCAAGTGCGGGA |
| GAPDH_R         | Metabion | AGTAGAGGCAGGATGATGT |
| ATF6_F          | Metabion | CCCGTATTTCTCAGGGGTGCT |
| ATF6_R          | Metabion | TCACCTCCCTAGTCTCCTGCT |
| ATF6_UTR_F      | Metabion | GAAACCTCCTCCCCGTGCT |
| ATF6_UTR_R      | Metabion | CAGAAGTGAAGGGGCAATCA |

Software and algorithms

| Software and algorithms | URL | DOI |
|-------------------------|-----|-----|
| ASEQ                    | https://bitbucket.org/CibioBCG/aseq/ | https://doi.org/10.1186/s12920-015-0084-2 |
| Ensembl rest API        | https://rest.ensembl.org/ | https://doi.org/10.1093/bioinformatics/btu613 |
| SHAPEIT                 | https://mathgen.stats.ox.ac.uk/ | https://doi.org/10.1016/j.ajhg.2013.09.002 |
| IMPUTE2                 | http://mathgen.stats.ox.ac.uk/impute/impute_v2.html | https://doi.org/10.1371/journal.pgen.1000529 |
| Survival R package      | https://cran.r-project.org/web/packages/survival/index.html | https://doi.org/10.1007/978-1-4757-3294-8 |
| TESS                    | https://www.cbil.upenn.edu/tess | https://doi.org/10.1002/0471250953.b0206s21 |
| VariantAnnotation R package | https://bioconductor.org/packages/release/bioc/html/VariantAnnotation.html | https://doi.org/10.1093/bioinformatics/btu168 |
METHOD DETAILS

SNP status in MCF7 cells
We retrieved from the literature genotype assays for MCF7. Two different public datasets (GEO and NCI-60) were used containing information about homozygosis or heterozygosis conditions for SNP alleles in MCF7, the unique identifier, the relative and alternative base in specific single positions. One SNP-array based file (GSM888366) contains 756,260 SNPs with the information about the reference and alternative bases (Barretina et al., 2012), while the second exome-based file (Reinhold et al., 2012) contains 85,593 SNPs. The SNPs in common (N = 6,454) were compared to verify the consistency among the two datasets: more than 91% of the SNP calls were consistent. This concordance suggested the possibility to merge the two files removing the data with opposite information. A first screening was performed through the elimination of SNPs called as homozygous in the two datasets, but with a different base call: i.e. reference “0/0” vs alternative “1/1”. Together, the final file presents 11,544 heterozygous SNPs of which 1,802 in 3’UTR and 729 in 5’UTR. Variants were annotated using VariantAnnotation R package (Obenchain et al., 2014).

Differential allele-specific expression analysis
To identify allelic imbalances between polysomal and total RNA fractions a computational approach that exploits SNP allelic fractions variability across biological replicates was implemented. Given an experimental condition C, SNP allelic fractions variability for condition C is calculated as:

\[
V_c = \text{mean}\left(\left\{\frac{\text{snp}_i^R - \text{snp}_i^A}{\text{snp}_i^A} \text{ such that } 1 \leq i \leq N \text{ and } 1 \leq j, k \leq M \text{ and } j < k\right\}\right)
\]

where N is the number of considered SNPs, M is the number of biological replicates, and snp_i^R is the allelic fraction of SNP _i_ computed from the RNA-seq BAM file representing the biological replicate _R_ using ASEQ pileup module (Romanel et al., 2015) and defined as:

\[
\text{snp}_i^R = \frac{\#\text{ALT}}{\#\text{ALT} + \#\text{REF}}
\]

where #ALT and #REF correspond to, respectively, the number of reads supporting the SNP _i_ alternative base and the number of reads supporting the SNP _i_ reference base. Of note, in the pileup computation only reads and bases with quality above 20 and a minimum of local depth of coverage of 10X were considered.

Now, given variabilities values _V_p_oly and _V_t_ot computed, respectively, from RNA-seq replicate samples of polysomal and total RNA fractions, we define as tranSNPs all SNPs satisfying the following condition:

\[
\forall_{j \in [1, \cdots, M]} \left(\text{snp}_{j}^{\text{poly}} + \text{V}_{\text{poly}} < \text{snp}_{j}^{\text{tot}} - \text{V}_{\text{tot}}\right) \lor \forall_{j \in [1, \cdots, M]} \left(\text{snp}_{j}^{\text{poly}} - \text{V}_{\text{poly}} > \text{snp}_{j}^{\text{tot}} + \text{V}_{\text{tot}}\right)
\]

Survival analysis using TCGA data
From our list of 147 tranSNPs an extended list of variants was built retrieving variants in strong linkage disequilibrium (LD). We queried the Ensembl rest API version GRCh37 (Yates et al., 2015) and retrieved all the variants with _r^2_ > 0.8 in a genomic window of 500kbp using general population data. For a total of 3,003 variants, distributed among 120 LD blocks, we were able to retrieve genotype calls from TCGA data. Specifically, raw TCGA genotype calls were downloaded from TCGA legacy data portal (portal.gdc.cancer.gov/legacy-archive) and only genotypes with confidence score larger than 0.1 were retained. Genotypes were analyzed with SHAPEIT v2 (Delaneau et al., 2013) to infer haplotype structure and optimize genotype content information for the imputation process. Variants were imputed using IMPUTE v2.3.2 (Howie et al., 2009) against a reference panel built from 1,000 Genomes Project data. The final extended list of variants was used to perform variant-specific survival analysis using TCGA breast cancer survival data (Liu et al., 2018). Overall Survival (OS), Disease-Specific Survival (DSS), Disease-Free Interval (DFI) and Progression-Free Interval (PFI) were considered in our analysis. Dominant and recessive models were built, respectively, grouping together patients with homozygous alternative and heterozygous genotypes (AA vs AB+BB) and homozygous reference and heterozygous genotypes (AA+AB vs BB). Kaplan-Meier survival curves were built for each variant and Cox proportional hazards regression models were used and inspected. To limit false positive results, for each considered outcome, association results demonstrating a p value <0.05 (minimum between Log rank test p value and Wald test p value was considered) were aggregated at the level of LD blocks and >5% of variants reproducing the association signal in a block were required to nominate the block as associated. Survival analyses were performed using survival R package (Therneau and Grambsch, 2000).
Identification of SNPs disrupting putative RBP consensus motifs

We retrieved 552 human RNA binding consensus motifs from RBPDB database (Cook et al., 2011). Motif analysis was performed on the list of 33 UTR tranSNPs resulting from the survival analysis. For each variant, two sequences of length 60bp with the SNP reference or alternative allele in the middle were built using GRCr37 reference genome. All sequences were then collected in a FASTA file and TESS software (Schug, 2008) was ran on this file to compute motifs scores. TESS tool provides a set of log-likelihood-ratio-based scores, among which we used the score \( L_{a} \) which represent the log-odds ratio of the match, and the score \( L_{m} \) which represents the maximum possible log-odds ratio for a match from a given consensus motif. For each motif overlapping a variant, results were collected only if the score \( L_{a} \) for the reference or the alternative allele was at least 3 and it was computed on the forward (transcribed) strand. Confident results were filtered by considering only entries having a \( L_{a}/L_{m} \) value greater than 0.5 in at least one of the two conditions. In addition, only variants presenting an absolute ratio between reference and alternative \( L_{a} \) scores greater than 10% or presence of a motif with the reference but not with the alternative (or vice-versa) were retained.

Polysomal profiling and RNA-sequencing

MCF7 parental cells were seeded in P100 dishes and treated at about 70% confluence by 1 \( \mu \)M doxorubicin or 10 \( \mu \)M Nutlin, for 16 hours. DMSO was used for treatment control. Cytoplasmic lysates were obtained and polysome fractionations performed as described in (Dassi et al., 2015; Provenzani et al., 2006; Rizzotto et al., 2020; Rossi et al., 2019; Zaccara et al., 2014). Briefly, cytoplasmic lysates were loaded on a 15 ml, 10-50% sucrose gradient, ultra-centrifuged (40k rpm for 100’) and fractionated with an automated fraction collector (1 ml per fraction Teledyne ISCO), tracing the 254\( _{\text{nm}} \) absorbance. All the lighter fractions containing subpolysomal fractions (from the top to the gradient up to the fractions corresponding to the 80S monosome) were pooled in a tube. Heavier fractions corresponding to polysomal RNAs (two or more ribosomes) were also pooled in a separate tube. RNA was purified by extraction with 1 volume of phenol-chloroform. The aqueous fraction was recovered after centrifugation, transferred to a new tube where RNA precipitation was induced by the addition of 1 volume of isopropanol and a subsequent centrifugation step. Pellets were washed once using 70\% V/V ethanol to remove phenol contaminations and resuspended in sterile, RNase free water. Total RNA samples were instead obtained by TRIzol (ThermoFisher) extraction of a separate cell population prepared in parallel. RNA preparations were quantified by NanoDrop microvolume spectrophotometer. Purity and integrity were checked by Agilent 2100 Bionalyzer electrophoretic runs exploiting the Agilent RNA 6000 Nano kit. Sampling with RIN over 6 were used to prepare libraries for RNA sequencing. Two biological replicates for each RNA type (total and polysomal) and condition (mock, doxorubicin, Nutlin) were sequenced. Libraries were prepared following the TruSeq RNA Library preparation kit v2 protocol (Illumina), starting from 1\( _{\text{pg}} \) of input RNA. Paired end (100 bp) sequencing was performed on a HiSeq 2500 (Illumina). FASTA files were aligned to the reference genome GRCr37 using TopHat v2.0.10 (Kim et al., 2013), resulting in BAM files with an average of ~32 million unique mapping (mapping quality >5) reads. Differential expression analysis was performed with Cuffdiff (Trapnell et al., 2013) comparing Nutlin and doxorubicin samples against mock samples, considering both total and polysomal RNA fractions. Pathway enrichment analysis of deregulated genes in all tested comparisons was performed using clusterProfiler R package (Yu et al., 2012) exploiting KEGG database (Kanehisa et al., 2020). Comparison of multiple gene lists obtained from condition-associated tranSNPs was performed with Metascape (Zhou et al., 2019).

Cloning 5’UTR or 3’UTR allelic variants for gene reporter assays

The commercial Firefly luciferase plasmid pGL4.13 was used to clone the two rs1055472 alleles BRI3BP 5’UTR as well as two regions of the ATF6 3’UTR containing different set of SNPs, using a PCR based approach that introduced the XbaI restriction enzyme site. MCF7 cDNA was used as template, given the heterozygous state for the SNPs of interest. The cloned regions correspond to genomic coordinates chr1:125509977-125510349 (hg19) for BRI3BP and chr1:161930870-161931403 or chr1:161931723-161932422 for ATF6. Primers used for the cloning are listed in the key resources table, where the lowercase bases are the restriction site sequence included in the primers after two initial adenines. The pGL3 basic vector was instead used to clone the two rs2395655 alleles of the CDKN1A 5’UTR exploiting the unique NcoI site that overlaps with the firefly start codon and using complementary primers containing the 5’UTR sequence of interest for each SNP allele that were annealed prior to ligation to generate NcoI sites at both ends. After in vitro ligation (T4 DNA ligase, NEB), E. coli competent cells were transformed and colonies picked to recover plasmids based first on a colony PCR approach, followed by plasmid recovery from
positive colonies followed by electrophoresis migration and digestion pattern analysis. The successful cloning of both SNP alleles for the two targets was then verified by Sanger sequencing (Eurofins Genomics) of independent candidate colonies.

**Dual-luciferase assays**

pGL4.13- or pGL3-derivative plasmids were transiently transfected along with the pRL-SV40 Renilla control luciferase in a 3:1 ratio (250 ng plus 50 ng for 24 well format) using Mirus-LTI transfection reagents following the manufacturer’s protocol. When needed, cells were treated with doxorubicin or Nutlin for 16 hours, starting 24 hours after transfection. Luciferase assays were performed using Dual-Luciferase™ Reporter (DLR™) Assay System (Promega), following the provided protocol. Light units were measured using a Tecan M200pro multilable plate reader through three technical replicates of well. Presented in the graphs are the average Relative Light Units (Firefly/Renilla) and standard deviations of three biological experiments, compared using a two-tailed, equal variance Student’s t-test.

**RIP assays**

Ribonucleoprotein particles immunoprecipitation (RIP) experiments were performed following a published protocol with some modifications described recently (Keene et al., 2006; Uroda et al., 2019), starting from $4 \times 10^7$ MCF7 cells. After lysis, the supernatants were collected and 1% of each sample was set aside as input while the remaining was incubated for 4 hours at 4°C with protein A magnetic beads (Thermo Fisher Scientific) coated either with 3 μg of an anti-PABPC1 antibody (Abcam) or with 3 μg of normal Rabbit IgG (Santa Cruz). After the washing steps, RNA was isolated from Input and IP samples using TRIzol (Thermo Fisher Scientific) and converted to cDNA using the RevertAid First Strand cDNA Synthesis Kit and standard protocol (Thermo Fisher Scientific). qPCR reactions were performed using the qPCRBIO SyGreen Mix (PCR Biosystems) master mix on a CFX384 thermal cycler (Biorad). GAPDH was tested as a comparison. Two pairs of primers were used for ATF6, of which one targeting exon 9, and the second targeting the 3UTR sequence comprising the rs2499847 SNP whose alleles are predicted to impact a PABPC1 binding site (Table S7). The primers sequences can be found in the key resources table.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Statistical tests applied throughout the study are specified in results, figure legends and in the methods accordingly. Given a list of variants, LD blocks were determined identifying the number of cliques in a graph were nodes are variants and edges are present between two nodes when the corresponding variants are in LD with an $r^2 > 0.8$. In all considered cases the number of cliques corresponded to the number of connected components. Luciferase assay data were analyzed using a 2-way ANOVA with Sidak’s multiple comparison test. RIP assays were analyzed using a two-tailed unpaired t-test.