Environmental perturbations lead to extensive directional shifts in RNA processing

A. L. Richards[1,†], D. Watza[1], A. Findley[1], A. Alazizi[1], X. Wen[2], A. A. Pai[3,†], R. Pique-Regi[1,4,†], F. Luca[1,4,†]

[1]Center for Molecular Medicine and Genetics, Wayne State University
[2]Department of Biostatistics, University of Michigan
[3]Department of Biology, Massachusetts Institute of Technology
[4]Department of Obstetrics and Gynecology, Wayne State University

[†]To whom correspondence should be addressed:
fluca@wayne.edu, rpique@wayne.edu, athma@mit.edu, allison.richards2@wayne.edu

March 23, 2017

Abstract

Environmental perturbations have large effects on both organismal and cellular traits, including gene expression, but the extent to which the environment affects RNA processing remains largely uncharacterized. We assessed changes in RNA processing events across 89 environments in five human cell types and identified 15,628 event shifts (FDR = 15%) comprised of eight event types in 4,567 genes. Many of these changes occur consistently in the same direction across conditions, indicative of global regulation by trans factors. Accordingly, we demonstrate that environmental modulation of binding for splicing factors predicts shifts in intron retention, while binding for transcription factors predicts shifts in AFE usage in response to specific treatments. Further, we validate these findings profiling chromatin accessibility with ATAC-seq in LCLs exposed to selenium. Together, these results demonstrate that RNA processing is dramatically changed in response to specific environmental perturbations through different mechanisms regulated by trans factors.

Introduction

Variation in gene expression has long been associated with cellular and organismal phenotypes. For example, studies have found that gene expression in blood and bronchial epithelial cells differs among individuals with asthma [1, 2, 3, 4]. Such differences in gene expression occur in specific cellular pathways, such as the glucocorticoid response pathway [1, 5, 6, 7], leading to the general usage of glucocorticoids to treat asthma. These studies, and others, have demonstrated that variation in gene expression plays a role in complex traits and cellular responses [8, 9, 10, 11, 12]. More recently, however, researchers have begun to assess the impact of alternative mRNA isoform usage on phenotypes. Previous studies have found that RNA processing, leading to differential isoform usage, is different in certain diseases such as Alzheimer's disease and several forms of cancer [13, 14, 15, 16, 17]. Furthermore, studies have identified global shifts in exon usage associated with developmental or diseased cellular states. For instance, shorter 3' untranslated region (UTR) isoforms are prevalent in proliferating or cancerous cells [18, 19]. Cancer is also associated with increased retention of introns [20, 21].

Li *et al.* recently identified genetic variants associated with inter-individual variation in mRNA splicing and identified almost 2,900 splicing Quantitative Trait Loci (QTLs). Further, they showed that splicing QTLs are also enriched for genetic variants associated with several complex traits in Genome-Wide Association Studies (GWAS), demonstrating the potential importance of splicing misregulation in complex traits [22]. Previous work from our lab and others have shown that gene-by-environment interactions can impact both gene expression and complex traits [23, 24, 25, 26, 27, 28]. While splicing QTLs have been identified both in humans and mice [22, 29, 30, 31], less is known about how gene-by-environment interactions may affect RNA processing. The first step to address this question is to characterize RNA processing in response to environmental perturbations.

RNA processing is regulated in response to certain environmental stimuli, such as cancer therapy drugs, nutrient starvation and infection [32, 33, 34, 35] some of which influence cell viability [36, 37, 38]. For example, UV exposure leads to differential isoform usage in the gene *BCL2L1*, which is involved in the regulation of apoptosis. UV leads to increased abundance of Bcl-x_s which favors apoptosis as opposed to Bcl-x_l which is anti-apoptotic [39]. Other studies have demonstrated widespread, directed changes in the regulation of RNA processing. Infection with *Listeria monocytogenes* and *Salmonella typhimurium* led to increased inclusion of cassette exons and shorter 3'UTRs genome-wide [35]. The longer versions of 3'UTRs that were shortened were found to be enriched with particular microRNA binding sites, suggesting that the RNA processing shift leading to shorter 3'UTRs may be a way for these genes to evade down-regulation following infection. Despite the fact that these studies have increased our understanding of factors that influence changes in RNA processing, they have investigated only a limited number of environments. Cataloguing and characterizing RNA processing changes across many environments, in a tightly controlled study using specific treatments, is necessary to increase our understanding of the cellular mechanisms leading to variation in RNA-processing, including which aspects are common across many environments and which are specific to certain perturbations.

Our study aimed to systematically assess the impact of a broad range of environmental perturbations on the regulation of RNA processing. We measured RNA processing patterns in five cell types across over 30 treatments, corresponding to a total of 89 cellular environments with at least 3 biological replicates

and additional technical replicates (297 RNA-seq libraries in total with 130M reads on average) [23]. The treatments represent compounds to which we are exposed in daily life, ranging from metal ions and vitamins to allergy medication. This work demonstrates the extent of alternative RNA processing in response to a wide range of specific environmental perturbations, indicating molecular mechanisms by which trans factors influence this process.

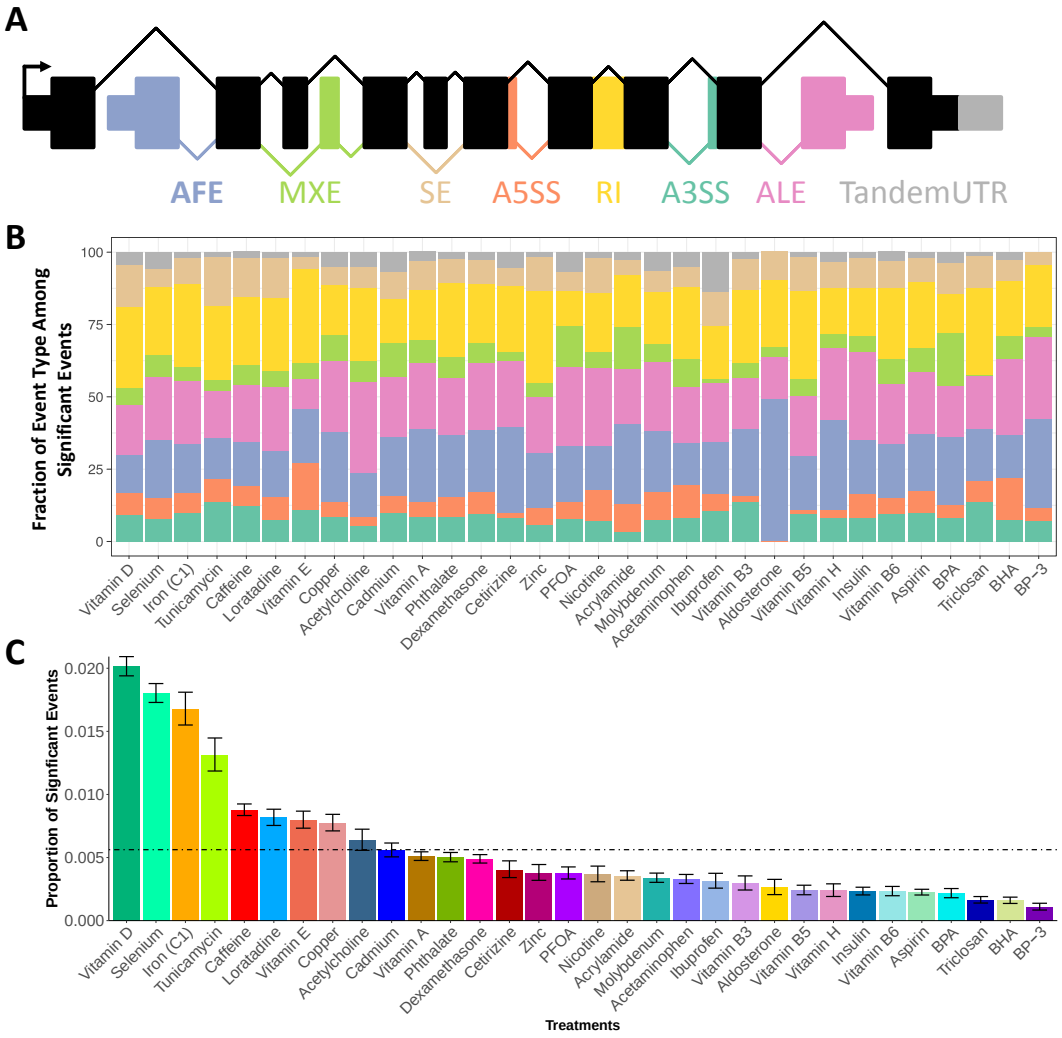

Figure 1: RNA Processing Events and Gene Expression Changes Following Treatment. A) Diagram depicting the 8 types of RNA splicing changes characterized in this work: alternative first exon, mutually exclusive exon, skipped exon, alternative 5' splice site, retained intron, alternative 3' splice site, alternative last exon, and tandem untranslated region. B) Graph showing the estimated proportion of each event type within a given treatment resulting from a logistic model. C) Proportion of significant changes in events over the total number of events that were tested in that treatment. Each bar combines all cell types treated with the compound. Error bars denote the standard error from a binomial test. The dotted line indicates the average proportion of significant events across all treatments.

Results

External stimuli induce environment-specific shifts in RNA processing

Using high-throughput RNA sequencing, we identified 32 compounds that induce gene expression changes in 32,451 genes in 5 different cell types (a total of 89 environments) [23] (Supplementary File 1 - Section 1, Table S1). In order to identify changes in RNA processing, we utilized the probabilistic framework implemented in the software Mixture of Isoforms (MISO) [40], which characterizes changes in exon usage by calculating a percent spliced in (PSI, Ψ) value. The Ψ value is calculated by quantifying the fraction of reads specific to an inclusion isoform, specifically, reads aligning to the alternative exon or its junctions. Instead of entire isoforms, which may involve multiple RNA processing mechanisms that are convolved together, we focused on exons that are tied to known RNA processing mechanisms. We focused on events that involve known curated isoforms (see methods), rather than novel isoforms, and characterized variation in RNA processing events across different environments. This allowed us to and learn about cis- and trans-acting mechanisms leading to the RNA processing response. Specifically, we characterized changes in eight event types: skipped exons (SE), retained introns (RI), alternative 3' or 5' splice sites (A3SS, A5SS), mutually exclusive exons (MXE), alternative first or last exons (AFE, ALE), and tandem untranslated regions (TandemUTR) (Figure 1A, Supplementary File 1 - Figure S1 shows a treatment color key used throughout the manuscript). Across all conditions, we identified 15,628 changes in RNA processing, representing a unique set of 9,064 events that significantly differ between at least one treatment and control conditions (Table 1, Supplementary File 2 and Supplementary File 1 - Figure S2). Each significant change in RNA processing event was identified based on RNA sequencing data across cell lines derived from three unrelated individuals (example in Figure 2 and at `http://genome.grid.wayne.edu/RNAprocessing`). Across all environments, the most abundant event types with shifts were RI, AFE and ALE (relative to the number of sites tested), while the least abundant was A5SS (Figure 1B).

Table 1: RNA processing events across 89 environments.

	SE	**A3SS**	**A5SS**	**RI**	**MXE**	**AFE**	**ALE**	**TandemUTR**
# of sig. changes	2286	813	522	2173	580	5193	3801	260
# of sig. events	1561	564	283	1196	363	2747	2162	188
# of events tested	19382	6316	4270	4009	4173	11505	6941	2358
% sig. events	8.1	8.9	6.6	29.8	8.7	23.9	31.1	8

As we studied events in a given cell type across conditions, we found treatment-specific shifts in RNA processing, resulting in vast differences in the number and type of event shifts (examples in Figure 3). We found a wide range in the number of significant shifts across treatments, with vitamin D producing the highest number (2,653 events) and BP3 leading to the lowest number of significant shifts in RNA processing (66 events) (average = 488 events, 0.5% of events tested) (Figure 1C, cell type-specific changes analyzed in Supplementary File 1 - Section 2). The number of RNA processing changes in each environment is not correlated to the sequencing depth of the library (Spearman's $\rho = 0.12$, $p = 0.07$; Supplementary File 1 - Section 3, Figure S3A), but it is correlated to the number of differentially expressed genes in each

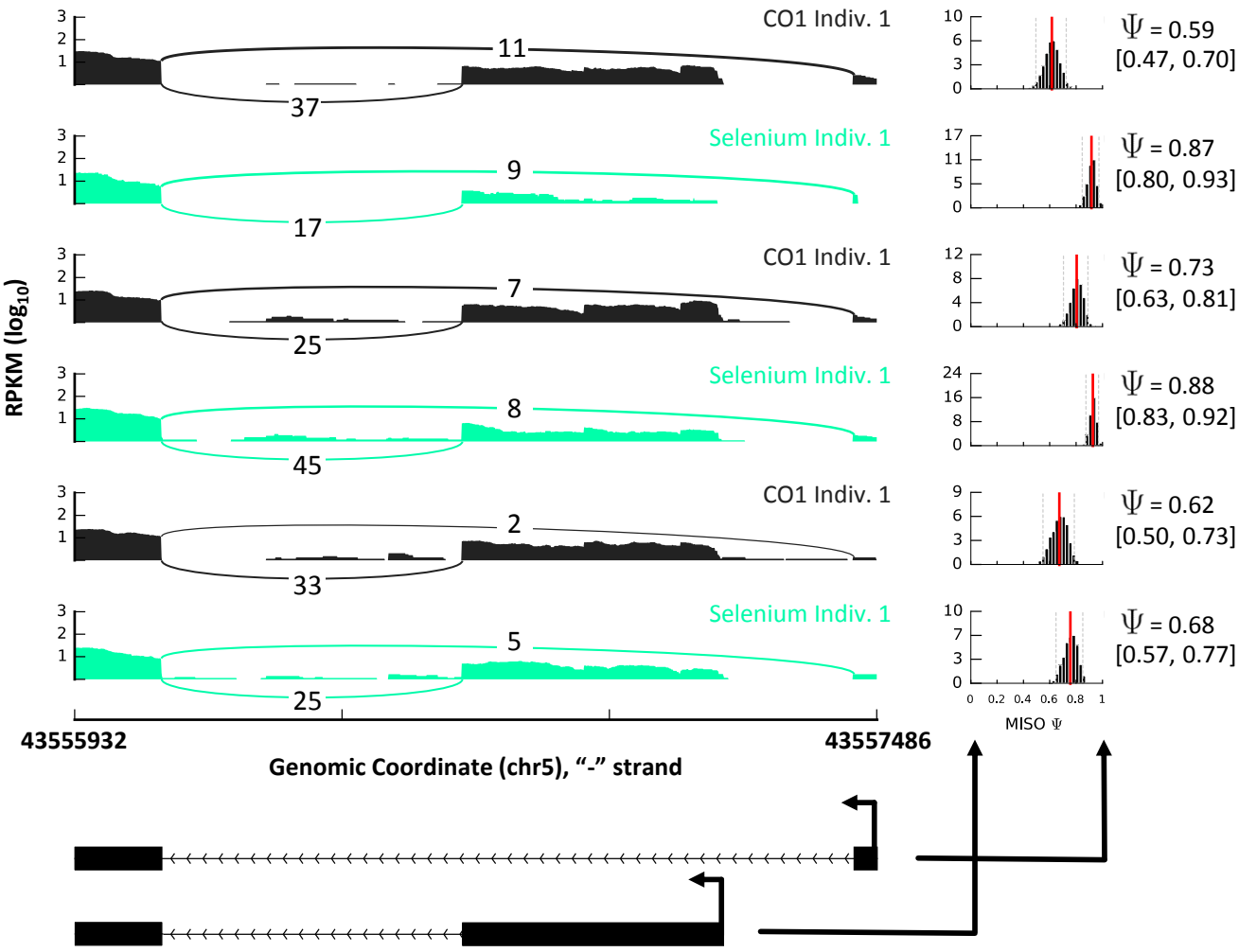

Figure 2: Sashimi plot showing AFE change following selenium treatment in LCLs across 3 unrelated individuals. The plots on the right show the PSI value for each sample with confidence intervals. The plots to the left show the read coverage in each exon with model of this region below. High Ψ indicates preference for the upstream AFE.

environment (Spearman's $\rho = 0.63$, $p < 0.01$; Supplementary File 1 - Figure S3B), suggesting the same underlying mechanism inducing changes in RNA processing and in overall gene expression.

In addition to differences in the overall number of splicing changes, we also found differences in relative number of changes in certain event types (Figure 1B). While changes in AFEs represent the greatest overall number of changes across environments, there is substantial variation in the extent to which each event type changes within each treatment (Figure 1B). We utilized a generalized linear model to determine the proportion of event types among significant event shifts in a given treatment. With this model, we identified 5 treatments that showed enrichment for an event type, including vitamin D, tunicamycin, caffeine, vitamin

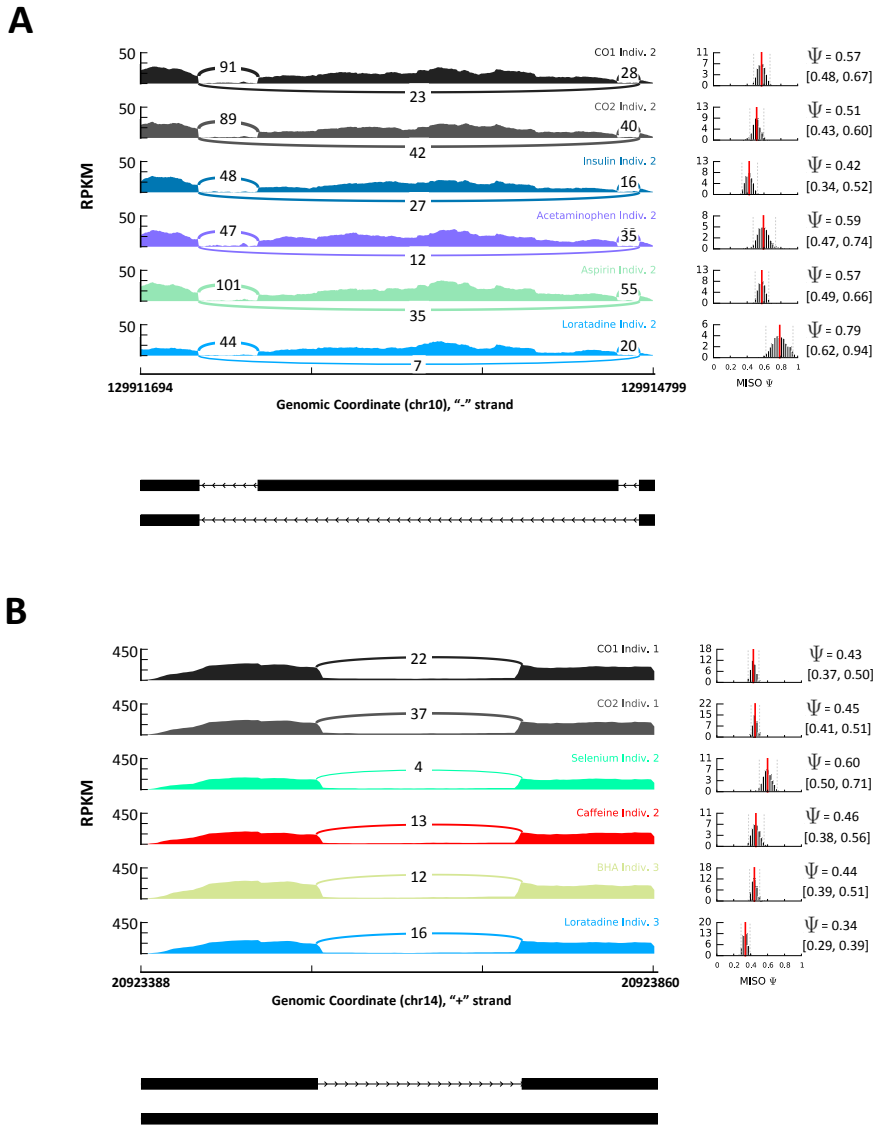

Figure 3: Example of RNA Processing Across Treatments in One Individual. The plots on the right show the Ψ value for each sample with confidence intervals. The plots to the left show the read coverage in each exon with a model of this region below each read coverage plot. A) Sashimi plot showing SE shift in *MKI67* in melanocytes following exposure to 4 treatments (insulin, acetaminophen, aspirin, and loratadine) and both controls. B) Sashimi plot showing SE shift in *APEX1* in melanocytes following exposure to 4 treatments (selenium, caffeine, BHA and loratadine) and both controls.

E and acrylamide. For example, vitamin E is depleted for ALE and SE while acrylamide is enriched for A5SS, AFE and MXE (Figure 1B).

For the most part, only one event type is found in a given gene under a certain condition. Specifically, an average of 3.7% of significant events occurred concurrently with another event type (Supplementary File

1 - Section 4). Together, these results demonstrate a widespread abundance of RNA processing changes in response to environmental perturbations and suggest that, similar to changes in gene expression, there are a large number of changes in RNA processing that are also cell type- and treatment-specific.

Direction of RNA Processing Shifts and Gene Expression Changes

Regulation of RNA processing events in response to environmental perturbations may be mediated by trans factors that impact many RNA processing events of the same type, or by cis-acting regulatory sequences that would impact each event separately. To investigate these two mechanisms we considered global shifts in RNA processing. Specifically, among the 8 event types with changes following treatment, 5 can be thought of directionally: SE, RI, AFE, ALE, and TandemUTR. Each event can be given a sign such that a positive $\Delta\Psi$ (same as positive Z-score) means an increase in usage of the skipped exon, upstream AFE, downstream ALE, longer TandemUTR or intron retention in the treatment sample as compared to control (Figure 4A).

When we focused on treatments with at least 30 significant RNA processing shifts of a certain event type, 23% of treatments showed a correlation ($p < 0.05$) between changes in RNA processing and changes in gene expression (examples in Supplementary File 1 - Figure S4, Table S2). For example, vitamin B6 induced a positive correlation between AFE and gene expression (Spearman's $\rho = 0.43$, $p = 0.005$) (Supplementary File 1 - Figure S4A). Specifically, genes shifting towards usage of the upstream AFE following vitamin B6 treatment also have increased expression in the treatment samples. On the other hand, tunicamycin leads to the opposite effect: increased expression following tunicamycin is found in genes that utilize the downstream AFE (Spearman $\rho = -0.34$, $p = 0.005$) (Supplementary File 1 - Figure S4B). These data suggest that cells respond to specific environmental perturbations with concerted shifts in RNA processing events and gene expression.

Coordinated RNA Processing Shifts Across Cellular Environments

To investigate the possibility of shared cis-acting regulatory mechanisms across various environments, we considered whether the same event changes in different environments. We found a greater proportion of shared event shifts in a cell type across different treatments (19% on average, Supplementary File 1 - Section 5 and Section 6, Figure S5A and S6A), as compared to the same treatment across cell types (1-6%, Supplementary File 1 - Figure S5B and S6B). In both cases, $> 71\%$ of sites showed consistent direction of shift. Furthermore, genes with shifts in RNA processing for multiple treatments in the same cell type were enriched for GO terms related to three cellular processes: translation, RNA processing and cellular metabolism (Supplementary File 1 - Table S3). These data suggest that these three functions are regulated by RNA processing as a common response to multiple environmental perturbations.

Though we did not find a large percentage of shared event shifts across treatments and cell types, we investigated whether the global shifts in events had consistent direction across environments suggesting a shared trans-acting mechanism of change. When studying RI across all environments, we identified 20 environments with a proportion of positive events significantly different from the expected proportion of 50% (Figure 4B), thus showing enrichment for intron retention as compared to the control. These results

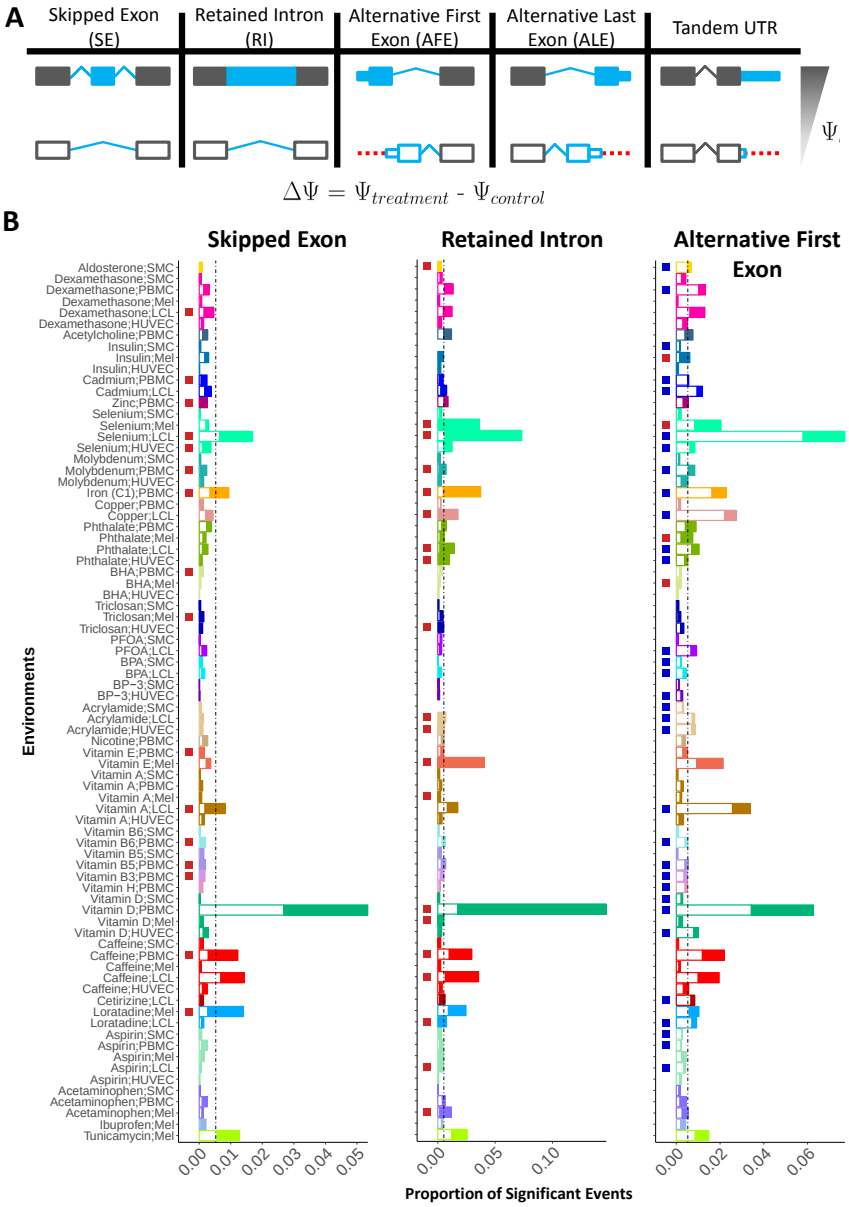

Figure 4: Direction of Shift in Events Following Treatment. A) Schematic of direction of event shifts for a given Ψ and $\Delta\Psi$. Shown for 5 event types. B)These plots indicate the direction of shift for 3 event types: SE (left), RI (middle) and AFE (right). Each plot shows 78 environments for which these events were tested. The height of each bar shows the proportion of significant event shifts for each environment. Each bar is then broken in two with the shaded region showing the proportion of the significant changes that shifted towards a positive $\Delta\Psi$ (inclusion of exon, intron or upstream AFE) while the white region of each bar is the proportion of sites with a negative $\Delta\Psi$. The column of boxes shows if there is a departure from the expected 50:50 for positive to negative $\Delta\Psi$ (tested using a binomial test). Red denotes enrichment for $\Delta\Psi > 0$ and blue for $\Delta\Psi < 0$).

suggest a common mechanism for intron retention in cells that respond to changes in the environment. For example, even though vitamin D causes many more changes in alternative splicing in PBMCs, all cell types trend towards retaining introns following vitamin D treatment. This can be more clearly seen when considering all events (not just significant events), where all 4 cell types show a shift toward more positive values in their ECDF (Kolmogorov-Smirnov (KS) test $p < 0.05$) denoting higher $\Delta\Psi$ values, retaining of introns, following vitamin D treatment (Supplementary File 1 - Figure S7A and B).

For RI and SE, we found that all treatments led to shifts in the same direction: inclusion of the skipped exon and intron (Figure 4B). However, for AFE, while most treatments led to usage of the downstream AFE (negative Z-scores), three treatments in melanocytes were significantly enriched for shifts to the upstream AFE (positive Z-scores) (Figure 4B). For example, insulin leads to a shift toward the downstream AFE in SMCs but a shift toward the upstream AFE in melanocytes. This is also apparent when we consider all event shifts in these environments (KS test $p < 0.05$) (Supplementary File 1 - Figure S7C and D). Only treatments in melanocytes show this trend while all other cell types that are represented in the 33 environments are enriched for the downstream AFE. These data suggest a melanocyte-specific mechanism that leads to greater usage of upstream TSS following environmental perturbations. Both ALE and TandemUTR also showed deviation from the expected 50:50 ratio of positive to negative events but the trend was less clear (Supplementary File 1 - Figure S8). These results demonstrate that global shifts in RNA processing events can be determined solely by the treatment or by the combined effect of treatment and cell type. Furthermore, these shifts are not dependent on splice site strength directly (Supplementary File 1 - Section 7), thus suggesting that trans-acting factors lead to global shifts in RNA processing across environmental perturbations.

Environmental shifts in SE and RI are mediated by changes in splicing factors expression and binding

In order to elucidate the specific factors involved in the global shifts in SE and RI events, we focused on factors likely to influence RNA processing, specifically splicing factors. We quantified the gene expression changes of splicing factors across all environments to determine if there was a correlation to the number of positive (inclusive) RNA processing shifts. The underlying hypothesis is that shifts in exon usage may be explained by splicing factors that: 1) have activity largely mediated by changes in gene expression, and 2) have the same influence over splicing in all treatments.

We identified 9 splicing factors (of 174 tested) with changes in gene expression correlated with percent positive events for TandemUTR, AFE or RI (BH FDR $< 5\%$, example in Figure 5A and B, Supplementary File 1 - Table S4). Notably, we identified *HNRNPF* which is negatively correlated with TandemUTR events. This suggests that the increased expression of *HNRNPF* under treatment conditions leads to shortening of TandemUTRs (more negative TandemUTR events). Previous work has shown that *HNRNPF* is an important factor in determining polyadenylation site [41]. Furthermore, our result is consistent with others who show that *HNRNPF* expression is correlated with shorter 3'UTRs following infection [35]. Our results suggest that the changes in *HNRNPF* expression is a mechanism of regulation of UTR length not only following infection, but more generally in response to a large number of environmental perturbations. For example, vitamin D decreased expression of *HNRNPF* and led to longer UTRs.

While we identified global correlations between splicing factor gene expression changes and splicing changes following treatment, many factors may influence RNA processing differently following various treatments and so we may miss an effect by investigating common expression patterns across environments. Also, some factors are known to have different effects depending on binding location and not necessarily on overall gene expression. For example, when SR proteins (serine-arginine proteins) bind upstream of 5' splice site, they induce splicing but do not have the same effect when bound in the intron [42]. With this in mind, we asked whether predicted binding sites of splicing factors may explain SE and RI, two of the most abundant splicing changes in our dataset. First, we characterized motifs that are present upstream, downstream or within the alternative unit (exon for SE or intron for RI). We, then, utilized an elastic-net regularized generalized linear model (GLM-NET) to predict splicing changes in 8 environments with greater than 100 significant event shifts (5 with SE, 3 with RI), based on the splicing factors binding motif occurrences. When studying the model as a whole, we found that area under the curve (AUC) for each environment ranges from 0.67 for PBMCs exposed to iron to 0.86 for PBMCs exposed to vitamin D, suggesting that binding of splicing factors is important for determining changes in splicing following treatment, but the impact differs across cellular environments (Figure 5C). We also found that the genomic location of a binding site, relative to the splicing event, is an important predictive feature. For example, a motif for *RBM8A* (M054_0.6 from RNAcompete [43]) is a part of the predictive model of SE in LCLs treated with vitamin D but only when the motif is located in the upstream intron. This demonstrates the positional effect of binding that others have characterized for some splicing factors [42, 44, 45, 46] and expands its importance across a large number of environmental perturbations.

Effect of transcription factor expression and binding on AFE

Only two splicing factors had gene expression changes correlated with shifts in AFE, which is the most abundant type of RNA processing shift in our dataset. This is likely due to different molecular mechanisms underlying each event type. To investigate the major mechanisms responsible for shifts in TSS usage, we focused on transcription factors. Similar to our analysis with splicing factors, we first hypothesized that shifts in AFE could be the consequence of changes in gene expression for transcription factors that promote usage of either the upstream or the downstream TSS and have similar effects in all environments.

We identified 10 (out of 1,343) transcription factors whose change in expression is correlated with shifts in AFE (BH FDR $< 5\%$) (example in Figure 5D and E, Supplementary File 1 - Table S5). Interestingly, when we analyzed all RNA processing event types, we also identified 2 transcription factors whose gene expression was correlated to shifts in TandemUTR usage (BH FDR $< 5\%$) suggesting a link between the start of transcription and poly-A processing, as has been suggested by previous work [48, 49, 50, 51]. Together, these results suggest that transcription factor binding influences the choice of TSS leading to a consequent shift in alternative first exon usage.

To directly determine the effect of TF binding on AFE shifts, we then utilized transcription factor footprints identified in DNase-seq data from ENCODE and the RoadMap Epigenomics [52, 47, 53] to predict shifts in AFE usage in 12 environments. We used footprints from more than 150 cell types to better capture a wider range of cellular environments, as determined by tissue of origin or culturing conditions. To predict AFE shifts, we considered the number of footprints present within 1000bp in either

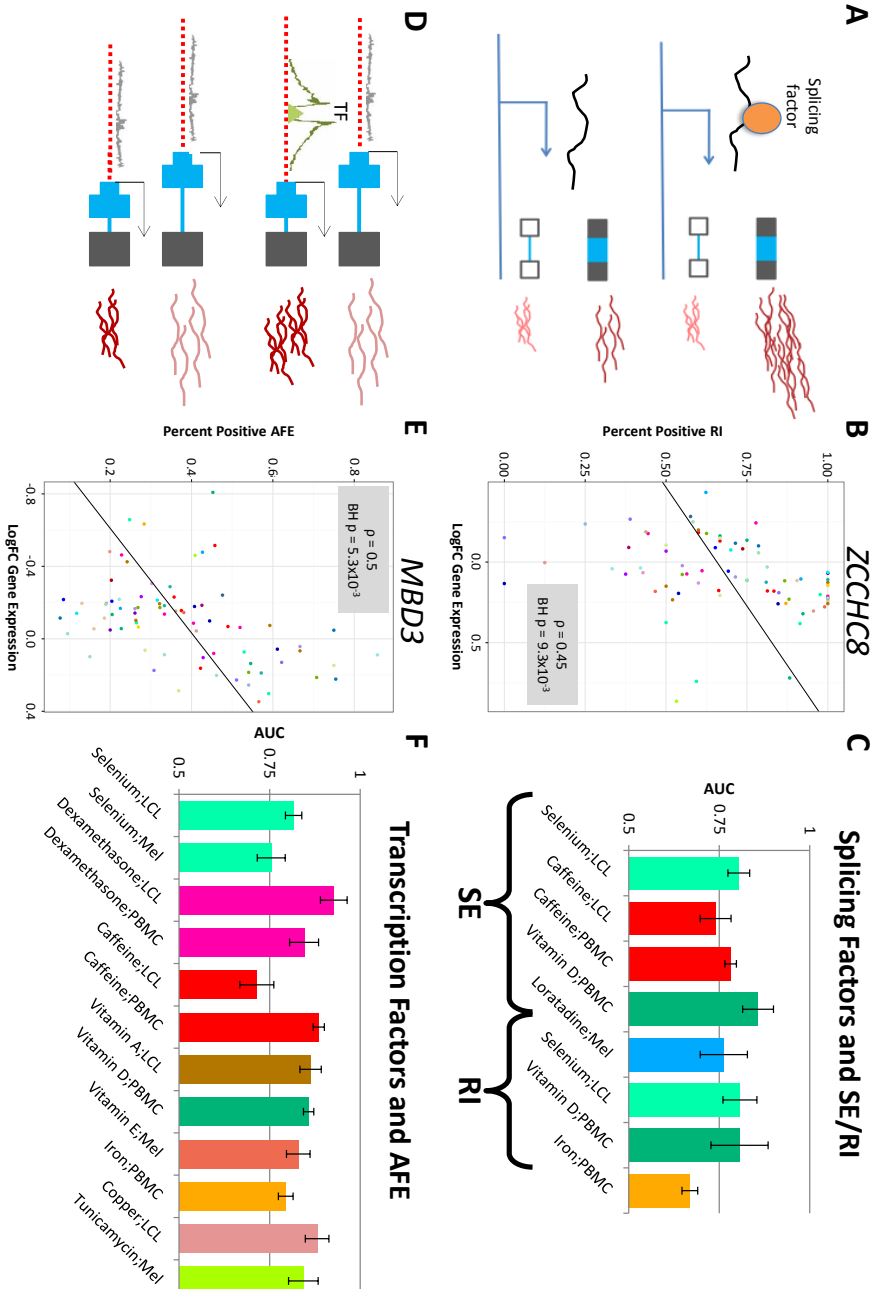

Figure 5: Effect of Trans-Factor Binding on RNA Processing Shifts. A) and D) show models of hypothesized mechanism of splicing or transcription factor influence on RNA processing and exon usage. B) An example of a correlation between the changes in gene expression of a transcription factor (*MBD3*) and the percent of AFEs that shift towards the upstream AFE across all environments for which gene expression could be assessed. E) An example of a correlation between the changes in gene expression of an RNA processing factor (*ZCCHC8*) and the percent of RIs that shift towards intron retention across all environments for which gene expression could be assessed. The correlation for B) and E) was tested using Spearman's rho and the p-value shown is Benjamini-Hochberg corrected while the trendline depict the best-fit line. C) Graph indicating the predictability (AUC as a proxy) of SE or RI shifts in a certain environment given predicted splicing factor binding sites (RNAcompete). F) Graph indicating the predictability (AUC as a proxy) of AFE shifts in a certain environment given transcription factor footprints [47].

direction of each transcription start site (defined as the beginning of each alternative first exon), and used GLM-NET (as we did in the splicing factor analysis). Across the 12 environments, the AUC ranges from 0.71 for caffeine in LCLs to 0.93 for dexamethasone in LCLs (Figure 5F). These data suggest that just as splicing factor binding predicts changes in SE and RI, transcription factor binding predicts changes in AFE following treatment.

Validation of the mechanism for AFE shifts

By inducing changes in transcription factor binding, specifically by perturbing the cellular environment, we can validate the effect of binding on AFE usage (Figure 6A). To this end, we performed ATAC-seq in LCLs following treatment with selenium and its vehicle control. First, we noticed that selenium leads to an overall reduction in chromatin accessibility near transcription start sites (Figure 6B). To determine how selenium influences binding of transcription factors near alternative TSS, we characterized chromatin accessibility following treatment with selenium, or control (with the footprints used in the prediction analysis).

We found significant differences in chromatin accessibility for 54 motifs, near the TSS that was preferentially used in the treatment versus the TSS preferred in the control condition. Of these 54 motifs, 23 are ETS transcription factor family members (or from motifs with similar sequence preferences). For example, we found a global decrease in chromatin accessibility for a motif for ELF2 (M01976) (Figure 6C), but there was a milder decrease in accessibility at the preferred TSS following selenium (Figure 6D), compared to the non-preferred TSS. These data suggest that at baseline ELF2 promotes transcription at both TSSs. However, following selenium treatment, though there is an overall decrease in ELF2 from both TSSs, there is a greater decrease from one TSS and this leads to a shift towards less usage of that TSS following treatment. All 23 motifs predicted from the ETS family of transcription factors show a similar change in binding as ELF2. More broadly, these results support a mechanism for changes in TSS usage driven by changes in chromatin accessibility and potentially transcription factor binding in response to perturbations of the cellular environment.

Discussion

We describe 15,628 event shifts following a wide range of environmental perturbations at 9,064 unique RNA processing event sites. We have provided a browsable web-resource cataloguing these RNA processing shifts. Researchers interested in a given gene, isoform, or treatment will be able to access our data to determine when RNA processing shifts occur and which other genes respond under similar environments. Mining of our results has the potential to inform on the mechanisms by which a cell responds to environmental perturbations and its genome-wide effect on RNA processing.

The majority of events could be characterized as AFE, ALE, RI or SE, suggesting that these are the most influenced by the environmental perturbations considered here. This is distinct from previous reports that TandemUTR events change most following infection [35] and suggests diverse mechanisms through which the cells respond to their environment.

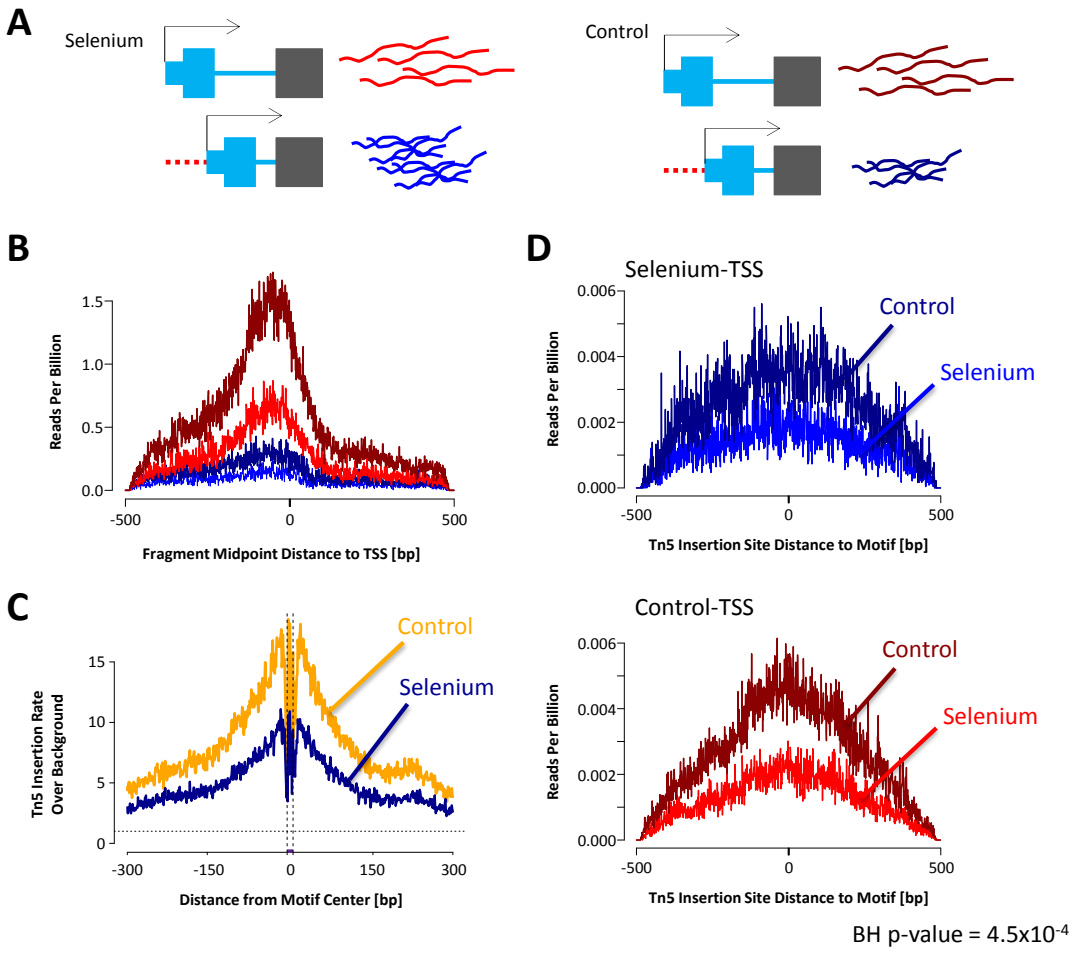

Figure 6: Binding of ELF2 Impacts AFE Shift Following Selenium Treatment. A) Model illustrating the situation where transcription is shifted towards the downstream AFE following selenium treatment and demonstrates an example of a TSS that would be included in our analysis. B) ATAC-seq denotes chromatin accessibility profile (read count normalized to total reads in the library) centered on TSS with AFE. Specifically, blue lines show accessibility at TSS to which the AFE shifts towards in selenium while red show the other TSS (as illustrated in A). Accessibility following selenium treatment is denoted by the bright blue and bright red lines while the control sample accessibility is denoted by the dark blue and dark red lines. C) ATAC-seq genome-wide profile centered on ELF2 motifs (M01976) (selenium- (blue) and control-treated (yellow) LCLs). D) ATAC-seq profiles centered on ELF2 motif locations within 1000bp of either TSS (colors are the same as in A and B). The difference in the ratio of treatment vs control read counts between the preferred and not preferred AFE is significant (BH p-value $= 4.5 \times 10^{-4}$).

Previous work has studied the role of splicing factors and transcription factors in RNA processing, in the absence of specific environmental perturbations. For example, others have shown that multiple splicing factors influence cassette exon usage, several of which fall into 2 protein families: hnRNPs and SRSFs. These 2 protein families often result in opposite splicing patterns [42]. These proteins may play a role in several of the RNA processing events that we study here, including SE, RI, A5SS, and A3SS. There are other studies that characterized proteins related to polyadenlyation site usage (which we study

as TandemUTR), including E2F, CSTF2, CSTF64 [54, 55]. Furthermore, recent studies have suggested that binding of transcription factors may influence differential use of transcription start sites in mice [56]. While these studies demonstrate the role of trans factors in RNA processing, we aimed to determine their role in global RNA processing changes in response to environmental perturbation.

Across 89 cellular environments, we found that binding sites for specific trans factors predict the shifts in events following treatment, thus demonstrating the importance of these factors and their binding locations for cellular response. We often find that not all binding sites for a given motif are predictive, but rather only binding sites in a certain location relative to the exon of interest. Furthermore, while previous studies have demonstrated the impact of binding location on RNA processing events at baseline, we demonstrate that the effect of binding in a certain location is treatment-specific. These results highlight the importance of studying trans factor binding across various environments. Further analysis of these binding sites will aid in understanding the details of the molecular mechanisms regulating RNA processing response to each cellular environment. For example, motifs associated with weaker binding of a trans factor may allow for more rapid changes in RNA processing and a more rapid cellular response.

Previous reports have characterized differences in transcription factor expression and binding across cellular environments [52, 57]. Here, we show that variation in transcription factor binding following environmental perturbations may determine TSS usage in addition to their function of influencing total gene expression. Feng *et al.* demonstrated the influence of transcription factors on TSS usage in mice [56]. Here, we expand on this knowledge by showing a similar function in human cells and in response to many environmental changes. Transcription factors are often regulated by environmental changes and are then responsible for impacting expression of many genes to promote re-establishment of cellular homeostasis (reviewed in [58]). Therefore, we suggest that TSS usage may also play a substantial role in cellular response and homeostasis.

Alternative RNA processing is predicted to occur in over 95% of multi-exon genes in humans across various tissues [59]. Our comprehensive catalog of genome-wide RNA processing changes can be utilized in future studies that aim to understand the role of RNA processing under various conditions and diseases as many of the treatments we used represent compounds to which individuals are commonly exposed. Furthermore, because RNA processing is associated with complex trait variation [22, 17], individual differences in RNA processing, specifically in response to environmental changes, could shed light on variation in organismal phenotypes.

Methods and Materials

RNA-seq data source

We used deep-sequenced RNA-seq data (fastq files) from Moyerbrailean *et al.*, 2016 [23]. Briefly, five cell types (LCL, PBMC, HUVEC, melanocyte and smooth muscle cells) were treated with 50 compounds to which humans are regularly exposed. Each environment (cell type and treatment) was represented in cell lines from three, unrelated individuals. We focused on 89 environments (Supplementary File 1 - Figure S1) that were sequenced to an average of 130M reads/library (297 RNA-sequencing libraries). These 89 environments include treatments and three vehicle controls (Supplementary File 1 - Table S1).

Alignment

In order to detect alternative splicing, we used Mixture of Isoforms (MISO) [40], which requires reads of the same length. Therefore, we selected reads with a length greater than or equal to 120bp. All reads were trimmed to 120bp. We also removed reads whose paired end was less than 120bp.

Reads were aligned to the hg19 human reference genome using STAR [60] (`https://github.com/alexdobin/STAR/releases`, version `STAR_2.4.0h1`), and the Ensemble reference transcriptome (version 75) with the following options:

```
STAR --runThreadN 12 --genomeDir <genome>
        --readFilesIn <fastqs.gz> --readFilesCommand zcat
        --outFileNamePrefix <stem> --outSAMtype BAM Unsorted
        --genomeLoad LoadAndKeep
```

where `<genome>` represents the location of the genome and index files, `<fastqs.gz>` represents that sample's fastq files, and `<stem>` represents the filename stem of that sample.

For each sample (individual cell line with a given treatment), we merged sequencing replicates (across lanes and runs on the same sequencer) using samtools (version 2.25.0). We further removed reads with a quality score of < 10 (equating to reads mapped to multiple locations).

Running MISO to detect splicing events

In order to detect alternative splicing, we used MISO on samples aligned as above. We utilized the events annotated and listed on `http://miso.readthedocs.io/en/fastmiso/index.html`. Specifically, we searched our data for 8 types of events with 5 from version 2 (SE, RI, A5SS, A3SS, and MXE, `http://miso.readthedocs.io/en/fastmiso/annotation.html`) and 3 from version 1 (AFE, ALE and TandemUTR, [61]). Two versions were used because AFE, ALE and TandemUTR were not annotated in version 2. We then ran `miso.py` on each of our samples for each of the 8 event types.

```
miso.py --run indexed_events/ my_sample1.bam --output-dir my_output1/
        --read-len 120
```

Then, we used `summarize_miso.py` to get the summary statistics for each event in each sample, including the percent spliced in value (PSI, Ψ).

```
summarize_miso --summarize-samples my_output1/ summaries/
```

To identify differential splicing, we used `compare_miso.py` which compares each event between treatment and control samples in the same individual cell line and experimental batch (plate).

```
compare_miso --compare-samples my_output1/  my_control1/ comparisons/
```

This script resulted in a $\Delta\Psi$, a Bayes factor and p-value for each comparison. We then focused on comparisons where both treatment and control contained 2 reads covering each isoform uniquely and a

total of 10 reads unique to either isoform for SE, RI, A5SS, A3SS, MXE, AFE and ALE. TandemUTR can only have reads specific to one isoform as the other isoform is simply a shorter version and completely overlaps the first. Therefore, we focused on comparisons of TandemUTR where both treatment and control contained 5 reads specific to the longer isoform and 10 total reads that covered either isoform.

Additionally, in order to inform on a cut-off for significant differential splicing, we performed comparisons between 2 controls (CO2 vs. CO1). Similar to treatments versus controls, we compared treatments performed in the same individual cell line and on the same plate. This generates an empirical null distribution that can be used to calibrate the statistical significance of the results. To this end, we used the same read requirements and filters as described above.

Detecting significant changes in splicing

Because we had samples from three individuals for each environment (cell type and treatment), we aimed to combine the differential splicing scores across individuals. In order to do this, first we constructed tables for each event type that have all comparisons for all events of that type and take the Bayes factor (BF) computed by MISO. Next, we used our comparison of CO2 to CO1 to estimate the empirical null distribution. This is equivalent to an empirical null distribution used in permutation-based approaches to correct p-values. The empirical p-values for each treatment versus control BF are calculated by the corresponding quantile in the empirical null distributions of BF in the CO1 versus CO2 comparisons. This was done separately for each event type, as they may have different underlying distributions under the null hypothesis of no changes, i.e. $\Delta\Psi = 0$. Then, we converted each empirical p-value to a Z-score while retaining the direction of the change from the $\Delta\Psi$ (calculated by MISO):

$$Z = sign(\Delta\Psi) \times |Q(p/2)| \tag{1}$$

where Q represents the quantile normal function `qnorm` in `R`. These Z-scores are then added across all individuals with enough reads for the specific event considered in a given environment and divided by the square root of the number of individuals. We required that for a given isoform, at least 2 of the 3 individuals had high enough coverage to be measured (see read minimums in previous section). Finally, we ranked these new Z-scores (which are a combined measure across 2 or 3 individuals in an environment) and calculated a Benjamini-Hochberg (BH) corrected p-value to control for the false discovery rate (FDR). We considered an event with a significant shift if the BH FDR $< 15\%$.

Assigning direction to AFE and ALE

Of the 8 event types, 3 are tested directionally by MISO. SE, RI and TandemUTR had the isoforms assigned such that a higher Ψ corresponded to more inclusion of the skipped exon, inclusion of the retained intron or longer UTR, respectively. In order to assess directionality of ALE and AFE events, we modified the Ψ signs such that higher Ψ values corresponded to the more upstream AFE or downstream ALE (on the transcribed strand, using the transcription start site or end site, respectively) (Figure 4A). This was done for all analysis steps considering directionality.

Gene expression changes

We calculated differential gene expression as described in Moyerbrailean *et al.* [23] using DESeq2 [62]. We analyzed the correlation between the number of significant shifts in exon usage (of any type) to genes that are differentially expressed in each environment using Spearman's ρ (Supplementary File 1 - Figure S3).

In order to make a reasonable comparison to changes in gene expression, which are relative to the relevant control sample from the same cell type, we used the $\Delta\Psi$, which was calculated compared to the same control samples. For each treatment (including all relevant cell types), we found genes that had a significant shift in an RNA processing event and determined the correlation between log-fold gene expression changes (measured over all 3 cell lines with DESeq2) and the average $\Delta\Psi$ across the same 3 individuals (Supplementary File 1 - Figure S4, and S2). In this analysis, we focused on treatment and event type combinations that had at least 30 RNA processing shifts in genes whose expression could be assessed in the same treatment across all cell types .

When we studied the correlation of a specific event to gene expression of either a splicing factor or transcription factor, we correlated the log-fold gene expression of that factor to the percent positive events (PPE) of a given event type across all environments:

$$\text{PPE}_A = \frac{\text{\# Positive Significant Shifts of Event Type A}}{\text{\# Significant AFE Shifts}} \tag{2}$$

The correlation between PPE and log-fold gene expression is calculated using Spearman's ρ and a best-fit line is added to each plot in Figures 5B and E.

Enrichment of event types among significant events in each treatment

In order to estimate the fraction for each event type while controlling for cell-type, and to identify enrichment of a specific event type among events with significant shifts in a given treatment, we used a generalized linear model:

$$log(\frac{p_l}{1-p_l}) \sim T_l \times E_l + C_l \tag{3}$$

where p_l is the probability that the event 'l' has a significant shift of a specific type E_l, for a given treatment T_l and cell-type C_l. This allowed us to study the interaction between treatment and event type (while removing the effect of cell type) on whether or not an event type is significantly enriched for a specific treatment.

The model incorporates an intercept which utilizes the information from one of each category in the equation such that every $\beta_{t,e}$ is relative to this baseline category. In order to create Figure 1B, we used the estimated fractions p_l from this logistic model, and conditional on each treatment t. We report the probability of each event type e among significant events. This allows us to compare proportion of event types across treatments. In order to determine enrichment of an event type among significant event shifts for a given treatment, we look at the p-value for the interaction term of treatment by event type ($\beta_{t,e} \neq 0$).

Identifying co-occurence of RNA processing changes

When searching for the co-occurrence of different event types, we focused on 58,787 events that could be mapped to one Ensembl gene ID, which included 7,264 events with significant changes in at least one environment. Events were said to co-occur when they could be found in the same gene in the same environment. For the overall numbers, this means that they could be tested in the same gene and the same environment. For the significant numbers, this means that both event types had at least one significant event change in that gene in the same environment.

Analysis of RNA processing changes across treatments and cell types

In order to compare RNA processing events across treatments and cell types, we focused only on events that could be tested in more than one environment. An event change was considered shared if there was a significant shift in the event in more than one treatment or cell type. When we looked for a consistent shift, we required that the event not only shifted in more than one treatment or cell type but that the $\Delta\Psi$ had the same sign in both environments.

Gene ontology analysis

We utilized GeneTrail [63] to find enrichment of gene ontology terms. We compiled a list of unique genes that had significant changes in RNA processing in at least 2 treatments in any of the 5 cell types (HUVEC, SMC, LCL, PBMC, and Mel). We then determined which GO categories were under/over-represented in 1,011 genes that were shared across treatments in any of the 5 cell types as compared to a list of all genes with significant changes in RNA processing in any environment irrespective of sharing (3,683 genes). We considered a category over/under-represented if the Benjamini-Hochberg FDR $< 5\%$.

Elastic-net regularized generalized linear model

In order to assess the predictive power of transcription factor binding on RNA processing changes following treatment, we utilized the 'glmnet' package in R [64]. This package uses an elastic-net regularized generalized linear model. We used this model to assess the role of transcription factor binding on AFE and splicing factor binding sites on SE and RI in environments with at least 100 significant event shifts (BH FDR $< 15\%$). In order to obtain a sufficient number of sites for modeling, we collected all event shifts with BH FDR $< 25\%$.

For our analysis of AFE, we used the transcription factor footprints derived from data collected by ENCODE and RoadMap Epigenomics [52, 47, 53]. We utilized footprints from all cell types because we expect binding to change following treatment and so did not want to be too restrictive on what we called a binding site. We then counted the number of footprints within 1000bp (upstream and downstream) of each TSS for each AFE that showed a significant shift following treatment. Next, for each AFE, we subtracted the number of footprints near the downstream TSS from the number of footprints near the upstream TSS. We then used these values, for each motif, as predictors in the model. The variable we attempted to predict using this model was the direction of the shift ($\Delta\Psi$) following treatment. From the

results of glmnet, we then used the lambda.1se, the lambda that was 1 standard error from the lambda that resulted in the highest AUC, to define an AUC for AFE in a given treatment.

For our analysis of SE and RI, we used the splicing factor binding sites predicted from RNAcompete [43]. We split each SE or RI with a significant shift following treatment into 5 regions around the event site. For SE, we had the region of the upstream intron, the skipped exon itself, the downstream intron, 100bp upstream of the 3' splice site, and 100bp downstream of the 3' splice site. For RI, we used the region of the upstream exon, the intron in question, the downstream exon, 100bp upstream of the 5' splice site, and 100bp downstream of the 5' splice site. We determined whether or not a splicing factor motif was found in any of these regions in the same direction of transcription and then separately considered these into the model as predictors. The variable we attempted to predict using this model was the direction of the shift ($\Delta\Psi$) following treatment. From the results of glmnet, we then used the lambda.1se, the lambda that was one standard error from the lambda that resulted in the highest AUC, to define an AUC for SE or RI in a given treatment.

ATAC-seq in LCLs exposed to selenium

The lymphoblastoid cell line (LCL) GM18508 was purchased from Coriell Cell Repository. LCLs were cultured in serum containing charcoal-stripped FBS and treated for 6 hours with 1μM selenium as described in [47]. Cells were also cultured in parallel with the vehicle control (water), to represent the solvent used to prepare the treatment. We then followed the protocol by [65] to lyse 25,000-100,000 cells and prepare ATAC-seq libraries, with the exception that we used the Illumina Nextera Index Kit (Cat #15055290) in the PCR enrichment step. Individual library fragment distributions were assessed on the Agilent Bioanalyzer and pooling proportions were determined using the qPCR Kapa library quantification kit (KAPA Biosystems). Library pools were run on the Illumina NextSeq 500 Desktop sequencer in the Luca/Pique-Regi laboratory. Barcoded libraries from three replicates (25,000, 50,000 and 75,000 cells each) were pooled and sequenced on multiple sequencing runs for 100M 38bp PE reads.

Reads were aligned to the reference human genome hg19 using `bwa mem` ([66] `http://bio-bwa.sourceforge.net`). Reads with quality <10 and without proper pairs were removed using `samtools` (`http://github.com/samtools/`).

To assess global shifts in accessibility, reads with different fragment length were partitioned into four bins: 1) [39-99], 2) [100-139], 3) [140-179], 4) [180-250]. For each fragment, the two Tn5 insertion sites were calculated as the position 4bp after the 5'-end in the 5' to 3' direction. Then for each candidate motif, a matrix $\boldsymbol{X}$ was constructed to count Tn5 insertion events: each row represented a sequence match to motif in the genome (motif instance), and each column a specific cleavage site at a relative bp and orientation with respect to the motif instance. We built a matrix $\{\boldsymbol{X}_l\}_{l=1}^{4}$ for each fragment length bin, each using a window half-size S=150bp resulting in $(2 \times S + W) \times 2$ columns, where W is the length of the motif in bp. The motif instances were scanned in the human reference genome hg19 using position weight (PWM) models from TRANSFAC and JASPAR as previously described [67]. Then we used CENTIPEDE and motif instances with posterior probabilities higher than 0.99 to denote locations where the transcription factors are bound.

Validating AFE mechanism with ATAC-Seq

First, we assessed chromatin accessibility within 1kb (in either direction) of each AFE by quantifying the number of reads (lengths 30-140bp, as this corresponds to lengths shorter than those that wrap around a nucleosome allowing us to focus on open chromatin). The read count was summed across AFEs that were either upstream or downstream and then normalized to the total reads in the library (either selenium- or control-treated).

In order to compare the effect of transcription factor binding changes on AFE shifts following LCL exposure to selenium, we started with the transcription factor footprints from [47] used for the prediction analysis. We split these footprints into those found within 500bp of the TSS towards which transcription shifted following selenium versus those found within 500bp of the TSS which was less preferred following treatment (termed Selenium-TSS or Control-TSS). For this analysis, we studied all AFE shifts with a BH FDR $< 25\%$. We used a more relaxed threshold for this analysis because we also must require the AFE to be within 500bp of a transcription factor footprint and would otherwise not have a sufficient number of sites to draw any conclusions. We then quantified the read counts in the selenium and control treated samples at these positions and normalized these counts to the overall read counts in each library. For each footprint, we calculated the ratio of normalized read counts in treatment versus control libraries. We then used these ratios to perform a Student's t-Test across all footprints of a specific transcription factor near the Selenium-TSSs compared to the Control-TSSs (2-tailed). Changes in transcription factor binding activity were considered significant if BH FDR $< 5\%$.

Additional Files

Supplementary File 1 — Supplemental Results.

Supplementary text for additional results, figures and tables.

Supplementary File 2 — 15,628 RNA Processing Shifts.

This table describes all 15,628 RNA processing shifts that we identify in our data. These sites can also be found on our browsable web-resource (`http://genome.grid.wayne.edu/RNAprocessing`). This table has 20 columns as follows: 1) Unique identifier for each event, 2) Plate name which is a key covariate among our samples, 3) Event name from MISO database, 4) Chromosome of event, 5) Strand of mRNA, 6) Start positions for exons, 7) End positions for exons, 8) Treatment ID, 9) Treatment name, 10) Cell Type, 11) Control ID for the tested treatment, 12) Control name, 13) Type of event, 14) Number of individuals that could be assessed, 15) Number of individuals that had p-value derived from the logBF < 0.05, 16) Number of individuals that had positive $\Delta\Psi$ and had p-value derived from logBF < 0.05, 17) Combined Z-score, 18) q-value, 19) Ensemble gene ID, and 20) Gene symbol.

Acknowledgments

We thank members of the Luca and Pique-Regi groups for helpful discussions and comments.

Funding

This work was supported by the National Institutes of Health [5R01GM109215 to F.L and R.P.]; and the American Heart Association [14SDG20450118 to F.L.].

Competing Interests

The authors declare that they have no competing interests.

References

[1] Mukherjee, M., Svenningsen, S., Nair, P.: Glucocortiosteroid subsensitivity and asthma severity. Curr Opin Pulm Med **23**(1), 78–88 (2017)

[2] Nieuwenhuis, M.A., Siedlinski, M., van den Berge, M., Granell, R., Li, X., Niens, M., van der Vlies, P., Altmuller, J., Nurnberg, P., Kerkhof, M., van Schayck, O.C., Riemersma, R.A., van der Molen, T., de Monchy, J.G., Bosse, Y., Sandford, A., Bruijnzeel-Koomen, C.A., Gerth van Wijk, R., Ten Hacken, N.H., Timens, W., Boezen, H.M., Henderson, J., Kabesch, M., Vonk, J.M., Postma, D.S., Koppelman, G.H.: Combining genomewide association study and lung eQTL analysis provides evidence for novel genes associated with asthma. Allergy **71**(12), 1712–1720 (2016)

[3] Li, X., Hastie, A.T., Hawkins, G.A., Moore, W.C., Ampleford, E.J., Milosevic, J., Li, H., Busse, W.W., Erzurum, S.C., Kaminski, N., Wenzel, S.E., Meyers, D.A., Bleecker, E.R.: eQTL of bronchial epithelial cells and bronchial alveolar lavage deciphers GWAS-identified asthma genes. Allergy **70**(10), 1309–1318 (2015)

[4] Bosse, Y.: Genome-wide expression quantitative trait loci analysis in asthma. Curr Opin Allergy Clin Immunol **13**(5), 487–494 (2013)

[5] Chang, P.J., Michaeloudes, C., Zhu, J., Shaikh, N., Baker, J., Chung, K.F., Bhavsar, P.K.: Impaired nuclear translocation of the glucocorticoid receptor in corticosteroid-insensitive airway smooth muscle in severe asthma. Am. J. Respir. Crit. Care Med. **191**(1), 54–62 (2015)

[6] Poon, A.H., Eidelman, D.H., Martin, J.G., Laprise, C., Hamid, Q.: Pathogenesis of severe asthma. Clin. Exp. Allergy **42**(5), 625–637 (2012)

[7] Christodoulopoulos, P., Leung, D.Y., Elliott, M.W., Hogg, J.C., Muro, S., Toda, M., Laberge, S., Hamid, Q.A.: Increased number of glucocorticoid receptor-beta-expressing cells in the airways in fatal asthma. J. Allergy Clin. Immunol. **106**(3), 479–484 (2000)

[8] Li, Y., Xiao, X., Ji, X., Liu, B., Amos, C.I.: RNA-seq analysis of lung adenocarcinomas reveals different gene expression profiles between smoking and nonsmoking patients. Tumour Biol. **36**(11), 8993–9003 (2015)

[9] Skj?rven, K.H., Jakt, L.M., Dahl, J.A., Espe, M., Aanes, H., Hamre, K., Fernandes, J.M.: Parental vitamin deficiency affects the embryonic gene expression of immune-, lipid transport- and apolipoprotein genes. Sci Rep **6**, 34535 (2016)

[10] Nicolae, D.L., Gamazon, E., Zhang, W., Duan, S., Dolan, M.E., Cox, N.J.: Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. **6**(4), 1000888 (2010)

[11] Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Frohlich, I., Imboywa, S., Von Korff, A., Okada, Y., Patsopoulos, N.A., Davis, S., McCabe, C., Paik, H.I., Srivastava, G.P., Raychaudhuri, S., Hafler, D.A., Koller, D., Regev, A., Hacohen, N., Mathis, D., Benoist, C., Stranger, B.E., De Jager, P.L.: Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. Science **344**(6183), 519–523 (2014)

[12] Grundberg, E., Small, K.S., Hedman, A.K., Nica, A.C., Buil, A., Keildson, S., Bell, J.T., Yang, T.P., Meduri, E., Barrett, A., Nisbett, J., Sekowska, M., Wilk, A., Shin, S.Y., Glass, D., Travers, M., Min, J.L., Ring, S., Ho, K., Thorleifsson, G., Kong, A., Thorsteindottir, U., Ainali, C., Dimas, A.S., Hassanali, N., Ingle, C., Knowles, D., Krestyaninova, M., Lowe, C.E., Di Meglio, P., Montgomery, S.B., Parts, L., Potter, S., Surdulescu, G., Tsaprouni, L., Tsoka, S., Bataille, V., Durbin, R., Nestle, F.O., O'Rahilly, S., Soranzo, N., Lindgren, C.M., Zondervan, K.T., Ahmadi, K.R., Schadt, E.E., Stefansson, K., Smith, G.D., McCarthy, M.I., Deloukas, P., Dermitzakis, E.T., Spector, T.D.: Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. **44**(10), 1084–1089 (2012)

[13] Caswell, J.L., Camarda, R., Zhou, A.Y., Huntsman, S., Hu, D., Brenner, S.E., Zaitlen, N., Goga, A., Ziv, E.: Multiple breast cancer risk variants are associated with differential transcript isoform expression in tumors. Hum. Mol. Genet. **24**(25), 7421–7431 (2015)

[14] Lai, M.K., Esiri, M.M., Tan, M.G.: Genome-wide profiling of alternative splicing in Alzheimer's disease. Genom Data **2**, 290–292 (2014)

[15] K?dzierska, H., Pop?awski, P., Hoser, G., Rybicka, B., Rodzik, K., Soko?, E., Bogus?awska, J., Ta?ski, Z., Fogtman, A., Koblowska, M., Piekie?ko-Witkowska, A.: Decreased Expression of SRSF2 Splicing Factor Inhibits Apoptotic Pathways in Renal Cancer. Int J Mol Sci **17**(10) (2016)

[16] Goehe, R.W., Shultz, J.C., Murudkar, C., Usanovic, S., Lamour, N.F., Massey, D.H., Zhang, L., Camidge, D.R., Shay, J.W., Minna, J.D., Chalfant, C.E.: hnRNP L regulates the tumorigenic capacity

of lung cancer xenografts in mice via caspase-9 pre-mRNA processing. J. Clin. Invest. **120**(11), 3923–3939 (2010)

[17] Paronetto, M.P., Passacantilli, I., Sette, C.: Alternative splicing and cell survival: from tissue homeostasis to disease. Cell Death Differ. **23**(12), 1919–1929 (2016)

[18] Lai, D.P., Tan, S., Kang, Y.N., Wu, J., Ooi, H.S., Chen, J., Shen, T.T., Qi, Y., Zhang, X., Guo, Y., Zhu, T., Liu, B., Shao, Z., Zhao, X.: Genome-wide profiling of polyadenylation sites reveals a link between selective polyadenylation and cancer metastasis. Hum. Mol. Genet. **24**(12), 3410–3417 (2015)

[19] Liaw, H.H., Lin, C.C., Juan, H.F., Huang, H.C.: Differential microRNA regulation correlates with alternative polyadenylation pattern between breast cancer and normal cells. PLoS ONE **8**(2), 56958 (2013)

[20] Dvinge, H., Bradley, R.K.: Widespread intron retention diversifies most cancer transcriptomes. Genome Med **7**(1), 45 (2015)

[21] Zhang, Q., Li, H., Jin, H., Tan, H., Zhang, J., Sheng, S.: The global landscape of intron retentions in lung adenocarcinoma. BMC Med Genomics **7**, 15 (2014)

[22] Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., Pritchard, J.K.: RNA splicing is a primary link between genetic variation and disease. Science **352**(6285), 600–604 (2016)

[23] Moyerbrailean, G.A., Richards, A.L., Kurtz, D., Kalita, C.A., Davis, G.O., Harvey, C.T., Alazizi, A., Watza, D., Sorokin, Y., Hauff, N., Zhou, X., Wen, X., Pique-Regi, R., Luca, F.: High-throughput allele-specific expression across 250 environmental conditions. Genome Res. **26**(12), 1627–1638 (2016)

[24] Tyrrell, J., Wood, A.R., Ames, R.M., Yaghootkar, H., Beaumont, R.N., Jones, S.E., Tuke, M.A., Ruth, K.S., Freathy, R.M., Davey Smith, G., Joost, S., Guessous, I., Murray, A., Strachan, D.P., Kutalik, Z., Weedon, M.N., Frayling, T.M.: Gene-obesogenic environment interactions in the UK Biobank study. Int J Epidemiol (2017)

[25] Joseph, P.G., Pare, G., Anand, S.S.: Exploring gene-environment relationships in cardiovascular disease. Can J Cardiol **29**(1), 37–45 (2013)

[26] Buil, A., Brown, A.A., Lappalainen, T., Vinuela, A., Davies, M.N., Zheng, H.F., Richards, J.B., Glass, D., Small, K.S., Durbin, R., Spector, T.D., Dermitzakis, E.T.: Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. Nat. Genet. **47**(1), 88–91 (2015)

[27] Zhernakova, D.V., Deelen, P., Vermaat, M., van Iterson, M., van Galen, M., Arindrarto, W., van 't Hof, P., Mei, H., van Dijk, F., Westra, H.J., Bonder, M.J., van Rooij, J., Verkerk, M., Jhamai, P.M., Moed, M., Kielbasa, S.M., Bot, J., Nooren, I., Pool, R., van Dongen, J., Hottenga, J.J., Stehouwer, C.D., van der Kallen, C.J., Schalkwijk, C.G., Zhernakova, A., Li, Y., Tigchelaar, E.F., de Klein,

N., Beekman, M., Deelen, J., van Heemst, D., van den Berg, L.H., Hofman, A., Uitterlinden, A.G., van Greevenbroek, M.M., Veldink, J.H., Boomsma, D.I., van Duijn, C.M., Wijmenga, C., Slagboom, P.E., Swertz, M.A., Isaacs, A., van Meurs, J.B., Jansen, R., Heijmans, B.T., 't Hoen, P.A., Franke, L.: Identification of context-dependent expression quantitative trait loci in whole blood. Nat. Genet. **49**(1), 139–145 (2017)

[28] Saha, A., Kim, Y., Gewirtz, A.D.H., Jo, B., Gao, C., McDowell, I.C., Consortium, G., Engelhardt, B.E., Battle, A.: Co-expression networks reveal the tissue-specific regulation of transcription and splicing. bioRxiv (2016). doi:10.1101/078741

[29] Hasin-Brumshtein, Y., Khan, A.H., Hormozdiari, F., Pan, C., Parks, B.W., Petyuk, V.A., Piehowski, P.D., Brummer, A., Pellegrini, M., Xiao, X., Eskin, E., Smith, R.D., Lusis, A.J., Smith, D.J.: Hypothalamic transcriptomes of 99 mouse strains reveal trans eQTL hotspots, splicing QTLs and novel non-coding genes. Elife **5** (2016)

[30] Ongen, H., Dermitzakis, E.T.: Alternative Splicing QTLs in European and African Populations. Am. J. Hum. Genet. **97**(4), 567–575 (2015)

[31] Gutierrez-Arcelus, M., Ongen, H., Lappalainen, T., Montgomery, S.B., Buil, A., Yurovsky, A., Bryois, J., Padioleau, I., Romano, L., Planchon, A., Falconnet, E., Bielser, D., Gagnebin, M., Giger, T., Borel, C., Letourneau, A., Makrythanasis, P., Guipponi, M., Gehrig, C., Antonarakis, S.E., Dermitzakis, E.T.: Tissue-specific effects of genetic and epigenetic variation on gene regulation and splicing. PLoS Genet. **11**(1), 1004958 (2015)

[32] Tsalikis, J., Pan, Q., Tattoli, I., Maisonneuve, C., Blencowe, B.J., Philpott, D.J., Girardin, S.E.: The transcriptional and splicing landscape of intestinal organoids undergoing nutrient starvation or endoplasmic reticulum stress. BMC Genomics **17**, 680 (2016)

[33] Solier, S., Barb, J., Zeeberg, B.R., Varma, S., Ryan, M.C., Kohn, K.W., Weinstein, J.N., Munson, P.J., Pommier, Y.: Genome-wide analysis of novel splice variants induced by topoisomerase I poisoning shows preferential occurrence in genes encoding splicing factors. Cancer Res. **70**(20), 8055–8065 (2010)

[34] Dutertre, M., Sanchez, G., Barbier, J., Corcos, L., Auboeuf, D.: The emerging role of pre-messenger RNA splicing in stress responses: sending alternative messages and silent messengers. RNA Biol **8**(5), 740–747 (2011)

[35] Pai, A.A., Baharian, G., Page Sabourin, A., Brinkworth, J.F., Nedelec, Y., Foley, J.W., Grenier, J.C., Siddle, K.J., Dumaine, A., Yotova, V., Johnson, Z.P., Lanford, R.E., Burge, C.B., Barreiro, L.B.: Widespread Shortening of 3' Untranslated Regions and Increased Exon Inclusion Are Evolutionarily Conserved Features of Innate Immune Responses to Infection. PLoS Genet. **12**(9), 1006338 (2016)

[36] Edmond, V., Moysan, E., Khochbin, S., Matthias, P., Brambilla, C., Brambilla, E., Gazzeri, S., Eymin, B.: Acetylation and phosphorylation of SRSF2 control cell fate decision in response to cisplatin. EMBO J. **30**(3), 510–523 (2011)

[37] Wang, L., Miura, M., Bergeron, L., Zhu, H., Yuan, J.: Ich-1, an Ice/ced-3-related gene, encodes both positive and negative regulators of programmed cell death. Cell **78**(5), 739–750 (1994)

[38] Zhao, S., Liu, W., Li, Y., Liu, P., Li, S., Dou, D., Wang, Y., Yang, R., Xiang, R., Liu, F.: Alternative Splice Variants Modulates Dominant-Negative Function of Helios in T-Cell Leukemia. PLoS ONE **11**(9), 0163328 (2016)

[39] Munoz, M.J., Perez Santangelo, M.S., Paronetto, M.P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J.J., Bird, G., Bentley, D., Bertrand, E., Kornblihtt, A.R.: DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. Cell **137**(4), 708–720 (2009)

[40] Katz, Y., Wang, E.T., Airoldi, E.M., Burge, C.B.: Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat. Methods **7**(12), 1009–1015 (2010)

[41] Veraldi, K.L., Arhin, G.K., Martincic, K., Chung-Ganster, L.H., Wilusz, J., Milcarek, C.: hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. Mol. Cell. Biol. **21**(4), 1228–1238 (2001)

[42] Erkelenz, S., Mueller, W.F., Evans, M.S., Busch, A., Schoneweis, K., Hertel, K.J., Schaal, H.: Position-dependent splicing activation and repression by SR and hnRNP proteins rely on common mechanisms. RNA **19**(1), 96–102 (2013)

[43] Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., Na, H., Irimia, M., Matzat, L.H., Dale, R.K., Smith, S.A., Yarosh, C.A., Kelly, S.M., Nabet, B., Mecenas, D., Li, W., Laishram, R.S., Qiao, M., Lipshitz, H.D., Piano, F., Corbett, A.H., Carstens, R.P., Frey, B.J., Anderson, R.A., Lynch, K.W., Penalva, L.O., Lei, E.P., Fraser, A.G., Blencowe, B.J., Morris, Q.D., Hughes, T.R.: A compendium of RNA-binding motifs for decoding gene regulation. Nature **499**(7457), 172–177 (2013)

[44] Kanopka, A., Muhlemann, O., Akusjarvi, G.: Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. Nature **381**(6582), 535–538 (1996)

[45] Ibrahim, E.C., Schaal, T.D., Hertel, K.J., Reed, R., Maniatis, T.: Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. Proc. Natl. Acad. Sci. U.S.A. **102**(14), 5002–5007 (2005)

[46] Wang, E., Mueller, W.F., Hertel, K.J., Cambi, F.: G Run-mediated recognition of proteolipid protein and DM20 5' splice sites by U1 small nuclear RNA is regulated by context and proximity to the splice site. J. Biol. Chem. **286**(6), 4059–4071 (2011)

[47] Moyerbrailean, G.A., Kalita, C.A., Harvey, C.T., Wen, X., Luca, F., Pique-Regi, R.: Which Genetics Variants in DNase-Seq Footprints Are More Likely to Alter Binding? PLoS Genet. **12**(2), 1005875 (2016)

[48] Martincic, K., Alkan, S.A., Cheatle, A., Borghesi, L., Milcarek, C.: Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. Nat. Immunol. **10**(10), 1102–1109 (2009)

[49] Nimura, K., Yamamoto, M., Takeichi, M., Saga, K., Takaoka, K., Kawamura, N., Nitta, H., Nagano, H., Ishino, S., Tanaka, T., Schwartz, R.J., Aburatani, H., Kaneda, Y.: Regulation of alternative polyadenylation by Nkx2-5 and Xrn2 during mouse heart development. Elife **5** (2016)

[50] Yang, Y., Li, W., Hoque, M., Hou, L., Shen, S., Tian, B., Dynlacht, B.D.: PAF Complex Plays Novel Subunit-Specific Roles in Alternative Cleavage and Polyadenylation. PLoS Genet. **12**(1), 1005794 (2016)

[51] Nagaike, T., Manley, J.L.: Transcriptional activators enhance polyadenylation of mRNA precursors. RNA Biol **8**(6), 964–967 (2011)

[52] ENCODE Project Consortium: An integrated encyclopedia of DNA elements in the human genome. Nature **489**(7414), 57–74 (2012)

[53] Roadmap Epigenomics Consortium, A. Kundaje, Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., *et al.*: Integrative analysis of 111 reference human epigenomes. Nature **518**(7539), 317–330 (2015)

[54] Elkon, R., Ugalde, A.P., Agami, R.: Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet. **14**(7), 496–506 (2013)

[55] Nazim, M., Masuda, A., Rahman, M.A., Nasrin, F., Takeda, J.I., Ohe, K., Ohkawara, B., Ito, M., Ohno, K.: Competitive regulation of alternative splicing and alternative polyadenylation by hnRNP H and CstF64 determines acetylcholinesterase isoforms. Nucleic Acids Res. (2016)

[56] Feng, G., Tong, M., Xia, B., Luo, G.Z., Wang, M., Xie, D., Wan, H., Zhang, Y., Zhou, Q., Wang, X.J.: Ubiquitously expressed genes participate in cell-specific functions via alternative promoter usage. EMBO Rep. **17**(9), 1304–1313 (2016)

[57] Yip, K.Y., Cheng, C., Bhardwaj, N., Brown, J.B., Leng, J., Kundaje, A., Rozowsky, J., Birney, E., Bickel, P., Snyder, M., Gerstein, M.: Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. Genome Biol. **13**(9), 48 (2012)

[58] Bahrami, S., Drabl?s, F.: Gene regulation in the immediate-early response process. Adv Biol Regul **62**, 37–49 (2016)

[59] Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J.: Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. **40**(12), 1413–1415 (2008)

[60] Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R.: STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**(1), 15–21 (2013)

[61] Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B.: Alternative isoform regulation in human tissue transcriptomes. Nature **456**(7221), 470–476 (2008)

[62] Love, M.I., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**(12), 550 (2014)

[63] Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N., Elnakady, Y.A., Muller, R., Meese, E., Lenhof, H.P.: GeneTrail–advanced gene set enrichment analysis. Nucleic Acids Res. **35**(Web Server issue), 186–192 (2007)

[64] Friedman, J., Hastie, T., Tibshirani, R.: Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw **33**(1), 1–22 (2010)

[65] Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J.: Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods **10**(12), 1213–1218 (2013)

[66] Li, H., Durbin, R.: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**(14), 1754–1760 (2009)

[67] Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., Pritchard, J.K.: Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. **21**(3), 447–455 (2011)