Performances of targeted RNA-sequencing for the analysis of fusion transcripts, gene mutation and expression in haematological malignancies.

Sandrine Hayette[1,2,7], Béatrice Grange[1], Maxime Vallee[3], Claire Bardel[3,4], Sarah Huet[1,2], Isabelle Mosnier[1], Kaddour Chabane[1], Thomas Simonet[3], Marie Balsat[5], Maël Heiblig[5], Isabelle Tigaud[1,7], Franck E. Nicolini[6,7], Sylvain Mareschal[2], Gilles Salles[2,5], Pierre Sujobert[1,2]

1. Hospices Civils de Lyon, Hôpital Lyon Sud, Service d'hématologie biologique, Pierre-Bénite, France
2. Cancer Research Center of Lyon, INSERM U1052 UMR CNRS 5286, Equipe labellisée Ligue Contre le Cancer, Université de Lyon, Lyon, France
3. Hospices Civils de Lyon, Plateforme de séquençage NGS - HCL, cellule bioinformatique, Bron,France
4. Université de Lyon, Université Lyon 1, CNRS, Laboratoire de Biométrie et Biologie Evolutive UMR 5558, Lyon,France
5. Hospices Civils de Lyon, Hôpital Lyon Sud, Service d'hématologie clinique, Pierre-Bénite, France
6. Hematology department and INSERM U1052, CRCL, Centre Léon Bérard, Lyon, France
7. French group of CML, Centre Léon Bérard, Lyon, France

Corresponding author:
Dr. Pierre Sujobert
pierre.sujobert@chu-lyon.fr

Service d'hématologie biologie, Hôpital Lyon Sud

165 chemin du grand revoyet. 69365 Pierre Bénite

phone: +33(0)6 64 77 59 12

fax: +33(0)478864104

Running title: Targeted RNA-seq in haematology

Number of words: Abstract: 184, main text: 2356

Number of Tables: 0

Number of Figures: 4

Supplemental Tables: 3

Conflict of interest disclosure: the authors do not have any conflict of interest to disclose

Abstract

RNA sequencing holds great promise to improve the diagnostic of haematological malignancies, because this technique enables to detect fusion transcripts, to look for somatic mutations in oncogenes, and to capture transcriptomic signatures nosological entities. However, the analytical performances of targeted RNA sequencing have not been extensively described in diagnostic samples. Using a targeted panel of 1385 cancer-related genes in a series of 100 diagnosis samples and 8 controls, we detected all the already known fusion transcripts, and also discovered unknown and/or unsuspected fusion transcripts in 12 samples. Regarding the analysis of transcriptomic profiles, we show that targeted RNA sequencing is performant to discriminate acute lymphoblastic leukemia entities driven by different oncogenic translocations. Additionally, we show that 86% of the mutations identified at the DNA level are also detectable at the mRNA level, except for nonsense mutations which are subjected to mRNA decay. We conclude that targeted RNA sequencing might improve the diagnosis of haematological malignancies. Standardization of the preanalytical steps and further refinements of the panel design and of the bioinformatical pipelines will be an important step towards its use in standard diagnostic procedures.

Introduction

In haematological malignancies as in cancer in general, the goal of the diagnosis procedures is not only to accurately classify the patient's disease according to the consensual WHO guidelines[1], but also to identify biomarkers of prognostic or predictive values. A part of this information can be captured by morphology and immunophenotyping, but it relies more and more on the analysis of the genomic alterations of the neoplastic cells[2]. Nowadays, conventional cytogenetics and targeted sequencing of relevant genes are still the standard procedures. However, technological outbreaks such as whole genome sequencing, ATAC-sequencing or RNA sequencing (RNA-seq) might refine the diagnosis by unravelling genomic alterations outside coding regions[3], epigenetic signatures[4,5] and gene expression profiles respectively[6].

In this study, we have chosen to assess the diagnostic value of RNA-seq, because this technique allows to explore three levels of genetic information: gene sequence, gene fusions, and gene expression. Interestingly, each of these different levels of analysis brings independent information about the neoplastic cell, and accordingly their integration should refine the precision of the diagnosis. For example, acute myeloid leukaemia (AML) patients prognosis is evaluated by cytogenetics (copy number abnormalities and structural variants), further refined by the analysis of the mutational status of a few genes, and could maybe be improved by transcriptomic signatures such as the LSC17 which is a proxy of the number of leukemic stem cells[7].

Different techniques of library preparation for RNA-seq have been described, enabling the analysis of all the RNA molecules of a sample, or using enrichment step to target genes of interest such as messenger RNA, or small RNA species. Of note, the choice of the

library preparation should optimize the balance between the number of targets of interest and the required depth of sequencing, in order to remain economically affordable in a routine setting. To date, most of the genes involved in cancer have been already identified by large programs of whole exome sequencing[8]. Based on these considerations, we have decided to evaluate the performances of a targeted RNA-seq panel of 1385 genes involved in cancer biology. We present here the analytical performances of targeted RNA-seq to detect fusion transcripts, to identify transcriptional profiles associated with clinically relevant entities, and to detect the recurrent mutations with clinical significance in haematological malignancies.

Material and methods

Samples

One hundred diagnosis samples from patients with the following haematological malignancies were included: acute leukaemia (acute myeloid leukaemia (AML), n=51 including 7 acute promyelocytic leukaemias (APL), B-cell acute lymphoblastic leukaemia (B-ALL, n=27), mixed phenotype acute leukaemia (n=1), T-cell acute lymphoblastic leukaemia (T-ALL, n=1)), myeloproliferative neoplasms (chronic myeloid leukaemia (CML, n=12) and other myeloproliferative neoplasms (n=2)), hypereosinophilic syndromes (HES, n=3)), chronic myelomonocytic leukemia (CMML, n=2) and myelodysplastic syndrome with multilineage dysplasia (n=1). These samples were chosen to enrich the cohort in fusion transcript due to chromosomal translocations, based on the results of conventional cytogenetics, in order to test the performances of targeted RNA-seq to detect fusion transcripts. Moreover, we used four controls (C1 to C4) prepared by pooling blood samples from 5 healthy donors for each, and four bone marrow samples from healthy donors. The characteristics of the samples are provided in supplemental Table 1. The procedures followed were in accordance with the Helsinki declaration, as revised in 2008.

Cytogenetic R and G-banding analyses were performed according to standard methods. The definition of a cytogenetic clone and description of karyotypes followed the current International System for Human Cytogenetic Nomenclature.

For a subset of samples (n=45), the analysis of a panel of 105 genes was already performed for routine diagnostic procedures, as already described[9].

RNA extraction

Three different protocols of RNA extraction were used (supplemental Table 1). For ALL samples and 3 bone marrow samples from healthy donors, RNA was extracted with

Trizol reagent (TRIZ: Invitrogen, Carlsbad, CA, USA). For AML samples RNA was extracted with NucleoSpin RNA kit (MN: Macherey Nagel, Düren, Germany). For CML, HES, CMML and MDS samples, RNA was extracted with MN or the Maxwell 16LEV simplyRNA Blood Kit (Max: Promega, Madison, WI, USA). For control samples C1 to C4, RNA was extracted after Ficoll enrichment with either Trizol or MN methods, in order to assess the effect of extraction protocol on transcriptomic analysis performances.

RNA sequencing

Library preparation for NGS sequencing was performed using TruSight RNA Pan-Cancer Panel (Illumina, SanDiego, CA, USA) targeting 1385 genes involved in cancer biology. Libraries from 16 samples were multiplexed and sequenced on a Nextseq 500 device with a 2x81 paired-end run on a mid-output flowcell according to the manufacturer's instructions (mean number of reads by sample: 32.10^6; range 20 to 59.10^6).

Bioinformatical analysis

After demultiplexing, adapter sequences were trimmed with Cutadapt and reads were mapped to the human genome (GRCh37). The detection of gene fusions was performed first with the commonly used STAR-Fusion pipeline (parameters: --FusionInspector validate) and STAR-2pass[10], and all the negative samples were reanalysed with the recently launched nf-Core[11] and Arriba (https://github.com/suhrig/arriba/) pipelines. Putative fusions were validated by reverse transcription and polymerase chain reaction (primers sequences are provided in supplemental Table 2). Gene expression analysis (after trimmed mean of M values (TMM) normalization), principal Component Analysis (PCA), k-means clustering, two tailed t-test and Heat Map generation followed by hierarchical clustering were performed using Omics Explorer software (Qlucore AB, Lund, Sweden). For gene mutation analysis on RNAseq data, we looked at all the mutations found at the DNA level by combining the same homemade workflow as for

144 DNA, and visual inspection of the BAM files in case of unfound mutation. In brief we first
145 gather the variant alleles called with Freebayes and Varscan2[12,13]. Among this raw set,
146 we kept alleles whose read frequency was either above 20%, or for those below, if their
147 frequency was more than 5 fold the median of the frequencies of all the samples from
148 the same run. A second filtering step was applied to get rid of variants whose occurrence
149 was above 1% in GnomAD mixed populations[14].
150
151
152

Results

Identification of fusion transcripts

Fusion transcript positivity threshold was determined by the detection of at least one junction read and one spanning read between two different genes. All putative new fusion transcripts have been validated by PCR. All of the 57 rearrangements identified by cytogenetics or molecular biology were identified by targeted RNA-seq (Figure 1). Notably, RNA-seq detected all the *BCR-ABL1* canonical and rare transcripts (e13a2 (n=2); e14a2 (n=5); e1a2 (n=4); e1a3 (n=2); e6a2 (n=1); e13a3 (n=3); and e19a2(n=2)), as well as all the *PML-RARA* transcripts (BCR1 (n=2) BCR2 (n=2); BCR3(n=3)) and *MLL (KMT2A)* fusions (n=19). Of note, two samples with *FIP1L1-PDGFRA* fusion transcripts and one with *KMT2A* duplication were missed when analysed with the STAR-fusion pipeline, but recovered with nf-core and Arriba bioinformatics pipelines.

Eighteen samples had a chromosomal translocation without detected fusion transcript based on routine molecular biology. In 5 patients, targeted RNA-seq identified a fusion transcript already described in the literature (*KAT6-CREBP, NPM1-MLF1, PCM1-JAK2, DEK-NUP214, ZMYND11-MBTD1*[15]) (Figure 1). In two additional patients, a new fusion transcript was identified and confirmed by RT-PCR and Sanger sequencing (*FUS-FEV; EEA1-PDGFRB*). These fusion transcript were in frame, probably leading to the expression of an abnormal fusion protein (Figure 2A and B). No evidence of fusion transcript was found in the 11 remaining samples.

Finally, we detected a fusion transcript in five samples without detectable translocation on conventional cytogenetics: *SET-NUP214, EP300-ZNF384, KMT2A-MLLT4, KMT2A-MLLT10, VWC2-IKZF1* (Figure 1). The *VWC2-IKZF1* fusion transcript (Figure 2C), never described so far, was detected in an acute lymphoblastic leukemia with a t(9;22) leading to the expression of the *BCR-ABL1*-transcript (patient 10, supplemental data Table 1).

We hypothesize that this fusion might represent a new mechanism of *IKZF1* gene inactivation recurrently identified in Phi+-ALL[16,17].

As it was previously described in non-cancer tissues and cells[18], several fusions with open reading frame were also detected in control and patients' samples. Some of them such as *TFG-GPR128*, *POLE-FUS* or *OAZ1-DOT1* were expressed at high level and have been also validated by RT-PCR and sequencing.

Finally, in order to assess the sensitivity threshold of RNA-seq to detect fusion transcripts, we analysed serial dilutions of two patients with *PML-RARA* and *BCR-ABL* fusion transcripts respectively. The detection threshold was below 6% for both fusion transcripts.

Transcriptome analysis

The analysis of transcriptome in the routine diagnosis procedure is technically challenging, because of interferences linked to the source of the samples analysed (e.g. bone marrow vs peripheral blood), the preparation of the samples (isolation of the mononucleated cells with Ficoll or not), the RNA extraction method, and the batch effect of library preparation and sequencing. Instead were developed signatures based on a limited number of transcripts analysed with technical platforms such as RT-MLPA[19] or Nanostring technology[7,20]. Here we assessed the feasibility of transcriptome analysis based on RNA-seq of a panel of 1385 genes.

First, we evaluated the magnitude of systematic biases in transcriptomic analysis introduced by the protocol of RNA extraction and the sequencing process. The same blood samples from healthy donors were extracted after Ficoll enrichment either with Trizol (n=4), or with Macherey Nagel kits (n=4). A supervised analysis based on extraction method identified 20 differentially expressed genes (fold change threshold 2, false discovery rate q<0,05) (Figure 3A). On the contrary, when we compared the

transcriptome of RNA extracted from blood samples from healthy donors, whose RNA-seq libraries and sequencing were not prepared and run the same day, there was no gene differentially expressed according to the batch of library preparation or sequencing (fold change threshold 2, false discovery rate q<0,05, data not shown).

Then, we analysed bone marrow samples extracted with the same method (Trizol) from three groups with at least 3 patients: ALL with *KMT2A-AFF1* (n=7), ALL with *TCF3-PBX1* (n=4), and normal bone marrow controls (n=3). Of note, these RNA were extracted at the time of diagnosis, over a period of 19 years, introducing a potential bias due to differences in RNA conservation. Clustering of these samples in three categories (by the k-means method) distinguishes the three groups of samples according to the diagnosis, with no misclassification (Figure 3B). The analysis of the 50 most differentially expressed genes between control and both type of ALL confirmed previously described features such as *HOXA3, HOXA9, HOXA10* and *FLT3* overexpression in *KMT2A-AFF1*[21] and *CD19, WNT16* and *PBX1* up-regulation in *TCF3-PBX1* (Figure 3C)[22].

Detection of gene mutations.

Forty five patients analysed with targeted RNA-seq were also analysed at the DNA level for a panel of 105 genes recurrently mutated in haematological malignancies[9]. Among the 95 genes captured in both panels, 122 mutations were detected at the DNA level in 39 different genes. As shown in Figure 4, 106/122 (87%) mutations identified at the DNA level were also found in the RNA seq data. Among the 16 mutations missed at the mRNA level, frameshift mutations were overrepresented (missed mutations 11/16 vs 12/106, Fischer's exact test p<0.0001). Two other missed mutations (I1897T and G218V from *TET2* and *U2AF1*, respectively) were in low coverage areas (<30x).

Discussion

This work reports the analytical performances of RNA-sequencing of a panel of 1385 genes to improve the diagnosis of haematological malignancies, based on a series of 100 diagnosis samples and 8 controls.

Overall, this technique detect 100% of fusion transcripts of these samples, including *FIP1L1-PDGFRA* fusions which often require nested PCR to be identified because of low levels of expression[23]. Of note, two fusion transcripts were found only by alternative bioinformatics pipelines, which highlights the major impact of the bioinformatics analysis on the performances of targeted RNA-seq. This might explain suboptimal detection of *KMT2A* and *PDGFRA* fusions in previous studies[24]. Interestingly, RNA-seq allowed the identification of 12 fusion transcripts which were not suspected with usual analysis recommended in the diagnosis of haematological malignancies[25]. As more and more case reports describe successful opportunistic use of targeted therapies in patients with fusion transcripts[26–28], the identification of unexpected fusion transcripts might offer interesting targets in relapsed/refractory patients. In the series reported here, the patient with an hypereosinophilic syndrome associated with the novel *EEA1-PDGFRB* fusion was treated with imatinib and achieved a durable haematological remission. Moreover, as translocations are most of the time drivers events which are stable during disease evolution[30], they can be used to track minimal residual disease with high sensitivity RT-qPCR, and adapt therapeutic intensity accordingly. However, it remains to be determined if the prognostic impact of minimal residual disease described for CBF AML[31], CML, or APL is also true for the less recurrent fusion transcripts.

Regarding the analysis of the transcriptomic profile, we show that targeted transcriptome analysis can be used for nosological purposes if the pre-analytical workflow is the same for the samples analysed. Larger series are needed to precise the

253 performances of targeted RNA-seq to resolve this task. Another interesting question
254 would be to assess the performances of targeted RNA-seq to measure clinically relevant
255 signatures such as the LSC17[7] or the more recently described pLSC6[32] signatures in
256 AML, but it will need an optimization of the design of the panel to capture all relevant
257 mRNAs.
258 The third clinical interest of targeted RNA-seq assessed here is the detection of acquired
259 somatic mutations. Even if most of the mutations identified at the DNA level were found
260 in RNA-seq data, the nonsense mutations were rarely detected. This is probably at least
261 in part due to the phenomenon of mRNA decay, which degrades preferentially truncated
262 mRNA[33], and this will remain a biological limitation of RNA-seq for mutation assessment.
263 Finally, given the growing importance of clonal architecture analysis based on variant
264 allele frequency (VAF) deconvolution[34], we should keep in mind that the VAF measured
265 at the mRNA level might not be good surrogate markers of clonal architecture, because it
266 takes into account allelic expression bias.
267 Altogether, RNA-seq of a targeted panel of genes might improve the diagnosis of
268 haematological malignancies, and highlight potential therapeutic targets. Some of the
269 limitations of this technique might be resolved with the optimization of the panel design
270 and the bioinformatics pipelines for haematological malignancies. However, because
271 some limitations have a biological explanation, such as poor performances to detect non-
272 sense mutations, RNA-seq should not replace the analysis of genomic DNA, but could be
273 rather a good orthogonal method for verifying genomic mutations and a powerful
274 complement to increase the molecular characterisation of haematologic malignancies at
275 diagnosis.
276
277

278 **Figure legends**

279 **Figure 1:** Description of the 72 fusion transcripts detected by targeted-RNAseq in the
280 whole cohort.

281

282 **Figure 2:** Description of the three new fusion transcripts discovered in this cohort.
283 Schematic representation of the 3 new fusion transcripts identified by targeted RNAseq :
284 *FUS-FEV* from t(2;16) (A), *EEA1-PDGFRB* from t(5;12) (B) and *VWC2-IKZF1* (C). For each
285 fusion transcript is provided a schematic representation of the translocation at the
286 genomic level, a graphical representation of the coverage depth in the targeted RNA-seq,
287 and a schematic representation of the protein fusion.

288

289 **Figure 3**: A: Heatmap representation of the 20 genes differentially expressed (fold
290 change >2, q value <0,05) between the same control blood samples after RNA extraction
291 with two different methods (trizol vs Macherey Nagel). B: Principal component analysis
292 and unsupervised k-means clustering of 14 samples processed with the same
293 preanalytical steps (*KMT2A-AFF1* ALL, white dots, n=7, *TCF3-PBX1* ALL, yellow dots,
294 n=4, normal bone marrow samples, purple dots, n=3). C: Heatmaps showing the 50 most
295 differentially expressed genes between normal samples and *KMT2A-AFF1* ALL samples
296 (left) and between normal samples and *TCF3-PBX1* ALL samples (right).
297 **Figure 4**: Performances of somatic mutations detection based on RNA-seq analysis. The
298 relative number of mutations correctly identified or undetected (non-sense, low
299 coverage or other) are presented.

300

301

302

Supplemental Table 1: Characteristics of the samples included in the study.

Supplemental Table 2: Sequences of the primers used to validate the fusion transcripts.

Supplemental Table 3: List of gene mutations identified by DNA sequencing which were analysed on RNA-seq data.

References

1. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, Advani R, Ghielmini M, Salles GA, Zelenetz AD, Jaffe ES. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood, 2016, 127:2375–90

2. Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, Dombret H, Ebert BL, Fenaux P, Larson RA, Levine RL, Lo-Coco F, Naoe T, Niederwieser D, Ossenkoppele GJ, Sanz M, Sierra J, Tallman MS, Tien H-F, Wei AH, Löwenberg B, Bloomfield CD. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood, 2017, 129:424–47

3. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature, 2020, 578:82–93

4. Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ, Majeti R, Chang HY. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet, 2016, 48:1193–203

5. Yi G, Wierenga ATJ, Petraglia F, Narang P, Janssen-Megens EM, Mandoli A, Merkel A, Berentsen K, Kim B, Matarese F, Singh AA, Habibi E, Prange KHM, Mulder AB, Jansen JH, Clarke L, Heath S, van der Reijden BA, Flicek P, Yaspo M-L, Gut I, Bock C, Schuringa JJ, Altucci L, Vellenga E, Stunnenberg HG, Martens JHA. Chromatin-Based Classification of Genetically Heterogeneous AMLs into Two Distinct Subtypes with Diverse Stemness Phenotypes. Cell Reports, 2019, 26:1059-1069.e6

6. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature, 2000, 403:503–11

7. Ng SWK, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, Arruda A, Popescu A, Gupta V, Schimmer AD, Schuh AC, Yee KW, Bullinger L, Herold T, Görlich D, Büchner T, Hiddemann W, Berdel WE, Wörmann B, Cheok M, Preudhomme C, Dombret H, Metzeler K, Buske C, Löwenberg B, Valk PJM, Zandstra PW, Minden MD, Dick JE, Wang JCY. A 17-gene stemness score for rapid determination of risk in acute leukaemia.

Nature, 2016, 540:433–7

8. Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Børresen-Dale A-L, Boyault S, Burkhardt B, Butler AP, Caldas C, Davies HR, Desmedt C, Eils R, Eyfjörd JE, Foekens JA, Greaves M, Hosoda F, Hutter B, Ilicic T, Imbeaud S, Imielinski M, Jäger N, Jones DTW, Jones D, Knappskog S, Kool M, Lakhani SR, et al. Signatures of mutational processes in human cancer. Nature, 2013, 500:415–21

9. Huet S, Paubelle E, Lours C, Grange B, Courtois L, Chabane K, Charlot C, Mosnier I, Simonet T, Hayette S, Tigaud I, Thomas X, Salles G, Subtil F, Sujobert P. Validation of the prognostic value of the knowledge bank approach to determine AML prognosis in real life. Blood, 2018, 132:865–7

10. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol, 2019, 20:213

11. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol, 2020. https://doi.org/10.1038/s41587-020-0439-x

12. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. ArXiv:12073907 [q-Bio], 2012

13. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res, 2012, 22:568–76

14. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. Genomics, 2019. https://doi.org/10.1101/531210

15. Plesa A, Sujobert P. Cannibalistic acute myeloid leukemia with ZMYND11-MBTD1 fusion. Blood, 2019, 133:1789–1789

16. Martinelli G, Iacobucci I, Storlazzi CT, Vignetti M, Paoloni F, Cilloni D, Soverini S, Vitale A, Chiaretti S, Cimino G, Papayannidis C, Paolini S, Elia L, Fazi P, Meloni G, Amadori S, Saglio G, Pane F, Baccarani M, Foà R. *IKZF1* (Ikaros) Deletions in *BCR-ABL1* –Positive Acute Lymphoblastic Leukemia Are Associated With Short Disease-Free Survival and High Rate of Cumulative Incidence of Relapse: A GIMEMA AL WP Report. JCO, 2009, 27:5202–7

17. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, Ma J, Liu W, Cheng C, Schulman BA, Harvey RC, Chen I-M, Clifford RJ, Carroll WL, Reaman G, Bowman WP, Devidas M, Gerhard DS, Yang W, Relling MV, Shurtleff SA, Campana D, Borowitz MJ, Pui C-H, Smith M, Hunger SP, Willman CL, Downing JR. Deletion of *IKZF1* and Prognosis in Acute Lymphoblastic Leukemia. New England Journal of Medicine, 2009, 360:470–80

18. Babiceanu M, Qin F, Xie Z, Jia Y, Lopez K, Janus N, Facemire L, Kumar S, Pang Y, Qi Y, Lazar IM, Li H. Recurrent chimeric fusion RNAs in non-cancer tissues and cells. Nucleic Acids Res, 2016, 44:2859–72

19. Mareschal S, Ruminy P, Bagacean C, Marchand V, Cornic M, Jais J-P, Figeac M, Picquenot J-M, Molina TJ, Fest T, Salles G, Haioun C, Leroy K, Tilly H, Jardin F. Accurate Classification of Germinal Center B-Cell–Like/Activated B-Cell–Like Diffuse Large B-Cell Lymphoma Using a Simple and Rapid Reverse Transcriptase–Multiplex Ligation-Dependent Probe Amplification Assay. The Journal of Molecular Diagnostics, 2015, 17:273–83

20. Huet S, Tesson B, Jais J-P, Feldman AL, Magnano L, Thomas E, Traverse-Glehen A, Albaud B, Carrère M, Xerri L, Ansell SM, Baseggio L, Reyes C, Tarte K, Boyault S, Haioun C, Link BK, Feugier P, Lopez-Guillermo A, Tilly H, Brice P, Hayette S, Jardin F, Offner F, Sujobert P, Gentien D, Viari A, Campo E, Cerhan JR, Salles G. A gene-expression profiling score for prediction of outcome in patients with follicular lymphoma: a retrospective training and validation analysis in three international cohorts. The Lancet Oncology, 2018, 19:549–61

21. Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR, Korsmeyer SJ. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet, 2002, 30:41–7

22. Diakos C, Xiao Y, Zheng S, Kager L, Dworzak M, Wiemels JL. Direct and Indirect Targets of the E2A-PBX1 Leukemia-Specific Fusion Protein. PLoS ONE, 2014, 9:e87602

23. and The French Eosinophil Network, Roche-Lestienne C, Lepers S, Soenen-Cornu V, Kahn J-E, Laï J-L, Hachulla E, Drupt F, Demarty A-L, Roumier A-S, Gardembas M, Dib M, Philippe N, Cambier N, Barete S, Libersa C, Bletry O, Hatron P-Y, Quesnel B, Rose C, Maloum K, Blanchet O, Fenaux P, Prin L, Preudhomme C. Molecular characterization of the idiopathic hypereosinophilic syndrome (HES) in 35 French patients with normal conventional cytogenetics. Leukemia, 2005, 19:792–8

24. Stengel A, Nadarajah N, Haferlach T, Dicker F, Kern W, Meggendorfer M, Haferlach C. Detection of recurrent and of novel fusion transcripts in myeloid malignancies by targeted RNA sequencing. Leukemia, 2018, 32:1229–38

25. SH S, E C, NL H, ES J, SA P, H S, J T. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. n.d.

26. Jan M, Grinshpun DE, Villalba JA, Dal Cin P, Sykes DB, Iafrate AJ, Ebert BL, Hobbs GS, Nardi V. A cryptic imatinib-sensitive G3BP1-PDGFRB rearrangement in a myeloid neoplasm with eosinophilia. Blood Advances, 2020, 4:445–8

27. Tanasi I, Ba I, Sirvent N, Braun T, Cuccuini W, Ballerini P, Duployez N, Tanguy-Schmidt A, Tamburini J, Maury S, Doré E, Himberlin C, Duclos C, Chevallier P, Rousselot P, Bonifacio M, Cavé H, Baruchel A, Dombret H, Soulier J, Landman-Parker J, Boissel N, Clappier E. Efficacy of tyrosine kinase inhibitors in Ph-like acute lymphoblastic leukemia harboring ABL-class rearrangements. Blood, 2019, 134:1351–5

28. Decool G, Domenech C, Grardel N, Plesa A, Raczkiewicz I, Ducourneau B, Ruminy P, Pages M-P, Girard S, Fenwarth L, Preudhomme C, Bertrand Y, Duployez N. Efficacy of Tyrosine Kinase Inhibitor Therapy in a Chemotherapy-refractory B-cell Precursor Acute Lymphoblastic Leukemia With ZC3HAV1-ABL2 Fusion: HemaSphere, 2019, 3:e193

29. Panagopoulos I, Brunetti M, Stoltenberg M, Strandabø RAU, Staurseth J, Andersen K, Kostolomov I, Hveem TS, Lorenz S, Nystad TA, Flægstad T, Micci F, Heim S. Novel GTF2I–PDGFRB and IKZF1–TYW1 fusions in pediatric leukemia with normal karyotype. Exp Hematol Oncol, 2019, 8:12

30. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N, Gundem G, Van Loo P, Martincorena I, Ganly P, Mudie L, McLaren S, O'Meara S, Raine K, Jones DR, Teague JW, Butler AP, Greaves MF, Ganser A, Döhner K, Schlenk RF, Döhner H, Campbell PJ. Genomic Classification and Prognosis in Acute Myeloid Leukemia. New England Journal of Medicine, 2016, 374:2209–21

31. Jourdan E, Boissel N, Chevret S, Delabesse E, Renneville A, Cornillet P, Blanchet O, Cayuela J-M, Recher C, Raffoux E, Delaunay J, Pigneux A, Bulabois C-E, Berthon C, Pautas C, Vey N, Lioure B, Thomas X, Luquet I, Terré C, Guardiola P, Béné MC, Preudhomme C, Ifrah N, Dombret H. Prospective evaluation of gene mutations and minimal residual disease in patients with core binding factor acute myeloid leukemia. Blood, 2013, 121:2213–23

32. Elsayed AH, Rafiee R, Cao X, Raimondi S, Downing JR, Ribeiro R, Fan Y, Gruber TA, Baker S, Klco J, Rubnitz JE, Pounds S, Lamba JK. A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid leukemia. Leukemia, 2020, 34:735–45

33. Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol, 2004, 5:89–99

34. Itzykson R, Duployez N, Fasan A, Decool G, Marceau-Renaut A, Meggendorfer M, Jourdan E, Petit A, Lapillonne H, Micol J-B, Cornillet-Lefebvre P, Ifrah N, Leverger G, Dombret H, Boissel N, Haferlach T, Preudhomme C. Clonal interference of signaling mutations worsens prognosis in core-binding factor acute myeloid leukemia. Blood, 2018, 132:187–96

Figure 1

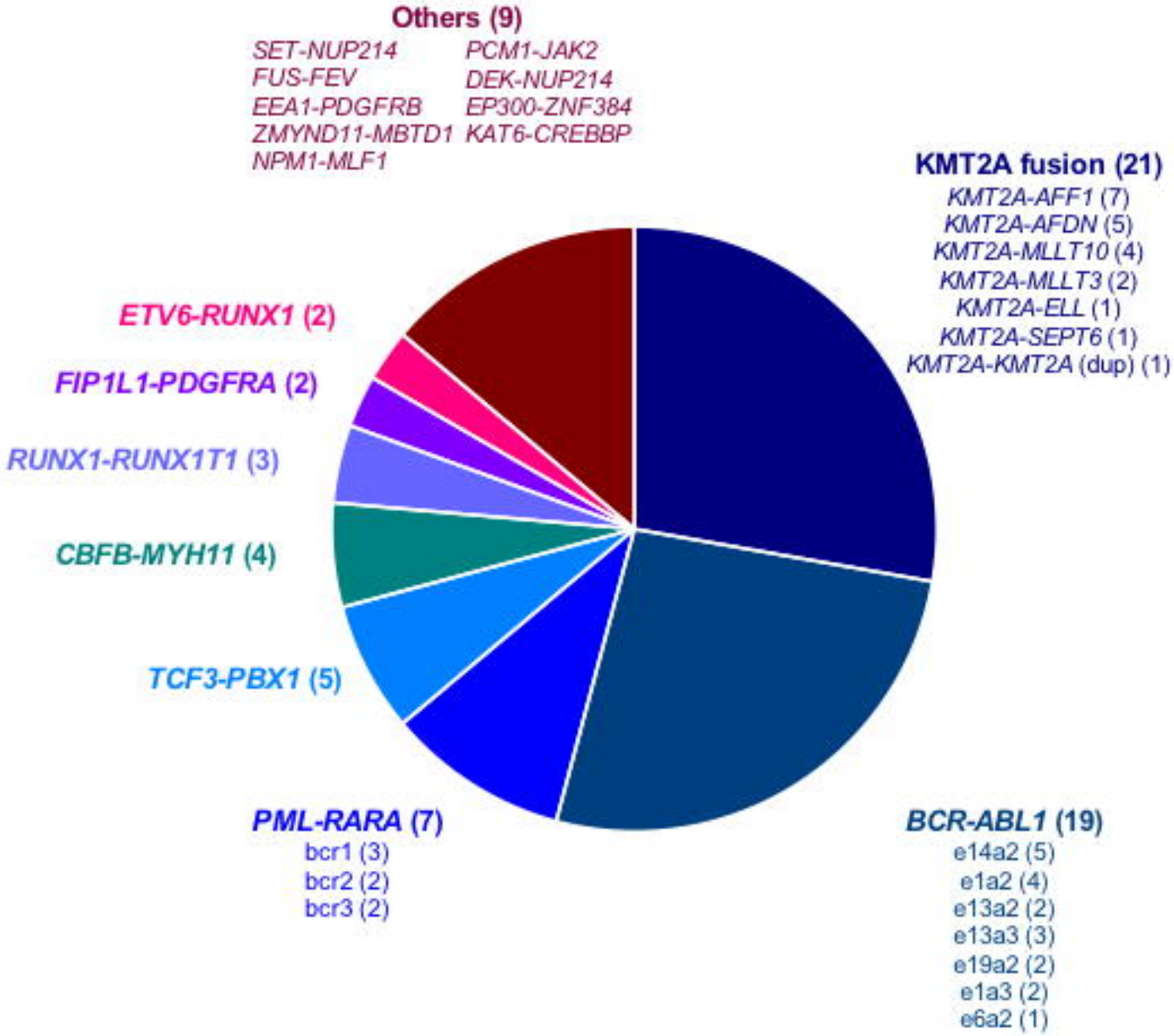
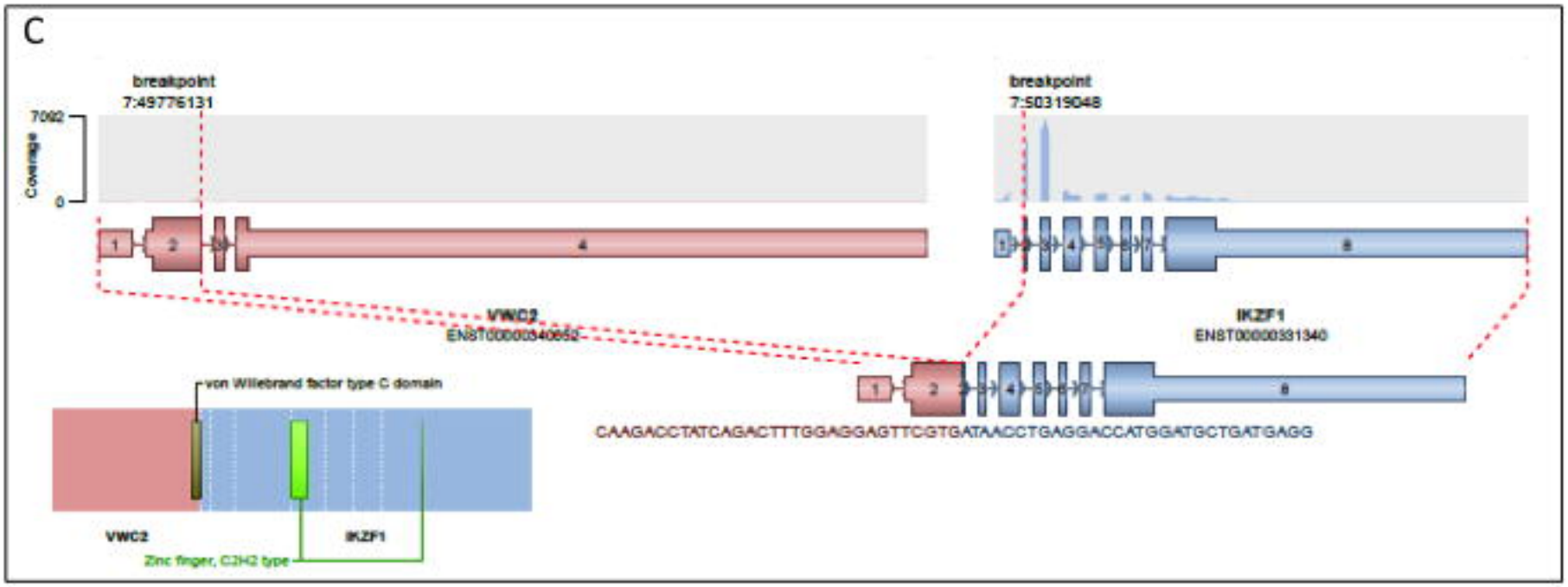

Others (9)
SET-NUP214
FUS-FEV
EEA1-PDGFRB
ZMYND11-MBTD1
NPM1-MLF1
PCM1-JAK2
DEK-NUP214
EP300-ZNF384
KAT6-CREBBP
KMT2A fusion (21)
KMT2A-AFF1 (7)
KMT2A-AFDN (5)
KMT2A-MLLT10 (4)
KMT2A-MLLT3 (2)
KMT2A-ELL (1)
KMT2A-SEPT6 (1)
KMT2A-KMT2A (dup) (1)
ETV6-RUNX1 (2)
FIP1L1-PDGFRA (2)
RUNX1-RUNX1T1 (3)
CBFB-MYH11 (4)
TCF3-PBX1 (5)
PML-RARA (7)
bcr1 (3)
bcr2 (2)
bcr3 (2)
BCR-ABL1 (19)
e14a2 (5)
e1a2 (4)
e13a2 (2)
e13a3 (3)
e19a2 (2)
e1a3 (2)
e6a2 (1)

Figure 2 continued



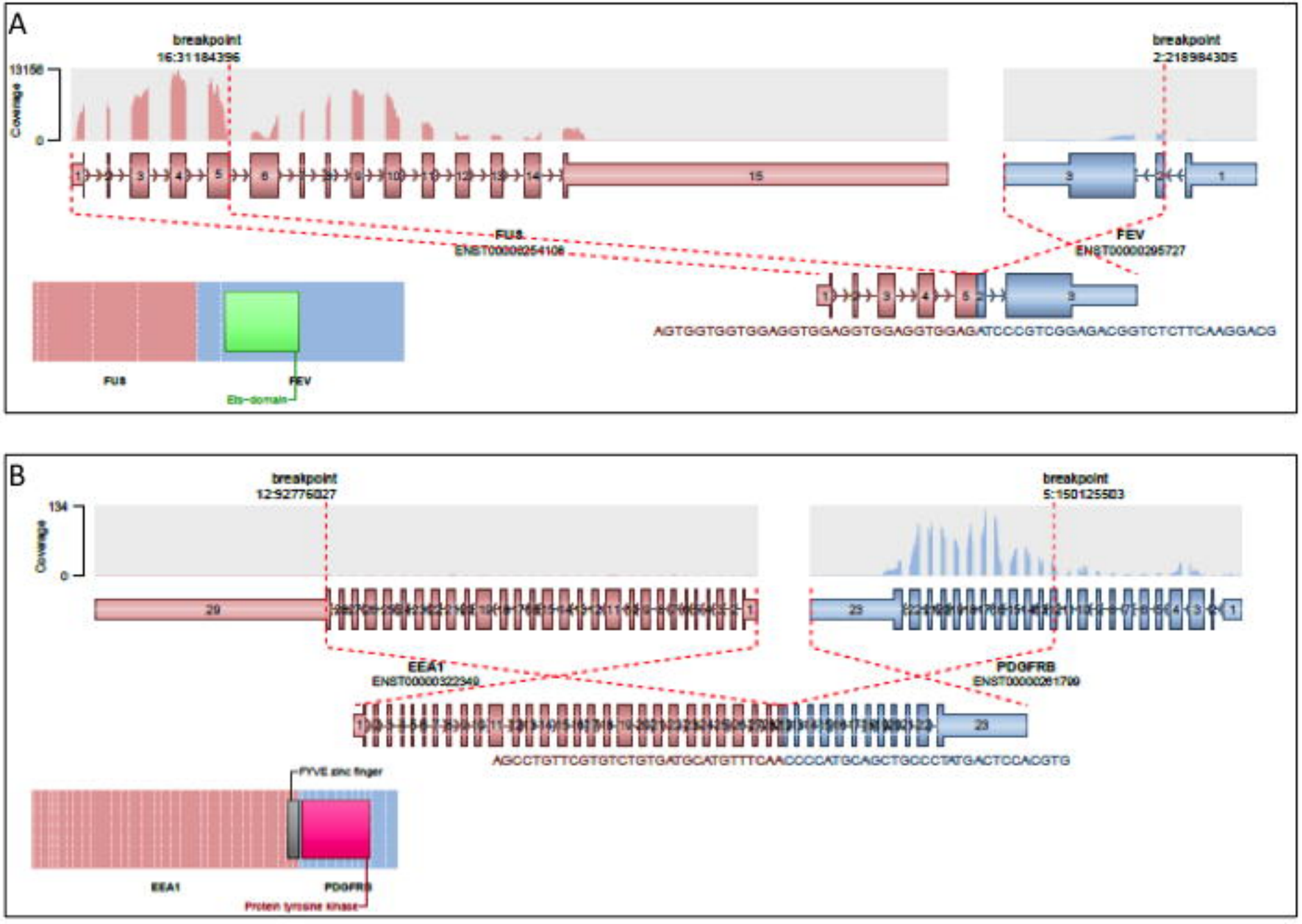
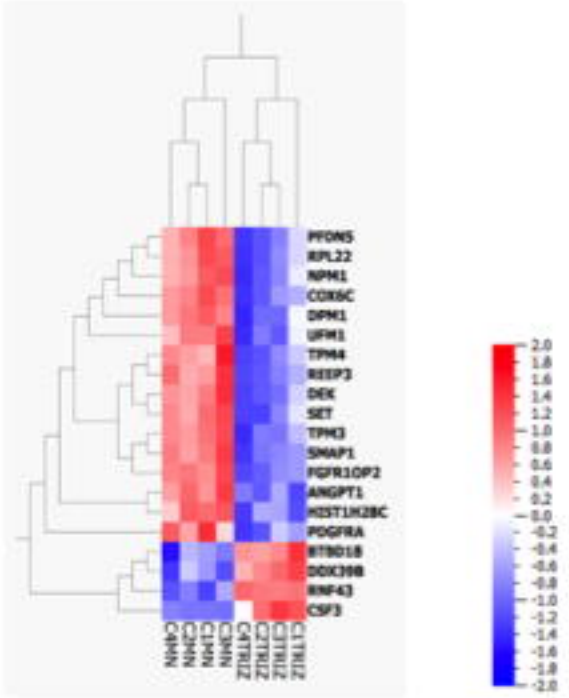
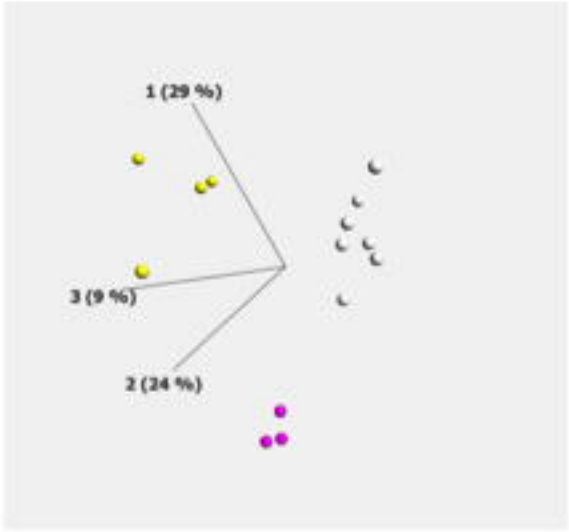

A
breakpoint
16:31184396
breakpoint
2:218984305
Coverage
13156
0
FUS
ENST00000254108
FEV
ENST00000295727
AGTGGTGGTGGAGGTGGAGGTGGAGGTGGAGATCCCGTCGGAGACGGTCTCTTCAAGGACG
FUS
FEV
Ets-domain
B
breakpoint
12:92776827
breakpoint
5:150125503
Coverage
134
0
EEA1
ENST00000322349
PDGFRB
ENST00000261700
AGCCTGTTCGTGTCTGTGATGCATGTTTCAACCCCATGCAGCTGCCCTATGACTCCACGTG
FYVE zinc finger
EEA1
PDGFRB
Protein tyrosine kinase

Figure 3

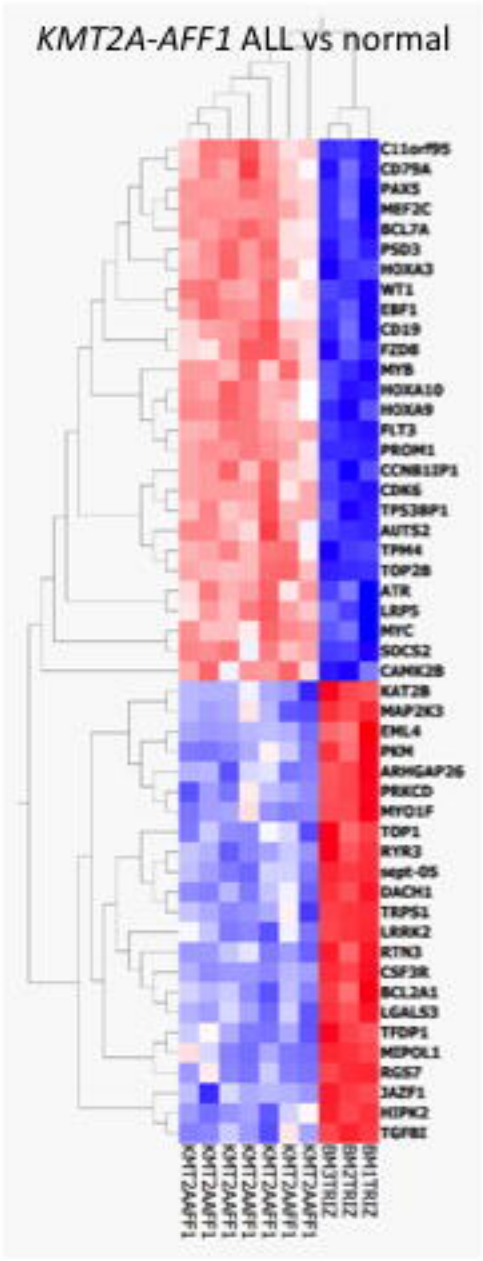
A


PFDN5
RPL22
NPM1
COX6C
DPM1
UFM1
TPM4
REEP3
DEK
SET
TPM3
SMAP1
FGFR1OP2
ANGPT1
H2ST1H2BC
PDGFRA
BTBD18
DDX39B
RNF43
CSF3
C4HN
C2HN
C3HN
C1HN
C5HN
C4TRIZ
C2TRIZ
C3TRIZ
C1TRIZ
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
-1.6
-1.8
-2.0

B


1 (29 %)
3 (9 %)
2 (24 %)

C


KMT2A-AFF1 ALL vs normal
C11orf95
CD79A
PAX5
MEF2C
BCL7A
PSD3
HOXA3
WT1
EBF1
CD19
FZD6
MYB
HOXA10
HOXA9
FLT3
PROM1
CCNB1IP1
CDK6
TP53BP1
AUTS2
TPM4
TOP2B
ATR
LRP5
MYC
SOCS2
CAMK2B
KAT2B
MAP2K3
EML4
PKM
ARHGAP26
PRKCD
MYO1F
TOP1
RYR3
sept-05
DACH1
TRPS1
LRRK2
RTN3
CSF3R
BCL2A1
LGALS3
TFDP1
MIPOL1
RGS7
JAZF1
HIPK2
TGFBI
KMT2AAFF1
KMT2AAFF1
KMT2AAFF1
KMT2AAFF1
KMT2AAFF1
KMT2AAFF1
KMT2AAFF1
BM3TRIZ
BM2TRIZ
BM1TRIZ

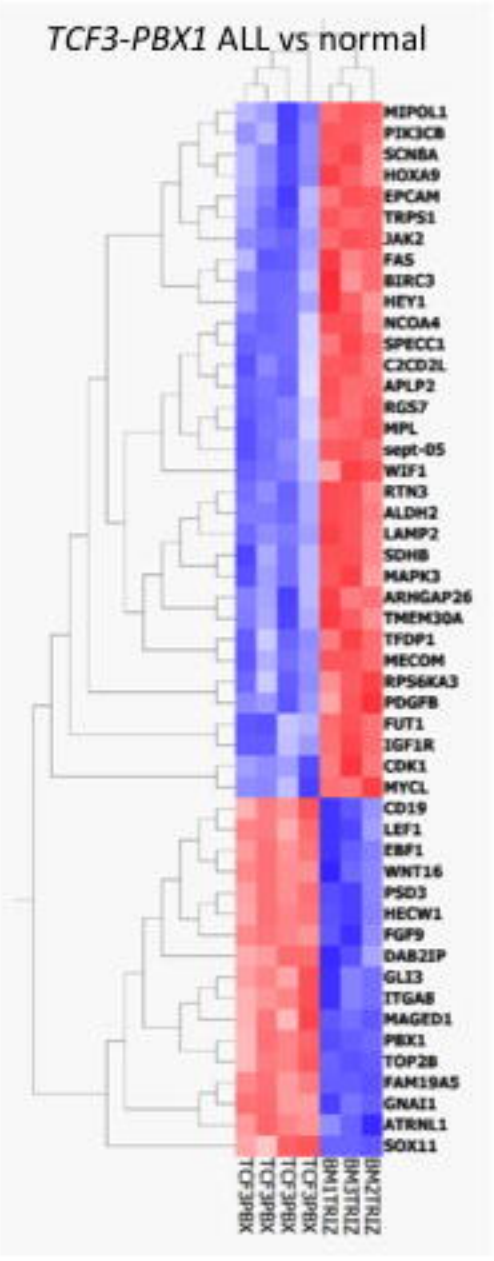
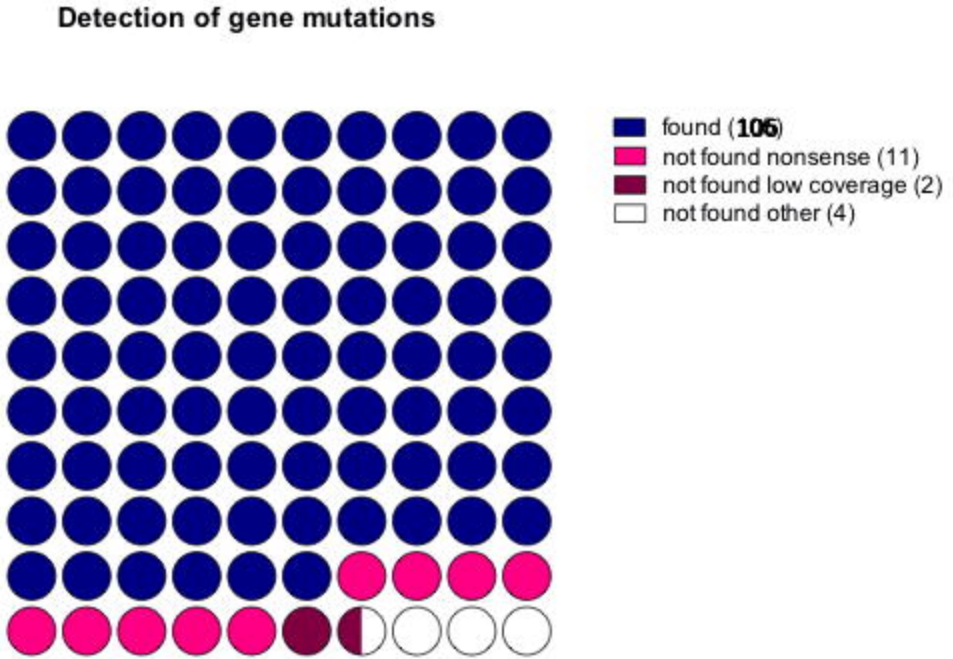

TCF3-PBX1 ALL vs normal
MIPOL1
PIK3CB
SCNBA
HOXA9
EPCAM
TRPS1
JAK2
FAS
BIRC3
HEY1
NCOA4
SPECC1
C2CD2L
APLP2
RGS7
MPL
sept-05
WIF1
RTN3
ALDH2
LAMP2
SDHB
MAPK3
ARHGAP26
TMEM30A
TFDP1
MECOM
RPS6KA3
PDGFB
FUT1
IGF1R
CDK1
MYCL
CD19
LEF1
EBF1
WNT16
PSD3
HECW1
FGF9
DAB2IP
GLI3
ITGA8
MAGED1
PBX1
TOP2B
FAM19A5
GNAI1
ATRNL1
SOX11
TCF3PBX
TCF3PBX
TCF3PBX
TCF3PBX
TCF3PBX
BM1TRIZ
BM3TRIZ
BM2TRIZ

Figure 4


Detection of gene mutations
found (106)
not found nonsense (11)
not found low coverage (2)
not found other (4)