

1 **Widespread aberrant alternative splicing despite molecular** 2 **remission in chronic myeloid leukemia patients**

3 Ulf Schmitz¹⁻³, Jaynish S. Shah^{2,3}, Bijay P. Dhungel^{2,3}, Geoffray Monteuiis², Phuc-Loi Luu⁴,
4 Veronika Petrova¹, Cynthia Metierre², Shalima S. Nair^{5,6}, Charles G. Bailey^{2,3}, Verity A.
5 Saunders⁷, Ali G. Turhan⁸, Deborah L. White^{7,9-12}, Susan Branford^{9,11,13,14}, Susan Clark^{4,16},
6 Timothy P. Hughes^{9,10,14,15}, Justin J-L. Wong^{3,17}, John E.J. Rasko^{2,3,18}

7 ¹Computational BioMedicine Laboratory Centenary Institute, The University of Sydney,
8 Australia; ²Gene & Stem Cell Therapy Program Centenary Institute, The University of
9 Sydney, Australia; ³Faculty of Medicine & Health, The University of Sydney, Australia;
10 ⁴Epigenetics Research Laboratory, Genomics and Epigenetics Division, Garvan Institute of
11 Medical Research, Sydney, Australia; ⁵Kinghorn Centre for Clinical Genomics Core Facility,
12 Garvan Institute of Medical Research, Sydney, Australia; ⁶St Vincent's Clinical School,
13 Faculty of Medicine, UNSW, Darlinghurst, Australia; ⁷Cancer Program, Precision Medicine
14 Theme, South Australian Health & Medical Research Institute, Adelaide, SA, Australia;
15 ⁸APHP, Division of Hematology, Paris Sud University Hospitals and Inserm U935
16 INGESTEM Pluripotent Stem Cell Infrastructure 78 Rue du Général Leclerc, 94275 Le
17 Kremlin Bicetre, France; ⁹School of Medicine, Faculty of Health and Medical Sciences,
18 University of Adelaide, Adelaide, SA, Australia; ¹⁰Australasian Leukaemia and Lymphoma
19 Group, Melbourne, VIC, Australia; ¹¹School of Biological Sciences, Faculty of Sciences,
20 University of Adelaide, Adelaide, Australia; ¹²South Australian Health and Medical Research
21 Institute, Adelaide, Australia; ¹³Department of Genetics and Molecular Pathology, Centre for
22 Cancer Biology, SA Pathology, Adelaide, Australia; ¹⁴School of Pharmacy and Medical
23 Science, Division of Health Sciences, University of South Australia, Adelaide, Australia; ¹⁵St
24 Vincent's Clinical School, UNSW, Sydney, Australia; ¹⁶Department of Haematology, Centre
25 for Cancer Biology, SA Pathology, Adelaide, Australia; ¹⁷Epigenetics and RNA Biology
26 Program Centenary Institute, The University of Sydney, Australia; ¹⁸Cell and Molecular
27 Therapies, Royal Prince Alfred Hospital, Camperdown, Australia

28

29 **Correspondence:** John E.J. Rasko, Locked Bag No. 6, Newtown 2042, NSW

30 Australia; e-mail: j.rasko@centenary.org.au; Tel: 02 95656160

31

32

33

34 **Abstract**

35 **Background:** Vast transcriptomics and epigenomics changes are characteristic of
36 human cancers including leukemia. At remission, we assume that these changes
37 normalise so that omics-profiles resemble those of healthy individuals. However, an
38 in-depth transcriptomic and epigenomic analysis of cancer remission has not been
39 undertaken. A striking exemplar of targeted remission induction occurs in chronic
40 myeloid leukemia (CML) following tyrosine kinase inhibitor (TKI) therapy.

41 **Methods:** Using RNA sequencing and whole-genome bisulfite sequencing, we
42 profiled samples from chronic-phase CML patients at diagnosis and remission, and
43 compared these to healthy donors.

44 **Results:** Remarkably, our analyses revealed that abnormal splicing distinguishes
45 remission samples from normal controls. This phenomenon is independent of the TKI
46 drug used and in striking contrast to the normalisation of gene expression and DNA
47 methylation patterns. Most remarkable are the high intron retention (IR) levels that
48 even exceed those observed in the diagnosis samples. Increased IR affects cell cycle
49 regulators at diagnosis and splicing regulators at remission. We show that aberrant
50 splicing in CML is associated with reduced expression of specific splicing factors,
51 histone modifications and reduced DNA methylation.

52 **Conclusions:** Our results provide novel insights into the changing transcriptomic and
53 epigenomic landscapes of CML patients during remission. The conceptually
54 unanticipated observation of widespread aberrant alternative splicing after remission
55 induction warrants further exploration. These results have broad implications for
56 studying CML relapse and treating minimal residual disease.

57

58 **Keywords:** transcriptomic complexity, alternative splicing, intron retention, DNA
59 methylation, epigenetics, BCR-ABL1, histone modifications, CML, cancer

60

61 **Background**

62 Characterisation of *BCR-ABL1* and its constitutive tyrosine kinase activity has
63 facilitated understanding of the pathogenesis and therapy of CML¹. Tyrosine kinase
64 inhibitor (TKI) treatment results in positive responses in 60-90% of patients with
65 chronic phase CML^{2,3}. However, molecular mechanisms that facilitate CML
66 maintenance and relapse remain elusive⁴. While the molecular causes of CML are
67 well understood, measures of therapy success based on *BCR-ABL1* expression and
68 white blood cell count are not reliable predictors of relapse.

69 *BCR-ABL1* overexpression perturbs the hematopoietic transcriptome^{4,5}. CML
70 progression, maintenance, and therapy may be affected by aberrant expression and
71 alternative splicing of cancer-causing genes^{6,7}. Alternative splicing is vital for cellular
72 homeostasis and dysregulated intron retention (IR) has been associated with
73 numerous human cancers including leukemia⁸. However, the role of alternative
74 splicing in CML remains unknown.

75 In addition to risk-associated genomic variants which predict poor outcome⁹, there is
76 evidence that epigenetic regulation affects CML pathology and therapy success¹⁰.

77 Epigenetic regulation via DNA methylation and nucleosome occupancy play key
78 roles in constitutive and alternative mRNA splicing regulation¹¹⁻¹³. Epigenetic
79 regulation via DNA methylation is known to regulate constitutive and alternative
80 mRNA splicing. To date, no systematic transcriptomics and epigenomics analyses in
81 CML remission samples have been conducted, neither have the effects of different
82 therapeutic regimens on gene expression and alternative splicing been assessed. CML

83 presents an excellent model to study epigenetically mediated transcriptomic
 84 alterations in myeloid cells that cause malignant transformation and the response to
 85 TKI treatment. New insights into this multilayered regulatory network could provide
 86 important clues for targeted therapies used in other malignancies as well.
 87 Here we report novel transcriptional and epigenetic patterns in a multi-omics analysis
 88 of CML patient samples before and after effective front-line TKI treatment. Using this
 89 model, we found abnormal splicing in chronic-phase CML patients, which we
 90 confirmed in a larger independent cohort. Surprisingly, we also observed aberrant
 91 splicing patterns in complete molecular remission. This is in contrast to the
 92 normalisation of gene expression and DNA methylation patterns driven by the
 93 reconfiguration of blood cell composition. More specifically, we observed a marked
 94 increase in IR and differential exon usage in CML diagnosis and remission compared
 95 to normal samples, as well as evidence that epigenetic factors modulate splicing in
 96 CML.

97 **Methods**

98 *Patient samples*

99 We retrieved nine diagnostic specimens (total leukocytes from peripheral blood
 100 collected in EDTA) from patients enrolled in the RESIST (1x imatinib;
 101 ACTRN12610000055000), ENESTxtnd (4x nilotinib, NCT01254188) and
 102 PINNACLE (4x nilotinib + pegylated interferon, ACTRN12612000851864) trials,
 103 with matched remission specimens (major or complete molecular response at 12 or 24
 104 months). In addition, eight CML patient samples (7 diagnostic and 1 matched
 105 remission) from the French Persistem study were obtained. For the validation of IR
 106 events and fusion transcripts we retrieved further matched diagnosis/remission
 107 samples from the TIDELII trial (20x imatinib, ACTRN12607000325404). RNA was

108 isolated from mononuclear cells (from peripheral blood collected in LiHep) which
109 had been cryopreserved and subsequently thawed into TRIzol. All studies were
110 approved by the institutional review boards (HREC protocol No: 131015, 070718c,
111 081211, and 101010 - Royal Adelaide Hospital) and are in accordance with the
112 Declaration of Helsinki. Patients provided written informed consent.

113

114 *RNA isolation and mRNA-seq*

115 Total RNA was isolated from diagnosis, remission, and control samples (peripheral
116 blood mononuclear cells - PBMCs) using Trizol according to the manufacturer's
117 protocol. For mRNA sequencing, poly-A enriched mRNA libraries were prepared
118 from 1 µg of total RNA using the TruSeq Stranded RNA sample prep kit (Illumina),
119 prior to paired-end sequencing using the HiSeq 2500 platform. RNA sequencing was
120 performed in triplicates for each sample. The raw sequencing data (fastq files) have
121 been deposited at GEO under accession GSE144119.

122

123 *mRNA sequencing data analysis*

124 Paired-end RNA-sequencing reads (125 nt) were trimmed and mapped to the human
125 reference genome hg38 using STAR¹⁴. Quality control of raw and mapped sequencing
126 reads was performed using FASTQC (github.com/s-andrews/FastQC), RSeQC¹⁵, and
127 multiqc¹⁶. Putative confounders and batch effects were excluded using principal
128 components analyses (Supplementary Figure 1). STAR-FUSION¹⁷ was used for the
129 identification of fusion genes and Fusion Inspector ([FusionInspector.github.io](https://github.com/FusionInspector)) for *in*
130 *silico* validation of the predicted gene fusions. Gene expression levels specified as
131 transcripts per million (TPM) were determined using Salmon 0.14.1¹⁸.

General statistics on alternative splicing events were determined using the rMATS software¹⁹. We used our IRFinder algorithm for the detection of IR events in introns extracted from an Ensembl gtf file (regions between two adjacent exons)²⁰. IRFinder estimates the abundance of IR by computing the ratio between gene transcripts retaining an intron and the sum of all transcripts of the respective gene (more information is provided in the Supplementary Materials). Differentially used exons were determined using the R Bioconductor package DEXSeq²¹. Differential exon usage (DEU) = changes in the relative usage of exons:

$$transcripts_{with_exon} / (transcripts_{with_exon} + transcripts_{without_exon})$$

IR validation

Extracted RNA (2 µg) was treated with DNase Turbo (Invitrogen) and reverse transcribed into cDNA using oligo(dT) priming and Superscript III (ThermoFisher) according to the protocol supplied by the manufacturer. For each sample, a corresponding control without reverse transcriptase was used to assess DNA contamination. For qPCR, cDNA templates were amplified, and the C_t values were quantified with SYBR Green Master Mix (ThermoFisher). Beta-2 microglobulin (B2M) was used as the normalization control for cDNA input. Experiments were performed using the CFX96 Real-Time PCR System (BIORAD). The list of primers is provided in Supplementary Table 1.

Whole genome bisulfite sequencing (WGBS)

WGBS libraries were prepared following Illumina's "Whole- Genome Bisulfite sequencing for Methylation Analysis" protocol. Briefly, 1 µg of genomic DNA was spiked with 0.5% unmethylated lambda DNA and sonicated to generate fragments of

size between 150 to 300 bp. Library preparation was performed using Illumina's Paired-end DNA Sample Prep Kit (discontinued, Illumina, CA, USA) according to the manufacturer's protocol. The size-selected libraries were then subjected to bisulfite conversion as previously described²². Adaptor-ligated bisulfite treated DNA was enriched by 10 cycles of PCR amplification using the PfuTurbo Cx Hotstart DNA Polymerase (Stratagene). Qualitative and quantitative checks of the libraries were performed using Agilent's High sensitivity DNA kit (Agilent) and KAPA Library quantification kit (KAPA Biosystems). Three lanes of paired-end 100 bp sequencing was performed for each of the library on the Illumina HiSeq2500 platform using the TruSeq v3 cluster kits and SBS kits to achieve coverage ranging between 25-30x.

166

167 *WGBS data analysis*

Reads were processed and aligned to the human (hg38) reference genome using Meth10X (github.com/luuloi/Meth10X)²³. In short, the Meth10X pipeline takes raw reads in fastq format and trims the adaptors, which are then aligned to the human reference genome using bwa-meth (github.com/brentp/bwa-meth). The generated bam files were marked with duplication and merged if necessary. Estimation of the duplication rate, coverage bias (genomic features) and methylation bias in reads was carried out to provide quality control. All the metrics such as percentage of unmapped/mapped read metrics, mapping quality distribution, GC content distribution, insert size distribution and coverage distribution were generated by Qualimap 2 for further evaluation of the alignment. Finally, a count table and bigwig files of methylated and coverage at each CpG site in the genome was constructed for all samples. Differentially methylated regions (DMRs) were identified using

180 MethPipe²⁴ with count table of all samples as input. The analysis of differential
181 methylation and IR is described in the Supplementary Materials.

182

183 *ChIP-seq data analysis*

184 ChIP-seq data of histone modifications in K562 cells was retrieved from ENCODE
185 (encodeproject.org) and further processed as described in the Supplementary
186 Materials.

187

188 *Statistical analyses*

189 All statistical analyses were performed in R v.3.6.2. The Wald test implemented in the
190 R package DESeq2²⁵ was used for the identification of statistically significant
191 differential alternative splicing events, and differential gene expression analysis. P-
192 values were adjusted for the false discovery rate using the Benjamini-Hochberg
193 procedure. Functional enrichment analysis of differentially expressed genes and genes
194 with differentially retained introns was performed using DAVID 6.7²⁶ with the human
195 genome and the set of expressed genes, respectively, used as a background.

196

197 **Results**

198 The goal of our work was to evaluate transcriptomic and epigenetic changes in CML
199 patients after major or complete molecular response. We retrieved sixteen peripheral
200 blood samples of Philadelphia-positive (Ph+) CML patients. For 10 of these patients,
201 we also retrieved matched samples after remission. In addition, we retrieved
202 peripheral blood from six healthy donors (Table 1). We subjected all samples to RNA
203 sequencing and all matched diagnosis/remission samples as well as control samples to
204 WGBS.

205

206 *Heterogeneous CML transcriptomes converge at remission*

207 RNA sequencing reads mapping to the *BCR-ABL1* locus were found in diagnosis
 208 samples only (Supplementary Figure 2A), confirming the results of clinical qRT-
 209 PCR-based *BCR-ABL1* quantification (Supplementary Data 1). Apart from the
 210 characteristic *BCR-ABL1* fusion we identified other recurring intra- and inter-
 211 chromosomal fusions (Supplementary Figure 2A&B; Supplementary Data 2) that
 212 occur in diagnosis, remission or control samples. A novel fusion transcript between
 213 the myeloid cell-specific transmembrane glycoprotein *CLEC12A* (12p13.31) and the
 214 microRNA *miR-223* host gene *MIR223HG* (Xq12) was expressed at higher levels
 215 ($FC = 7.16$) at diagnosis compared to remission (Supplementary Figure 2C&D).
 216 For most patients we observed a global reduction of gene expression in remission
 217 compared to diagnosis (Supplementary Figure 3A). This can be attributed to the
 218 reconfigured blood cell composition, which resembles lymphoid cell-enriched normal
 219 controls (Figure 1A&B). A significant downregulation was observed in genes
 220 encoding kinases other than *BCR-ABL1* and kinase-like proteins such as *CENPE*,
 221 *CDK1*, *AURKB*, *MELK*, and *BUB1B* (Supplementary Figure 4). Reduced expression
 222 at remission, similar to levels observed in control samples, also affected cyclins such
 223 as *CCNA1*, *CCNB2* and other cell cycle regulators as well as genes related to DNA
 224 replication and repair (Figure 1A & Supplementary Figure 5A). In this context, no
 225 consistent differences were noted between patients treated with imatinib or nilotinib
 226 (Supplementary Figure 1).
 227 A set of genes (n=81), that are functionally associated with oxygen and bicarbonate
 228 transport (GO:0015671; GO:0015701) as well as the hemoglobin's chaperone
 229 pathway (Biocarta), was significantly differentially expressed in all three comparisons

(diagnosis vs control, remission vs control, and diagnosis vs remission; Supplementary Figure 5B). Overall, we observed that patients diagnosed with CML have highly heterogeneous transcriptomes. In contrast, remission transcriptomes are more alike (Figure 1C) and exhibit fewer differences in comparison to healthy control samples (Supplementary Figure 3B).

DNA methylation profiles return to normal at remission

To characterise putative epigenetic causes for differential gene expression in CML patients prior to and after successful TKI treatment, all matched diagnosis/remission samples as well as all control samples were subjected to WGBS. The analysis revealed almost 25,000 differentially methylated regions (DMRs) in the genomes of diagnosis and remission samples (Figure 2A, left). Vast DNA methylome differences were found in every matched patient sample (Supplementary Figure 6) and across all chromosomes (Figure 2B). In contrast, only 710 DMRs were observed between remission and control samples (Figure 2A, right). Indeed, multidimensional scaling showed that DNA methylation profiles of remission samples are closer to those from healthy donors rather than to matched diagnosis samples (Supplementary Figure 7). DMRs between diagnosis and remission samples contain, on average, a larger number of differentially methylated CpG sites (Figure 2C). This is consistent with previous reports that suggest an increase in CpG site methylation during CML progression that affects tumour-suppressor genes and regulators of cell proliferation¹⁰. DMRs are associated with processes of the innate immune response, apoptosis, cell differentiation, and cell migration (Figure 2D).

Our results show that DNA methylation returns to normal levels after successful TKI treatment. Most DMRs can be found in intronic regions, however, the highest density

of DMRs was observed in 5' untranslated regions (Figure 2E), suggesting a key role for differential DNA methylation in orchestrating CML gene expression regulation.

256

Aberrant alternative splicing distinguishes remission samples from healthy controls

To assess whether major or complete molecular remission also affects gene isoform expression, we conducted a systematic analysis of the five major forms of alternative splicing (Supplementary Figure 8A). In a pairwise comparison of alternative splicing event frequencies between matched diagnosis/remission samples, no consistent trend towards an increase or decrease in either of the two conditions was found (Supplementary Figure 8B). This suggests, that in contrast to gene expression profiles, splicing patterns remain atypical after major or complete molecular remission.

Alternative splicing analysis of the Cancer Genome Atlas (TCGA) cohort by Dvinge et al. (2015) revealed that the frequency of IR events is consistently higher in acute myeloid leukemia compared to normal controls⁸. Although mean IR frequencies are also higher in CML samples (diagnosis, $\bar{x} = 3\,382$; remission, $\bar{x} = 3\,484$) compared to healthy donors ($\bar{x} = 2\,822$), high inter-donor variability led to differences in IR frequencies that did not reach significance (Supplementary Figure 8C). However, when we subsampled the RNA sequencing data to facilitate a uniform sequencing depth for comparison, we observed a clear separation with significantly increased IR frequencies at diagnosis and remission compared to controls (Supplementary Figure 9). Moreover, we found that the IR ratios are consistently higher in diagnosis and remission samples compared to healthy controls (Figure 3A). Differences between diagnosis and remission samples were less pronounced (Figure 3A; Supplementary

Figure 10). While IR profiles are unsuitable for patient stratification (Supplementary Figure 11), differentially retained introns between diagnosis and remission samples are, similarly to differentially expressed genes, associated with cell cycle regulators as well as genes related to DNA replication and repair (Supplementary Figure 12). However, in remission IR primarily affects genes involved in splicing processes (Supplementary Figure 12) suggesting the aberrant splicing persists in remission due to auto-regulatory processes. Overall, IR profiles are highly dynamic in CML, which becomes apparent when considering the large overlap of differentially retained introns ($n = 419$) in all three comparisons (Figure 3B). We confirmed these results with an independent dataset of 59 CML patients from Branford et al.⁹, which includes samples from various stages of disease progression, and from 4 healthy controls (Figure 3C). The number of IR events in this cohort was highly variable. While there are on average more IR events in chronic phase compared to healthy controls, the number drops significantly at blast crisis (Figure 3C, right). Examining differential exon usage in our samples we observed a similar trend although not as marked as observed in differential IR, i.e. both diagnosis and remission samples exhibit a larger number of differentially used exons (DUEs) compared to normal controls. Most of these DUEs are upregulated (diagnosis, 78%; remission, 81%; Figure 3D). A large fraction of the DUEs at diagnosis (72%) remain differentially expressed at remission as well. Some genes of these mutual DUEs are associated with DNA repair mechanisms (top-enriched GO term; Figure 3E). Only 26 exons in total are differentially used in diagnosis vs remission samples (Figure 3D).

301 302 *Modulation of IR levels supports lineage-specific gene expression in CML*

303 IR is accepted as a mechanism of post-transcriptional gene regulation triggering
304 nonsense-mediated decay due to the presence of premature termination codons within

the retained introns^{27,28}. We analysed the relationship between IR and gene expression changes but overall the two variables did not significantly correlate (Figure 4A). However, many of the genes, whose expression does anti-correlate with the retention of their introns, are involved in cell cycle regulation (e.g. *CCL3*, *CKS2*, and *SPAG5*), and DNA replication/repair (e.g. *FANCI/FANCD2*, and *PCNA*) (Figure 4A). We selected three differentially retained introns of cell cycle regulator genes for experimental validation (*CKS2* intron 1, *CCL3* intron 1, and *SERPINB1* intron 2). Based on the RNA sequencing data analysis, all three introns have an increased expression and higher IR ratio at diagnosis compared to healthy controls (Figure 4A). We confirmed these observations via qRT-PCR (Figure 4B). The IR ratios of *CKS2* intron 1, and *SERPINB1* intron 2 increase further at remission resulting in a reduction of host gene expression, while the IR ratio of *CCL3* intron 1 is reduced at remission and associated with an increase in gene expression (Figure 4A & B). Interestingly, some instances of inverse relationships between IR and gene expression affect lineage-specific genes (Figure 4C). For example, the chemokine Interleukin-8 (*CXCL8*), which is predominantly expressed in neutrophils, is upregulated at diagnosis. This could be attributed to the increased myeloid progenitor cell abundance (Figure 1B), but also to the reduced IR levels in *CXCL8* mRNA (Figure 4A). In control samples, the IR levels increase from 12.9% to 26.7% in *CXCL8* (intron 2), while gene expression is reduced 4-fold. *CXCL8* promotes CML cell proliferation²⁹ and previous reports suggest that *CXCL8* expression is modulated both by *BCR-ABL1* expression (causing increase) and TKI treatment (causing inhibition of *CXCL8*)³⁰. In contrast, *PRKCH* is downregulated at CML diagnosis along with increased IR levels (Figure 4A). *PRKCH* is a serine-threonine kinase that regulates hematopoietic stem cell function and is specifically expressed in lymphoid cells. Notably, high

330 *PRKCH* expression has been associated with poor prognosis in acute myeloid
331 leukemia³¹.

332 These observations suggest that lineage-specific expression of drivers of cell
333 proliferation and hematopoietic cell differentiation can be mediated through IR level
334 changes in CML.

335 *Epigenetic features modulate intron retention in CML*

336 Characteristics of the retained introns in our CML cohort are in accordance with
337 previously described attributes, such as their relatively short length (Supplementary
338 Figure 13), and higher GC content (Supplementary Figure 14). While these are
339 intrinsic features of IR, we sought potential *trans*-regulators and found that the gene
340 that most strongly correlates ($\rho = 0.64$) with IR frequencies is *MED21* (Mediator
341 Complex Subunit 21; Figure 5A). *MED21* regulates gene transcription by interacting
342 with RNA polymerase II³². RNA polymerase II elongation rates influence splice site
343 recognition and IR¹³. Interestingly, the expression of *ZNF160*, encoding a Zinc Finger
344 Protein and repressor of transcription, strongly anti-correlates ($\rho = -0.62$) with IR
345 frequencies (Figure 5A).

346 We have previously shown that DNA methylation regulates IR in myeloid cells¹³. To
347 confirm this mode of regulation in our CML cohort, we compared CpG methylation
348 around the splice sites (± 200 bp) and in the centre (200 bp) of retained and non-
349 retained introns. We confirmed decreased DNA methylation levels near the 3' splice
350 junction and within the body of retained introns (Figure 5B). Similar patterns were
351 observed in remission and control samples suggesting that this mode of IR regulation
352 is also present in lymphoid cells (Supplementary Figure 15).

353 To further investigate the role of epigenetics in the regulation of IR in CML, we
354 retrieved RNA-seq and ChIP-seq data of the K562 CML cell line from ENCODE. We

355 analysed transcriptomics data of RNA binding protein knockdown experiments and
 356 observed significantly increased IR occurrences with the most drastic IR increase (~5
 357 fold) following Poly(RC) Binding Protein 2 (*PCBP2*) knockdown (Figure 5C). All
 358 genes with enriched binding motifs near frequently retained introns²⁰ were
 359 downregulated at CML diagnosis, except for *SRSF5* (Supplementary Figure 16).
 360 Three of these RNA binding proteins, including *PCBP2*, remained downregulated in
 361 remission (Figure 5D).
 362 To identify additional potential explanations for the dysregulation of IR in CML, we
 363 analysed histone modifications in K562 cells. While higher levels of H3K36
 364 trimethylation have previously been associated with IR^{33,34}, we found only marginal
 365 differences in H3K36me3 levels in retained versus non-retained introns. Increasing
 366 numbers of H3K36me3 peaks around introns were observed in more highly expressed
 367 genes (Supplementary Figure 17). However, we found that monomethylation and
 368 acetylation of histone H3 lysines 4 and 9, respectively (H3K4me1 and H3K9ac) were
 369 significantly enriched near splice sites of retained introns and within their intron
 370 bodies in contrast to non-retained introns (Figure 5E). While a significant enrichment
 371 of H3K9ac has been found to co-occur at many gene regulatory elements in mouse
 372 embryonic stem cells³⁵, we here established a strong association of H3K9ac with IR.
 373 Elevated levels of H3K4 methylation in association with IR have been previously
 374 reported by Zhou et al.³³.
 375 In summary, our analysis of CML patient transcriptomes revealed aberrant alternative
 376 splicing both at diagnosis and remission. Elevated IR ratios in CML can be explained
 377 by multiple regulatory mechanisms including the regulation of RNA Pol II
 378 elongation, demethylation of CpG sites, and histone modifications (Figure 5F). The
 379 TKI treatment itself seems to have no effect on IR (Supplementary Figure 18).

380

381 **Discussion**

382 In this study, we analysed epigenomics and transcriptomics data from patients
383 diagnosed with chronic phase Ph⁺ CML and matched remission samples after
384 effective treatment with TKIs. Thereby, we generated novel insights into the
385 molecular responses to TKI therapy beyond the known reduction in *BCR-ABL1*
386 expression.

387 We have identified recurring fusion transcripts in patient samples including an
388 uncharacterised fusion between the cell surface marker C-type lectin domain family
389 12 member A (*CLEC12A*; 12p13.31) and the *MIR223* gene (Xq12). Other non-
390 specific intra-chromosomal fusions include *EEF1DP3-FRY*, *EIF4E3-FOXP1*,
391 *KANSL1-ARL17B*, and the mitochondrial fusion *ND6-TE* (Supplementary Figure 2A).
392 An increasing number of mitochondrial mutations and fusions have been described as
393 a consequence of increased reactive oxygen species during ageing³⁶, however, the age-
394 range (21-70 yrs) of *ND6-TE*-positive patients in our cohort does not support an age-
395 related occurrence of this fusion.

396 Gene expression profiles of our CML remission samples are more similar to those of
397 healthy donors, likely due to the change in cell composition and the loss of *BCR-*
398 *ABL1* clones. Consistent with previous studies, we found cell cycle regulators among
399 the most aberrantly expressed genes in the diagnosis samples. For example, cyclins
400 A1 and B2 (*CCNA1*, *CCNB2*) were among the most differentially expressed genes,
401 with overexpression at diagnosis and low expression at remission (comparable to
402 expression levels observed in healthy controls). While cyclin D2 (*CCND2*) has
403 previously been characterized as a *BCR-ABL*-dependent mediator of cell proliferation
404 in hematopoietic cells³⁷, marginal changes in *CCND2* (FC= 0.87, p=.04) expression

405 suggest that other cyclins such as *CCNA1* (FC=12.3, p=5.4e-53) and *CCNB2*
406 (FC=7.2, p= 9.2e-55) were among the more potent regulators of cell proliferation in
407 this cohort of CML patients.

408 The most surprising results was that, in contrast to gene expression, alternative
409 splicing patterns did not return to normal in CML remission samples. Among other
410 forms of alternative splicing, IR affects transcriptomic complexity and leads to the
411 inclusion of premature termination codons in mature transcripts causing nonsense-
412 mediated decay^{38,39,27,39,40}. Normal granulocytic blood cells make use of this
413 conserved mechanism of gene regulation by increasing the number of intron-retaining
414 transcripts of hundreds of genes that are essential to granulocytic differentiation^{27,38}.

415 Shen *et al.* have shown that alternative splicing events can be used to construct
416 predictors for patient survival, that outperform gene expression-based predictors in
417 multiple cancers⁴¹. Splicing-based prognostic markers have already been found for
418 several cancer types^{41,42}. Therefore, a splicing-based predictor for TKI cessation
419 success is not just desirable but an achievable prospect, once clinical data of trial TKI
420 cessation together with remission transcriptomics data become available.

421 Our results were consistent across patients and different specific TKI treatments. It is
422 known that the cell composition of the peripheral blood changes in CML - with
423 myeloid progenitor cell-enrichment at diagnosis and mature lymphoid cell-enrichment
424 at remission. The latter is very similar to the cell composition observed in the
425 peripheral blood of healthy individuals (Supplementary Figure 5). Therefore,
426 epigenomic and transcriptomic differences between diagnosis and remission may be
427 expected. However, remarkable are the persistent alternative splicing perturbations in
428 remission that do not resemble patterns observed in healthy controls.

429 We found new evidence for epigenetic regulators of alternative splicing, such as
 430 reduced DNA methylation, and increased H3K4me1 and H3K9ac marks around
 431 retained introns. We have previously shown that DNA methylation can regulate IR in
 432 myeloid cells and during terminal granulopoiesis^{13,43}. In CML reduced DNA
 433 methylation in the intron body and 3' splice site is associated with IR (Figure 5B). A
 434 role for the promoter and enhancer mark H3K4 mono-methylation in gene repression
 435 has been identified previously⁴⁴, however, H3K4me1-dependent alternative splicing
 436 has not been described before. Both H3K4me1 and H3K9ac are present at the
 437 promoters of intron-retaining and non-retaining genes but are predominantly found
 438 near splice sites and within bodies of retained introns in contrast to non-retained
 439 introns (Figure 5E). While both histone marks seem indicative of active transcription
 440 (Figure 5E), by regulating Pol II recruitment and elongation rates^{44,45} around splice
 441 sites they could promote IR in CML. IR regulation via H3.3K36me3-specific readers,
 442 such as previously shown in HeLa cells³⁴, seems less relevant in CML patients.

443 **Conclusions**

444 In summary, our study has provided new insights into the epigenomic and
 445 transcriptomic landscapes of CML patients at diagnosis and remission. We have
 446 shown that, despite a changing cell composition at major or complete molecular
 447 remission, alternative splicing remains aberrantly regulated with high IR levels
 448 affecting multiple cell cycle regulators. Further research is required to test whether
 449 sustained RNA processing alterations in CML remission facilitate relapse following
 450 TKI cessation. The new knowledge about epigenetic links to alternative splicing in
 451 CML could facilitate the development of epigenetic therapeutic adjuvants to increase
 452 the likelihood of successful TKI cessation. Modulation of spliceosome function has

recently been proposed as a new therapeutic avenue in leukemia⁴⁶, which is an idea reinforced by our results.

List of abbreviations

CML - chronic myeloid leukemia
 DEU - Differential exon usage
 DUEs - differentially used exons
 DMRs - Differentially methylated regions
 FC - fold change
 GO - gene ontology
 IR - intron retention
 PBMCs - peripheral blood mononuclear cells
 Ph+ - Philadelphia-positive
 TCGA - The Cancer Genome Atlas
 TPM - transcripts per million
 TKI - tyrosine kinase inhibitor
 WGBS - Whole genome bisulfite sequencing

Declarations

Ethics approval and consent to participate

All studies were approved by the institutional review boards (HREC protocol No: 131015, 070718c, 081211, and 101010 - Royal Adelaide Hospital) and are in accordance with the Declaration of Helsinki. Patients provided written informed consent.

476 *Consent for publication*

477 Not applicable.

478 *Availability of data and materials*

479 The dataset supporting the conclusions of this article is available in the GEO
480 repository, [GSE144119](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144119).

481 *Competing interests*

482 S.B. is member of the advisory board of Qiagen, Novartis and Cepheid, received
483 honoraria from Qiagen, Novartis, Bristol-Myers Squibb and Cepheid, and research
484 support from Novartis. D.W. received RS funding and honoraria from BMS and
485 honoraria from Amgen. J.E.J.R. has received honoraria or speakers fees (GSK,
486 Miltenyi, Takeda, Gilead, Pfizer, Spark, Novartis, Celgene, bluebird bio); Director of
487 Pathology (Genea); equity ownership (Genea, Rarecyte); consultant (Rarecyte,
488 Imago); chair, Gene Technology Technical Advisory, OGTR, Australian
489 Government. The remaining authors declare no competing financial interests.

490 *Funding*

491 This work was supported by the National Health and Medical Research Council
492 (Investigator Grant #1177305 to J.E.J.R., Project #1128175 and #1129901 to J.E.J.R.
493 and J.J.-L.W., #1126306 to J.J.-L.W., APP1027531 and APP1104425 to S.B., and
494 APP1135949 to T.P.H.; the NSW Genomics Collaborative Grant (J.E.J.R. and J.J.-
495 L.W.); Cure the Future (J.E.J.R.); and an anonymous foundation (J.E.J.R.). U.S. and
496 J.J.-L.W. hold Fellowships from the Cancer Institute of New South Wales. U.S. also
497 received support from the Australian Academy of Science in form of an Australia-

498 India Early and Mid-Career Fellowship. Financial support was provided by Tour de
499 Cure (Scott Canner Research Fellowship) to C.G.B. and for research grants to C.G.B.
500 and J.E.J.R. This research was funded by the Cancer Council NSW Project Grants
501 (RG11-12, RG14-09, RG20-12) to J.E.J.R., C.G.B., and U.S. Support was also
502 received from the Ray and Shirl Norman Cancer Research Trust (S.B.).

503 *Authors' contributions*

504 J.E.J.R. and J.J.-L.W. designed the research. U.S., J.S.S., P.L.L., and V.P. performed
505 bioinformatic analyses and interpreted the data. D.L.W., V.A.S., S.B., A.G.T., and
506 T.P.H. provided CML patient samples, clinical data and further insight. B.D., G.M.,
507 S.N., C.M., and C.G.B. performed and/or supervised experiments. U.S. wrote the
508 manuscript with help from J.J.-L.W., J.S.S., G.M., B.D., and V.P. All authors
509 reviewed and approved the manuscript.

510 *Acknowledgements*

511 We would like to acknowledge Dr. Andreas Schreiber (University of South Australia)
512 for retrieving the sequencing data from the Branford et al. study⁹.

513

514

515 **References**

- 516 1. Peng C, Li S. Chronic Myeloid Leukemia (CML) Mouse Model in
517 Translational Research. *Methods Mol Biol.* 2016;1438:225-243.
- 518 2. Talpaz M, Silver RT, Druker BJ, et al. Imatinib induces durable hematologic
519 and cytogenetic responses in patients with accelerated phase chronic myeloid
520 leukemia: results of a phase 2 study. *Blood.* 2002;99(6):1928-1937.
- 521 3. Apperley JF. Part I: mechanisms of resistance to imatinib in chronic myeloid
522 leukaemia. *Lancet Oncol.* 2007;8(11):1018-1029.
- 523 4. Perrotti D, Jamieson C, Goldman J, Skorski T. Chronic myeloid leukemia:
524 mechanisms of blastic transformation. *J Clin Invest.* 2010;120(7):2254-2264.
- 525 5. Yong AS, Melo JV. The impact of gene profiling in chronic myeloid
526 leukaemia. *Best Pract Res Clin Haematol.* 2009;22(2):181-190.
- 527 6. Ma W, Giles F, Zhang X, et al. Three novel alternative splicing mutations in
528 BCR-ABL1 detected in CML patients with resistance to kinase inhibitors. *Int J Lab*
529 *Hematol.* 2011;33(3):326-331.
- 530 7. Radich JP, Dai H, Mao M, et al. Gene expression changes associated with
531 progression and response in chronic myeloid leukemia. *Proc Natl Acad Sci U S A.*
532 2006;103(8):2794-2799.
- 533 8. Dvinge H, Bradley RK. Widespread intron retention diversifies most cancer
534 transcriptomes. *Genome Med.* 2015;7(1):45.
- 535 9. Branford S, Wang P, Yeung DT, et al. Integrative genomic analysis reveals
536 cancer-associated mutations at diagnosis of CML in patients with high-risk disease.
537 *Blood.* 2018;132(9):948-961.
- 538 10. Heller G, Topakian T, Altenberger C, et al. Next-generation sequencing
539 identifies major DNA methylation changes during progression of Ph⁺ chronic
540 myeloid leukemia. *Leukemia.* 2016;30(9):1861-1868.
- 541 11. Wong ACH, Rasko JEJ, Wong JJ. We skip to work: alternative splicing in
542 normal and malignant myelopoiesis. *Leukemia.* 2018;32(5):1081-1093.
- 543 12. Tilgner H, Nikolaou C, Althammer S, et al. Nucleosome positioning as a
544 determinant of exon recognition. *Nat Struct Mol Biol.* 2009;16(9):996-1001.
- 545 13. Wong JJ, Gao D, Nguyen TV, et al. Intron retention is regulated by altered
546 MeCP2-mediated splicing factor recruitment. *Nat Commun.* 2017;8:15134.
- 547 14. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq
548 aligner. *Bioinformatics.* 2013;29(1):15-21.
- 549 15. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments.
550 *Bioinformatics.* 2012;28(16):2184-2185.
- 551 16. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis
552 results for multiple tools and samples in a single report. *Bioinformatics.*
553 2016;32(19):3047-3048.
- 554 17. Haas BJ, Dobin A, Stransky N, et al. STAR-Fusion: Fast and Accurate Fusion
555 Transcript Detection from RNA-Seq. *bioRxiv.* 2017:120295.
- 556 18. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast
557 and bias-aware quantification of transcript expression. *Nature methods.*
558 2017;14(4):417.
- 559 19. Shen S, Park JW, Lu ZX, et al. rMATS: robust and flexible detection of
560 differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci U S*
561 *A.* 2014;111(51):E5593-5601.
- 562 20. Middleton R, Gao D, Thomas A, et al. IRFinder: assessing the impact of
563 intron retention on mammalian gene expression. *Genome Biol.* 2017;18(1):51.

- 564 21. Anders S, Reyes A, Huber W. Detecting differential usage of exons from
565 RNA-seq data. *Genome Res.* 2012;22(10):2008-2017.
- 566 22. Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M. DNA methylation:
567 bisulphite modification and analysis. *Nature protocols.* 2006;1(5):2353.
- 568 23. Nair SS, Luu PL, Qu W, et al. Guidelines for whole genome bisulphite
569 sequencing of intact and FFPE DNA on the Illumina HiSeq X Ten. *Epigenetics*
570 *Chromatin.* 2018;11(1):24.
- 571 24. Song Q, Decato B, Hong EE, et al. A reference methylome database and
572 analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One.*
573 2013;8(12):e81148.
- 574 25. Love MI, Huber W, Anders S. Moderated estimation of fold change and
575 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
- 576 26. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis
577 of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44-
578 57.
- 579 27. Wong JJ, Ritchie W, Ebner OA, et al. Orchestrated intron retention regulates
580 normal granulocyte differentiation. *Cell.* 2013;154(3):583-595.
- 581 28. Edwards CR, Ritchie W, Wong JJ, et al. A dynamic intron retention program
582 in the mammalian megakaryocyte and erythrocyte lineages. *Blood.* 2016;127(17):e24-
583 e34.
- 584 29. Corrado C, Raimondo S, Saieva L, Flugy AM, De Leo G, Alessandro R.
585 Exosome-mediated crosstalk between chronic myelogenous leukemia cells and
586 human bone marrow stromal cells triggers an interleukin 8-dependent survival of
587 leukemia cells. *Cancer Lett.* 2014;348(1-2):71-76.
- 588 30. Hantschel O, Gstoettenbauer A, Colinge J, et al. The chemokine interleukin-8
589 and the surface activation protein CD69 are markers for Bcr-Abl activity in chronic
590 myeloid leukemia. *Mol Oncol.* 2008;2(3):272-281.
- 591 31. Porter SN, Magee JA. PRKCH regulates hematopoietic stem cell function and
592 predicts poor prognosis in acute myeloid leukemia. *Exp Hematol.* 2017;53:43-47.
- 593 32. Plaschka C, Lariviere L, Wenzek L, et al. Architecture of the RNA
594 polymerase II–Mediator core initiation complex. *Nature.* 2015;518(7539):376.
- 595 33. Zhou Y, Lu Y, Tian W. Epigenetic features are significantly associated with
596 alternative splicing. *BMC Genomics.* 2012;13:123.
- 597 34. Guo R, Zheng L, Park JW, et al. BS69/ZMYND11 reads and connects histone
598 H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA
599 processing. *Mol Cell.* 2014;56(2):298-310.
- 600 35. Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L. H3K9 and
601 H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks
602 a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC*
603 *Genomics.* 2012;13:424.
- 604 36. Westermann B. Mitochondrial fusion and fission in cell life and death. *Nat*
605 *Rev Mol Cell Biol.* 2010;11(12):872-884.
- 606 37. Jena N, Deng M, Sicinska E, Sicinski P, Daley GQ. Critical role for cyclin D2
607 in BCR/ABL-induced proliferation of hematopoietic cells. *Cancer Res.*
608 2002;62(2):535-541.
- 609 38. Schmitz U, Pinello N, Jia F, et al. Intron retention enhances gene regulatory
610 complexity in vertebrates. *Genome Biol.* 2017;18(1):216.
- 611 39. Wong JJ, Au AY, Ritchie W, Rasko JE. Intron retention in mRNA: No longer
612 nonsense: Known and putative roles of intron retention in normal and disease biology.
613 *Bioessays.* 2016;38(1):41-49.

614 40. Kurosaki T, Maquat LE. Nonsense-mediated mRNA decay in humans at a
615 glance. *J Cell Sci.* 2016;129(3):461-467.
616 41. Shen S, Wang Y, Wang C, Wu YN, Xing Y. SURVIV for survival analysis of
617 mRNA isoform variation. *Nat Commun.* 2016;7:11548.
618 42. Li Y, Sun N, Lu Z, et al. Prognostic alternative mRNA splicing signature in
619 non-small cell lung cancer. *Cancer Lett.* 2017;393:40-51.
620 43. Gao D, Pinello N, Nguyen TV, et al. DNA methylation/hydroxymethylation
621 regulate gene expression and alternative splicing during terminal granulopoiesis.
622 *Epigenomics.* 2019;11(1):95-109.
623 44. Cheng J, Blum R, Bowman C, et al. A role for H3K4 monomethylation in
624 gene repression and partitioning of chromatin readers. *Mol Cell.* 2014;53(6):979-992.
625 45. Gates LA, Shi J, Rohira AD, et al. Acetylation on histone H3 lysine 9
626 mediates a switch from transcription initiation to elongation. *J Biol Chem.*
627 2017;292(35):14456-14472.
628 46. Lee SC, Dvinge H, Kim E, et al. Modulation of splicing catalysis for
629 therapeutic targeting of leukemia with mutations in genes encoding spliceosomal
630 proteins. *Nat Med.* 2016;22(6):672-678.
631 47. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular
632 heterogeneity landscape. *Genome Biol.* 2017;18(1):220.
633 48. Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for
634 comparative genomics. *Genome Res.* 2009;19(9):1639-1645.
635 49. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based Human
636 Protein Atlas. *Nat Biotechnol.* 2010;28(12):1248-1250.
637

Figure Legends

Figure 1 Molecular remission is reflected in transcriptomic profiles. (A) The Top 200 differentially expressed genes with lowest adjusted p -value in likelihood ratio test. **(B)** Predicted cell type enrichment in peripheral blood samples. The stacked bar graph shows normalized cell type enrichment scores for patient and control transcriptomes based on bulk RNA sequencing data. Predictions were made with the *xCell* R package (github.com/dviraran/xCell)⁴⁷. HSC – hematopoietic stem cell, CLP – common lymphoid progenitor, CMP – common myeloid progenitor, GMP - granulocyte macrophage progenitor, MEP - megakaryocytic erythroid progenitor, NK – natural killer, DC – dendritic cell. **(C)** Principal component analysis separates diagnosis samples from the common features of remission samples and healthy controls.

Figure 2 DNA methylation in CML. (A) Frequency of differentially methylated regions (DMRs) based on delta values between pooled diagnosis and remission samples (left) as well as remission and control samples (right). **(B)** Circos plot illustrating genome-wide differences in the frequency of DMRs (central ring with blue scatter plot: diagnosis vs remission; inner ring with red scatterplot: remission vs control). This plot was generated using the Circos software⁴⁸. **(C)** Scatterplot illustrating the number of significantly differentially methylated CpG sites per DMR and associated delta values (red points, diagnosis vs remission; blue points, remission vs control). **(D)** Enriched GO terms associated with DMRs (remission vs diagnosis) in gene promoter regions (red dashed line marks $p = 0.05$). **(E)** Distribution and density of DMRs across gene features. RC – remission vs control; DR – diagnosis vs remission; DC – diagnosis vs control.

Figure 3 Global alternative splicing patterns are similar despite major molecular remission. (A) Scatterplots illustrating differential IR events between pooled diagnosis/remission/control samples binned into hexagons to avoid overplotting (yellow: $p_{adj} \leq 0.05$; red: $p_{adj} \leq 0.01$). **(B)** Venn diagram showing intersections of mutually differentially retained introns. **(C)** Dynamics of IR across CML phases. CML patient data from Branford et al.⁹ was analysed for the occurrence of IR events. Left: Sample sizes for Control, CP: Chronic Phase, AP: Accelerated Phase, MBC: Myeloid Blast Crisis, and LBC: Lymphoid Blast Crisis samples. Middle: Sample numbers matched from different phases. Right: Boxplot of IR frequencies in samples from different CML phases. Mann-Whitney U test; ** ($p < 0.01$), ns ($p > 0.05$). **(D)** Differential exon usage in pooled diagnosis vs control (left), diagnosis vs remission (middle), and remission vs control samples. **(E)** Intersection of differentially used exons in diagnosis and remission samples compared to controls (the top-enriched GO term and associated genes are listed in the middle).

669

670 **Figure 4 Differentially retained introns inhibit the expression of cell cycle regulators following major**
671 **molecular remission. (A)** Scatterplots illustrating the relationship between differentially retained introns (ΔIR)
672 and differential gene expression ($\log_2$ expression fold change). Introns ($|\Delta IR| \geq 0.1$) with inverse relationship to
673 host gene expression ($|\log(FC)| \geq 1$) are highlighted in red. **(B)** Coverage plots of selected differential IR events
674 (left) and qRT-PCR validation of differential IR between diagnosis and control (right). qRT-PCR validation was
675 performed in triplicates on independent diagnosis samples and results were normalised to *B2M* expression. D -
676 diagnosis; C - control. **(C)** Blood cell type expression of genes with inverse IR/gene expression relationships
677 (diagnosis v control; blue - myeloid, magenta - lymphoid cells). The data was retrieved from the Human Protein
678 Atlas (v19.3; <http://www.proteinatlas.org>)⁴⁹.

679

680 **Figure 5 Regulation of IR in CML. (A)** Two genes that most strongly correlate/anti-correlate with IR frequencies
681 (Spearman correlation). **(B)** CpG methylation around intron splice sites (± 200 bp) and in the centre (200 bp) of
682 retained- (red) and non-retained introns (blue) in CML diagnosis samples. General linear hypothesis test; p -value <
683 $2.2e-16$ (adj.- $p < 1e-10$). **(C)** The effect of RNA binding proteins on IR frequency. RNA-seq data of shRNA-
684 mediated knockdown of indicated RNA binding proteins in K562 cells was retrieved from the ENCODE project
685 (encodeproject.org). WT - wild type. **(D)** Expression of RNA binding proteins with enriched binding motifs near
686 frequently retained introns²⁰. TPM - transcripts per million mapped reads; C - control; D - diagnosis; R -
687 remission; n.s. - non-significant, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **(E)** Histone modifications with increased
688 numbers of peaks near splice sites of retained introns (red: retained; blue: non-retained introns). The analysis was
689 repeated for each expression quartile to exclude the possibility that increased histone modifications are associated
690 with gene expression. **(F)** Possible mechanisms of IR regulation in CML discussed in the text.

691

692 Tables

693 **Table 1 Patient and healthy donor samples used in RNAseq and WGBS experiments.** Sequencing was
694 performed on PBMCs.

	Diagnosis	Remission	Control
Number of samples	16	11	6
Sex (F/M)	11/5	8/3	2/4
Median age yrs (range)	54 (21-69)	56 (23-70)	37 (30-42)
Trial (RESIST/ENESTxtnd/PINNACLE/ PERSISTEM)	1/4/4/7	1/4/4/2	
Timepoint	Screening (x12) /Day1	6-35 months	
First line TKI (imatinib/nilotinib)		8/8	
RNAseq and WGBS (RNAseq only)	10 (+6)	10 (+1)	6

695

Figure 1

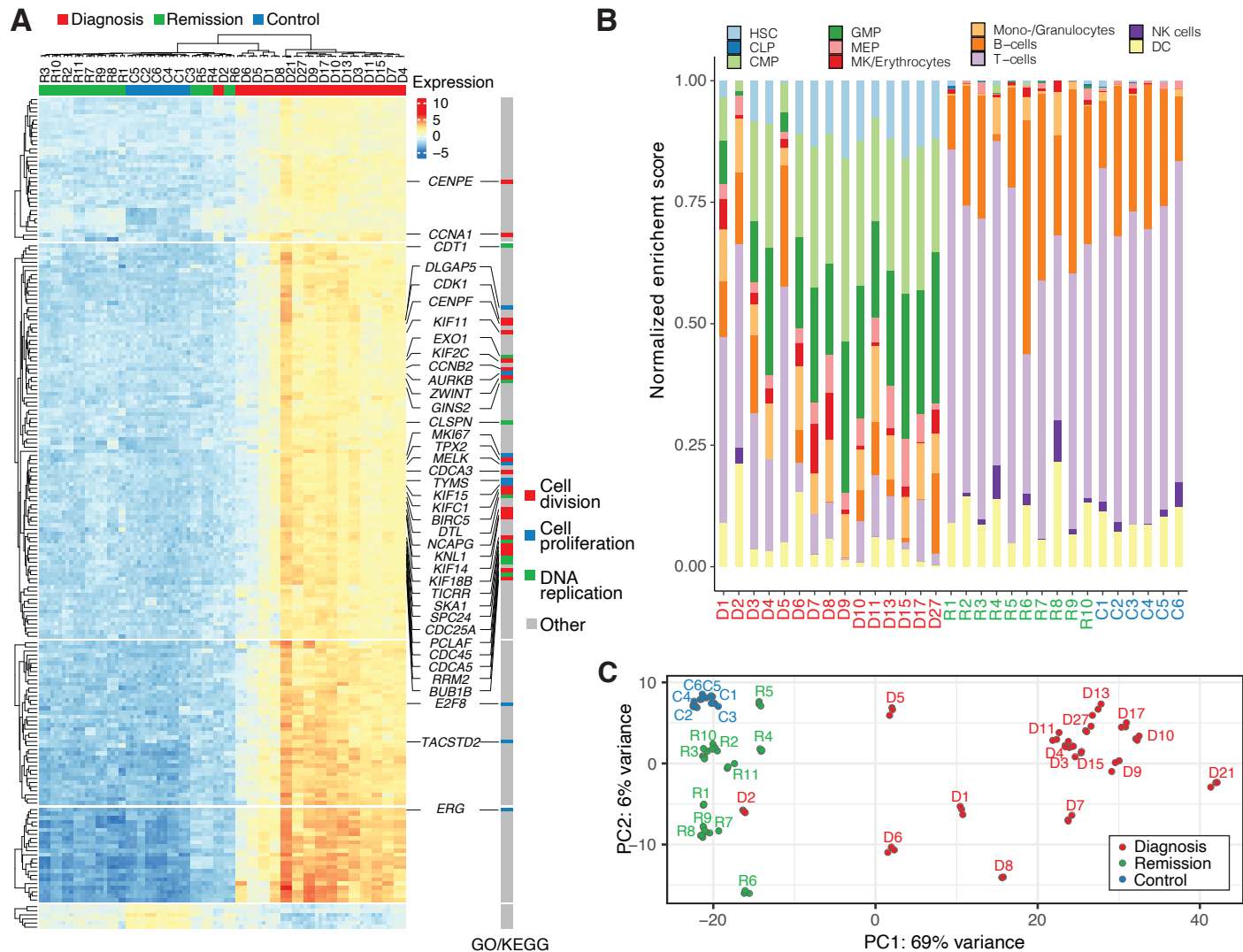


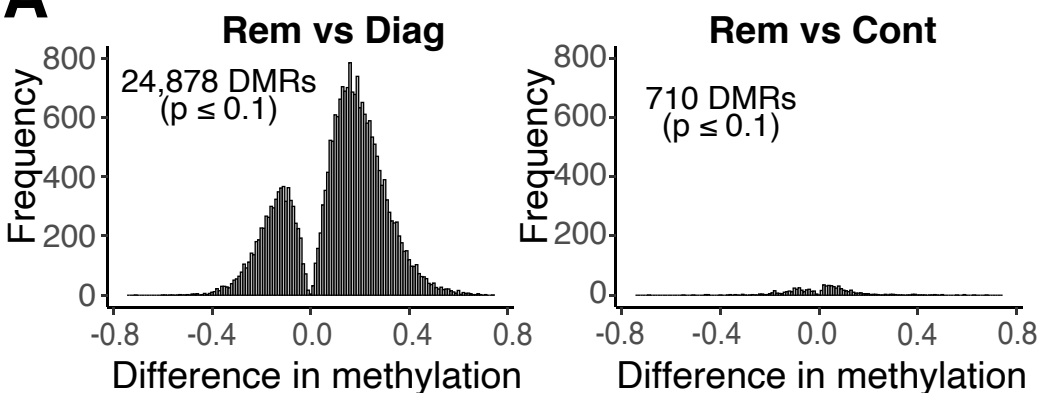
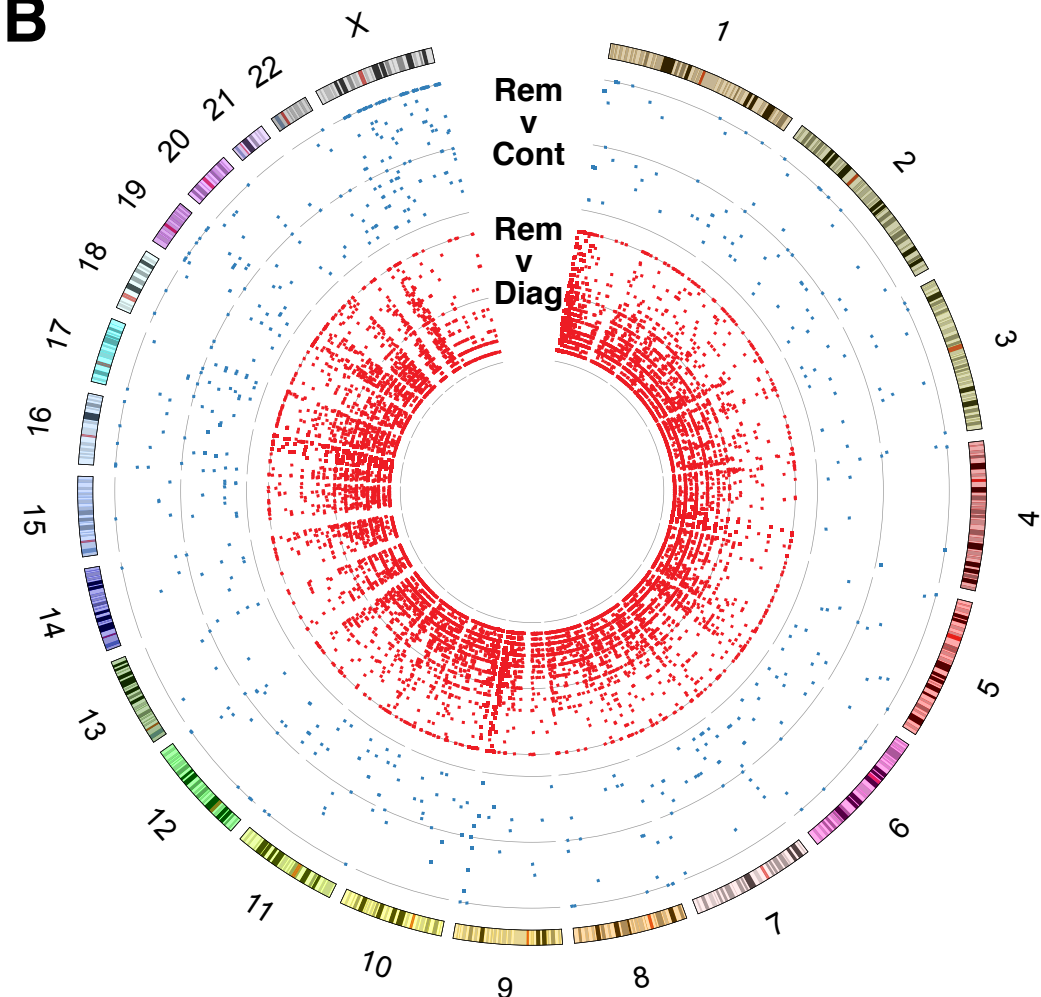
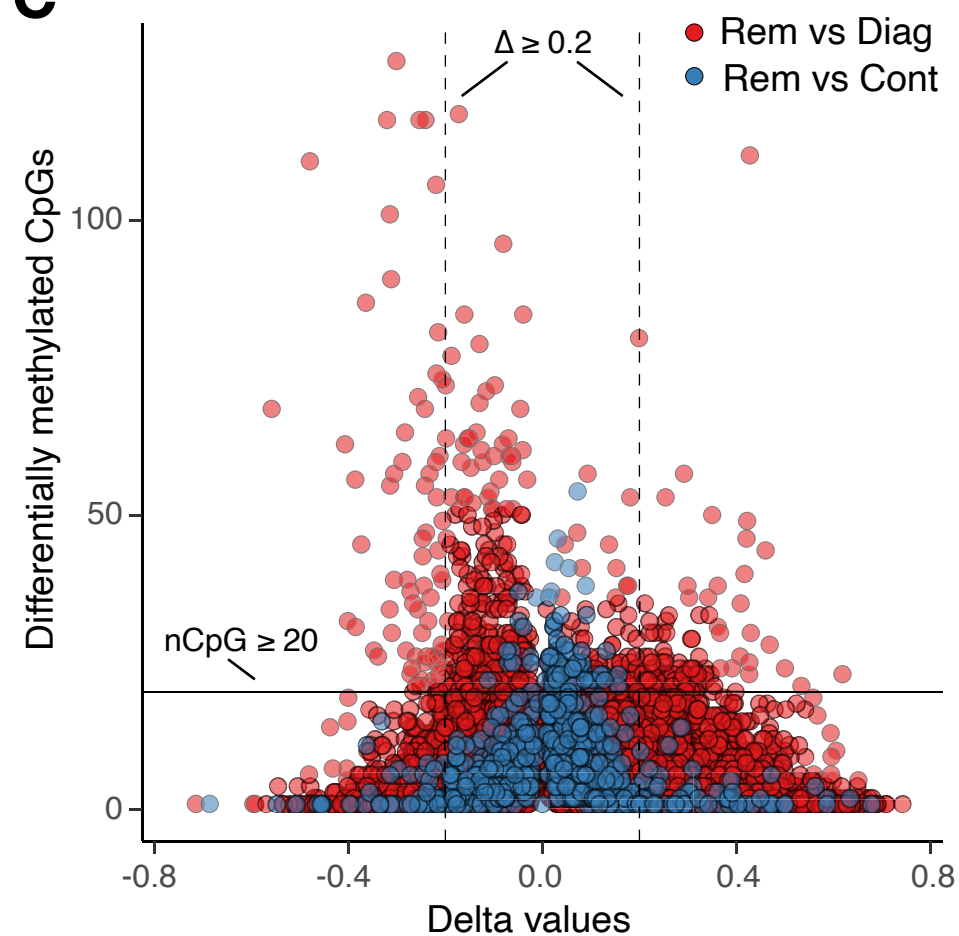
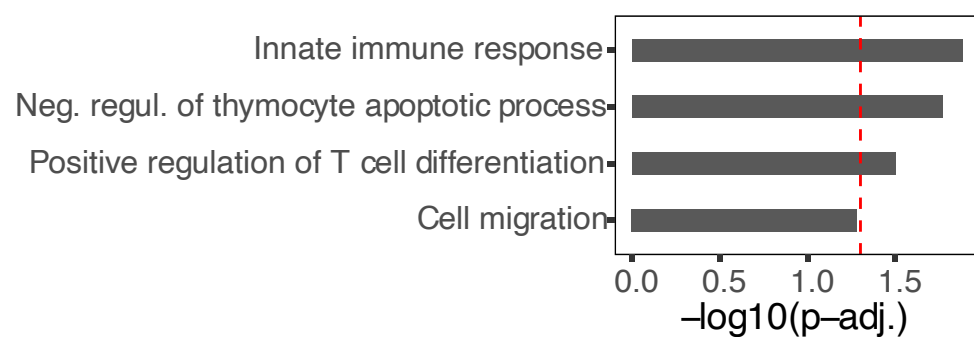
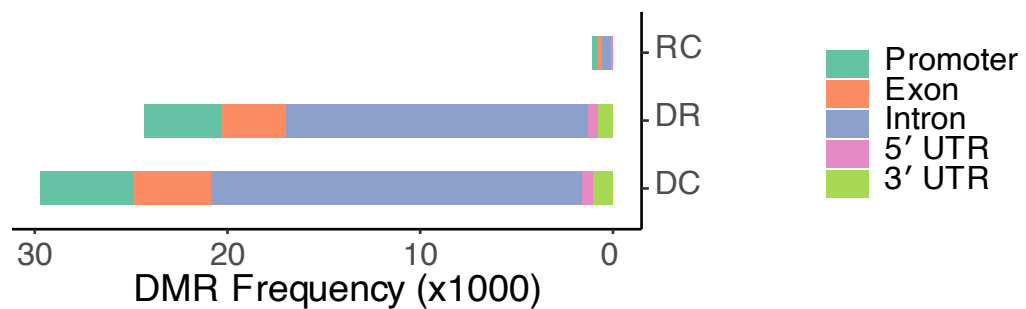
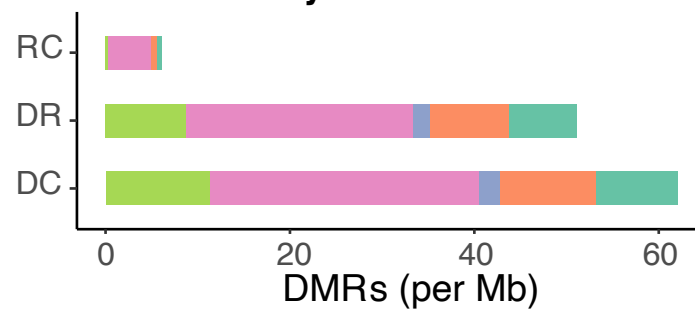
Figure 2**A****B****C****D****E****DMR Distribution****DMR Density**

Figure 3

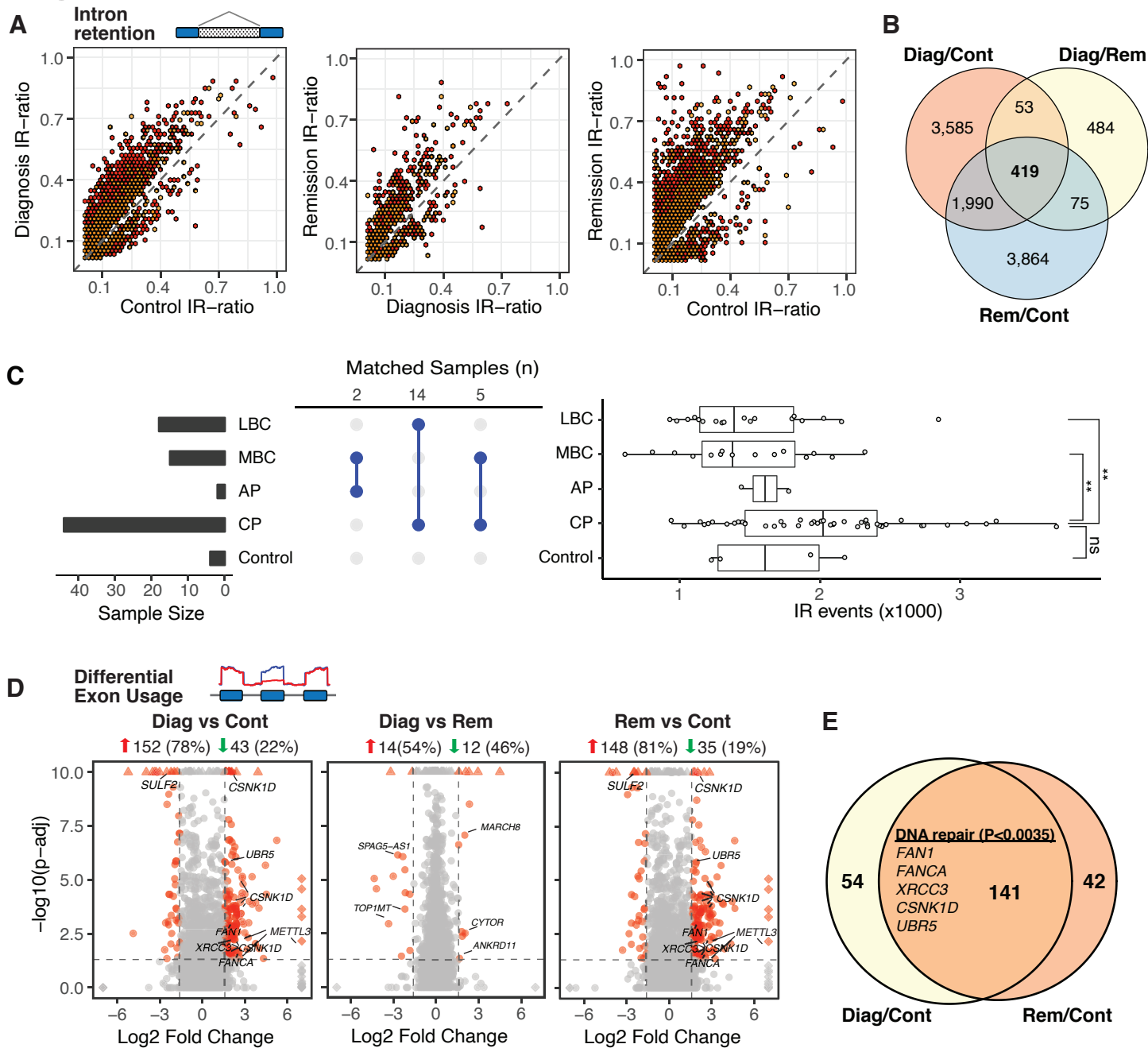


Figure 4

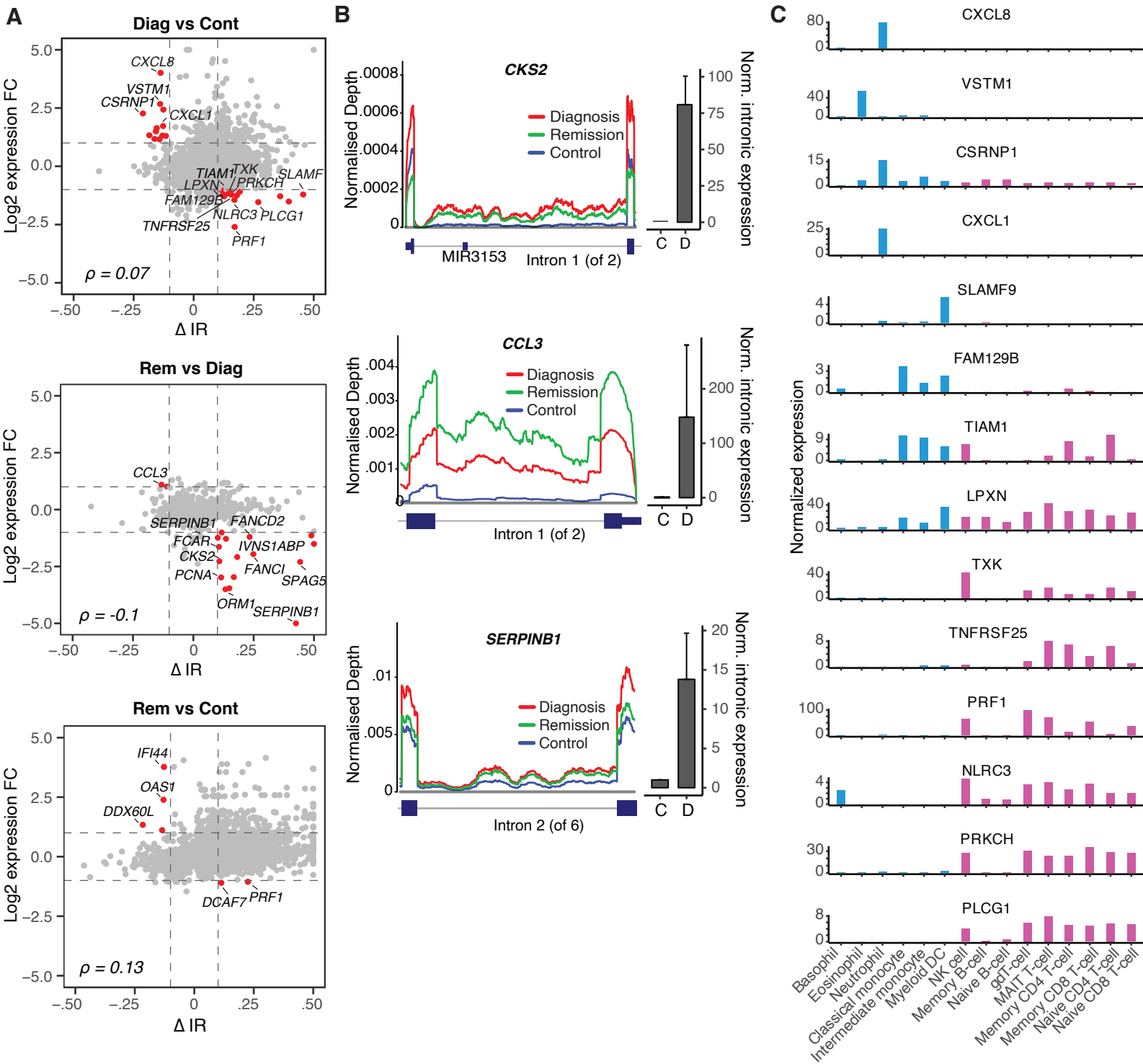


Figure 5

