Transcriptional readthrough precedes alternative splicing programs triggered in CML cells by imatinib

Paulina Podszywałow-Bartnicka*, Morgan Shine, Jing Lin, and Karla M. Neugebauer*

Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT, USA

*Co-correspondence: karla.neugebauer@yale.edu

paulina.podszywalow-bartnicka.@yale.edu

Short Title: Transcriptional readthrough is triggered by imatinib

Keywords

Transcriptional readthrough, Chronic Myelogenous Leukemia (CML), imatinib, 3' end cleavage, alternative splicing, intron retention, hematopoiesis, chemotherapy resistance, readthrough chimera

Abstract

Cellular stresses regulate transcriptional readthrough, whereby RNA polymerase II elongates past a gene's polyadenylation cleavage site without RNA cleavage. Readthrough has been reported in several cancer types. Here, we use long-read sequencing of nascent RNA to quantify transcriptional readthrough in chronic myelogenous leukemia (CML) cells and characterize early responses to the targeted therapeutic, imatinib. We show that the amount, length, and gene-specificity of readthrough increase within 1 hour, before gene expression and alternative splicing alterations emerge. Strikingly, imatinib-dependent mRNA isoform changes involved "readthrough chimeras", in which exons from an upstream gene are alternatively spliced to exons in a downstream gene. Altered mRNA isoforms and chimera levels were detected in imatinib-resistant K562 as well as CML patient cells. Thus, imatinib can provoke a cascade of early changes to transcription and splicing fidelity that may lead to longer-term adjustments in gene expression, cancer cell differentiation, and the development of therapy resistance.

Teaser

Precision RNA sequencing was used to discover the earliest response of leukemia cells treated with targeted therapy: transcriptional readthrough.

INTRODUCTION

From prokaryotes to eukaryotes and from yeast to man, nascent RNA is the substrate of numerous RNA processing steps, including 5' end capping, splicing, and 3'end cleavage [1]. Specific RNA sequences recruit the corresponding machineries – namely the capping enzymes, the spliceosome, and the Cleavage and Polyadenylation Complex (CPC) – that act quickly to modify the nascent transcript as it emerges from RNA polymerase II (Pol II). Human genes harbor an average of 8 introns each, of which approximately 75% are co-transcriptionally removed by the spliceosome [1-4]. The recent introduction of long-read sequencing of nascent RNA as a tool for analyzing co-transcriptional RNA processing has enabled researchers to associate multiple RNA processing steps with each other and with the position of Pol II. Emerging evidence points to tight coordination of intron removal at the level of individual transcripts [5-8]. For example, unprocessed nascent transcripts – with all introns retained and a lack of cleavage at the polyadenylation site (PAS) – were observed in budding and fission yeasts, showing that RNA splicing and cleavage efficiencies are controlled at the individual transcript level. When 3' end cleavage fails, Pol II proceeds past the PAS and continues downstream of the parent gene [6, 9, 10]. A simple interpretation of this phenotype is that the activity of CPC is either delayed or blocked when a transcript has not been co-transcriptionally spliced. The consequences of transcriptional readthrough are the subject of intense investigation.

Numerous studies have identified transcriptional readthrough as a common, rapid response to cellular stress. Hypoxia, heat shock, starvation, and viral infection each lead to increases in the level of intergenic transcription, as measured by RNA sequencing experiments that identify "downstream of genes" (DoG) RNAs [11, 12]. Increased intergenic transcription can encompass many kilobases (kb) past the most distal gene boundary and is believed to be contiguous with the upstream "parent" gene rather than due to activation of new downstream transcription start sites [13, 14]. When transcription continues into the next gene on the DNA strand, transcription and/or splicing of the "read-in gene" are often disrupted [10, 15, 16]. Although DoG RNAs were initially considered to be stress-induced, their presence has since been broadly documented in normal tissues, cancer cells, and aging cells and tissues [17-19]. Transcriptional readthrough is a hallmark of clear cell renal carcinoma cells and prostate cancer cells, where a marked increase in transcription downstream of genes leads to the formation of chimeric RNA transcripts comprised of exons from the parent gene spliced together with exons

from the read-in gene [20, 21]. Chimeric transcripts can encode truncated proteins, newly fused together functional globular domains or, alternatively, proteins with intrinsically disordered extensions [19]. Thus, the induced expression of chimeric transcripts and encoded proteins in response to stress may be physiologically relevant.

Transcriptional readthrough has been vastly understudied in the context of cancer. By contrast, pre-mRNA splicing regulation is often disrupted in cancer and even drives cancer initiation in some blood cancers, making splicing a target for clinical therapies [22]. We are interested in the possibility that transcriptional readthrough and associated changes in RNA processing play roles in the development and treatment of chronic myeloid leukemia (CML), where alternative splicing has been identified in patients [23]. Activation of cellular stress responses is detected in stem cells and accompanies carcinogenesis [24, 25]. These evolutionarily conserved pathways are detected in primary cells and hijacked by different cancer types, supporting survival and chemotherapy resistance. In CML, the unfolded protein response and phosphorylation of eIF2α accompany cancer progression and reduce the effectiveness of the targeted therapeutic drug imatinib [26, 27]. The activity of Bcr-Abl1 onco-kinase affects protein translation and induces stress granules in CML cells [28, 29]. Similar cellular phenotypes were detected in acute myeloid leukemia (AML), linking condensation of stress granules to changes in splicing regulation [30]. Two chemotherapeutic agents used to treat AML, among other cancers, JTE-607 and camptothecin, cause transcriptional readthrough by directly inhibiting the enzymatic activity of the CPC and topoisomerase 1, respectively [31-34]. These chemotherapeutic compounds directly and indirectly impact transcription and pre-mRNA splicing [35-37], making it unclear if transcriptional readthrough is purely related to the biochemical activities of these drugs or, alternatively, if it is related to the stress induced in cancer cells by therapeutic treatment.

Here, we have specifically addressed the possibility that transcriptional readthrough could be induced by therapeutic targeting of a cytoplasmic pro-oncogenic protein, Bcr-Abl1. We used imatinib treatment of CML cells as a well-studied model, and assayed the cellular response to exposure times (1 and 18 hours), within one cell cycle period. Because imatinib directly inhibits the activity of BCR-Abl1 onco-kinase and *not* the transcription or RNA processing machinery, our experiment enables us to determine the transcriptional effects of cancer cells

stressed by targeted cancer therapy. To achieve this, we have also implemented long-read sequencing of nascent RNA together with short-read sequencing datasets to quantify and rigorously characterize the occurrence of transcriptional readthrough at the gene and single transcript level, enabling us to detect associated events like alternative splicing and changes in the formation of chimeric transcripts. The early response of CML cells to imatinib is not well characterized. Remarkably, our approach detects some of the long-term alterations characteristic of month-long treatment that arise already within these short treatment periods. Taken together, our findings show that transcriptional readthrough is the first detectable effect of imatinib on gene expression, which precedes and likely facilitates later gene expression changes.

RESULTS

The targeted chemotherapeutic imatinib induces transcriptional readthrough in K562 cells

To study the effects of imatinib on RNA processing, we chose the K562 cell line, which originates from a patient with chronic myeloid leukemia in blast crisis (CML-BC) [38], as a model system. Transcriptional readthrough is typically studied by short-read sequencing of nuclear RNA [34]. In this study, we implemented long-read sequencing of nascent RNA to detect and quantify the number and length of reads that extend past the PAS (Fig. 1A). Short-time (1h) imatinib treatment increased coverage within the intergenic region relative to the number that have 3' ends within the gene body. We reasoned that levels of transcriptional readthrough might be altered by imatinib treatment compared to control and that the chemical inhibitor of CPC, JTE-607, could serve as a positive control for increased transcriptional readthrough of at least a subset of genes [34, 39]. Nascent RNA long-read libraries were prepared by isolating chromatin fractions (Fig. S1A) from three biological replicates each for four different treatments: control (CTRL), 1-hour treatment with JTE-607 (JTE 1h), 1-hour treatment with imatinib (IM 1h), and 18-hour treatment with imatinib (IM 18h) [40]. Read numbers, mapping statistics, and read lengths are available in Tables S1 and S2; the read length distributions (Fig. S1B), similar to previous nascent RNA long reads, allow us to analyze RNA processing early in gene transcription and additionally capture longer reads that might reflect transcriptional readthrough for approximately 11,0000 genes per dataset (Fig. S1C). Normalization indicates a similar read length distribution and number of representative genes across all datasets (Figs. S1D,E). Meta-analysis of the resulting read coverage across the gene body and extending 4 kb upstream of the transcription start site (TSS) and downstream of the last PAS revealed that JTE-607 treatment led to transcriptional readthrough, as expected (Fig. 1C). Note that the decrease in transcriptional readthrough approaching +4 kb is in part determined by the length limitations of long-read sequencing (Fig. S1F, Fig. S1B, D). Taken together, IM 1h and 18h treatments were characterized by higher levels of readthrough compared to CTRL and lower levels than JTE. These data raise the question of which genes are affected by imatinib.

To generate parameters from our data that would enable gene-specific analysis of nascent RNAs sequenced in our libraries, we defined readthrough reads as those that originate at either the TSS or within the gene but have 3' ends ≥100 nucleotides downstream of the last annotated

PAS (Fig. 1C). We defined reads with 3' ends within the transcribed gene (i.e. between the TSS and the PAS) as "gene body" reads (gb). This allowed us to derive a gene-specific metric, the Readthrough Index (RTI), which signifies the proportion of long reads that are not cleaved efficiently. Different treatment groups (Fig. 1D, S2), had similar RTI distributions (Fig. S2A, B); however, the IM 1h dataset displayed significantly higher RTIs (Fig. 1D, S2B). JTE 1h treatment was characterized by overall more genes (~1,450/million reads) with RTI>0 compared with CTRL, IM 1h, and IM 18h, which each had 1,000-1,100 genes/million reads with readthrough (Fig. 1E). These data show that JTE inhibits the cleavage of some but not all genes, consistent with the findings of others [34, 39], and that transcriptional readthrough is detected and quantifiable in nascent RNA long-read sequencing libraries.

To address whether imatinib treatment induces readthrough in different sets of genes, we analyzed unique and overlapping genes with RTI≥0.2 (Fig. 1F). 528 genes were common among CTRL and imatinib treatments. The sample with the most (819) unique genes with readthrough is IM 18h, while 449 genes in IM 1h and 18h were shared and not present in CTRL. To determine how the extent of readthrough varies between CTRL, IM 1h, and IM 18h, we created a heatmap with CTRL RTIs sorted from low to high by gene; corresponding RTI values for both imatinib treatments were aligned accordingly (Fig. 1G). The data show a range of altered intensities across the genes analyzed for IM 1h and 18h, with many similarities between the two timepoints. Thus, even though IM 18h has the largest number of unique genes with RTI≥0.2 (Fig. 1F), our data indicate a striking induction of transcriptional readthrough within 1 hour of imatinib treatment.

Intron retention is associated with readthrough in all conditions

Long reads also reveal splicing information that can be correlated with other features of each read within the same RNA molecule [6, 9, 10]. Therefore, we sought to determine if transcripts with transcriptional readthrough displayed elevated levels of intron retention and, if true, whether imatinib treatment altered this result. To do so, we quantified co-transcriptional splicing efficiency (CoSE) of CTRL, IM 1h, and IM 18h, by aligning our long reads to gene annotations (Fig. 2A, top panel). Relatively few introns showed substantial changes in CoSE in either the positive (more splicing) or negative (intron retention) direction after 1 or 18 h IM treatment

(Fig. 2A, bottom panel) or after JTE treatment (Fig. S3A). Only ~10% of introns were retained in full-length reads that begin at the TSS and have 3' ends within the gene body, consistent with efficient co-transcriptional splicing and unaffected by IM 1h or 18h (Fig. 2B). As expected, more introns (~35%) were retained in readthrough reads, and this was similarly unaffected by imatinib. Notably, JTE treatment reduced intron retention in the fraction of readthrough genes that could be caused by its sequence-dependent impact on splicing [34, 39]. Because of read length limitations, focusing on full-length reads decreases the number of intron-containing readthrough reads we can analyze. Therefore, we separately analyzed readthrough reads that did not start at the TSS and found ~40% and ~75% retained introns in gene body and readthrough categories, respectively. Again, neither parameter was significantly affected by imatinib (Fig. 2B). However, 1h treatment with imatinib significantly increased intron retention detected by CoSE in transcripts with lower RTI (Fig. S3B). A partial correlation between transcriptional readthrough and gene-specific CoSE was reinforced by global analysis, reflecting the increased intron retention present in readthrough transcripts (Fig. S3B,C). We conclude that readthrough transcripts are correlated with intron retention in human K562 cells and that imatinib does not increase intron retention levels independently of transcriptional readthrough.

In a previous study, we found that transcripts from the β-globin gene and other erythroid genes displayed "all-or-none" RNA processing, in which half of the transcripts were efficiently spliced and cleaved at the 3' end [6]. The other half contained all of their introns and displayed transcriptional readthrough. We searched for genes displaying "all-or-none" processing (RTI≥0.2 and CoSE≤0.7) and found 44 genes exhibiting this behavior, one of which is UROD (Fig. 2C). Alignment of subsampled UROD transcripts shows partially spliced, all spliced, and all unspliced transcripts displayed in the order of Pol II elongation from TSS to PAS. As expected, the majority of readthrough transcripts are unspliced, and both IM 1h and 18h treatments increase the RTI. The UROD gene encodes the enzyme uroporphyrinogen decarboxylase, which is crucial for heme production. Interestingly, GO term analysis for the genes with RTI≥0.2 in IM 1h and 18h highlights metal ion binding and enzymatic activities characteristic of erythropoiesis (Fig. S4A-B), suggesting that imatinib is able to direct K562 cells into the erythroid lineage on shorter timescales than previously appreciated [41].

Erythroid differentiation accompanies transcriptional readthrough upon imatinib treatment

To determine the cellular significance of these early gene expression changes in response to imatinib, we generated datasets from short-read polyA+ RNA-seq libraries to corroborate and extend our findings (Table S1). To ask if the genes with high RTI (see Fig. 1G) belonged to previously or newly expressed genes, we analyzed 1818 genes with RTI≥0.2 in IM 18h samples; RTI values per gene were sorted according to the IM 1h dataset to visualize how readthrough and gene expression change over time (Fig. 3A, upper panel). The same 1818 genes were sorted in the same order and then expressed in terms of fold change in mRNA level compared to control, showing that only minor changes in mRNA levels were detected at 1h (Fig. 3A, lower panel). Strikingly, the data at 6h and 18h clearly indicate that 78% of the genes with high RTI values at 1h are downregulated over the time period. In contrast, the genes with low RTI at 1h are mostly upregulated and increase in RTI by 18h, accounting for the increased numbers of readthrough genes by the later timepoint. By comparison, only 9% and 5% respectively of the 1491 genes with readthrough upon 1h JTE treatment (133 genes at 6h and 73 genes at 18h) were downregulated; ≤1% were upregulated (18 and 8 genes at 6h and 18h respectively; Fig. S5). We conclude that genes with significant readthrough at 1h of imatinib treatment are on a trajectory of decreasing gene expression, which could be caused by either repressed transcription and/or a relative decrease in transcript stability owing to the failure to cleave at the PAS.

Analysis of differential gene expression revealed an increase in some erythroid genes, such as hemoglobins, even after 1h of imatinib treatment (Fig. S6A). To determine if genes with transcriptional readthrough identified upon IM 1h or 18h in K562 cells belong to the constellation of genes (erythroid or non-erythroid) that are up- or down-regulated during normal erythropoiesis, we utilized a dataset from human $CD34^+$ hematopoietic stem cell progenitors (HSPCs) undergoing differentiation into poly- and ortho-erythroid cells [42]. 334 of the identified genes with RTI≥0.2 (16%) overlapped with the normal erythropoiesis dataset, increasing in number from 1h to 18h of imatinib (Fig. 3B). Mean RTI for these overlapping genes reveals a trend to higher RTI in IM 1h (Fig. S6B). We used RT-qPCR to validate two examples – hemoglobin ε (*HBE1*) and ferritin heavy chain (*FTH1*) – that displayed increased transcriptional readthrough and mRNA levels in IM 18h (Fig. 3C). Interestingly, neither of two

other chemotherapeutics with distinct mechanisms of action – Talazoparib (BMN) or Homoharringtonine (HHT) – induced transcriptional readthrough, indicating the specificity of imatinib's effects on 3' end cleavage, at least for these two genes. To further validate the specific, early induction of erythropoietic genes by imatinib, we performed western blotting and detected elevated globins γ and ε as well as Kell endopeptidase (*KEL*) in IM 18h samples (Fig. 3D). Finally, flow cytometry analysis of K562 cells treated with imatinib for 48h confirmed their differentiation to CD45+ CD235a+ erythroblasts (Fig. 3E), validating and extending previous findings [41].

Imatinib leads to alternative mRNA isoforms with and without transcriptional readthrough

It is well known that alternative splicing and intron retention are regulatory features of erythropoiesis [42-44]. Since readthrough and intron retention are early responses to imatinib treatment, we wondered if and when alternative splice isoforms might be detected. To rigorously address this, we generated short-read sequencing datasets from total nuclear RNA (Table S3); our rationale was (*i*) changes in splicing will first be detected in the nucleus before achieving steady-state mRNA representation in the cytoplasm and (*ii*) intron retention events are better detected in the nuclear fraction. Using splicing per intron (SPI) as a statistically powerful metric, we detected a few intron retention events in IM 1h and an increase in positive and negative splicing changes at 18h (Fig. 4A), in agreement with the long-read nascent RNA sequencing results (see Fig. 2A). Interestingly, reduced mRNA levels for proteins involved in the regulation of splicing (Fig. S7A), including SRSF1 as previously described [45], were detected after imatinib treatment. We confirmed downregulation Gemin5 and Serine–arginine protein kinase 1 (SRPK1) levels by immunoblotting (Fig. S7B). We observed reduced level of WD repeat-containing protein 43 (WDR43), representative of a pool of imatinib-downregulated proteins involved in ribosome biogenesis (Fig. S7B), that was preceded by an increase in RTI score upon 1h of imatinib treatment (from 0.37 to 0.49). In contrast, increased *RBM5* mRNA was *not* reflected at the protein level, presumably due to intron retention and nuclear degradation of these transcripts (Fig. S7C). We conclude that the mRNA and, in some cases, protein levels of stimulatory RNA processing factors decrease upon imatinib treatment.

Imatinib, by impacting the expression of factors involved in splicing regulation, should affect alternative splicing and thereby influence cell response to treatment. SR protein kinase 1 modulates the activity of splicing factors with serine/arginine domain, and its downregulation offers a direct explanation of the imatinib link to the stress-induced changes in alternative splicing [46]. Suppression of SRPK1 by imatinib induces cell death in K562 cells [47]. Also, altered level of Gemin5 was found to modulate alternative splicing of a spectrum of transcripts [48], and its upregulation causes changes primarily in exon skipping [49]. Analysis of our mRNA-seq dataset was conducted, comparing CTRL to IM and JTE 18h using rMATS. This revealed ~500 alternative splicing events in IM 18h, greater than 10 times the number of changes in IM 1h (Fig. S8A), including a large number of cassette exon events, both inclusion and exclusion (Fig. 4B). As a control for possible direct or indirect effects of transcriptional readthrough on alternative splicing, a dataset was collected for JTE 1h and 18h; approximately 40 and 150 significant alternative splicing changes were detected, respectively (Fig. S8A). By contrast, another stress condition associated with transcriptional readthrough – hyperosmotic shock – led to almost 400 alternative splicing events, including many cassette exons and retained introns, after only 1h of 110 mM KCl treatment (Fig. S8A-B). We conclude that direct inhibition of 3' end cleavage by JTE itself leads to modest but significant numbers of alternative splicing events, suggesting that some of the alternative mRNA isoforms induced by IM could be due to readthrough. But, in addition, alternative splicing is a robust and specific response to imatinib treatment that differs from JTE treatment and hyperosmotic shock.

To determine if alternative mRNA isoforms induced by imatinib are linked to transcriptional readthrough, we performed two types of analyses. First, we asked if retained introns (RI) and cassette exons (CE) detected in nuclear RNA by alternative splicing analysis with rMATS were correlated with genes with RTI≥0.2. Only 95/648 genes had both changes in CE inclusion and high RTI, and only 24/167 genes had both RI and high RTI (Fig. 4C). Interestingly, this set of RI genes includes erythropoietic genes, like UROD, and other genes that will be highlighted below. Intron retention in *DDX17* mRNA is displayed in detail and validated by RT-PCR in Figs. S5C-D. Increased intron retention in *EIF4A2* is likely driven by the presence of five snoRNAs in different introns [50]. Second, we compared sets of alternative mRNA isoforms detected in IM and JTE 18h (see Fig. 4B) to determine the extent of gene overlap. For intron retention (Fig. 4D, left panel), we find that only 4 genes are common between IM and

JTE, while 45 intron-retained isoforms are unique to IM. Similarly, 324 cases of cassette exon usage were not present in the JTE dataset (Fig. 4D, right panel), in agreement with data discussed above that transcriptional readthrough does not directly cause intron retention in these genes.

To explore a functionally important example of alternative splicing induced by JTE (of 93 cases of CE detected), we selected the mRNA encoding the eukaryotic translation initiation factor 4H (*EIF4H*) harboring a cassette exon (exon 5) that defines two different protein isoforms (Fig. 4E, left). Inclusion of short peptide sequence in eIF4H longer isoform induces conformational change that could potentiate a protein-protein interaction with eIF4A helicase thereby increasing translation of selected transcripts and contributing to cancerogenesis [51-53]. Quantification of splicing events (Fig. 4E, right) and RT-PCR validation (Fig. S8E) revealed that both IM and JTE significantly increase the proportion of isoform 1 and decrease the proportion of isoform 2, containing exon 5. Because 27 of the CE events induced by JTE overlap with those induced by IM, including *EIF4H*, it is therefore possible that these alternative splicing events are directly linked to transcriptional readthrough in both cases. A summary of the observed differential splicing changes attributable to the erythroid gene expression program reveals extensive alternative splicing regulation of numerous up- and down-regulated genes (Fig. S8F).

Given our early detection of imatinib-induced transitions to cell fates (e.g. erythropoiesis), we considered the possibility that some of the observed alternative splicing could be related to the development of imatinib resistance. We analyzed an mRNA-seq dataset from K562 cells treated with a progressively increasing concentration of imatinib over a period of 10 months [54]. These cells are imatinib-resistant and display a large number of alternative splicing changes: 97 mutually exclusive exons, 808 cassette exons, and 162 retained introns (Fig. 5A). Although none of the identities of retained introns overlap with our IM 18h dataset, 94 cassette and 3 mutually exclusive exons splicing events did overlap. Thus, alternative splicing may contribute to early steps towards long-term imatinib resistance, provided at the cellular adaptation level and independent of genomic mutation-driven resistance. This includes *NFE2L1*, which encodes an endoplasmic reticulum-localized protein that translocates to the nucleus upon stress, where it is involved in the regulation of erythroid differentiation as a transcription factor (*NRF1*) (for review see [55]). Results of RT-PCR across the exon validate the change in isoforms (Fig. S8G-H). Gene Ontology analysis for these overlapping alternatively spliced genes revealed

many changes consistent with progression to other cell fates, such as transcription factors that promote myeloid differentiation and factors that control proper chromosome segregation (Fig. 5B). Indeed, many of the overlapping genes are related to erythroid differentiation, showing that development of imatinib resistance involves at least a partial shift to that gene expression program (Fig. 5C). To determine if these changes are comparable to a patient treated with imatinib, we performed mRNA-seq on CD34+ cells isolated at diagnosis and after 6 months of treatment (Fig. 5D). Comparing the different categories of numerous alternative splicing changes present in the patient samples after imatinib treatment, 26 were common with our IM 18h sample, of which 6 genes belonged to the erythroid gene expression program. One example is the translation initiation factor *EIF4A2*, which contains a "poison cassette" exon that harbors premature stop codons, triggering nonsense-mediated decay [56]. Elevated inclusion of this exon is visible in the nascent RNA long-read sequencing dataset and validated by RT-PCR (Figs. 5E, S8I). The relative abundances of the NMD and expressed isoforms are negatively correlated (Fig. 5F), indicating that potentially more EIF4A2 protein is expressed in response to imatinib treatment.

Imatinib modulates the formation of chimeric mRNAs

Transcriptional readthrough can lead to the formation of RNA chimeras and even produce fusion proteins from those spliced mRNA products [20, 21]. We used bioinformatic tools to search for these "readthrough chimeras" from K562 total and nuclear RNA sequencing datasets as well as CD34+ cells and sought to validate them with our long-read data. Remarkably, the dataset with the most chimeric reads was imatinib-resistant K562 cells and CD34+ from CML patients (Fig. S9A,B). A striking number of chimeras were also detected by short- and long-reads in K562 cells at 18 hours, including the *RBM14-RBM4* chimera in which the 5'SS at the end of *RBM14* exon 1 is spliced to the 3' splice site of the RBM4 exon 2 (Fig. 6A, 6B, S9A,B). These chimeras correlated with elevated RTI values from the upstream-most gene (Fig. S9C), and we see that transcription of the downstream gene may increase and/or be coordinated with the upstream gene over a time course of imatinib treatment (Fig. S9D,E).

To quantify expression and the relative abundance of the single RNA and chimeric RNA isoforms in physiologically relevant cells, total RNA from CD34+ cells from three healthy and

two CML patient donors was prepared. Specific primers were designed to amplify either the first exon (or a region of the upstream parental gene) to which the abundance of the chimera – detected with primers across the exon-exon junction – would be normalized (Fig. 6C, Table S4), using reverse transcription followed by quantitative PCR (RT-qPCR). Analysis of four chimeras showed significant expression of *RBM14-RBM4* and *ZNF649-ZNF577* chimeric mRNAs in healthy cells that increased substantially in CD34+ cells from CML patients. In contrast, variable amounts of *ARPC4-TTL3* and *DMC1-DDX17* chimeras were detected in these primary cells. Analysis of our K562 datasets revealed that chimeras are expressed in CTRL (48), IM 1h (60), and IM 18h (31) with significant overlaps in detection (Fig. 6E). Quantification of chimeras by RT-qPCR, using the above method, revealed that some – namely, *ARPC4-TTL3*, *CECR7-IL17RA*, *RBM14-RBM4*, and *ZNF649-ZNF577* – increase significantly by 18 hours of imatinib treatment with little or no increase at 1h (Figs. 6F, S10A), placing them in the category of alternative splicing changes seen at the later timepoint. In contrast, *COX6A1-GATC* is significantly downregulated, while other chimeras do not change. Interestingly, treatment of healthy and CML donor cells with IM 18h did not markedly change absolute levels of chimeras relative to parental genes (Fig. S10B). However, *ARPC4-TTL3* and *RBM14-RBM4* did increase by up to 3-fold relative to CTRL when healthy donor cells were treated for 18 hours and *DMC1-DDX17* increased when CML donors received 1h of imatinib, suggesting the potential for physiological relevance for these cells. Therefore, we conclude that expression of readthrough chimeras is part of a response to imatinib treatment, enabled by both transcriptional readthrough and alternative splicing changes.

DISCUSSION

Here, we have used precision RNA sequencing methods to analyze the earliest changes in gene expression that take place in K562 and CML patient cells upon chemotherapy. Overall, initial responses to imatinib, which directly targets the activity of the *BCR-ABL1* oncogene, are multi-faceted within the first 24 hours. We depict the relative timing and importance of this cascade of gene expression changes as a series of steps and cellular options (Fig. 7). First, long-read sequencing of nascent RNA identified immediate activation of transcriptional readthrough, a blockage or delay in the cleavage of nascent RNA 3' ends [11, 12]. To our knowledge, this is the first identifiable change in gene expression caused by imatinib, which is known to cause cell death, erythroid differentiation, and chemotherapy resistance at much later time points. Second, RNA signatures of longer-term imatinib treatment, such as elevation in erythroid gene expression, were detected as early as 1 hour. Previous studies identified development of imatinib resistance in K562 cells [41] involving protection from cell death by ferroptosis [57] as a simultaneous occurrence with erythropoiesis. These unexpectedly early changes in transcriptional readthrough and gene expression may have prognostic value because higher expression of the CD235a erythroid marker is consistent with high efficacy of imatinib treatment and the acquisition of major molecular remission [58]. Third, changes in alternative splicing at 18h of imatinib treatment were also detected in imatinib-resistant K562 cells (10 months) as well as samples from a CML patient collected at diagnosis and after imatinib treatment (6 months), further suggesting potential early roles in the development of chemotherapy resistance. Fourth, we discovered chimeric RNAs that were induced by imatinib and elevated in CML patients versus healthy donors, emphasizing the outcome of two coordinated events: ***(i)*** repression of splicing in the upstream gene and ***(ii)*** transcriptional readthrough into the downstream gene; these two events enable the elevated expression of chimeric RNAs and potentially novel protein products in CML cells. We discuss these features in more depth below.

The identification of readthrough as an early event was facilitated by its quantification from long-read sequencing of nascent RNA, revealing the proportion of transcripts that are uncleaved at the last annotated polyA cleavage site. The long reads establish the molecular integrity of each RNA molecule, ruling out the possibility that transcripts initiating in downstream intergenic space would be counted. This differs from previous determinations of transcriptional readthrough, which quantify short-read density downstream of gene ends [15, 16,

18, 32, 59]. The mechanisms that lead to early readthrough upon imatinib treatment are not currently known and require further investigation. Importantly, we show that the induction of readthrough is specific to imatinib as a therapeutic, and not seen with homoharringtonin or talazoparib. Imatinib's triggering mechanism is also unlikely to be a direct inhibition of the CPC, because JTE – a drug that does directly suppress cleavage – generates a larger number of readthrough genes that is largely non-overlapping with those triggered by imatinib. Intron retention, which is associated with transcriptional readthrough, could be a cause or an effect of transcriptional readthrough; indeed prior studies in budding yeast showed that intron retention is caused by transcriptional readthrough and vice versa [10, 60].

Our time-resolved study provides a rigorous metric – the ReadThrough Index (RTI) – that enables transcriptional readthrough is an underlying mechanism related to splicing regulation and erythropoiesis. For example, transcriptional readthrough was associated with inefficient splicing in some of the induced erythroid genes (e.g. *UROD*, *HBE1,* and *FTH1*), even though these genes are overall strongly expressed. Indeed, we showed that splicing factor mRNA and protein levels encoding splicing factors decrease following the early detection of readthrough in these genes. Thus, mechanistically, readthrough may lead to down regulation of their gene of origin and ultimately determine the fraction of cells able to attain the erythroid fate. One demonstrated mechanism of downregulation includes intron retention or inclusion of poison exons, both of which reduce mRNA levels by activating nonsense-mediated decay [56], consistent with our data. Transcriptional readthrough correlated with other forms of alternative splicing; DDX17 and EIF4H underwent alternative splicing upon both JTE and imatinib treatments, suggesting that readthrough alone– as opposed to a cellular stress response activated by imatinib treatment– drives alternative splicing in some cases. Because JTE does not induce erythropoiesis, these events have to be seen as differing from the red cell differentiation program activated by imatinib. Long-term responses to imatinib included changes in cassette exon inclusion and intron retention, consistent with previous findings [61]. Considering the potential cascade of molecular events over time (Fig. 7), changes in the expression of splicing factors could lead to downstream changes in mRNA isoform levels [62]. Further studies will be needed to identify the specific relationships between changes in individual splicing factors and mRNA isoforms expressed.

From a pool of alternative isoform changes induced by IM 18h, we identified encoded factors that define erythroid differentiation, such as *GATA1* and *NFE2L1/NRF1* (transcription factors) and *UROD* (a cytoplasmic enzyme involved in the heme biosynthesis pathway). In K562 cells, erythroid differentiation manifested by increased expression of hemoglobin can be induced by hemin [63], and increases cell survival upon imatinib [64-66]. Increased level of TCF11/NRF1 induces expression of proteasome, thereby maintaining proteostasis and playing a cytoprotective role [67, 68] in CML cells with increased activation of the unfolded protein response [26, 27]. Imatinib caused increased expression of the longer *NFE2L1* isoform, also known as *TCF11*. The Nhe4l domain, encoded by the alternative exon, plays a role in NRF1/TCF11 dimerization with other transcription factor MAFG [69] and alters the spectrum of genes regulated by NRF1/TCF11 [61]. Increased expression of *UROD* was observed in *MYCN*-expressing acute lymphoblastic leukemia, and imbalanced synthesis of porphyrins determined cancer cell survival [70]. As postulated recently [60], transcriptional readthrough might facilitate alternative splicing decisions in such cases. We show that alternative mRNA isoforms expressed in imatinib-resistant K562 cells (10 months) occurred much earlier, at 18h of treatment, demonstrating these RNAs as potential molecular signatures of chemotherapy resistance. Interestingly, we identified reduced poison exon inclusion in *EIF4A2* transcripts already within 18h of imatinib treatment. Inclusion of the poison cassette exon between exons 10 and 11 produced a truncated protein that was rapidly degraded in differentiating cardiomyoblasts [71]. Therefore, we infer that reduced inclusion of the *EIF4A2* poison exon would greatly increase protein expression. Given that the promotion of cell growth and proliferation by Bcr-Abl1 involves stimulation of translation [56], this could serve as a rescue mechanism that promotes the progression of cancer despite treatment. This agrees with an earlier observation that alternative splicing of *EIF4A2* was essential for increased protein levels in the development of AML [72].

Effects of transcriptional readthrough and alternative splicing encompass the generation of chimeric transcripts [10, 16, 20, 21]. Chimeric transcripts may be hallmarks of RNA processing in oncogenesis and response to therapy [73]. The mechanism for generating RNA chimeras between two genes in tandem must not only involve transcription from the upstream gene into the downstream gene, but, in addition, intron retention in the upstream gene is required to provide a 5' splice site in the upstream gene that can be spliced to a 3' splice site in the

downstream gene [20]. Thus, the modification of alternative splicing patterns accompanying transcriptional readthrough upon imatinib would drive the generation of new and/or elevated chimeric transcripts. Reduced levels of SRSF1, SRPK1 and/or Gemin5 could impact the readthrough chimera cis-splicing. By far, SRRM1 and SF3B1 were identified as possible regulators of the chimeric transcript cis-splicing [74]. Their activity is modulated by SRPK1-mediated phosphorylation of SRRM1 or SRSF1. If translated, the resulting chimeric proteins might display gain or loss of function of the original proteins, have different stability, cellular localization, or a modified spectrum of interactors. This has been observed in the case of *RBM14-RBM4* [75]. Apart from protein-protein fusions, we also observed lncRNA-protein chimeric transcripts, like lncRNA *CECR7-IL17RA*. Interestingly, stimulation of *IL17RA* cell surface receptor plays a pro-survival role in leukemia, and targeting of IL-17A cytokine increases imatinib's effectiveness in Bcr-Abl1-positive acute lymphoblastic leukemia [76]. In addition, so-called 'ripples' in Pol II transcriptional activity in intergenic regions are associated with chromatin changes that could lead to altered protein-coding gene expression [77]. Epigenetic changes accompanying erythroid differentiation [78] could perhaps support CML resistance [79] by broadly adjusting gene expression.

Taken together, the early modifications in RNA processing we observed upon imatinib treatment – including transcriptional readthrough, intron retention, alternative splicing, and readthrough chimera formation – could be exploited as a potential vulnerability after further study. Fusion proteins synthesized from chimeric transcripts can serve as diagnostic markers or neoepitopes for targeted T cell-mediated therapy [80] or anti-cancer vaccines [81]. Similarly, peptides originating from non-sense mediated decay of translated transcripts with retained introns were found to be presented by MHC I [82]. This underscores the significance of the detection of RNA processing changes as early as 1 hour and progressing through an early timeframe of gene expression changes in response to imatinib. Identified changes in RNA processing and detection of isoforms can deliver robust hallmarks of cell responses to therapeutics, like induction of cell differentiation, possibly allowing for early detection of arising therapy resistance.

MATERIALS AND METHODS

Cell treatments and fractionation

K562 cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% FBS, 100 Units/ml Penicillin, and 100 ug/ml Streptomycin. Cells were treated with 110 mM KCl, 5 μM JTE, or 1 μM Imatinib for 45 min, 1h, 3h, 6h or 18h, as indicated in the figure legends. Each experiment was performed in triplicate to obtain three biological replicates. After washing in cold PBS, $8x10^7$ cells were subjected to fractionation to obtain chromatin-associated nascent RNA, as described [6, 40]. The only modification was an increase in the number of times the pellet was washed with PBS between fractionation steps from one wash to three washes. Fractionation was performed in 4 technical replicates, with $2x10^7$ cells for each replicate. One of the replicates was used to assay the isolated cytoplasmic, nuclear, and chromatin fractions by western blotting. To this end, cold PBS was added to each of the cell and nuclei lysis supernatants to achieve the same final volume. The corresponding volume of PBS was also added to the chromatin pellet. Then, the samples were sonicated on ice at 15% amplitude for 20 sec, and centrifuged for 10 min at 20,000 x g, 4°C. Aliquots of the same volume from each supernatant were incubated with a NuPAGE sample buffer (Thermo #NP0007) at 95°C for 5 min before subjecting them to NuPAGE electrophoresis in a 4-12% Bis-Tris gel (Thermo #NP0322PK2) and MOPS-SDS buffer (Thermo #NP0001). Following transfer, the nitrocellulose membranes were incubated at 4ºC overnight with primary antibodies against: GAPDH (SantaCruz #sc-25778), Pol II (4H8) (SantaCruz #sc-47701), Histone H3 (Novus #NB500-171), and alpha-tubulin (Sigma #T6074).

Primary CD34+ cells were enriched from CML patient samples using magnetic beads with anti-CD34 antibodies (Miltenyi #130-046-702) and separated on columns (Miltenyi #130-042-401), following the manufacturer's protocol. Enrichment up to approx. 70% of CD45+CD34+ cells was confirmed by flow cytometry analysis of the cell surface markers staining. Primary CD34+ cells from healthy donors were expanded in STEM SF II medium (StemSpan #09655) supplemented with the essential growth factors (StemSpan #02691).

Western blotting analysis of protein level in whole cell lysates

After treatment, the cell pellets were washed in cold PBS followed by lysis at 95°C in the SDS lysis buffer and analyzed by western blotting, as described elsewhere [83]. The following

antibodies from Proteintech were used to detect proteins on the nitrocellulose membranes: anti-KEL (#67393-1-Ig), anti-c-Myc (#10828-1-AP), anti-NDC80 (#66960-1-Ig), anti-ATF4 (#10835-1-AP), anti-HBG1 (#25728-1-AP), anti-HBE1 (#12361-1-AP), anti-FTL (#10727-1-AP).

Library preparation for long-read sequencing on the Pacific Biosciences platform

Nascent RNA was isolated from a chromatin pellet in Trizol (Thermo #15596026) after incubating at 50ºC for 10 min with shaking at 1400 rpm, followed by extraction in chloroform and RNA purification using RNeasy Mini kit (Qiagen #74104). DNA was removed by on-column digestion using the RNase-Free DNase set (Qiagen #79254). RNA samples were subjected to library preparation for PacBio long-read sequencing according to the procedure described before [40]. Briefly, rRNA and poly(A+) RNA were depleted using RiboMinus Eukaryote System v2 (Thermo #A15026) and DynaBeads mRNA DIRECT Micropurification (Thermo #61021) kits, respectively, followed by ligation of an adapter with UMI to the 3' end of RNA by T4 RNA ligase (NEB #M0351L), then reverse transcription by SMARTer PCR cDNA Synthesis kit (Takara/Clontech #634925), and library amplification using Advantage 2 PCR kit (Takara/Clontech #639137). Libraries were subjected to size selection using AMPure XP beads (Beckman Coulter #A63880), to select for fragments longer than 500nt. The PacBio HiFi library was prepared using SMRTbell Express Template Prep Kit 2.0 from PacBio following the manufacturer's instructions. Libraries were sequence on PacBio RS II and Revio instruments at Yale Center for Genome Discovery.

Alongside the many strengths of long read sequencing are a few limitations, which are well-known and important to acknowledge [1, 2]. Library preparation for long-read sequencing, including amplification by limited cycles of PCR, might bias the results towards increased splicing owing to more robust amplification of shorter reads. This is visible when comparing the fraction of spliced transcripts based on reads subsampled to start at the TSS versus anywhere in the gene (e.g. Fig 2B). To reduce this bias in the analyses focused on splicing, we included control conditions from the same library preparation as a reference. Moreover, the classification of alternative splicing events as 'significant' is limited by the required level of coverage and significance. Changing cutoffs and thresholds thus modify the list of transcripts detected as "hits" in different analyses; for example, transcripts significant for a given RTI have a

requirement for the indicated number of long reads (e.g. 20 or 30); this limits the number of genes we can analyze for alternative splicing that co-occurs with readthrough.

Nuclear and total mRNA preparation for short-read sequencing analysis

RNA from control and treated cells, whole cells (total mRNA selected for poly(A+) RNA during library preparation) or nuclei (nuclear RNA depleted of rRNA using RiboMinus Eukaryote System v2) obtained as an intermediate step of the cell fractionation procedure, was isolated following the procedure described above. All treatments were repeated three times as separate biological experiments. Samples were submitted for library preparation at Yale Center for Genome Analysis core facility and sequencing analysis using Illumina Nova-seq HiSeq of paired-end 150 bp fragments.

Reverse transcription, real-time PCR, and Sanger sequencing

Reverse transcription of total RNA was performed using the SuperScript III Reverse Transcriptase kit (Thermo #18080085), Universal Spike I RNA (TATAA Biocenter #RS25SI), and anchored oligo(dT_{18}) (Merck #D2773). Reverse transcription was carried out 50°C for 60 minutes, followed by inactivation at 70°C for 15 minutes. The final cDNA solution was diluted 1:3 for chimeric junction primers and 1:10 for parental exon primers.

The junction sequence of readthrough chimeras predicted by SOAPfuse informed the design of junction-specific primers. Real-time semi-quantitative PCR (RT-qPCR) was performed using the iTaq Universal SYBR Green Master Mix (Bio-Rad #1725121) on a CFX Connect Real-Time PCR Detection System (Bio-Rad), performing 40 cycles of denaturation at 95ºC for 10 sec followed by annealing/extension at 60ºC for 10 sec. RT-qPCR amplicons were quantified using the ΔΔCt Method, with universal spike-in RNA as the internal control. RT-qPCR products were electrophoresed and resolved on a 1.75% agarose gel. Bands of interest were excised and purified from the gel using the NucleoSpin Gel and PCR Clean-up kit (Takara Bio #740609) according to the manufacturer's instructions. For junction sequence validation, pre-mixed primer and cDNA samples were sent to Quintara Bioscience for Sanger sequencing.

Short-read sequencing analysis

Short-read sequencing pre-processing. The SRA Toolkit was used to download FASTQ files from the study by Bai J et al. [54] (control samples: SRR29032856, SRR29032857, SRR29032858; Imatinib-resistant samples: SRR29032862, SRR29032863, SRR29032864). For

Illumina datasets generated in this study, Illumina adapters were trimmed using FASTP with the following settings: -l 15 -c --detect_adapter_for_pe --trim_poly_g --dedup --trim_front2 0 -w 15. Quality control reports for all datasets were generated using FASTQC. All Illumina datasets were mapped to the hg38 genome using the STAR (v2.7.7a) aligner with settings: --runThreadN 6 --runMode alignReads --outFilterMultimapNmax 1 --alignSJoverhangMin 8 --alignIntronMin 21 --alignIntronMax 10000 --outSAMtype BAM SortedByCoordinate. The featureCounts program (Subread/2.0.3) was used to quantify the number of reads mapping to each gene with the settings: -T 6 -p --countReadPairs -t exon -g gene_id.

Splicing per intron calculation. The Python tool SPLICE-q was used to quantify coverage over the 5' and 3' splice sites of each intron [84]. Splicing per intron (SPI) values were calculated based on the SPLICE-q outputs by dividing the number of spliced counts (sum of split coverage over 5' and 3' splice sites) by the total number of counts (sum of unsplit and split coverage over 5' and 3' splice sites) mapping to that intron. To identify introns with significant changes in splicing, chi-square tests were performed by comparing the number of spliced reads and the number of unspliced reads for pairs of replicates (i.e., replicate one of untreated versus replicate one of treated, etc.). A single, adjusted p-value for each intron was determined using the Benjamini-Hochberg, or False-Discovery Rate, procedure. To be classified as significant, the adjusted p-value had to be less than 0.01. To be considered for SPI analysis, an intron had to have at least 500 reads pooled across all three replicates for both conditions.

Alternative splicing analysis. Changes in alternative splicing events for treated samples relative to the control were detected using rMATS-turbo [85] in a Conda environment with the following settings: python=3.7 rmats.py -t paired --readLength 150 --variable-read-length --novelSS --nthread 6. To be classified as significant, a change in an alternative splicing event had to have an absolute value of inclusion level difference greater than 0.1 and an adjusted p-value less than or equal to 0.05. A read count cutoff was also implemented to keep only events with at least 20 inclusion and skipping reads across per condition, pooled across all three replicates. Only junction reads were considered in the analysis of alternative splicing events. Rmats2sashimiplot was used to quantify and visualize selected changes in splicing for particular genes.

Gene expression analysis with DESeq2. Differential gene expression analysis was conducted using RStudio (v4.3.1) and the R/Bioconductor package DESeq2 (v1.40.2). A variance

stabilizing transformation was applied using the vst function (with the blind argument set to FALSE). Genes were classified as differentially expressed between control and treated cells if they exhibited a $\log_2$(Fold Change) greater than 0.6 (upregulated) or lower than -0.6 (downregulated) with an adjusted P-value less than 0.05.

Gene ontology term enrichment analysis. Functional annotation analysis of selected genes was performed using g:GOSt (g:Profiler RRID: SCR_006809, https://biit.cs.ut.ee/gprofiler/gost) with Benjamin-Hochberg False Discovery Rate (FDR) set to 0.05. Annotation enrichment was analyzed using Panther (https://www.pantherdb.org) with Fisher's Exact test requiring $p < 0.05$. Analysis of protein network enrichment was performed by searching in the ClueGO (RRID: SCR_005748) with GeneOntology (Nov2024) at CytoScape (v 3.10.3) (RRID: SCR_003032) with Benjamini-Hochberg FDR set to 0.05.

Identification of readthrough chimera transcripts with SOAPfuse. RNA-sequencing data were preprocessed using FASTP and mapped to the hg38 genome using the STAR aligner with settings: --runThreadN 16 --runMode alignReads --outFilterMultimapNmax 1 --alignSJoverhangMin 1 --alignIntronMin 21 --alignIntronMax 200000 --readFilesCommand zcat --outSAMtype BAM SortedByCoordinate --outFileNamePrefix ../results/bam/ --sjdbGTFfile {input.gtf}--genomeDir {input.genome_dir} --readFilesIn {input.R1_P} {input.R2_P} --chimOutType WithinBAM Junctions --chimMultimapNmax 2 --chimSegmentMin 15 --chimOutJunctionFormat 1 --chimMultimapScoreRange 3 --chimJunctionOverhangMin 15 --chimScoreJunctionNonGTAG -4 --chimNonchimScoreDropMin 10 --alignMatesGapMax 200000 --chimScoreSeparation 5 --chimScoreMin 0. Resulting bam files were used for alternative splicing analysis by rmats2Sashimiplot.

SOAPfuse (v1.26) was used to detect fusion transcripts from preprocessed paired-end RNA-seq data [86]. Alignment database was build using the hg38 human genome and gene annotation files from ENCODE. Reads pooled from biological replicates of the same condition were aligned to the database using the BWA aligner. SOAPfuse analysis settings required minimum 3 span reads and 3 junction covering reads, and a minimal intra-chromosome distance between genes of 500 nt. For comparative analysis, the output list was filtered to include only fusions with at least a total of 6 supporting reads and removed any fusions with ribosomal, small nucleolar, and small

nuclear RNAs. To specifically focus on potential readthrough chimeras in each RNA-seq dataset, only chimeras between adjacent genes sharing the same strand and orientation were considered.

Long-read sequencing analysis

Long-read sequencing pre-processing. Adapter sequences were removed using cutadapt (v3.4) with the following settings: cutadapt --cores=8 --discard-untrimmed --no-indels --revcomp -e 0.15. Long reads were mapped to the GRCh38.p14/hg38 genome using minimap2 (v2.22) with settings -ax splice:hq –secondary=no -t 12 -a -u f –seed 14. SAM outputs from minimap2 were converted to BAM and BED files using SAMtools (v1.21) and BEDTools (v2.30.0), respectively. UMICollapse (v1.0.0) was used for deduplication with the following settings: umicollapse -Xmx60g bam --umi-sep. Polyadenylated reads and non-unique reads (i.e., entries with duplicate read names) were removed using a custom script. Pre-processing steps and custom scripts to filter poly(A)+ reads were consolidated in Snakemake.

Genome annotation filtering. The "Convert GTF to BED12" feature of the Galaxy web server (https://usegalaxy.org/) was used to generate a BED12 version of the hg38 genome with entries for all transcripts. For splicing analysis, transcript entries were filtered to keep only the canonical transcript as defined by Ensembl for each gene. For readthrough analysis, the transcript with the distal-most 3' end was used to prevent false-positive readthrough identification.

Comparison of gene coverage. Analysis of gene region coverage was performed using deepTools2 [87] in a Conda environment. First, bamCoverage tool was used with the following options: --binSize 40 --normalizeUsing RPKM --effectiveGenomeSize 2913022398. Coverage from 3 replicates was averaged using bigwigAverage. The resulting bigWig files were used to compare coverage of reads within the gene body and 4 kb before the annotated transcription start site (TSS) or after the annotated transcription end site. A coverage matrix was prepared using computeMatrix scale-regions with the following settings: -R hg38_genes.bed (a reference containing only protein-coding genes was used) -bs 40 -b 4000 -a 4000 --regionBodyLength 7000 --skipZeros --missingDataAsZero. For plotting, the coverage traces were aligned within the gene body region.

Splicing status classification and co-transcriptional splicing efficiency calculation. Splicing status classification and co-transcriptional splicing efficiency (CoSE) calculation were performed as described [6]. The splicing status of each long read was assigned based on the

number of spliced introns it contained relative to the total number of introns it overlapped. Reads for which all possible introns were removed were classified as “all spliced”, reads for which all possible introns were not yet removed were classified as “all unspliced”, and reads with a combination of spliced and unspliced introns were classified as “partially spliced”. CoSE values were calculated for all introns from canonical transcripts. CoSE represents the number of spliced reads spanning an intron divided by the total number of reads spanning an intron.

Readthrough status classification and readthrough index calculation. Long reads were intersected with the last exon of the representative transcript for each gene using bedtools. For reads overlapping multiple genes, only the overlap with the greatest coverage was used for analysis. For each gene, the readthrough boundary was defined as 100 nucleotides downstream of the PAS. A read was classified as “readthrough” if its 3’ end extended beyond the readthrough boundary or classified as “gene body” if its 3’ end did not extend beyond the readthrough boundary. The readthrough index (RTI) was calculated for each gene by dividing the number of readthrough reads by the total number of reads mapping to the last exon of the gene. To be classified as a readthrough gene, the RTI had to be greater than or equal to 0.2.

The controls used for splicing analysis (CoSE, % of spliced reads) represent the control samples prepared concurrently with the imatinib-treated samples (GSM8673083, GSM8673084, GSM8673085); selecting this subset of controls corrects for potential differences in library preparation. The controls used for comparison of CoSE vs RTI values represent all control samples (GSM8673080, GSM8673081, GSM8673082, GSM8673083, GSM8673084, GSM8673085).

PAS coverage calculation. Long-read coverage in the region downstream of PASs was quantified using a custom Python script. Coverage was normalized to the position 200 nucleotides upstream of each PAS.

Data availability

Long and short read sequencing raw data are available through the National Center for Biotechnology Information (NCBI) under GEO accession number GSE283849 and GSE 283813.

Code availability

Code for RNA-seq data processing and downstream analyses is publicly available at https://github.com/NeugebauerLab/Cancer-stress.

Acknowledgements

The authors would like to thank members of the Neugebauer and Steitz labs for helpful discussions and Joan Steitz for comments on the manuscript. The authors are grateful for support from the National Institutes of Health (R01 GM140735 and R01 HL167071 to KMN). MS was supported by the training grant (1T32GM145469-01A1) and an F31 predoctoral fellowship (1F31ES037182-01) from NIEHS. PPB received additional support from a Polish National Science Center grant (2018/30/M/NZ3/00274) and a Senior Fulbright Fellowship (2022/23). PPB is a Forbeck Scholar (2024/27) and a member of the 'TRANSLACORE' Cost Action CA21154. The Yale Cooperative Center for Excellence in Hematology providing CD34+ cells was supported by NIDDK Grant U54DK106857-10S1. Data acquisition at Yale Center for Genomic Analysis was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number 1S10OD030363-01A1. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Author Contributions

P.P.B., M.S., J.L. and K.M.N. conceptualized the study. P.P.B. and J.L. carried out all wet lab experiments, P.P.B., M.S., and J.L. analyzed the data. The original manuscript was written by P.P.B. and K.M.N. and was subsequently edited by P.P.B., M.S., J.L., and K.M.N. The work was supervised by P.P.B. and K.M.N.

Declaration of Interests

The authors declare no conflict of interest.

REFERENCES

1. Shine, M., et al., *Co-transcriptional gene regulation in eukaryotes and prokaryotes.* Nat Rev Mol Cell Biol, 2024. **25**(7): p. 534-554.
2. Merens, H.E., et al., *Timing is everything: advances in quantifying splicing kinetics.* Trends Cell Biol, 2024. **34**(11): p. 968-981.
3. Ameur, A., et al., *Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain.* Nat Struct Mol Biol, 2011. **18**(12): p. 1435-40.
4. Tilgner, H., et al., *Microfluidic isoform sequencing shows widespread splicing coordination in the human transcriptome.* Genome Res, 2018. **28**(2): p. 231-242.
5. Drexler, H.L., K. Choquet, and L.S. Churchman, *Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores.* Mol Cell, 2020. **77**(5): p. 985-998 e8.
6. Reimer, K.A., et al., *Co-transcriptional splicing regulates 3' end cleavage during mammalian erythropoiesis.* Mol Cell, 2021. **81**(5): p. 998-1012 e7.
7. Kim, H.K., et al., *Alternative splicing isoforms in health and disease.* Pflugers Arch, 2018. **470**(7): p. 995-1016.
8. Kim, S.W., et al., *Widespread intra-dependencies in the removal of introns from human transcripts.* Nucleic Acids Res, 2017. **45**(16): p. 9503-9513.
9. Herzel, L., K. Straube, and K.M. Neugebauer, *Long-read sequencing of nascent RNA reveals coupling among RNA processing events.* Genome Res, 2018. **28**(7): p. 1008-1019.
10. Alpert, T., et al., *Widespread Transcriptional Readthrough Caused by Nab2 Depletion Leads to Chimeric Transcripts with Retained Introns.* Cell Rep, 2020. **33**(4): p. 108324.
11. Rosa-Mercado, N.A. and J.A. Steitz, *Who let the DoGs out? - biogenesis of stress-induced readthrough transcripts.* Trends Biochem Sci, 2022. **47**(3): p. 206-217.
12. Morgan, M., et al., *It's a DoG-eat-DoG world-altered transcriptional mechanisms drive downstream-of-gene (DoG) transcript production.* Mol Cell, 2022. **82**(11): p. 1981-1991.
13. Vilborg, A., et al., *Widespread Inducible Transcription Downstream of Human Genes.* Mol Cell, 2015. **59**(3): p. 449-61.
14. Cardiello, J.F., J.A. Goodrich, and J.F. Kugel, *Heat Shock Causes a Reversible Increase in RNA Polymerase II Occupancy Downstream of mRNA Genes, Consistent with a Global Loss in Transcriptional Termination.* Mol Cell Biol, 2018. **38**(18).
15. Rosa-Mercado, N.A., et al., *Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression.* Mol Cell, 2021. **81**(3): p. 502-513 e4.
16. Hadar, S., et al., *Stress-induced transcriptional readthrough into neighboring genes is linked to intron retention.* iScience, 2022. **25**(12): p. 105543.
17. Debes, C., et al., *Ageing-associated changes in transcriptional elongation influence longevity.* Nature, 2023. **616**(7958): p. 814-821.
18. Abe, K., et al., *Downstream-of-gene (DoG) transcripts contribute to an imbalance in the cancer cell transcriptome.* Sci Adv, 2024. **10**(27): p. eadh9613.
19. Caldas, P., et al., *Transcription readthrough is prevalent in healthy human tissues and associated with inherent genomic features.* Commun Biol, 2024. **7**(1): p. 100.
20. Zhang, Y., et al., *Chimeric transcript generated by cis-splicing of adjacent genes regulates prostate cancer cell proliferation.* Cancer Discov, 2012. **2**(7): p. 598-607.

21. Grosso, A.R., et al., *Pervasive transcription read-through promotes aberrant expression of oncogenes and RNA chimeras in renal carcinoma.* Elife, 2015. **4**.
22. Anczukow, O., et al., *Steering research on mRNA splicing in cancer towards clinical translation.* Nat Rev Cancer, 2024. **24**(12): p. 887-905.
23. Schmitz, U., et al., *Widespread Aberrant Alternative Splicing despite Molecular Remission in Chronic Myeloid Leukaemia Patients.* Cancers (Basel), 2020. **12**(12).
24. van Galen, P., et al., *Integrated Stress Response Activity Marks Stem Cells in Normal Hematopoiesis and Leukemia.* Cell Rep, 2018. **25**(5): p. 1109-1117 e5.
25. Podszywalow-Bartnicka, P. and K.M. Neugebauer, *Multiple roles for AU-rich RNA binding proteins in the development of haematologic malignancies and their resistance to chemotherapy.* RNA Biol, 2024. **21**(1): p. 1-17.
26. Kusio-Kobialka, M., et al., *The PERK-eIF2alpha phosphorylation arm is a pro-survival pathway of BCR-ABL signaling and confers resistance to imatinib treatment in chronic myeloid leukemia cells.* Cell Cycle, 2012. **11**(21): p. 4069-78.
27. Dudka, W., et al., *Targeting integrated stress response with ISRIB combined with imatinib treatment attenuates RAS/RAF/MAPK and STAT5 signaling and eradicates chronic myeloid leukemia cells.* BMC Cancer, 2022. **22**(1): p. 1254.
28. Prabhu, S., et al., *A novel mechanism for Bcr-Abl action: Bcr-Abl-mediated induction of the eIF4F translation initiation complex and mRNA translation.* Oncogene, 2007. **26**(8): p. 1188-200.
29. Podszywalow-Bartnicka, P., et al., *Downregulation of BRCA1 protein in BCR-ABL1 leukemia cells depends on stress-triggered TIAR-mediated suppression of translation.* Cell Cycle, 2014. **13**(23): p. 3727-41.
30. Biancon, G., et al., *Precision analysis of mutant U2AF1 activity reveals deployment of stress granules in myeloid malignancies.* Mol Cell, 2022. **82**(6): p. 1107-1122 e7.
31. Veloso, A., et al., *Genome-wide transcriptional effects of the anti-cancer agent camptothecin.* PLoS One, 2013. **8**(10): p. e78190.
32. Bhola, M., et al., *RNA interacts with topoisomerase I to adjust DNA topology.* Mol Cell, 2024. **84**(17): p. 3192-3208 e11.
33. Liu, Y., et al., *O-GlcNAcylation promotes topoisomerase IIalpha catalytic activity in breast cancer chemoresistance.* EMBO Rep, 2023. **24**(7): p. e56458.
34. Cui, Y., et al., *Elevated pre-mRNA 3′ end processing activity in cancer cells renders vulnerability to inhibition of cleavage and polyadenylation.* Nat Commun, 2023. **14**(1): p. 4480.
35. Listerman, I., A.K. Sapra, and K.M. Neugebauer, *Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells.* Nat Struct Mol Biol, 2006. **13**(9): p. 815-22.
36. Juge, F., et al., *The SR protein B52/SRp55 is required for DNA topoisomerase I recruitment to chromatin, mRNA release and transcription shutdown.* PLoS Genet, 2010. **6**(9): p. e1001124.
37. Boswell, S.A., et al., *Total RNA-seq to identify pharmacological effects on specific stages of mRNA synthesis.* Nat Chem Biol, 2017. **13**(5): p. 501-507.
38. Lozzio, C.B. and B.B. Lozzio, *Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome.* Blood, 1975. **45**(3): p. 321-34.
39. Liu, L., et al., *The anticancer compound JTE-607 reveals hidden sequence specificity of the mRNA 3′ processing machinery.* Nat Struct Mol Biol, 2023. **30**(12): p. 1947-1957.

40. Reimer, K.A. and K.M. Neugebauer, *Preparation of Mammalian Nascent RNA for Long Read Sequencing.* Curr Protoc Mol Biol, 2020. **133**(1): p. e128.
41. Jacquel, A., et al., *Imatinib induces mitochondria-dependent apoptosis of the Bcr-Abl-positive K562 cell line and its differentiation toward the erythroid lineage.* FASEB J, 2003. **17**(14): p. 2160-2.
42. An, X., et al., *Global transcriptome analyses of human and murine terminal erythroid differentiation.* Blood, 2014. **123**(22): p. 3466-77.
43. Edwards, C.R., et al., *A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages.* Blood, 2016. **127**(17): p. e24-e34.
44. Pimentel, H., et al., *A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis.* Nucleic Acids Res, 2014. **42**(6): p. 4031-42.
45. Sinnakannu, J.R., et al., *SRSF1 mediates cytokine-induced impaired imatinib sensitivity in chronic myeloid leukemia.* Leukemia, 2020. **34**(7): p. 1787-1798.
46. Zhong, X.Y., et al., *Regulation of SR protein phosphorylation and alternative splicing by modulating kinetic interactions of SRPK1 with molecular chaperones.* Genes Dev, 2009. **23**(4): p. 482-95.
47. Wang, H., et al., *SRPK1-siRNA suppresses K562 cell growth and induces apoptosis via the PARP-caspase3 pathway.* Mol Med Rep, 2018. **17**(1): p. 2070-2076.
48. Lee, J.H., et al., *Alterations in Gemin5 expression contribute to alternative mRNA splicing patterns and tumor cell motility.* Cancer Res, 2008. **68**(3): p. 639-44.
49. Francisco-Velilla, R., et al., *Alternative splicing events driven by altered levels of GEMIN5 undergo translation.* RNA Biol, 2024. **21**(1): p. 23-34.
50. Bergeron, D., et al., *Intronic small nucleolar RNAs regulate host gene splicing through base pairing with their adjacent intronic sequences.* Genome Biol, 2023. **24**(1): p. 160.
51. Zhang, S., et al., *RNA binding motif protein 10 suppresses lung cancer progression by controlling alternative splicing of eukaryotic translation initiation factor 4H.* EBioMedicine, 2020. **61**: p. 103067.
52. Kim, J., et al., *Autophagy-dependent splicing control directs translation toward inflammation during senescence.* Dev Cell, 2025. **60**(3): p. 364-378 e7.
53. Wu, D., et al., *An alternative splicing isoform of eukaryotic initiation factor 4H promotes tumorigenesis in vivo and is a potential therapeutic target for human cancer.* Int J Cancer, 2011. **128**(5): p. 1018-30.
54. Bai, J., et al., *Lycorine attenuated proliferation and induced apoptosis on imatinib-resistant K562 cell by inhibiting autophagy.* Discov Oncol, 2024. **15**(1): p. 217.
55. Liu, X., et al., *Unravelling the role of NFE2L1 in stress responses and related diseases.* Redox Biol, 2023. **65**: p. 102819.
56. Maquat, L.E., *Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics.* Nat Rev Mol Cell Biol, 2004. **5**(2): p. 89-99.
57. Trombetti, S., et al., *Over-Expressed GATA-1(S), the Short Isoform of the Hematopoietic Transcriptional Factor GATA-1, Inhibits Ferroptosis in K562 Myeloid Leukemia Cells by Preventing Lipid Peroxidation.* Antioxidants (Basel), 2023. **12**(3).
58. Krishnan, V., et al., *A single-cell atlas identifies pretreatment features of primary imatinib resistance in chronic myeloid leukemia.* Blood, 2023. **141**(22): p. 2738-2755.
59. Vilborg, A., et al., *Comparative analysis reveals genomic features of stress-induced transcriptional readthrough.* Proc Natl Acad Sci U S A, 2017. **114**(40): p. E8362-E8371.

60. Barr, K., et al., *Transcription termination promotes splicing efficiency and fidelity in a compact genome.* bioRxiv, 2025.
61. Wang, M., et al., *TCF11 Has a Potent Tumor-Repressing Effect Than Its Prototypic Nrf1alpha by Definition of Both Similar Yet Different Regulatory Profiles, With a Striking Disparity From Nrf2.* Front Oncol, 2021. **11**: p. 707032.
62. Rogalska, M.E., et al., *Transcriptome-wide splicing network reveals specialized regulatory functions of the core spliceosome.* Science, 2024. **386**(6721): p. 551-560.
63. Kawasaki, N., et al., *Control of hemoglobin synthesis in erythroid differentiating K562 cells. I. Role of iron in erythroid cell heme synthesis.* Arch Biochem Biophys, 1996. **328**(2): p. 289-94.
64. Nagai, T., et al., *Hemin reduces cellular sensitivity to imatinib and anthracyclins via Nrf2.* J Cell Biochem, 2008. **104**(2): p. 680-91.
65. Bonovolias, I.D. and A.S. Tsiftsoglou, *Hemin counteracts the repression of Bcl-2 and NrF2 genes and the cell killing induced by imatinib in human Bcr-Abl(+) CML cells.* Oncol Res, 2009. **17**(11-12): p. 535-47.
66. Schaefer, B. and S. Behrends, *Translocation of heme oxygenase-1 contributes to imatinib resistance in chronic myelogenous leukemia.* Oncotarget, 2017. **8**(40): p. 67406-67421.
67. Sotzny, F., et al., *TCF11/Nrf1-Mediated Induction of Proteasome Expression Prevents Cytotoxicity by Rotenone.* Antioxid Redox Signal, 2016. **25**(16): p. 870-885.
68. Nedomova, M., et al., *DDI2 protease controls embryonic development and inflammation via TCF11/NRF1.* iScience, 2024. **27**(10): p. 110893.
69. Ho, D.V., et al., *Characterization of NFE2L1-616, an isoform of nuclear factor-erythroid-2 related transcription factor-1 that activates antioxidant response element-regulated genes.* Sci Rep, 2023. **13**(1): p. 19900.
70. Fukuda, Y., et al., *Upregulated heme biosynthesis, an exploitable vulnerability in MYCN-driven leukemogenesis.* JCI Insight, 2017. **2**(15).
71. Murata, K., et al., *PRMT1 Deficiency in Mouse Juvenile Heart Induces Dilated Cardiomyopathy and Reveals Cryptic Alternative Splicing Products.* iScience, 2018. **8**: p. 200-213.
72. Liu, L., et al., *The splicing factor RBM17 drives leukemic stem cell maintenance by evading nonsense-mediated decay of pro-leukemic factors.* Nat Commun, 2022. **13**(1): p. 3833.
73. Jia, Y., Z. Xie, and H. Li, *Intergenically Spliced Chimeric RNAs in Cancer.* Trends Cancer, 2016. **2**(9): p. 475-484.
74. Chwalenia, K., et al., *A cell-based splicing reporter system to identify regulators of cis-splicing between adjacent genes.* Nucleic Acids Res, 2019. **47**(4): p. e24.
75. Chang, S.H., et al., *Alanine repeats influence protein localization in splicing speckles and paraspeckles.* Nucleic Acids Res, 2014. **42**(22): p. 13788-98.
76. Wang, F., et al., *Targeting IL-17A enhances imatinib efficacy in Philadelphia chromosome-positive B-cell acute lymphoblastic leukemia.* Nat Commun, 2024. **15**(1): p. 203.
77. Engreitz, J.M., et al., *Local regulation of gene expression by lncRNA promoters, transcription and splicing.* Nature, 2016. **539**(7629): p. 452-455.
78. Ryzhkova, A. and N. Battulin, *Genome Reorganization during Erythroid Differentiation.* Genes (Basel), 2021. **12**(7).

79. Krishnan, V., et al., *Integrating genetic and epigenetic factors in chronic myeloid leukemia risk assessment: toward gene expression-based biomarkers.* Haematologica, 2022. **107**(2): p. 358-370.
80. Huang, P., et al., *Neoantigens in cancer immunotherapy: focusing on alternative splicing.* Front Immunol, 2024. **15**: p. 1437774.
81. Shen, L., et al., *RNA Transcription and Splicing Errors as a Source of Cancer Frameshift Neoantigens for Vaccines.* Sci Rep, 2019. **9**(1): p. 14184.
82. Smart, A.C., et al., *Intron retention is a source of neoepitopes in cancer.* Nat Biotechnol, 2018. **36**(11): p. 1056-1058.
83. Wolczyk, M., et al., *TIAR and FMRP shape pro-survival nascent proteome of leukemia cells in the bone marrow microenvironment.* iScience, 2023. **26**(4): p. 106543.
84. de Melo Costa, V.R., et al., *SPLICE-q: a Python tool for genome-wide quantification of splicing efficiency.* BMC Bioinformatics, 2021. **22**(1): p. 368.
85. Wang, Y., et al., *rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data.* Nat Protoc, 2024. **19**(4): p. 1083-1104.
86. Jia, W., et al., *SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data.* Genome Biol, 2013. **14**(2): p. R12.
87. Ramirez, F., et al., *deepTools2: a next generation web server for deep-sequencing data analysis.* Nucleic Acids Res, 2016. **44**(W1): p. W160-5.

Figure Captions

Figure 1. Imatinib treatment induces transcriptional readthrough detected by long-read sequencing. (A) Schematic showing transcriptional readthrough, which occurs upon impaired 3' end cleavage. JTE-607 is a chemical inhibitor of CPSF73, and imatinib is a targeted therapeutic that inhibits the Bcr-Abl1 onco-kinase. PAS - polyadenylation signal, TSS - transcription start site, Pol II - RNA polymerase II. **(B)** Metagene analysis of RNA-Seq profiles for control (CTRL) and treated (IM 1h, IM 18h, JTE 1h) cells. The y-axis represents the read coverage combined from three biological replicates and analyzed in 40-nucleotide bins. RPKM – Reads Per Kilobase per Million mapped reads, TSS - transcription start site, PAS - polyadenylation signal. **(C)** Schematic representation of transcriptional readthrough index (RTI) calculation. **(D)** Distribution of RTI values per gene for each condition presented for quantiles. One-way ANOVA with Dunnett's test was used to compare impact of IM1h to other conditions (***$p<0.00001$, *$p=0.002$); n=number of genes in a quantile. **(E)** Fraction of genes with more or less than 20% of long reads displaying readthrough (RTI = 0 in beige; RTI < 0.2 in yellow; RTI ≥ 0.2 in maroon). In panels D and E, genes were required to have ≥ 10 reads/replicate to be included in these analyses. **(F)** Upset plot comparing genes with RTI ≥ 0.2 in CTRL, IM 1h, and IM 18h cells. **(G)** Heat map of RTI values for each expressed gene (n=1817 genes from panel **E**), ordered by RTI value in CTRL.

Figure 2. Transcriptional readthrough is associated with higher incidence of intron retention. K562 cells were untreated (CTRL) or treated with Imatinib (IM) or JTE-607 (JTE) for 1h or 18h. **(A)** Analysis of co-transcriptional splicing efficiency (CoSE) based on long-read sequencing of nascent RNA. *Upper panel*: Schematic explanation of CoSE metric. *Lower panel:* Change in CoSE value for IM 1h or 18h IM relative to CTRL plotted for individual introns (n=4662 total introns for 1h IM; 4846 for 18h IM analyses). Introns with significant changes in splicing efficiency (|Change in CoSE| ≥ 0.1 and adjusted p-value < 0.05) are marked in blue. An adjusted p-value was determined for each intron using chi-square tests for pairs of replicates and the Benjamini-Hochberg procedure. **(B)** *Upper panel*: schematic explanation of classification of long reads as "gene body" or "readthrough". *Lower panel:* Percent nascent RNA long reads with all introns retained ("All Unspliced"). Reads were either filtered for those beginning within 100 bp of the (TSS) or not filtered (noTSS). The black line represents the mean of 3 biological

replicates. **(C)** Alignment of nascent RNA long reads to the gene uroporphyrinogen decarboxylase (*UROD*). Reads are colored according to their splicing status and downsampled to 250 reads total. RTI values (mean ± S.E.M) for *UROD* from three independent experiments are shown. The black dotted line represents the beginning of the readthrough region, and the 3′ end of each read corresponds to Pol II position. PAS – polyadenylation signal.

Figure 3. Imatinib induced transcriptional readthrough is present in genes associated with erythropoiesis. (A) Changes in RTI and mRNA levels upon IM treatment for genes with RTI ≥ 0.2 in IM18h (n=1818), ordered by RTI in IM1h. Log2FC - log2 of fold change relative to CTRL. **(B)** Prevalence of genes with RTI ≥ 0.2 after IM 1h or 18h overlapping with genes that are up- or down-regulated upon progenitor differentiation into poly- and ortho-erythroid cells (Erythroid diff. UP or DOWN, respectively). Values represent numbers of genes. **(C)** Changes in transcriptional readthrough and mRNA levels upon 18-hour treatment with Imatinib (IM), Talazoparib (BMN), or Homoharringtonine (HHT) quantified for hemoglobin e (*HBE1*) and ferritin heavy chain (*FTH1*) using RT-qPCR. Significance based on two-tailed Student's t-test (*** $p < 0.005$). **(D)** Western blot detection of proteins encoded by selected up- and down-regulated genes important in erythroid differentiation and cell cycle regulation. Equal amounts of protein from whole cell lysates were loaded in each lane; a- Tubulin for reference. **(E)** Erythroid differentiation markers analyzed by flow cytometry in CTRL and cells treated with IM for 48 hours (IM 48h). Antibodies conjugated to fluorophores BV421 and BB515 were used for detection of CD45 and CD235a proteins, respectively. (**Upper panel**) Representative plot showing percent cells positive for each CD marker; (**lower panel**) black line represents mean value of three replicates (points). Significance based on two-tailed Student's t-test (*** $p < 0.005$).

Figure 4. Widespread alternative splicing in response to imatinib correlates with readthrough in some cases. (A) Change in SPI (splicing per intron) values from short-read sequencing of mRNA in IM 1h (upper) or 18h (lower) relative to CTRL. Introns (n = ~35300) detected by at least 500 reads and with significant changes in SPI (adjusted p-value < 0.05) are marked in blue (20 in IM1h and 193 in IM18h). Adjusted p-values were determined using chi-square tests for pairs of replicates and the Benjamini-Hochberg procedure. **(B)** Significant changes in alternative splicing in cells treated with JTE or IM 18h. CE - cassette exon, RI – retained intron, A3SS or A5SS - alternative 3′ or 5′ splice site, MXE - mutually exclusive events.

Number of significant events (FDR < 0.05) detected by at least 20 reads and with absolute inclusion level difference to CTRL > 0.1. **(C)** Comparison of genes with RTI ≥ 0.2 (RTI) and genes with significant changes in RI and CE at the level of nuclear RNA in IM 18h. Only single intersections with RTI are presented. **(D)** Comparison of genes with significant changes in RI (left panel) or CE (right panel) in IM or JTE 18h. **(E)** Sashimi plot for *EIF4H* (left) and quantification of isoform usage (right) in CTRL versus JTE or IM 18h. Isoforms 1 and 2 represent the skipping and inclusion, respectively, of exon 5 (beige highlight). Percent transcripts belonging to each isoform in 3 replicates (mean ± S.E.M.). Significance based on two-tailed Student's t-test (*, p<0.05; ***, p<0.001).

Figure 5. Early changes in alternative splicing control response to Imatinib in a longer time course. (A) Comparison of IM-resistant K562 cells (GSE267522) and K562 cells treated with IM for 18h (IM18h). (**Upper panel**) Significant changes (FDR < 0.05) in alternative splicing (AS), presented events detected by at least 20 reads. MXE – mutually exclusive events, CE – cassette exon, RI – intron retention. Number of significant events with absolute inclusion level difference to control cells > 0.1. (**Lower panel**) Upset plots comparing events induced by each condition. **(B)** Enriched Gene Ontology Biological Process terms (blue) for genes that are differentially spliced in both IM-resistant and IM18h treated cells (MXE and CE from A, Lower panel). **(C)** Comparison of genes that are upregulated upon erythroid differentiation (Erythroid diff. UP) and have RTI values ≥ 0.2 (RTImore02) or have significant AS changes upon IM 18h treatment (AS) with genes that have significant changes in CE or MXE for both IM-resistant cells and IM18h (Imatinib AS). **(D)** Comparison of significant AS changes for both IM-resistant and IM18h K562 cells (Imatinib AS) and changes for CD34+-enriched cells from a CML patient at diagnosis and after six months of IM treatment. (**E**) (**Upper panel**) Architecture of eukaryotic translation initiation factor 4A2 (*EIF4A2*). SnoRNAs indicated by black boxes with stars. (**Lower panel**) Downsampled nascent RNA long reads colored according to splicing status, with shaded box over poison exon. RTI values (mean ± S.E.M) for EIF4A2 from three independent experiments are shown. **(F)** Quantification of isoform usage for *EIF4A2*. The NMD and expressed isoforms represent inclusion and skipping, respectively, of the poison exon. Percent transcripts belonging to each isoform in 3 replicates (mean ± S.E.M.). Significance based on two-tailed Student's t-test (*, p<0.05).

Figure 6. Imatinib-induced formation of readthrough chimeras. (A) Examples of reads aligned to RBM14 and RBM4 from long- and short-read sequencing of IM-treated cells. Chimeric reads marked in blue. **(B)** Sashimi plot for RBM14 and RBM4 (chromosome 11: 66616624- 66644972) based on a single representative CTRL replicate. Only splicing events detected more than 5 times are shown. Black represents read coverage density in RPKM, gray lines represent splicing events within gene bodies, and the red line represents chimeric splicing. **(C)** Scheme of primer design for real-time PCR semi-quantification (qPCR) of chimera expression. **(D)** Expression of selected chimeras (red primers in C) in CD34+-enriched cells from three healthy donors (HD) and two CML patients (CML), normalized to spike-in controls and presented relative to the expression level of the last exon in the upstream gene (beige primers in C). **(E)** Comparison of readthrough chimeras identified by SOAPfuse analysis of mRNA sequencing for CTRL or IM-treated K562 cells. **(F)** Changes in chimera transcript expression normalized to spike-in control referenced to the last exon in the upstream gene and expressed as fold change upon IM treatment. Level in CTRL as 1.0, marked with the dashed line. Significance based on two-way Anova test (*, $p<0.05$; **, $p<0.005$; ***, $p<0.001$).

Figure 7. Working model: a cascade of molecular mechanisms triggered by imatinib treatment of CML cells. Transcriptional readthrough and intron retention are the first major changes detected at 1h. At 18 h, increasing intron retention, alternative splicing, and readthrough chimera formation may be associated with induced cell fates, such as erythropoiesis, and the development of resistance to therapy.

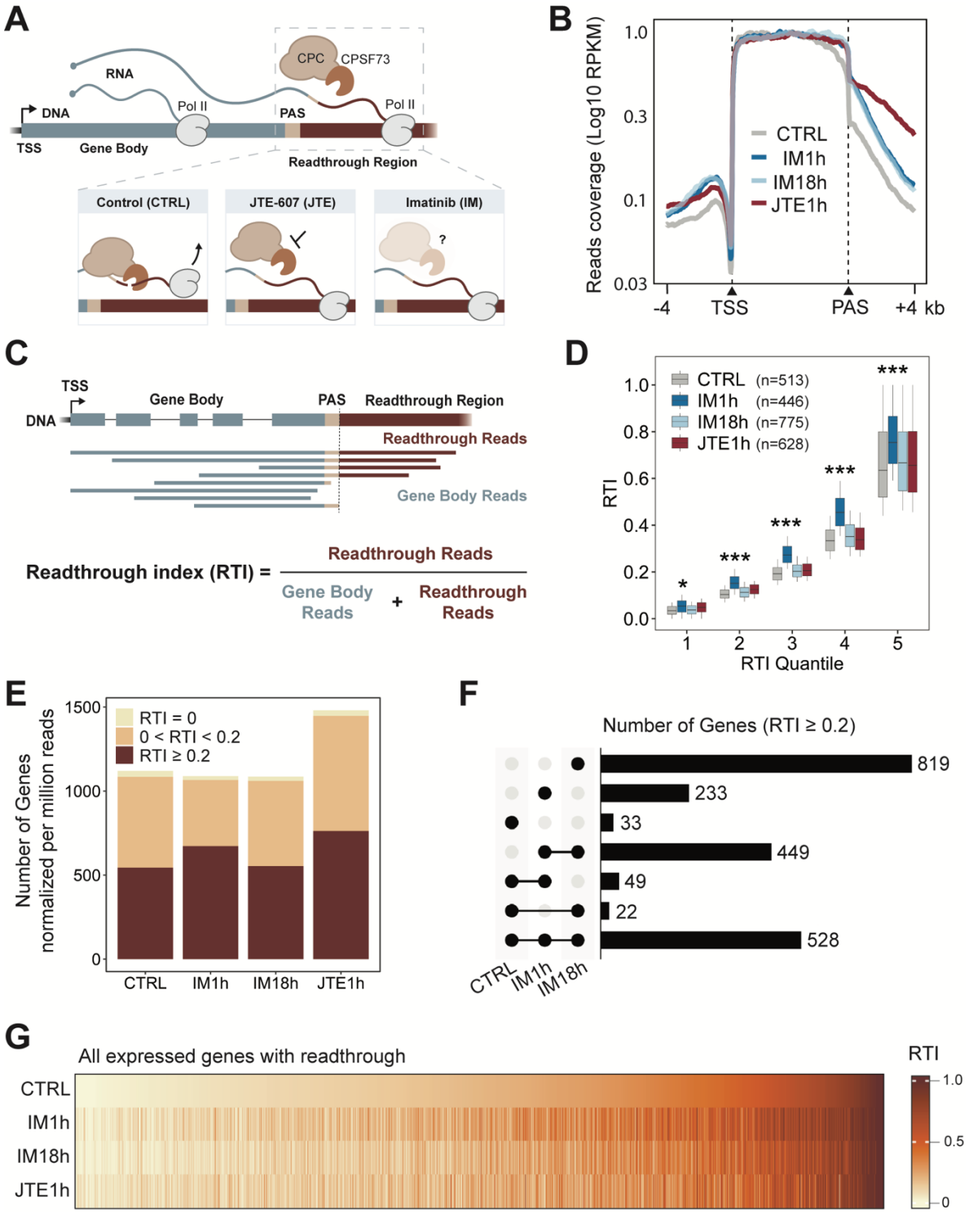
A
RNA
CPC
CPSF73
DNA
Pol II
PAS
Pol II
TSS
Gene Body
Readthrough Region
Control (CTRL)
JTE-607 (JTE)
Imatinib (IM)
B
Reads coverage (Log10 RPKM)
1.0
0.3
0.1
0.03
CTRL
IM1h
IM18h
JTE1h
-4
TSS
PAS
+4 kb
C
TSS
DNA
Gene Body
PAS
Readthrough Region
Readthrough Reads
Gene Body Reads
Readthrough index (RTI) = Readthrough Reads / (Gene Body Reads + Readthrough Reads)
D
CTRL (n=513)
IM1h (n=446)
IM18h (n=775)
JTE1h (n=628)
RTI
RTI Quantile
E
Number of Genes normalized per million reads
RTI = 0
0 < RTI < 0.2
RTI ≥ 0.2
CTRL
IM1h
IM18h
JTE1h
F
Number of Genes (RTI ≥ 0.2)
819
233
33
449
49
22
528
CTRL
IM1h
IM18h
G
All expressed genes with readthrough
CTRL
IM1h
IM18h
JTE1h
RTI
1.0
0.5
0

Figure 1.

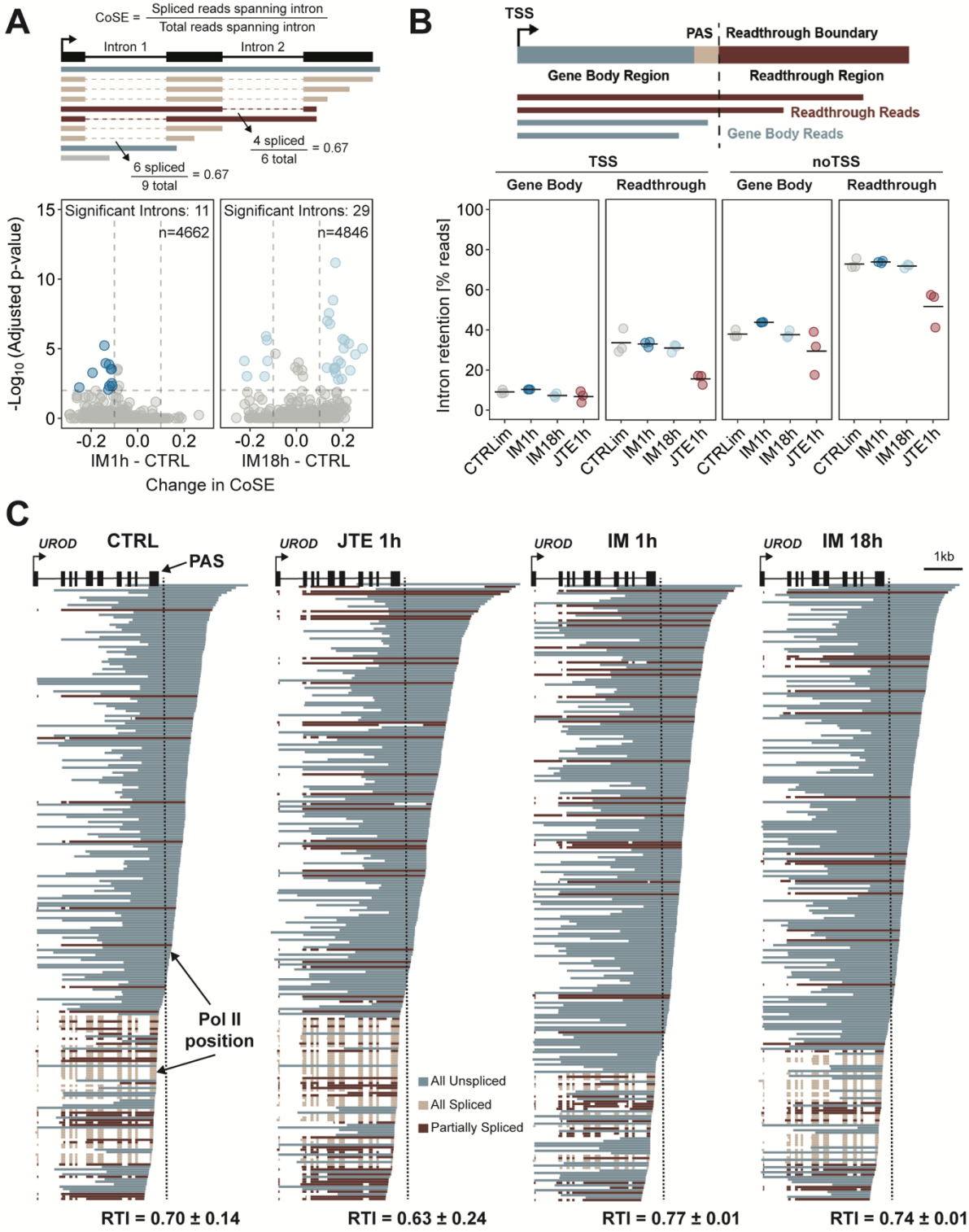
A
CoSE = Spliced reads spanning intron / Total reads spanning intron
Intron 1
Intron 2
4 spliced / 6 total = 0.67
6 spliced / 9 total = 0.67
Significant Introns: 11
n=4662
Significant Introns: 29
n=4846
-Log10 (Adjusted p-value)
IM1h - CTRL
IM18h - CTRL
Change in CoSE
B
TSS
PAS
Readthrough Boundary
Gene Body Region
Readthrough Region
Readthrough Reads
Gene Body Reads
TSS
noTSS
Gene Body
Readthrough
Intron retention [% reads]
CTRLim
IM1h
IM18h
JTE1h
C
UROD
CTRL
PAS
JTE 1h
IM 1h
IM 18h
1kb
Pol II position
All Unspliced
All Spliced
Partially Spliced
RTI = 0.70 ± 0.14
RTI = 0.63 ± 0.24
RTI = 0.77 ± 0.01
RTI = 0.74 ± 0.01

Figure 2.

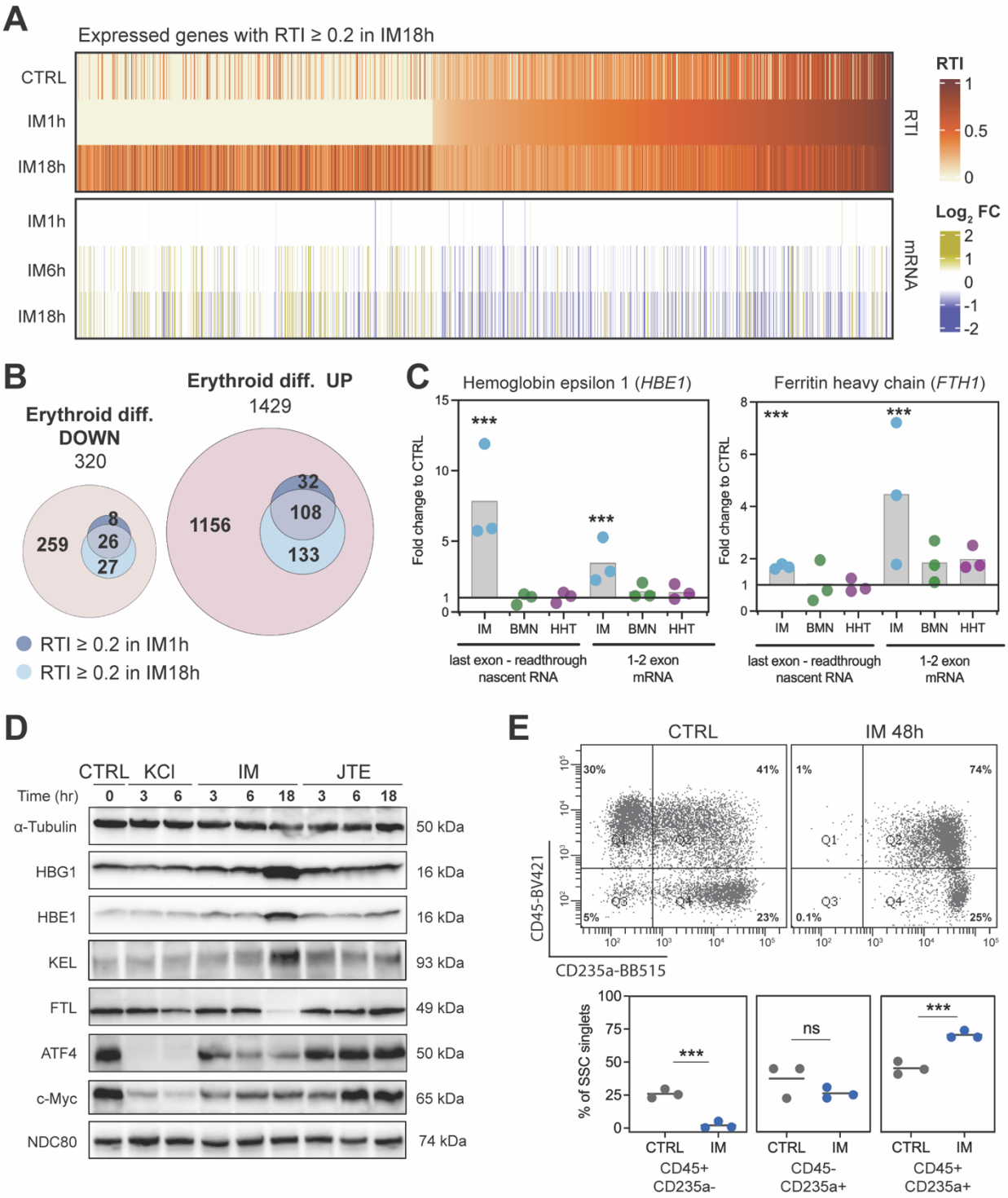
A
Expressed genes with RTI ≥ 0.2 in IM18h
CTRL
IM1h
IM18h
RTI
IM1h
IM6h
IM18h
mRNA
RTI
1
0.5
0
Log2 FC
2
1
0
-1
-2
B
Erythroid diff. DOWN
320
Erythroid diff. UP
1429
259
8
26
27
1156
32
108
133
RTI ≥ 0.2 in IM1h
RTI ≥ 0.2 in IM18h
C
Hemoglobin epsilon 1 (HBE1)
Ferritin heavy chain (FTH1)
Fold change to CTRL
***
IM
BMN
HHT
last exon - readthrough nascent RNA
1-2 exon mRNA
D
CTRL
KCl
IM
JTE
Time (hr)
0
3
6
18
α-Tubulin
HBG1
HBE1
KEL
FTL
ATF4
c-Myc
NDC80
50 kDa
16 kDa
16 kDa
93 kDa
49 kDa
50 kDa
65 kDa
74 kDa
E
CTRL
IM 48h
30%
41%
5%
23%
1%
74%
0.1%
25%
Q1
Q2
Q3
Q4
CD45-BV421
CD235a-BB515
% of SSC singlets
ns
CTRL
IM
CD45+ CD235a-
CD45- CD235a+
CD45+ CD235a+

Figure 3.

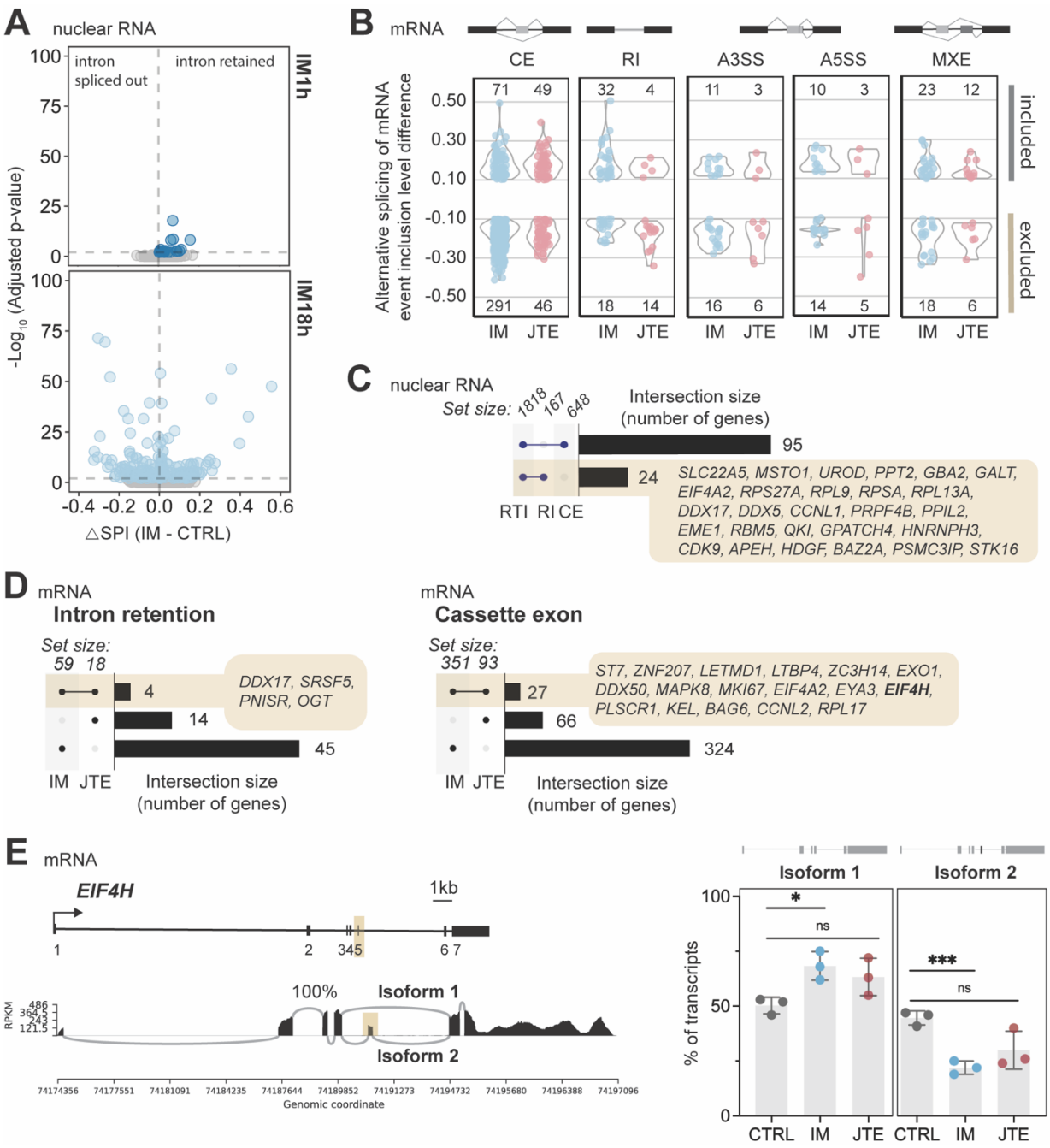
A
nuclear RNA
intron spliced out
intron retained
IM1h
IM18h
-Log10 (Adjusted p-value)
△SPI (IM - CTRL)
B
mRNA
CE
RI
A3SS
A5SS
MXE
Alternative splicing of mRNA event inclusion level difference
included
excluded
IM
JTE
C
nuclear RNA
Set size: 1818 167 648
Intersection size (number of genes)
95
24
RTI RI CE
SLC22A5, MSTO1, UROD, PPT2, GBA2, GALT, EIF4A2, RPS27A, RPL9, RPSA, RPL13A, DDX17, DDX5, CCNL1, PRPF4B, PPIL2, EME1, RBM5, QKI, GPATCH4, HNRNPH3, CDK9, APEH, HDGF, BAZ2A, PSMC3IP, STK16
D
mRNA
Intron retention
Set size: 59 18
4
14
45
DDX17, SRSF5, PNISR, OGT
IM JTE
Intersection size (number of genes)
mRNA
Cassette exon
Set size: 351 93
27
66
324
ST7, ZNF207, LETMD1, LTBP4, ZC3H14, EXO1, DDX50, MAPK8, MKI67, EIF4A2, EYA3, EIF4H, PLSCR1, KEL, BAG6, CCNL2, RPL17
IM JTE
Intersection size (number of genes)
E
mRNA
EIF4H
1kb
100%
Isoform 1
Isoform 2
RPKM
Genomic coordinate
Isoform 1
Isoform 2
% of transcripts
ns
***
CTRL IM JTE CTRL IM JTE

Figure 4.

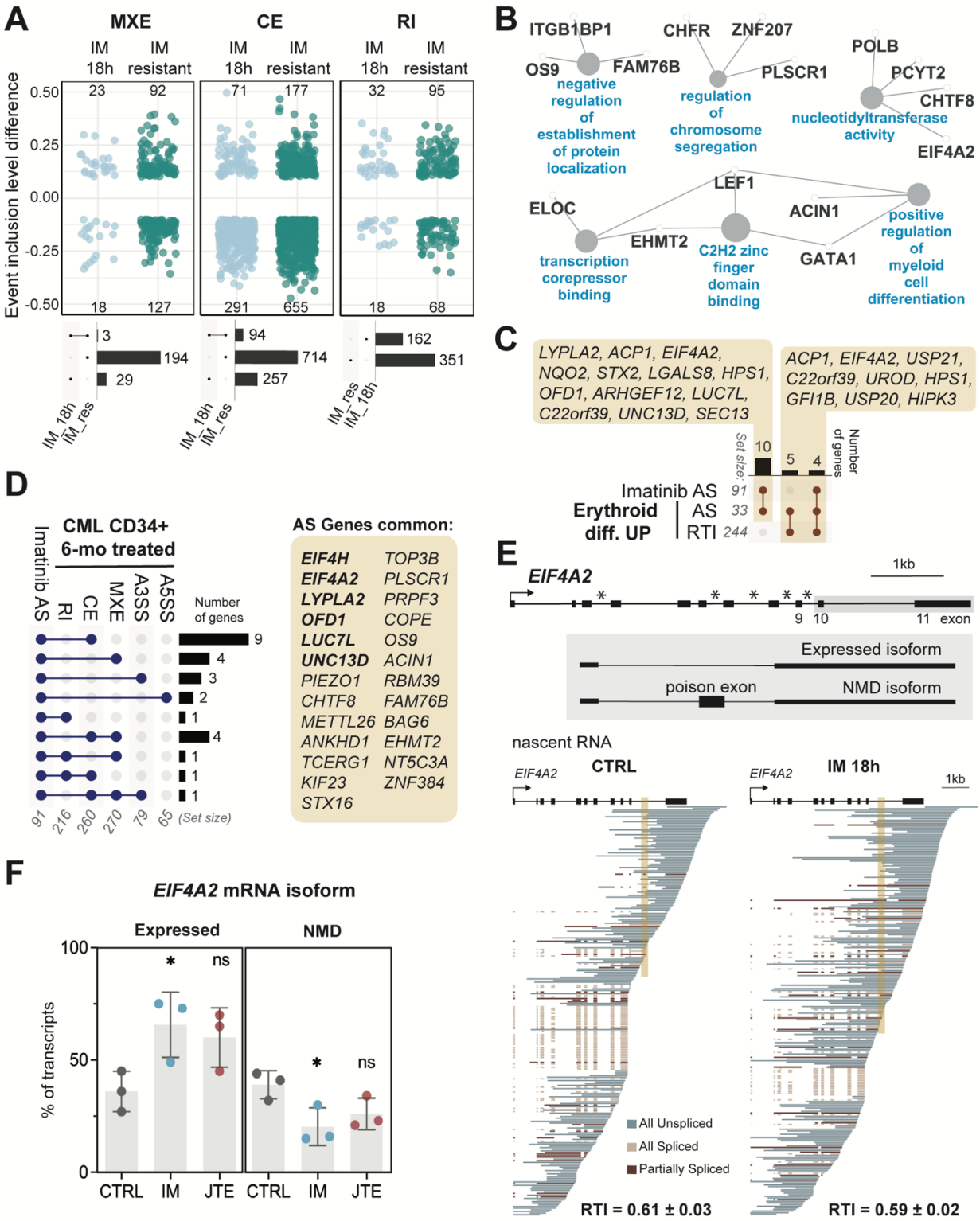
A
MXE
CE
RI
IM 18h
IM resistant
Event inclusion level difference
0.50
0.25
0.00
-0.25
-0.50
23
92
71
177
32
95
18
127
291
655
18
68
3
194
29
94
714
257
162
351
IM_18h
IM_res
B
ITGB1BP1
OS9
FAM76B
negative regulation of establishment of protein localization
CHFR
ZNF207
PLSCR1
regulation of chromosome segregation
POLB
PCYT2
CHTF8
EIF4A2
nucleotidyltransferase activity
LEF1
ELOC
EHMT2
transcription corepressor binding
C2H2 zinc finger domain binding
ACIN1
GATA1
positive regulation of myeloid cell differentiation
C
LYPLA2, ACP1, EIF4A2, NQO2, STX2, LGALS8, HPS1, OFD1, ARHGEF12, LUC7L, C22orf39, UNC13D, SEC13
ACP1, EIF4A2, USP21, C22orf39, UROD, HPS1, GFI1B, USP20, HIPK3
Set size
Number of genes
10
5
4
Imatinib AS
Erythroid diff. UP
AS
RTI
91
33
244
D
CML CD34+ 6-mo treated
Imatinib AS
RI
CE
MXE
A3SS
A5SS
Number of genes
9
4
3
2
1
4
1
1
1
91
216
260
270
79
65
(Set size)
AS Genes common:
EIF4H
EIF4A2
LYPLA2
OFD1
LUC7L
UNC13D
PIEZO1
CHTF8
METTL26
ANKHD1
TCERG1
KIF23
STX16
TOP3B
PLSCR1
PRPF3
COPE
OS9
ACIN1
RBM39
FAM76B
BAG6
EHMT2
NT5C3A
ZNF384
E
EIF4A2
1kb
9
10
11
exon
Expressed isoform
poison exon
NMD isoform
nascent RNA
EIF4A2
CTRL
IM 18h
1kb
All Unspliced
All Spliced
Partially Spliced
RTI = 0.61 ± 0.03
RTI = 0.59 ± 0.02
F
EIF4A2 mRNA isoform
Expressed
NMD
% of transcripts
100
50
0
*
ns
*
ns
CTRL
IM
JTE
CTRL
IM
JTE

Figure 5.

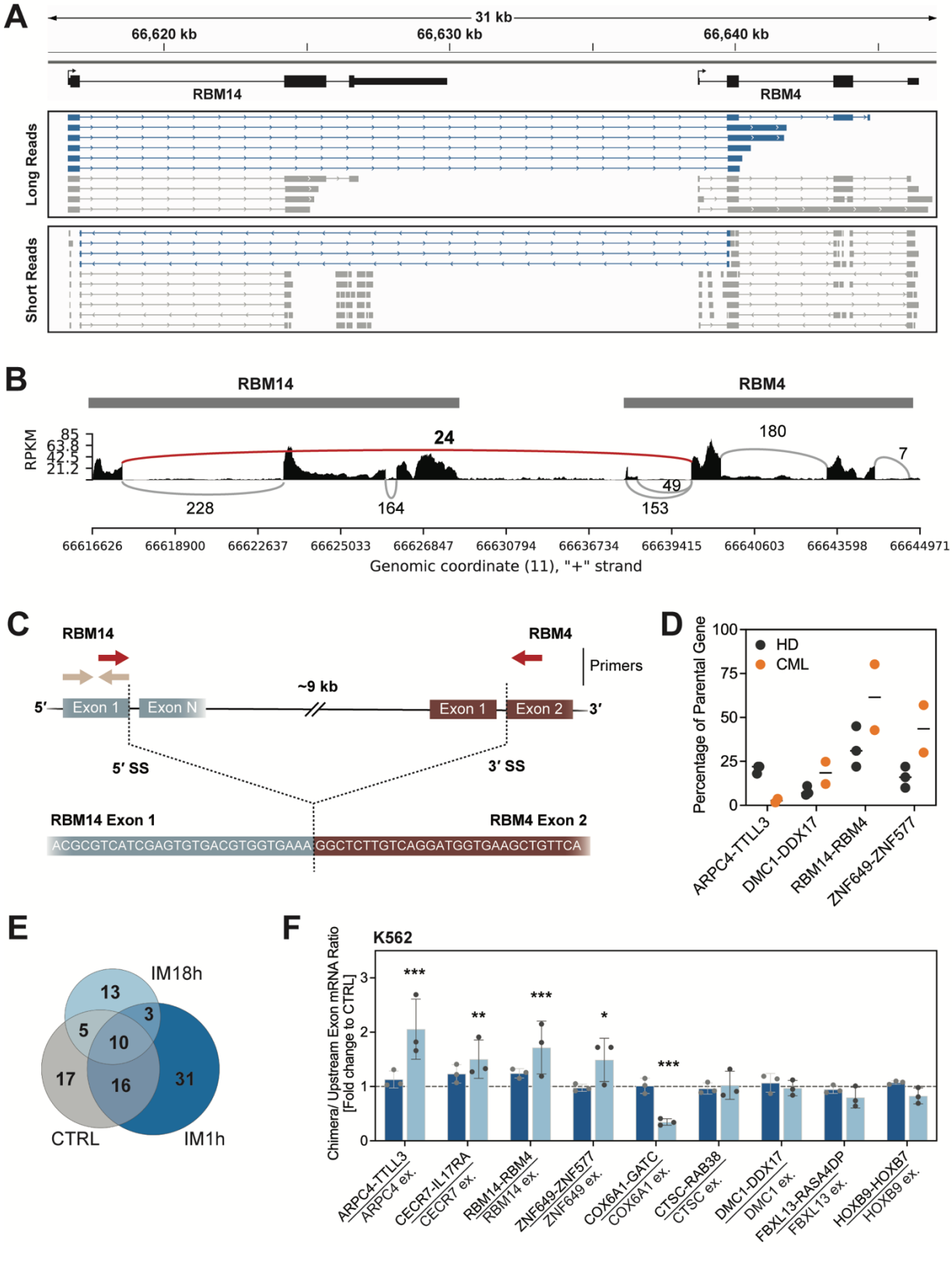
A
31 kb
66,620 kb
66,630 kb
66,640 kb
RBM14
RBM4
Long Reads
Short Reads
B
RBM14
RBM4
RPKM
85
63.8
42.5
21.2
24
180
7
228
164
49
153
66616626
66618900
66622637
66625033
66626847
66630794
66636734
66639415
66640603
66643598
66644971
Genomic coordinate (11), "+" strand
C
RBM14
RBM4
Primers
~9 kb
5′
Exon 1
Exon N
Exon 1
Exon 2
3′
5′ SS
3′ SS
RBM14 Exon 1
RBM4 Exon 2
ACGCGTCATCGAGTGTGACGTGGTGAAAGGCTCTTGTCAGGATGGTGAAGCTGTTCA
D
Percentage of Parental Gene
100
75
50
25
0
HD
CML
ARPC4-TTLL3
DMC1-DDX17
RBM14-RBM4
ZNF649-ZNF577
E
IM18h
13
3
5
10
17
16
31
CTRL
IM1h
F
K562
Chimera/ Upstream Exon mRNA Ratio
[Fold change to CTRL]
3
2
1
0
***
**
***
*
***
ARPC4-TTLL3
ARPC4 ex.
CECR7-IL17RA
CECR7 ex.
RBM14-RBM4
RBM14 ex.
ZNF649-ZNF577
ZNF649 ex.
COX6A1-GATC
COX6A1 ex.
CTSC-RAB38
CTSC ex.
DMC1-DDX17
DMC1 ex.
FBXL13-RASA4DP
FBXL13 ex.
HOXB9-HOXB7
HOXB9 ex.

Figure 6.

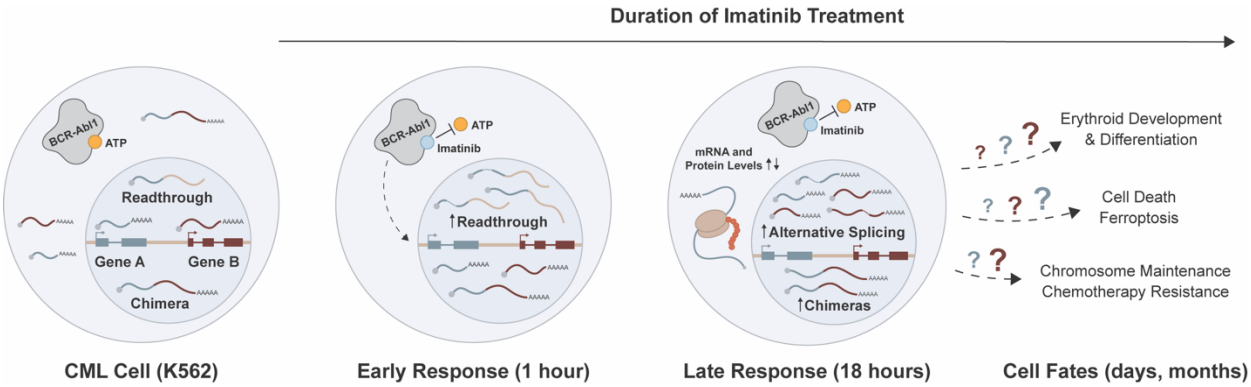
Duration of Imatinib Treatment
BCR-Abl1
ATP
Readthrough
Gene A
Gene B
Chimera
CML Cell (K562)
BCR-Abl1
ATP
Imatinib
↑Readthrough
Early Response (1 hour)
BCR-Abl1
ATP
Imatinib
mRNA and
Protein Levels ↑↓
↑Alternative Splicing
↑Chimeras
Late Response (18 hours)
Erythroid Development
& Differentiation
Cell Death
Ferroptosis
Chromosome Maintenance
Chemotherapy Resistance
Cell Fates (days, months)

Figure 7.