

Engineered model of t(7;12)(q36;p13) AML recapitulates patient-specific features and gene expression profiles

Running title: CRISPR/Cas9-engineered model of t(7;12) leukaemia

**Denise Ragusa^{1,2,3}, Ylenia Cicirò^{1,4}, Concetta Federico⁵, Salvatore Saccone⁵,
Francesca Bruno⁵, Reza Saeedi¹, Cristina Sisu¹, Cristina Pina^{*1,2}, Arturo Sala^{*1,4},
Sabrina Tosi^{*1,2,3}**

¹ College of Health, Medicine and Life Sciences, Division of Biosciences, Brunel University London, UB8 3PH, Uxbridge, UK

² Genome Engineering and Maintenance Network (CenGEM), Institute of Environment, Health and Societies, Brunel University London, Kingston Lane, UB8 3PH, UK

³ Leukaemia and Chromosome Research Laboratory, College of Health, Medicine and Life Sciences, Brunel University London, Kingston Lane, Uxbridge, UB8 3PH, UK

⁴ Centre for Inflammation Research and Translational Medicine (CIRTM), Brunel University London, UB8 3PH, Uxbridge, UK

⁵ Department of Biological, Geological and Environmental Sciences, University of Catania, via Androne 81, 95124 Catania CT, Italy

*corresponding author email sabrina.tosi@brunel.ac.uk

Abstract

Acute myeloid leukaemia carrying the translocation t(7;12)(q36;p13) is an adverse-risk leukaemia uniquely observed in infants. Despite constituting up to 30% of cases in under 2-year-olds, it remains poorly understood. Known molecular features are ectopic overexpression of the *MNX1* gene and generation of a fusion transcript in 50% of patients. Lack of research models has hindered understanding of t(7;12) biology, which has historically focused on *MNX1* overexpression rather than the cytogenetic entity itself. Here, we employed CRISPR/Cas9 to generate t(7;12) in the human K562 cell line, and in healthy CD34+ haematopoietic progenitors where the translocation was not sustained in long-term cultures or through serial replating. In contrast, in K562 cells, t(7;12) was propagated in self-renewing clonogenic assays, with sustained myeloid bias in colony formation and baseline depletion of erythroid signatures. Nuclear localisation analysis revealed repositioning of the translocated *MNX1* locus to the interior of t(7;12)-harbouring K562 nuclei - a known phenomenon in t(7;12) patients which associates with ectopic overexpression of *MNX1*. Crucially, the K562-t(7;12) model successfully recapitulated the transcriptional landscape of t(7;12) patient leukaemia. In summary, we engineered a clinically-relevant model of t(7;12) acute myeloid leukaemia with the potential to unravel targetable molecular mechanisms of disease.

Introduction

The t(7;12)(q36;p13) translocation is a recurrent chromosomal rearrangement uniquely associated with infant acute myeloid leukaemia (AML) (1,2). T(7;12) is the second most common cytogenetic abnormality in AML patients below the age of 2 years (3), and associates with a dismal prognosis. Structurally, t(7;12) is a balanced translocation disrupting the long arm of chromosome 7 (7q36.3) and the short arm of chromosome 12 (12p13.2), producing two derivative chromosomes, der(7) and der(12). The 7q36.3 breakpoint lies proximal to the *MNX1* gene, which is entirely relocated to der(12). The breakpoint on 12p13, on the contrary, disrupts the *ETV6* gene on its 5' portion (4-6).

Mechanisms of leukaemogenesis associated with t(7;12) remain poorly understood. A *MNX1/ETV6* fusion mRNA transcript is produced from der(12) in approximately half of patients with t(7;12) (7-9), but a translated *MNX1/ETV6* protein has never been detected. Introduction of the chimaeric transcript did not transform mouse bone marrow cells, suggesting that it may not contribute to the leukaemic phenotype (10).

A common feature among all t(7;12) patients is overexpression of the *MNX1* gene (7,9,11), which is proposed to result from disruption of its genomic positioning (7). The majority of reports on t(7;12)-mediated leukaemogenesis have focused on *MNX1* overexpression (10,12,13). *In vitro*, overexpression of *MNX1* in mouse haematopoietic stem cells (HSCs) did not result in expansion or increased survival capacity (10), and was shown to induce senescence (13). In human cord blood CD34+ haematopoietic stem and progenitor cells (HSPC), *MNX1* overexpression induced erythroid transcriptional programmes and reduced colony-forming capacity. *In vivo*, transplantations of *MNX1* overexpressing bone marrow cells did not cause leukaemia in the recipients, but resulted in an accumulation of *MNX1*-overexpressing cells in the megakaryocytic-erythrocyte fraction, but not in the granulocytic-monocyte nor in the mature lymphoid compartments (13).

Genome editing technologies can be harnessed for generation of chromosomal abnormalities (14-16). The CRISPR/Cas9 gene editing tool employs the endonuclease activity of the Cas9 enzyme to induce DNA double stranded breaks (DSB) at specific loci directed by guide RNAs (gRNAs). The induction of two simultaneous DSBs can lead to erroneous break repair and rejoining of the 'wrong' DNA ends, thereby forming a chromosomal translocation (17,18). Here, we sought to recreate the t(7;12) rearrangement *in vitro* using CRISPR/Cas9 in the attempt to uncover biological cues to its mechanisms in AML.

Results and Discussion

By delivery of CRISPR/Cas9 ribonucleoprotein (RNP) complexes targeting clinically accurate breakpoints described by Tosi *et al.* (5) and Simmons *et al.* (4) (**Figure 1A; Supplementary Table 1**), we were able to achieve the t(7;12) rearrangement in the leukaemia cell line K562, as confirmed by fluorescence in situ hybridisation (FISH) (**Figure 1B-C; Supplementary Figure 2A-C**) and amplification of genomic fusion junctions by polymerase chain reaction (PCR) (**Figure 1D; Supplementary Figure 1**). While *MNX1* is already expressed in K562 cells (19), the translocation further upregulated its expression (**Figure 1E**), consistent with ectopic *MNX1* overexpression in patients (7,9,11). Altered nuclear positioning has been proposed as a

significant mechanism of *MX1* overexpression in t(7;12) leukaemia, with relocalisation of the der(12) containing the translocated *MX1* gene to a more internal transcriptionally active position within the nucleus, where expression of *MX1* can be triggered (7). We used nuclear localisation analysis to understand whether this phenomenon was recapitulated in the K562 model. FISH using a custom 3-colour probe allowed to distinguish both derivatives in interphase nuclei (**Figure 1F; Supplementary Figure 2D-F**) in order to calculate radial nuclear location (RNL) values for each locus with respect to their positioning towards the interior or peripheral nuclear space (**Figure 1F-G**). As described in patients, the introduction of t(7;12) in K562 cells resulted in the altered positioning of the der(12) towards the nuclear interior, consistent with the relative gain in *MX1* expression (**Figure 1B**). Conversely, the der(7) containing the 5' of *ETV6* was repositioned towards the periphery (**Figure 1G**), also in agreement with the patient analysis by Ballabio *et al.* (7). Nuclear organisation as a mechanism of regulation of *MX1* expression was also identified in leukaemias harbouring different interstitial deletions of chromosome 7q (20). Specifically, the localisation of *MX1* was shown to be dependent on the GC-content of the chromosomal band harbouring the proximal breakpoint, with *MX1* being expressed when breakpoints affected GC-rich bands, which resulted in relocation of *MX1* towards the nuclear interior; in contrast, GC-poor breakpoints resulted in localisation at the periphery and *MX1* was not expressed (20).

While the transformed nature of the K562 cell line did not allow assessment of the transforming capacity of t(7;12) in this system, we observed a distinct phenotype associated with the K562-t(7;12). In clonogenic colony-forming assays on methylcellulose-based medium, K562-t(7;12) produced a significantly higher proportion of colonies with a diffuse phenotype, at the expense of the more prevalent compact colonies of K562 control (**Figure 2A-B**). K562 produces mostly erythroid-like compact colonies, while the diffuse phenotype associates with an immature myeloid identity (CFU-G or CFU-GM-like) (21), suggesting that t(7;12) changes the erythroid differentiation bias of K562. Importantly, the phenotype was maintained through replating, indicating that the translocation can be perpetuated through self-renewal. This is in contrast with the effects of *MX1/ETV6* and *MX1* overexpression in adult haematopoietic tissues, for which phenotypic changes could not be maintained through self-renewal (10,13).

We further confirmed the depletion of erythroid identity by subjecting both K562-t(7;12) and K562 control to induced erythroid differentiation assay by adding DMSO to the culture medium and assessing the expression of erythroid marker genes (22) (**Figure 2C; Supplementary Figure 3**). At the beginning of the assay at erythroid day 0 (E0), we observed a lower expression of *HBB* and *TAL1* in K562-t(7;12) compared to K562 control (**Supplementary Figure 3A**), indicating an attenuation of erythroid signatures, compatible with the myeloid bias of K562-t(7;12) colony formation. Control K562 underwent differentiation and plateaued at E2, as seen by upregulation of *HBB*, *TAL1*, and *EPOR*. In K562-t(7;12), the differentiation capacity was not blocked by the presence of the translocation as seen by the upregulation of the three genes, albeit displaying higher fold changes compared to the control (**Figure 2C; Supplementary Figure 3**). In this light, the erythroid differentiation assay confirmed that the erythroid differentiation programme was still achievable in K562-t(7;12), but at higher fold gene expression changes, suggesting retainment of erythroid cells at an earlier differentiation state in the absence of external differentiation cues, and/or relative expansion of myeloid-

affiliated cells as a result of the translocation. Together with the colony forming capacity suggesting an imbalance between erythroid and granulocytic/monocyte-like colonies (**Figure 2A**), we demonstrate that t(7;12) can affect erythroid differentiation programmes. In earlier models of t(7;12), *MNX1*-overexpressing bone marrow cells transplanted into recipient mice were primarily found in myeloid-committed progenitor populations and particularly within megakaryocytic/erythroid progenitors (13). In human cord-blood CD34+ HSCs, *MNX1* overexpression induced erythroid transcriptional programmes and reduced overall colony-forming capacity (13), indicating that t(7;12) encompasses but may exceed the observed effects of *MNX1* overexpression on erythroid differentiation.

In order to understand t(7;12) effects in untransformed haematopoietic cells, we introduced the translocation in adult CD34+ HSCs. Despite being able to confirm its presence by PCR (**Supplementary Figure 4A**), the frequency of t(7;12)-harbouring cells decreased through serial replating in methylcellulose-based colony-forming assays (**Figure 2D**), as well as in liquid culture (**Supplementary Figure 4B**), as determined by semi-quantitative PCR. We used Long-Term Culture Initiating Cell (LTC-IC) assays to test for leukaemia-initiating effect. This 2-step assay involves a 4-week co-culture with MS5 stroma, which selects for self-renewing or slowly proliferating HSPC, followed by assessment of HSPC functional potential in colony-forming assays. Similarly to short-term clonogenic assays, edited CD34+ cells did not show any advantage in persistence of LTC-IC, as measured by the lack of sustained colony formation following co-culture in MS5 and serial methylcellulose replating (**Figure 2E**). Taken together, our observations are in line with the previously reported inability of *MNX1* overexpression to transform adult haematopoietic cells (10,13). This may suggest a requirement for additional genetic events to achieve full *MNX1*/t(7;12)-associated transformation, or otherwise suggest that t(7;12) selectively affects a cell type or developmental stage not efficiently represented in adult mouse or human haematopoietic tissues. The transformed nature of K562 cells, with or without contribution from the persistence of foetal developmental programmes, as evidenced by the nature of haemoglobins expressed (23), may provide a more suitable substrate for persistence of t(7;12).

In order to understand if K562-t(7;12) recapitulated molecular signatures identified in t(7;12) AML patients, we performed RNA sequencing analysis of K562-t(7;12) compared to K562 control, and identified 436 upregulated and 116 downregulated genes by p value lower than 0.01, of which 196 under the threshold of 1% false discovery rate (FDR) (**Figure 3A**). Gene Ontology (GO) analysis revealed that differentially expressed genes (filtered by FDR) were involved in functions related to cell adhesion, immune response, and transport-related biological processes (**Figure 3B**). Genes associated with cell adhesion and the structural component of the matrix included collagen genes (*COL*), integrins (*ITG*), and the surface markers CD24 and CD84, which were also identified by further dissection of the upregulated genes by molecular function (**Figure 3D**). The upregulation of genes associated with cell adhesion and cell-cell interactions was previously reported by Wildenhain *et al.* (10), from the differential expression of adhesion-related genes in t(7;12) patients compared to *MLL* patients (i.e. *EDIL3*, *CNTNAP5*, *ANGPT1*, *DSG2*, *ITGA9*, *ITGAV*, *KDR*, and *SIGLEC6*). This gene expression profile was highly consistent with K562-t(7;12) (50% of upregulated genes matching), including an upregulation of contactin genes *CNTN1* and *CNTN4*, *ANGPT1*, a number of integrins (*ITGA2B*, *ITGA4*, *ITGA8*, *ITGA9*, *ITGAM*, *ITGAX*, *ITGB2*), and sialic acid-binding immunoglobulin-type lectins

(*Siglecs*; *SIGLEC6*, *SIGLEC10*, *SIGLEC14*, *SIGLEC17P*, *SIGLEC8*, *SIGLEC22P*) (**Figure 3C**; **Supplementary Figure 5**). Indeed, the involvement of cell adhesion and extracellular matrix modelling in bone marrow niche interactions has been proposed as an important leukaemic mechanism of t(7;12) (10,12,13,24). Our data capture a number of these genes and can provide a model for future studies of individual and coordinated gene contributions to t(7;12) biology. The signature from Wildenhain *et al.* (10) also included the downregulation of *HOXA* genes in t(7;12) compared to *MLL* patients, which we did not find as differentially expressed in K562-t(7;12) (**Supplementary Figure 5**). This reflects the specificity of *HOXA* dysregulation in *MLL* leukaemia (25), but not necessarily the relevance in t(7;12). Balgobind *et al.* (26) also described a distinctive gene expression signature of t(7;12) leukaemia compared to a broader set of AML subtypes including *MLL* rearrangements, t(8;21), inv(16) and t(15;17), identifying 12 discriminative genes (*TP53BP2*, *DNAAF4*, *EDIL3*, *LIN28B*, *BAMBI*, *MAF*, *FAM171B*, *AGR2*, *CRISP3*, *CTTNBP2*, *KRT72*, and *MMP*). Our model matched a differential expression of *TP53BP2*, *LIN28B*, *CTNNBP2*, and *MPP2*, with half of this signature with matched or partially matched genes (**Figure 3C**; **Supplementary Figure 5**). Downregulated genes were enriched for molecular functions of macromolecule binding (**Figure 3E**). Several haemoglobin genes were found to be downregulated (*HBZ*, *HBG1-2*, *HBA1-2*, *HBB*, and *HBE1*) (**Supplementary Figure 6**), which are related to as functional categories of haptoglobin, haemoglobin, and oxygen binding processes (**Figure 3E**). The concomitant downregulation of peroxiredoxins (*PRDX1*, *PRDX4*, and *PRDX6*) also highlighted peroxidase and antioxidant activities, which are consistent with biological functions of erythrocytes, and therefore accompany the relative depletion of erythroid identity (**Figure 2A-C**).

Overall, observed gene expression changes were consistent with individual gene signatures reported in analyses of t(7;12) patients (**Figure 3C**). We sought to further validate the clinical relevance of the K562-t(7;12) model by a systematic comparison with microarray and RNA sequencing data of t(7;12) patients from Balgobind *et al.* (26) and the TARGET database (28). We derived a unique 177-gene signature by comparison of t(7;12) with other individual paediatric subtypes (inv16, *MLL* rearrangements, normal karyotype = NK, t(8;21), and other) and identification of the common differential gene set (**Figure 4A**). Similarly, we defined a 122-gene list uniquely differentially expressed between t(7;12), but not other AML subtypes, and normal paediatric bone marrow (**Figure 4B**). Consistent with previous reports (10), both gene sets were enriched for cell adhesion (collagen genes, *SIGLEC8*, contactins), cellular transport and lipid metabolism gene ontologies (**Figure 4C-D**). We also note that *HOXA* genes, previously identified by comparing t(7;12) AML with *MLL* rearrangements (10), were not present in these signatures as a result of comparisons with multiple paediatric AML subtypes. Importantly, we used the 177 and 122 t(7;12) genes as custom gene sets for Gene Set Enrichment Analysis (GSEA), and showed significant enrichment in the K562-t(7;12) model (**Figure 4E-F**), thus demonstrating that engineering of the translocation in K562 cells resulted in a clinically-relevant model. Extraction of the core enriched genes, or leading edge of GSEA identifies individual patient-specific genes recapitulated in our model (**Figure 4G**), which can be regarded as future targets in mechanistic studies of t(7;12)-associated phenotypes. Crucially, *MXN1* was identified as a leading-edge gene, asserting its centrality to t(7;12) K562 biology (**Figure 4G**).

During the preparation of this manuscript, another group successfully recreated the t(7;12) translocation in human induced pluripotent stem cell (iPSC) using CRISPR/Cas9 (29), with comparable features to our model. The authors observed an increased frequency of erythroid and myeloid progenitors in colony forming assays, coherent with an increase in myeloid transcriptional programmes. Notably, the iPSC model reproduced the overexpression of *MX1* and production of a *MX1/ETV6* fusion transcript, albeit at a low transcriptional level that questions its biological relevance. Despite extensive FISH analyses in early studies (4-6), the exact breakpoints required to produce the fusion transcripts from t(7;12) have not been elucidated. In Nilsson et al. (29), the expression of the *MX1/ETV6* mRNA was achieved by targeting a similar genomic region to us on 7q36, but a distinct intronic region of *ETV6*. In our K562 model, a chimaeric transcript was not detected by potential gene fusion search by TopHat-Fusion on the RNA sequencing reads; an interesting observation given the 50% occurrence of chimaeric transcripts in patients (2). With the continued development of genome editing technologies, we foresee that these *in vitro* models will allow a more precise dissection of the breakpoint regions required for the formation of the yet uncharacterised *MX1/ETV6* chimaera. Both the iPSC and K562 models of t(7;12) showed similarities to patient-specific gene expression patterns, which strengthens the validity of CRISPR/Cas9 to generate disease models based on clinical genomic information.

In summary, we generated cellular model of t(7;12) translocation which recapitulates *MX1* overexpression and captures cell adhesion signatures putatively core to t(7;12) oncogenic programmes. The model displays a granulocytic-monocytic phenotype compatible with the lineage affiliation of clinical t(7;12) AML, and disrupts erythroid lineage programmes as described in *MX1* overexpression. We confirmed the inability of t(7;12) to transform adult haematopoietic tissues, highlighting the requirement for additional oncogenic hits, or a specific cellular background which may explain the unique association of t(7;12) with infants. Complementary to earlier *MX1*-based models (10,13), and more recent genome edited tools (29), our t(7;12) K562 model will provide a framework for the molecular elucidation of this unique subtype of infant AML.

Acknowledgements: Dr Giorgia Santilli is acknowledged for providing adult CD34+ cells. DR is the recipient of a Kidscan funded PhD studentship and partly supported by Brunel University London. YC and AS are supported by a grant from the Oracle Cancer Trust.

Author Contributions: DR was responsible for carrying out all experiments, extracting and analysing data, interpreting results, producing tables and figures and writing the manuscript. YC performed FISH experiments, captured and analysed fluorescent microscopy images, participated in editing and reviewing the manuscript. CF, SS, and FB contributed the gene positioning analysis and provided feedback on the manuscript. RS participated in gene expression analysis of K562 erythroid differentiation. CS supervised the computational analysis of transcriptomic data and reviewed the manuscript. CP designed and supervised functional analysis of K562 and CD34+ cells, participated in design and interpretation of transcriptomic data analysis, and reviewed the manuscript. AS was responsible for designing and supervising the project, with specific input on the gene editing, and reviewed the manuscript. ST was responsible for designing and

supervising the project, with specific input on the chromosome biology aspects including interpretation of FISH results, and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Competing Interests: Authors declare no competing interests.

Data Availability Statement: The RNA sequencing datasets generated and analysed during the current study have been deposited in the ArrayExpress repository under accession number E-MTAB-11851. The results published here are partly based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) (<https://ocg.cancer.gov/programs/target>) initiative, of the Acute Myeloid Leukemia (AML) cohort phs000465. The data used for this analysis are available at <https://portal.gdc.cancer.gov/projects>.

References:

- (1) Espersen ADL, Noren- Nyström U, Abrahamsson J, Ha S, Pronk CJ, Jahnukainen K, et al. Acute myeloid leukemia (AML) with t (7; 12)(q36; p13) is associated with infancy and trisomy 19: Data from Nordic Society for Pediatric Hematology and Oncology (NOPHO- AML) and review of the literature. *Genes, Chromosomes and Cancer* 2018;57(7):359-365.
- (2) Tosi S, Kamel YM, Owoka T, Federico C, Truong TH, Saccone S. Paediatric acute myeloid leukaemia with the t (7; 12)(q36; p13) rearrangement: a review of the biological and clinical management aspects. *Biomarker research* 2015;3(1):1.
- (3) Masetti R, Vendemini F, Zama D, Biagi C, Pession A, Locatelli F. Acute myeloid leukemia in infants: biology and treatment. *Frontiers in pediatrics* 2015;3:37.
- (4) Simmons H, Oseth L, Nguyen P, O'Leary M, Conklin K, Hirsch B. Cytogenetic and molecular heterogeneity of 7q36/12p13 rearrangements in childhood AML. *Leukemia* 2002;16(12):2408.
- (5) Tosi S, Harbott J, Teigler-Schlegel A, Haas OA, Pirc-Danoewinata H, Harrison CJ, et al. t (7; 12)(q36; p13), a new recurrent translocation involving ETV6 in infant leukemia. *Genes, Chromosomes and Cancer* 2000;29(4):325-332.
- (6) Tosi S, Hughes J, Scherer SW, Nakabayashi K, Harbott J, Haas OA, et al. Heterogeneity of the 7q36 breakpoints in the t (7; 12) involving ETV6 in infant leukemia. *Genes, chromosomes and cancer* 2003;38(2):191-200.
- (7) Ballabio E, Cantarella C, Federico C, Di Mare P, Hall G, Harbott J, et al. Ectopic expression of the HLXB9 gene is associated with an altered nuclear position in t (7; 12) leukaemias. *Leukemia* 2009;23(6):1179-1183.
- (8) Beverloo HB, Panagopoulos I, Isaksson M, van Wering E, van Drunen E, de Klein A, et al. Fusion of the homeobox gene HLXB9 and the ETV6 gene in infant acute myeloid leukemias with the t(7;12)(q36;p13). *Cancer Res* 2001 Jul 15;61(14):5374-5377.
- (9) von Bergh AR, van Drunen E, van Wering ER, van Zutven LJ, Hainmann I, Lönnerholm G, et al. High incidence of t (7; 12)(q36; p13) in infant AML but not in infant ALL, with a dismal outcome and ectopic expression of HLXB9. *Genes, Chromosomes and Cancer* 2006;45(8):731-739.

- (10) Wildenhain S, Ruckert C, Röttgers S, Harbott J, Ludwig W, Schuster F, et al. Expression of cell–cell interacting genes distinguishes HLXB9/TEL from MLL-positive childhood acute myeloid leukemia. *Leukemia* 2010;24(9):1657.
- (11) Park J, Kim M, Lim J, Kim Y, Han K, Lee J, et al. Three-way complex translocations in infant acute myeloid leukemia with t (7; 12)(q36; p13): the incidence and correlation of a HLXB9 overexpression. *Cancer Genet Cytogenet* 2009;191(2):102-105.
- (12) Wildenhain S, Ingenhag D, Ruckert C, Degistirici O, Dugas M, Meisel R, et al. Homeobox protein HB9 binds to the prostaglandin E receptor 2 promoter and inhibits intracellular cAMP mobilization in leukemic cells. *J Biol Chem* 2012 Nov 23;287(48):40703-40712.
- (13) Ingenhag D, Reister S, Auer F, Bhatia S, Wildenhain S, Picard D, et al. The homeobox transcription factor HB9 induces senescence and blocks differentiation in hematopoietic stem and progenitor cells. *Haematologica* 2019 Jan;104(1):35-46.
- (14) Cheong T, Blasco RB, Chiarle R. The CRISPR/Cas9 system as a tool to engineer chromosomal translocation in vivo. *Chromosome Translocation* 2018:39-48.
- (15) Choi PS, Meyerson M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nature communications* 2014;5(1):1-6.
- (16) Torres R, Martin M, Garcia A, Cigudosa JC, Ramirez J, Rodriguez-Perales S. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR–Cas9 system. *Nature communications* 2014;5:3964.
- (17) Maddalo D, Manchado E, Concepcion CP, Bonetti C, Vidigal JA, Han Y, et al. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* 2014;516(7531):423-427.
- (18) Brunet E, Jasin M. Induction of chromosomal translocations with CRISPR-Cas9 and other nucleases: understanding the repair mechanisms that give rise to translocations. *Chromosome Translocation* 2018:15-25.
- (19) Gulino GM, Bruno F, Sturiale V, Brancato D, Ragusa D, Tosi S, et al. From FISH to Hi-C: The Chromatin Architecture of the Chromosomal Region 7q36. 3, Frequently Rearranged in Leukemic Cells, Is Evolutionary Conserved. *International Journal of Molecular Sciences* 2021;22(5):2338.
- (20) Federico C, Owoka T, Ragusa D, Sturiale V, Caponnetto D, Leotta CG, et al. Deletions of Chromosome 7q Affect Nuclear Organization and HLXB9Gene Expression in Hematological Disorders. *Cancers* 2019;11(4):585.
- (21) Arya D, Sachithanandan SP, Ross C, Palakodeti D, Li S, Krishna S. MiRNA182 regulates percentage of myeloid and erythroid cells in chronic myeloid leukemia. *Cell death & disease* 2018;8(1):e2547-e2547.
- (22) Arede L, Foerner E, Wind S, Kulkarni R, Domingues AF, Giotopoulos G, et al. KAT2A complexes ATAC and SAGA play unique roles in cell maintenance and identity in hematopoiesis and leukemia. *Blood advances* 2022;6(1):165-180.
- (23) Rutherford T, Clegg JB, Higgs DR, Jones RW, Thompson J, Weatherall DJ. Embryonic erythroid differentiation in the human leukemic cell line K562. *Proc Natl Acad Sci U S A* 1981 Jan;78(1):348-352.
- (24) Schwaller J. Novel insights into the role of aberrantly expressed MNX1 (HLXB9) in infant acute myeloid leukemia. *Haematologica* 2019 Jan;104(1):1-3.

- (25) Chen S, Qin Z, Hu F, Wang Y, Dai Y, Liang Y. The role of the HOXA gene family in acute myeloid leukemia. *Genes* 2019;10(8):621.
- (26) Balgobind BV, Van den Heuvel-Eibrink MM, De Menezes RX, Reinhardt D, Hollink IH, Arentsen-Peters ST, et al. Evaluation of gene expression signatures predictive of cytogenetic and molecular subtypes of pediatric acute myeloid leukemia. *Haematologica* 2011 Feb;96(2):221-230.
- (27) Hauer J, Tosi S, Schuster FR, Harbott J, Kolb H, Borkhardt A. Graft versus leukemia effect after haploidentical HSCT in a MLL-negative infant AML with HLXB9/ETV6 rearrangement. *Pediatric blood & cancer* 2008;50(4):921-923.
- (28) Therapeutically Applicable Research to Generate Effective Treatments (TARGET). Available at: <https://ocg.cancer.gov/programs/target>.
- (29) Nilsson T, Waraky A, Östlund A, Li S, Staffas A, Asp J, et al. An induced pluripotent stem cell t (7; 12)(q36; p13) acute myeloid leukemia model shows high expression of MNX1 and a block in differentiation of the erythroid and megakaryocytic lineages. *International Journal of Cancer* 2022.
- (30) Kamel YM, Naiel A, Alshehri A, Vetter M, Saccone S, Anderson R, et al. Fluorescence in situ hybridisation assays designed for del (7q) detection uncover more complex rearrangements in myeloid leukaemia cell lines. 2014.
- (31) Federico C, Cantarella CD, Di Mare P, Tosi S, Saccone S. The radial arrangement of the human chromosome 7 in the lymphocyte cell nucleus is associated with chromosomal band gene density. *Chromosoma* 2008;117(4):399-410.

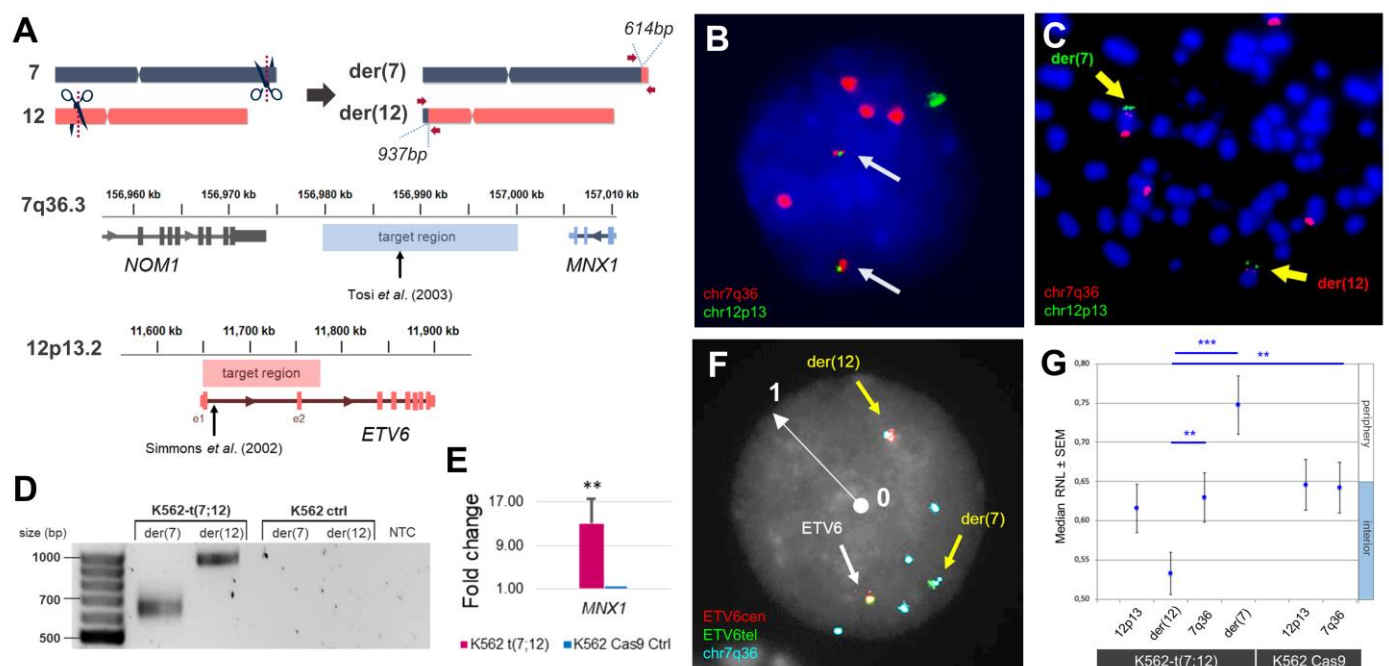


Figure 1 – Generation of t(7;12)(q36;p13) in K562 cells. **A)** Schematic representation of target regions for CRISPR/Cas9-directed cleavage on chromosomes 7 and 12 used for guide RNA (gRNA) design. Fusion junctions are flanked by arrows representing PCR primers used for confirmation, yielding products of 614 bp and 937 bp. Below, detailed location of the targeted regions on chromosome 7q36.3 and 12p13.2 with reference to known t(7;12) breakpoints used to generate the translocation. **B)** FISH using the specific t(7;12) probe XL t(7;12) MNX1/ETV6 (MetaSystems GmbH, Altlußheim, Germany, **Supplementary Table 3**) hybridising chromosome 7q36 in red and chromosome 12p13 in green. Two yellow fusion signals, pointed by arrows, indicate the presence of t(7;12) in a representative interphase nucleus of K562-t(7;12). The karyotype of K562 is nearly tetraploid and harbours complex rearrangements, including a duplication of the 7q36 locus within the short arm of chromosome 7, hence showing five 7q36 signals and two 12p13 signals (30) (**Supplementary Figure 2**). **C)** Metaphase spread of K562-t(7;12) hybridised with the same probe shows the presence of derivative chromosomes der(7) and der(12), pointed by arrows. **D)** Confirmation of the presence of t(7;12) in K562-t(7;12) but not K562 control ('ctrl') cells (K562 electroporated with Cas9 only) by PCR amplification of fusion junctions and product separation on agarose gel; bands correspond to the predicted sizes shown in panel A. NTC = no template control. **E)** qRT-PCR validation of overexpression of *MNX1* in K562-t(7;12) compared to K562 control. The fold change was calculated using the $\Delta\Delta C_t$ method by normalisation to the endogenous gene *HPRT1*. Error bars represent standard deviation (SD) of n=3. Primers are reported in **Supplementary Table 4**. **F)** A custom-made 3-colour probe (MetaSystems dual-colour ETV6 + PAC-derived RP5-1121A15) hybridising *ETV6* portions in red (centromeric) and green (telomeric), and the *MNX1* locus in cyan (**Supplementary Figure 2**; **Supplementary Table 3**), allowed visualisation of both derivative chromosomes in interphase nuclei (pointed by yellow arrows). The white radius arrow represents the distance between nuclear interior (value=0) and nuclear periphery (value=1), which was used in the calculation of radial nuclear locations (RNL). **G)** RNL of der(7) and der(12) signals in K562-t(7;12) interphase nuclei. The RNL values are expressed as median values of individual localisations of FISH signals (200 nuclei analysed per condition). Errors bars represent standard error of the mean (SEM). Values closer to 0 indicate an internal position within the nucleus (described in detail in Federico *et al.* (31)).

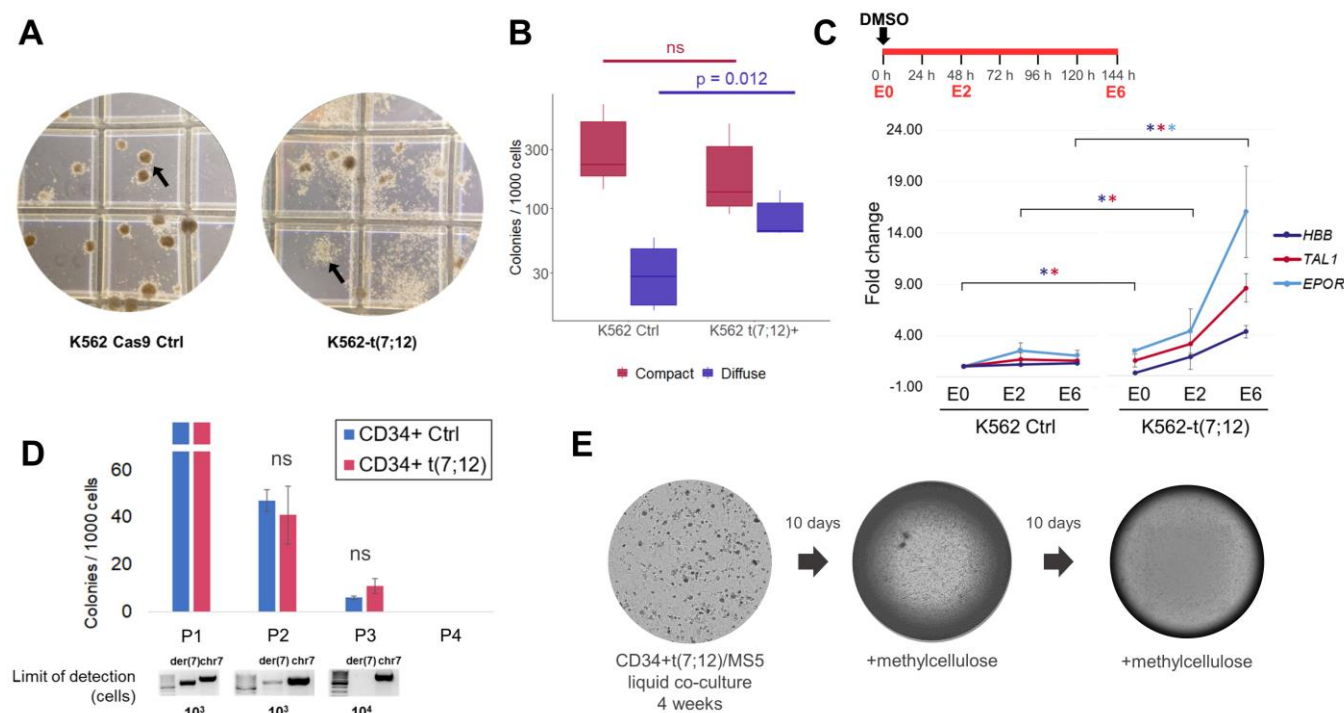


Figure 2 – Functional effects of t(7;12) in haematopoietic cells. **A)** Representative images of colony-forming assays of K562 control ('K562 Cas9 Ctrl') (left) and K562-t(7;12) (right) on Methocult H4434 Classic (StemCell Technologies). Arrows indicate distinct morphologies of compact (left) and diffuse (right) colonies. **B)** Frequency of colony formation per 1000 K562 Cas9 Ctrl and K562-t(7;12) plated cells. Mean and 95% CI of n=3 independent experiments. Statistical significance determined by 2-tailed Student's t-test. **C)** Quantitative RT-PCR gene expression analysis of erythroid differentiation of K562-t(7;12) cells; schematic representation of differentiation assay (top left). 100 000 cells were plated in RPMI medium supplemented with foetal bovine serum (FBS) with 1.5 μ M DMSO added at time point 0 h. 'E' indicates the day of erythroid differentiation, from E0 (0 h) to E6 (144 h). Expression of erythroid markers *EPOR*, *TAL1*, and *HBB* was quantified in K562-t(7;12) through E0 to E6, calibrated to K562-Ctrl E0. Fold changes were calculated using the $\Delta\Delta$ Ct method by normalisation to the endogenous gene *HPRT1*. Asterisks indicate statistically significant fold changes of K562-t(7;12) compared to K562 control for each gene. Full statistical tests on all conditions are shown in **Supplementary Figure 3**. Primers are reported in **Supplementary Table 4**. **D)** Introduction of the t(7;12) translocation in healthy CD34+ haematopoietic progenitors enriched from mobilised peripheral blood by magnetic separation. Serial replating colony-forming assays in Methocult H4434 Classic medium with multi-lineage cytokines. 10 000 CD34+ HSPCs were plated in methylcellulose-based medium and allowed to form colonies. Colonies were scored, collected and dissociated after each plating, and replated over several rounds. HSPCs are expected to originate progressively less colonies, unless a survival advantage is gained through transformation. 'P0-P4' indicate the order of plating. Progressively less colonies were produced, without statistically significant differences between the two conditions. Agarose gel electrophoresis of semi-quantitative PCR is shown underneath each colony count, indicating the limit of detection of the der(7) fusion junction. A non-translocated region of chromosome 7 was used as a control ('chr 7') to confirm the presence of amplifiable template DNA. **E)** MS5 stroma with the addition of CD34+t(7;12) cells. Two colonies were visible following methylcellulose addition after 4 weeks of co-culture. No colonies visible following a second round of co-culture and methylcellulose addition.

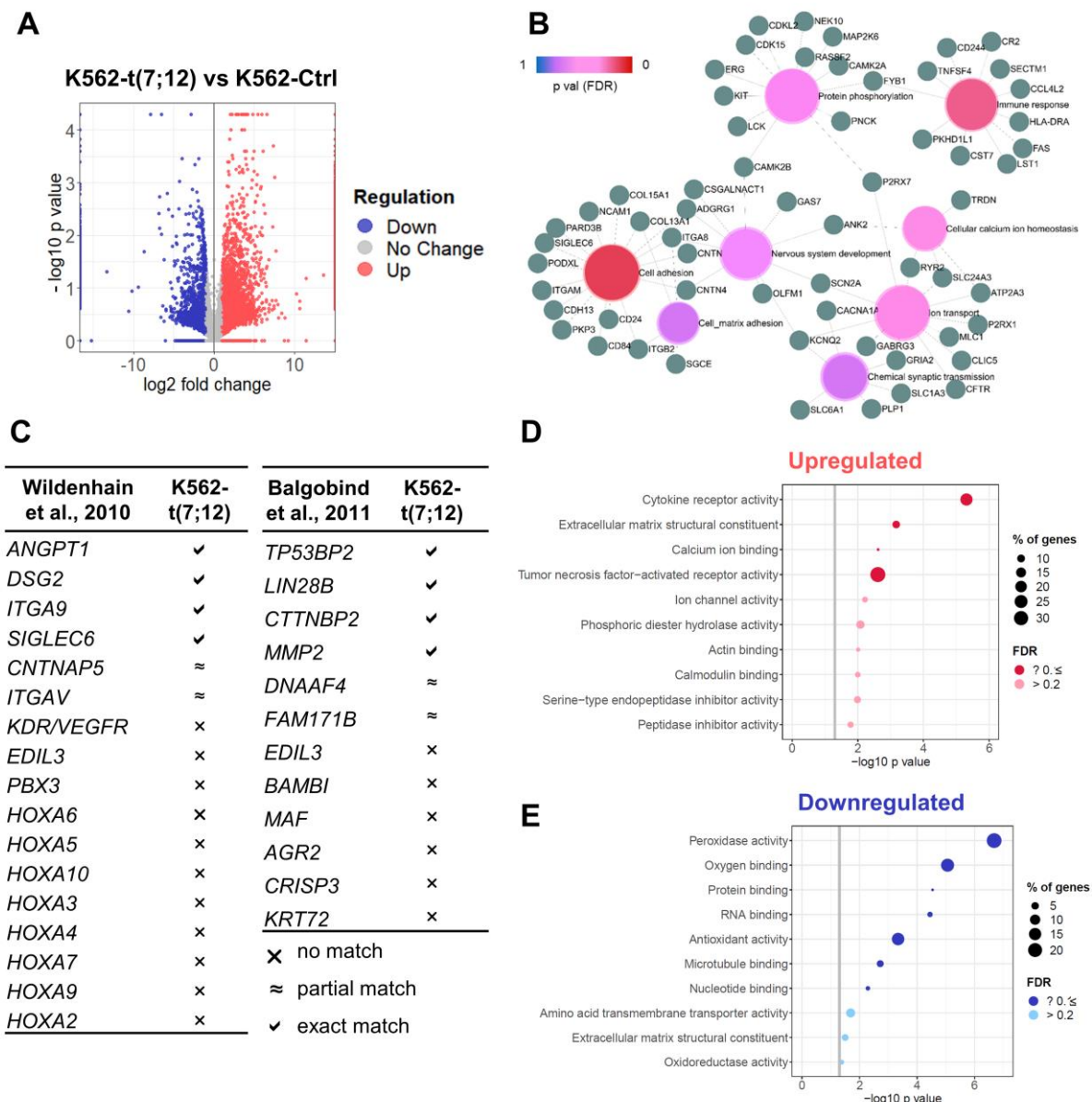


Figure 3 – Transcriptional analysis of K562-t(7;12). **A**) Volcano plot showing differentially expressed genes in K562-t(7;12) compared to K562 control, depicting up- (in red) and downregulated (in blue) genes and their statistical significance by $-\log_{10} p$ value. Genes with an absolute \log_2 fold change lower than 1 are shown in grey ('No change'). **B**) Gene Ontology network plot of differentially expressed genes by false discovery rate (FDR) < 0.1 showing significant biological processes (PANTHER) and genes associated to each GO term, constructed on ExpressAnalyst (available at www.expressanalyst.ca) **C**) Comparison of differentially expressed genes in K562-t(7;12) and published reports of genes associated with t(7;12) patients. Symbols indicate difference in expression between K562-t(7;12) and K562 control by statistical significance (threshold $p \leq 0.05$), defined as an exact gene match, a partial match (i.e. dysregulation of a gene belonging to the same family), or no match. **D-E**) GO terms relating to molecular functions (PANTHER) of upregulated (**D**) and downregulated (**E**) genes in K562-t(7;12). The grey intercept marks the $p=0.05$ threshold in $-\log_{10}$. The size of the dots indicates the percentage of genes mapping to a specific term. Dark red or blue indicates an enrichment with an FDR ≤ 0.1 .

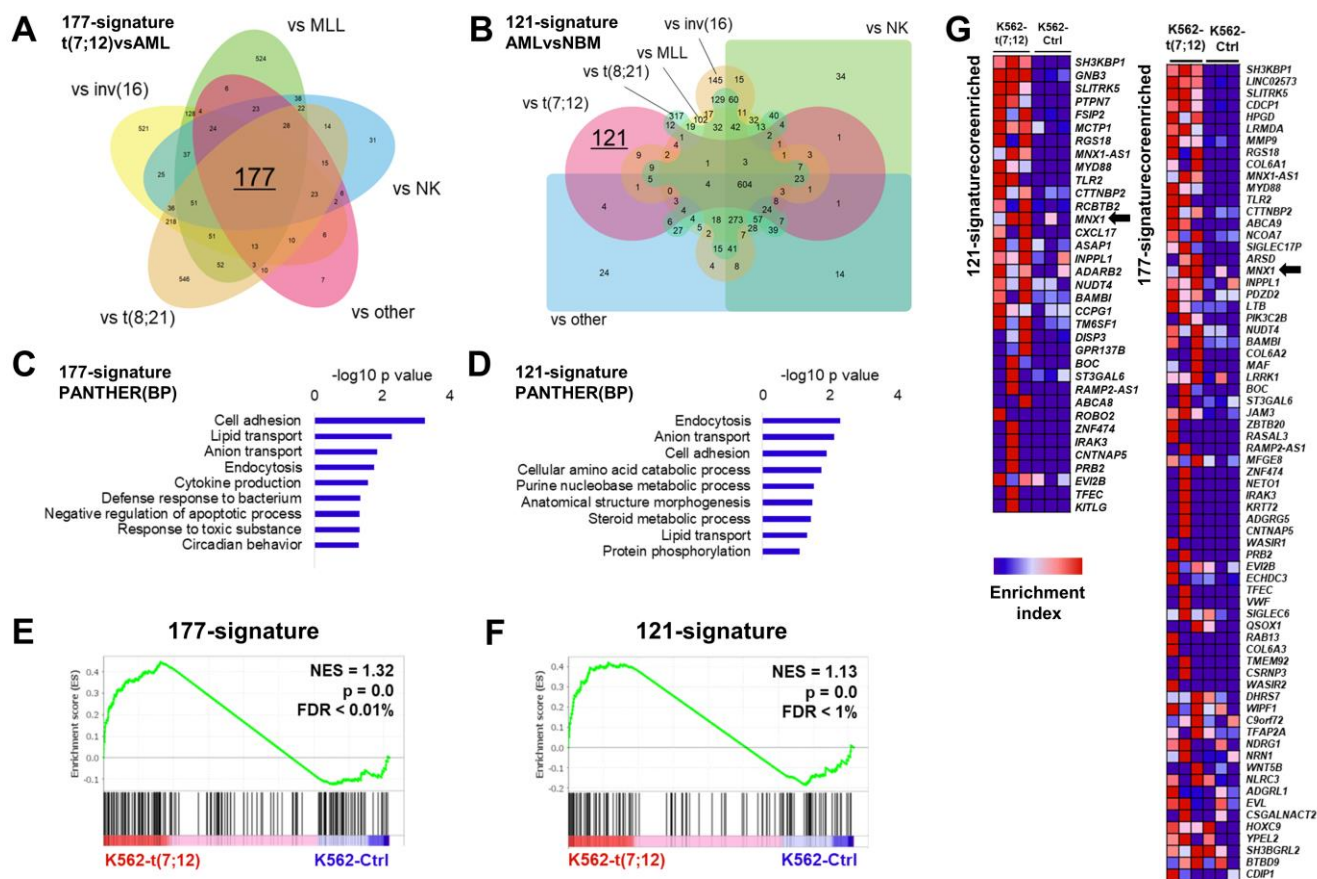
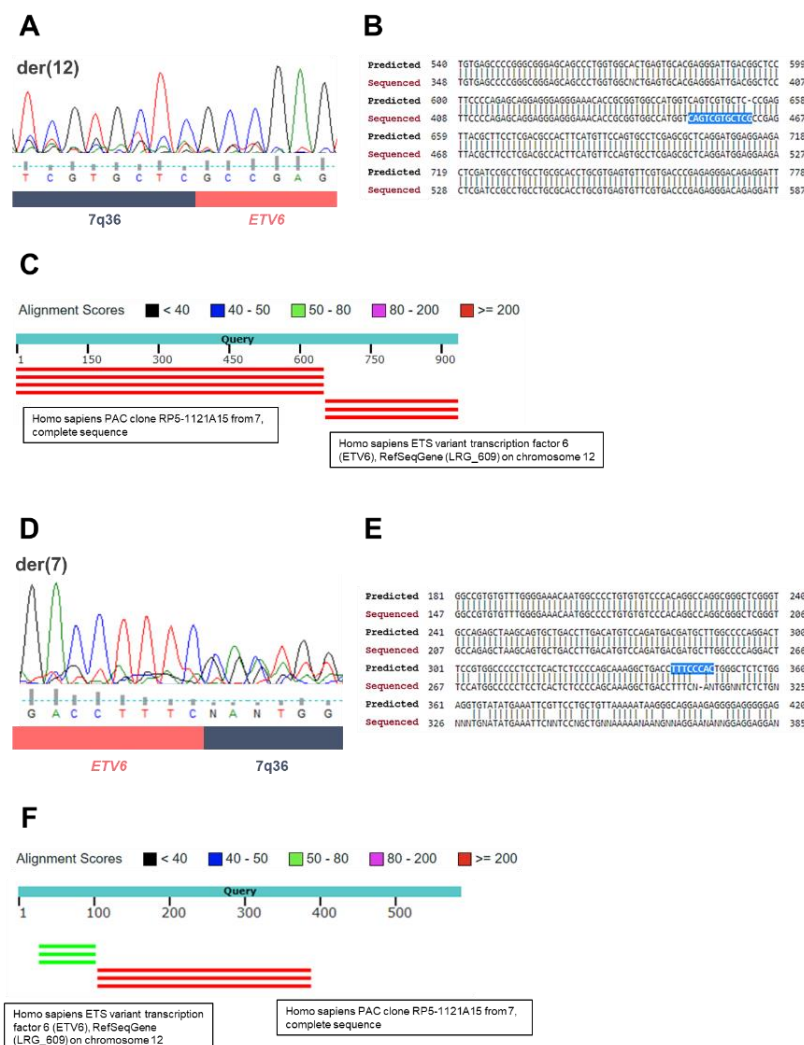
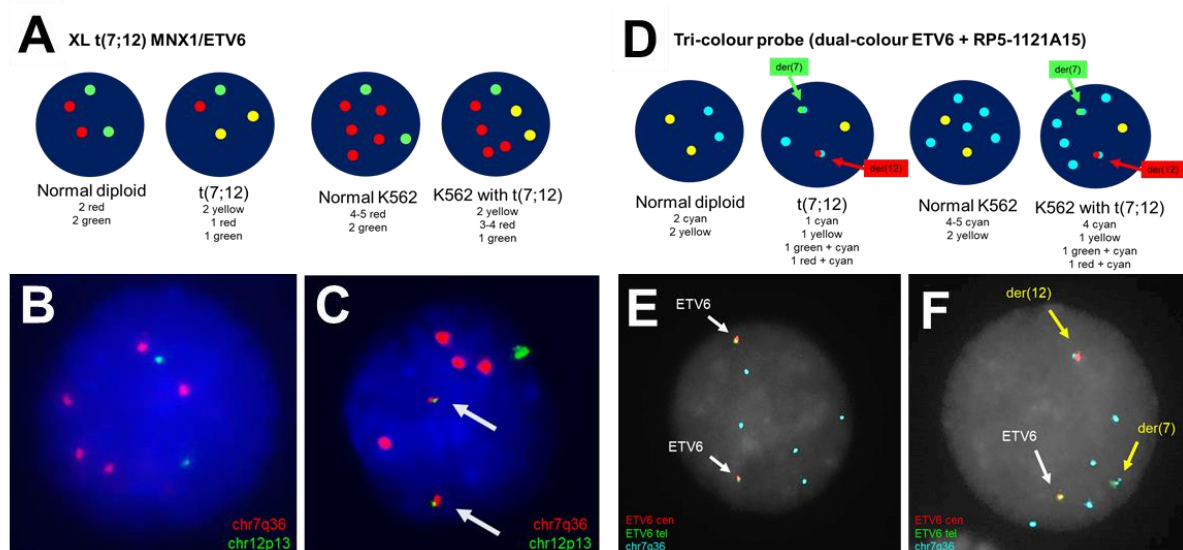


Figure 4 – Comparison of K562-t(7;12) transcriptional landscape with t(7;12) patient signatures. A) 177-gene signature of t(7;12)-patient gene expression extrapolated from published microarray and RNA sequencing datasets by comparison with other paediatric AML subtypes. The Venn diagram shows the 177 intersect genes. **B)** 121-gene t(7;12)-specific signature inferred by comparisons of gene expressions of paediatric AML and normal bone marrow (NBM) samples. Edward's Venn diagram highlights the 121 exclusive genes to t(7;12). **C-D)** Gene Ontology analysis of the 177-signature and 121-signature by the PANTHER annotation repository of biological processes (BP). **E-F)** GSEA enrichment plots of K562-t(7;12) gene expression profile using the 177- and 121-signatures. NES = normalized enrichment score; FDR = false discovery rate. **G)** Core enriched genes from the GSEA using the 177- and 121-signatures, shown by their enrichment index in K562-t(7;12) against K562 control. *MX1* is highlighted by the arrows.

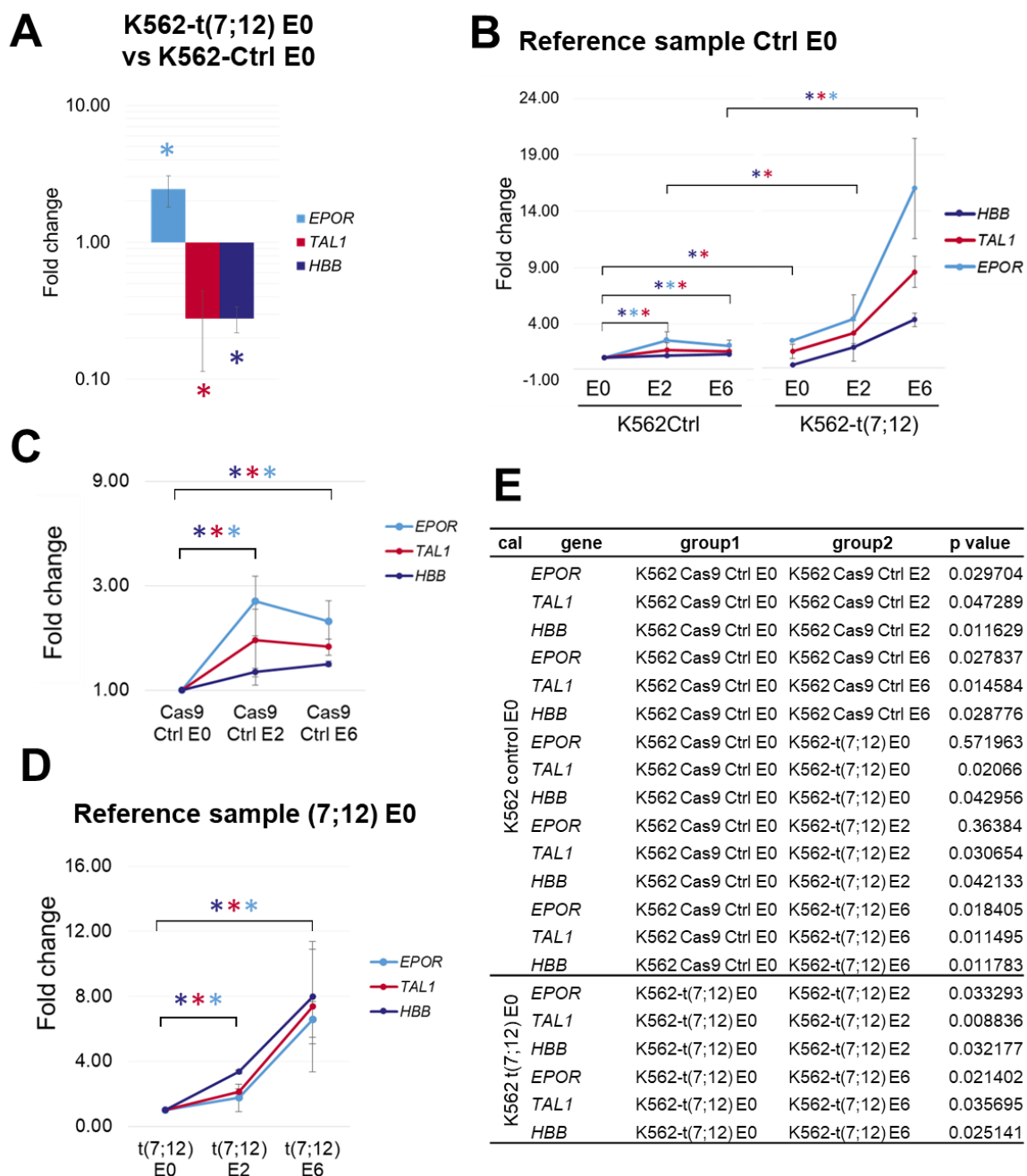
Supplementary Figures



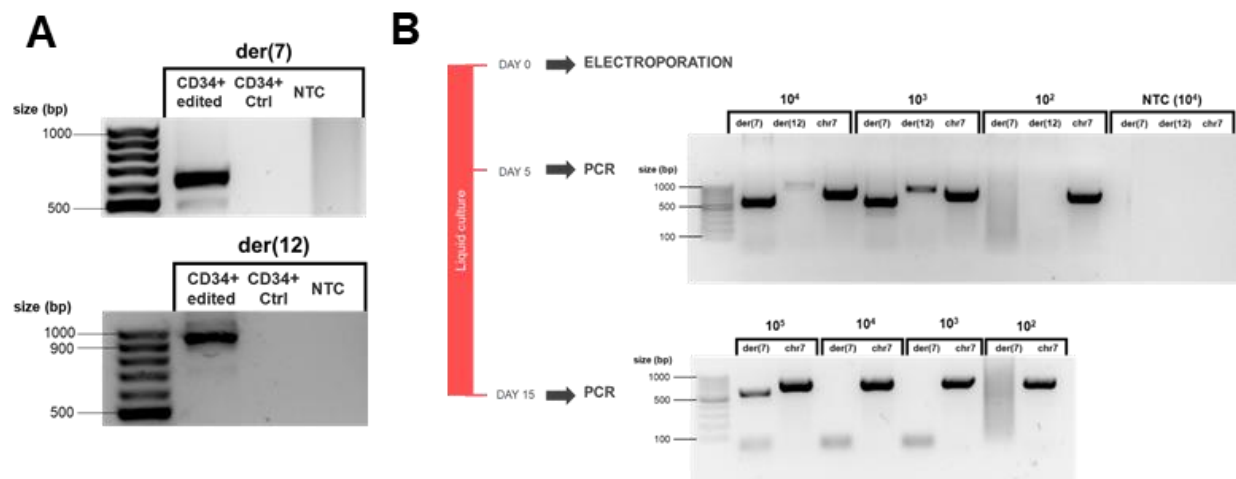
Supplementary Figure 1 - Sanger sequencing of der(7) and der(12) fusion junctions. Chromatograms showing the nucleotide sequence of amplified fusion junctions in der(12) (**A**) and der(7) (**D**). Sequence alignments of der(12) (**B**) and der(7) (**E**) corresponded to the predicted sequence. Highlighted nucleotides in blue correspond to the fusion junction. The identity of the sequences was confirmed by BLAST, showing matching sequence identity from chromosome 7q36 region (RP5-112A15) and the ETV6 gene in both derivatives (**C,F**).



Supplementary Figure 2 – FISH on K562-t(7;12) and K562 control nuclei using two probe sets. A) The t(7;12)-specific FISH probe XL t(7;12) MNX1/ETV6 (MetaSystems) patterns in normal diploid nuclei and t(7;12) on the left, and normal K562 and K562 with t(7;12) on the right. The presence of the t(7;12) is confirmed by two yellow fusion signals, by splitting of the red and green signals due to the translocation. The karyotype of K562 is nearly tetraploid and harbours complex rearrangements, including a duplications of the 7q36 locus containing MNX1 within the short arm of chromosome 7, hence showing five 7q36 signals and two 12p13 signals (1) **B)** A normal pattern indicates the expected signal pattern in non-edited K562, with 5 red signals corresponding to the 7q36 locus, and 2 green signals hybridising to 12p13. **C)** Two yellow fusion signals, pointed by arrows, are indicative of the presence of the t(7;12). **D)** Tri-colour probe (dual colour ETV6 probe + RP5-1121A15) patterns in normal diploid nuclei and t(7;12) on the left, and normal K562 and K562 with t(7;12) on the right. The two derivatives are distinguishable by the fusion of the green and cyan (der7), and red and cyan (der12). **E)** Representative nucleus and metaphase of K562 Cas9 Ctrl showing the hybridisation patterns of the 3-colour probe with non-translocated *ETV6* and 7q36 regions, indicated by 2 yellow fusion signals and 5 cyan signals, respectively. **F)** Representative K562-t(7;12) nucleus showing the presence of two derivative chromosomes indicated by yellow arrows, the non-translocated *ETV6* allele (white arrow), and 4 non-translocated 7q36 signals. The der(7) is characterised by the fusion of the green and cyan signals, while the der(12) is distinguishable by red and cyan overlapping signals.

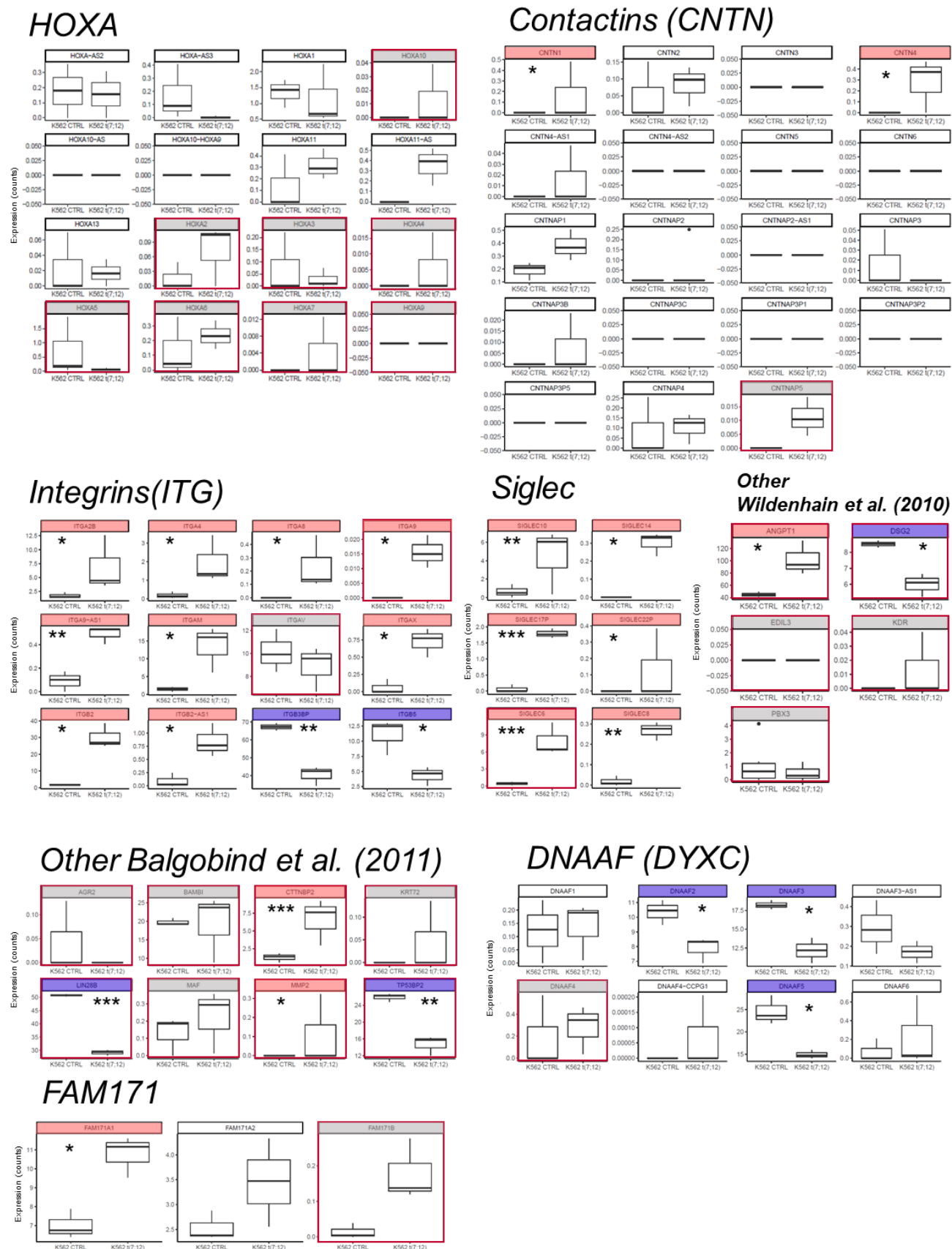


Supplementary Figure 3 – Erythroid differentiation assay of K562-t(7;12). **A)** Gene expression quantification of HBB, TAL1 and EPOR in K562-t(7;12) expressed in fold change compared to K562-Ctrl at the beginning of the erythroid differentiation assay (0 h; E0). Asterisks indicate a p value ≤ 0.05 with the colour corresponding to the gene as shown in the legend. Error bars indicate standard deviation on n=3. **B)** Full comparison of statistically significant gene expression changes, by fold changes calibrated to K562 control E0. **C)** Zoomed in view of K562 control with expanded fold change y axis. **D)** Fold changes in K562-t(7;12) calibrated to K562-t(7;12) E0, instead of K562 control. **E)** Table of p values determined by Student's T-test in all combinations for expression changes across E0 – E2 – E6. "Cal" refers to the calibrator/reference sample of choice, either K562-Ctrl E0 or K562-t(7;12) E0. Group 1 and group 2 indicate the two conditions compared.



Supplementary Figure 4 - PCR amplification of fusion junctions of t(7;12) in edited CD34+ HSPCs.

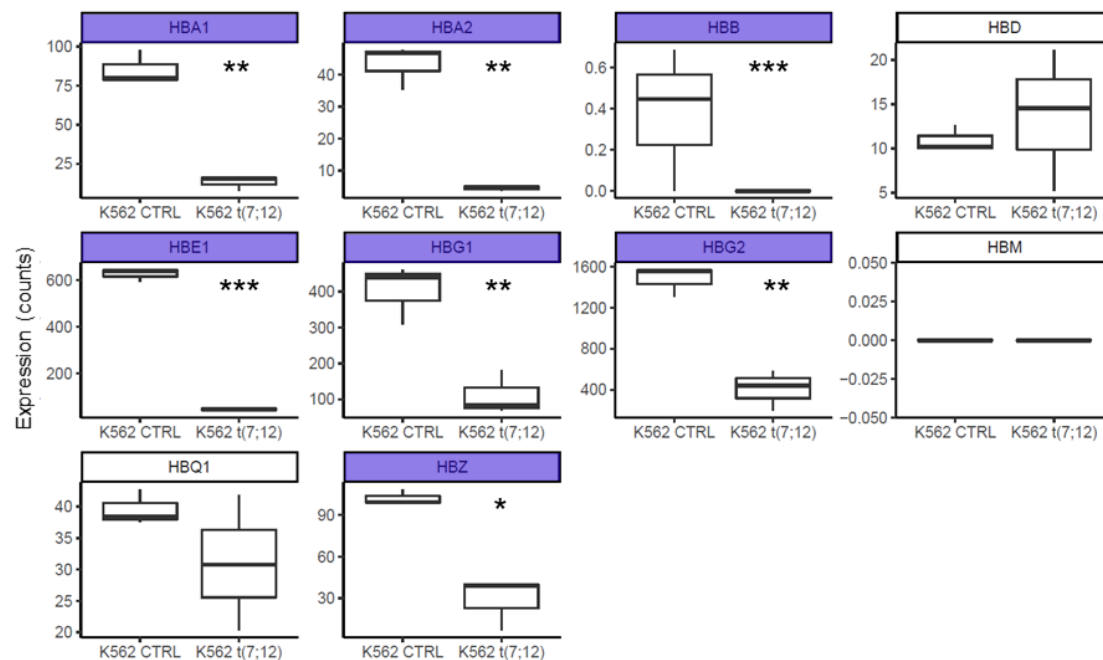
Agarose gel electrophoresis of amplified fusion junctions of der(7) (expected size 614 bp) and der(12) (expected size 937 bp) Ctrl = Cas9-only Control; NTC = no template control. Molecular marker: Gene Ruler 100 bp. Following electroporation, edited CD34+ HSPCs were grown in liquid culture for 15 days and screened for the presence of t(7;12) fusion junctions by semi-quantitative direct PCR. Varying amounts of cells were directly lysed and used as template for PCR to detect der(7) (614 bp) or der(12) (937 bp) fusion junctions. 'Chr7' indicates an unedited region of chromosome 7 used as control to confirm the presence of sufficient DNA template (735 bp). The lowest cell number at which a band is visible marks the limit of detection.



Supplementary Figure 5 –Expression differences in K562-t(7;12) compared to K562 control of known t(7;12)-associated genes. Gene expressions were extracted from RNA sequencing counts, and grouped by gene families. Red squares around the plot indicate a gene previously reported by Wildenhain *et al.* (2010)

or Balgobind *et al.* (2011) as specific for the t(7;12) subtype. Grey shading indicates that the gene was not significantly dysregulated in K562-t(7;12) determined by T-test, while red indicates a significant upregulation, and blue a significant downregulation. Asterisks symbolise p value thresholds of 0.05 (*), 0.001 (**), 0.0001 (***), and 0.00001 (****).

Haemoglobins



Supplementary Figure 6 – Differentially expressed haemoglobin genes in K562-t(7;12). Gene expressions were extracted from RNA sequencing counts of K562-t(7;12) and K562 control. Blue shading indicates a significant downregulation determined by T-test. Asterisks symbolise p value thresholds of 0.05 (*), 0.001 (**), 0.0001 (***), and 0.00001 (****).

Supplementary Tables

Supplementary Table 1 – gRNA sequences

NAME	SEQUENCE (5'-TO-3')	PAM
7q36 gRNA 4	ATGGTCAGTCGTGCTCGGAA	AGG
ETV6 gRNA 4	GAGGAAGCGTAACTCGGCAC	TGG

Supplementary Table 2 – PCR primer sequences

PRIMER NAME	SEQUENCE 5'-TO-3'
7q36 4 F	GTTTGGGGAAACAATGGCCC
7q36 4 R	GAGGCTACCAAGTGAAGCGT
Chr7 F	GCCACATCATCCAATGCTGG
Chr7 R	CTCCTGGGCCACAAGACAAT
Der7 F-3	CTGCACTGATGAAGCCGATG
Der12 R-4	GAAACAAGGAACAGGAACCTCAA
Der7 R-4	GGCTGTTTTCAACGATGGCA

Supplementary Table 3 – FISH probes ideograms

NAM E	IDEOGRAM	FLUOROPHO RES	MANUFACTU RER
XL t(7;12) MNX1/ETV6		FITC (ETV6) Cy3 (MNX1)	Metasystems
Dual-colour ETV6		FITC (ETV6 5') Cy3 (ETV6 3')	Metasystems
RP5-1121A15		Cy5 (cyan false colour)	PAC-clone

Supplementary Table 4 – qPCR primer sequences

GENE	FORWARD PRIMER 5'-TO-3'	REVERSE PRIMER 5'-TO-3'	REFERENCE
<i>HPRT1</i>	CCTGGCGTCGTGATTAGTGAT	TCGAGCAAGACGTTTCAGTCC	Arede <i>et al.</i> , 2022 (2)
<i>HBB</i>	AGGAGAAGTCTGCCGTTACTG	CCGAGCACTTTCTTGCCATGA	Arede <i>et al.</i> , 2022 (2)
<i>EPOR</i>	TGGCACATAGCGAACATTCCA	GGCTGGGAAGAGAATGCTGATT	This study
<i>TAL1</i>	TGATGGTCCCCACACCAAAG	AAAGGCCCGTTTCACATTCT	This study
<i>MNX1</i>	GTTCAAGCTCAACAAGTACC	GGTTCTGGAACCAAATCTTC	Gulino <i>et al.</i> , 2021 (3)

Supplementary Table 5 – 122-signature gene list

122 - signature				
<i>MXN1</i>	<i>SH3KBP1</i>	<i>NUDT4</i>	<i>RCBTB2</i>	<i>STK39</i>
<i>TTC39C</i>	<i>CSN1S1</i>	<i>TES</i>	<i>WSB1 ACP3</i>	<i>NUDT4B</i>
<i>CNTNAP5</i>	<i>RORB</i>	<i>ASAP1</i>	<i>PON2</i>	<i>ABCA6</i>
<i>ACOT9</i>	<i>PITPNM3</i>	<i>DISP3</i>	<i>ENOSF1</i>	<i>B3GALNT1</i>
<i>TLR2</i>	<i>PHACTR2</i>	<i>SPSB4</i>	<i>TP53BP2</i>	<i>KITLG</i>
<i>WIPF1</i>	<i>TRPC4</i>	<i>KIAA1217</i>	<i>ENOX1</i>	<i>CAV3</i>
<i>MXN1-AS1</i>	<i>KRT2</i>	<i>CREB5</i>	<i>METTL9</i>	<i>CSGALNACT2</i>
<i>CTTNBP2</i>	<i>MOCS2</i>	<i>BAMBI</i>	<i>CDX2</i>	<i>KCNIP1</i>
<i>OGFRL1</i>	<i>FOXB1</i>	<i>MBD2</i>	<i>FER</i>	<i>PPM1L</i>
<i>GALNT18</i>	<i>TMEM217</i>	<i>MCTP1</i>	<i>ST3GAL6</i>	<i>PPIL6</i>
<i>PRB2</i>	<i>PTGER2</i>	<i>NLRP11</i>	<i>FOXA2</i>	<i>KDR</i>
<i>MEIS2</i>	<i>FGF7</i>	<i>OR7E38P</i>	<i>PTPN7</i>	<i>CPQ</i>
<i>MYD88</i>	<i>RAPGEF4</i>	<i>TARS3</i>	<i>FSIP2</i>	<i>GNB3</i>
<i>LBR</i>	<i>CXCR4</i>	<i>NECTIN2</i>	<i>DENND10P1</i>	<i>COL6A3</i>
<i>IRAK3</i>	<i>ISOC1</i>	<i>CXCL17</i>	<i>PHF12</i>	<i>GPR137B</i>
<i>CYP2C18</i>	<i>COBL</i>	<i>TENT4B</i>	<i>ZNF114</i>	<i>NMNAT3</i>
<i>POU3F3</i>	<i>KIF13B</i>	<i>DHRS7</i>	<i>CCPG1</i>	<i>CYP2C9</i>
<i>SLITRK5</i>	<i>MYOF</i>	<i>TEC</i>	<i>GRIK1-AS1</i>	<i>PRSS1</i>
<i>UNC13C</i>	<i>TMX4</i>	<i>PIK3CG</i>	<i>ZNRF1</i>	<i>PPFIA3</i>
<i>SGMS2</i>	<i>FAM83B</i>	<i>TSPO</i>	<i>OR7E11P</i>	<i>ATF7IP2</i>
<i>SIGLEC8</i>	<i>SLC4A7</i>	<i>EVI2B</i>	<i>ADARB2</i>	
<i>RGS18</i>	<i>BOC</i>	<i>ABCA8</i>	<i>FAM126A</i>	
<i>TFEC</i>	<i>TFAP2A</i>	<i>TM6SF1</i>	<i>ROBO2</i>	
<i>DMRTA1</i>	<i>LIN28B</i>	<i>RAMP2-AS1</i>	<i>PELI2</i>	
<i>PRB1</i>	<i>ZNF474</i>	<i>INPPL1</i>	<i>ADA2</i>	

Supplementary Table 6 – 177-signature gene list

177 - signature				
COL6A3	CYP2C18	TSP0	EVL	SLC4A7
COL6A2	SPSB4	P3H3	STARD9	LINC01016
COL6A1	RAPGEF4	NLRC3	C9orf72	RAMP2-AS1
DNMT3B	LINC00887	FTCDNL1	LRRK1	LTK
CTBS	ZNF474	AGR2	OGFRL1	LTB
NLRP11	TFAP2A	PI16	HLX	SPRY2
TLR2	EVI2B	POU3F3	FAM30A	PIK3CG
HOXC-AS3	BTG3	TMX4	KDR	DMRTA1
RARA-AS1	TLE4	FOXB1	ANO2	MNX1-AS1
MAF	RASAL3	EFCC1	ADGRA3	VWF
DHRS7	HEXA	LINC00299	ZBTB20	ARSD
CCDC71L	CSRNP3	LINC00298	DSG2	TTC39C
WNT5B	ASL	ACOT9	NUDT4B	CRTAP
PAPSS1	SFTPA2	DAP	PDE6G	KRT2
NCOA7	SFTPA1	INPPL1	FAM171B	WIPF1
NETO1	MYD88	AGA	UNC13C	C2CD2
NRN1	TES	TMEM92	ADA2	JAM3
SH3BGR2	YPEL2	ADGRL1	LIN28B	NETO1-DT
SPRR2G	SIGLEC17P	TFEC	ATF7IP2	SIGLEC1
HERPUD2	MEIS2	LINC01694	LRRC28	GIMAP6
FKBP9	LBR	IRAK3	QSOX1	CSGALNACT2
ERMP1	RAB13	SH3KBP1	BOC	TFAP2A-AS1
FER	NDRG1	LINC00240	ISOC1	CTTNBP2
STAR	ST3GAL6	PPP1R13L	BAMBI	PRNP
SGMS2	PDZD2	TAS2R10	OAZ2	MFGE8
TARS3	SIGLEC8	WDFY1	GSTA4	TP53BP2
LONRF3	SIGLEC6	ACVR1B	TWSG1	CYP2C9
PIK3C2B	RGS18	ADGRG5	ABCA9	PRSS2
PKM	PTGER2	BTBD9	ABCA6	MMP9
MEGF9	HOXC9	NUDT4	CDCP1	
SH2D1A	LINC02573	KRT74	LPIN1	
ECHDC3	PGLYRP4	KRT73	RASGEF1A	
FAM83B	PGLYRP3	KRT72	CRYZL2P	
LRMDA	CDIP1	ABCD3	H2BC4	
CNTNAP5	PRB2	WASIR2	MAPK10	
GALNT18	PRB1	WASIR1	KIF13B	
SLITRK5	ZNF366	MNX1	HPGD	

Methods

1. Cell cultures

The K562 leukaemia cell line was grown in RPMI1640 (Gibco, Paisley, UK), 10% foetal bovine serum (FBS) (Gibco) and 1% penicillin/streptomycin antibiotics (100 U/mL/100 µg/mL, Gibco). Peripheral blood mobilised CD34+ haematopoietic stem and progenitor cells (HSPCs) were obtained from UCL, Institute of Child Health, Infection, Immunity and Inflammation Programme (London, UK), purchased from commercially available apheresis products. CD34+ HSPCs were cultured in StemSpan medium (StemCell Technologies, Cambridge, UK) supplemented with 100 ng/ml SCF, 100 ng/ml FLT3-L, 20 ng/ml TPO, and 20 ng/ml IL-3 (Peprotech, London, UK). Cell lines were tested for Mycoplasma contamination using MycoSensor PCR Assay Kit (Agilent, Didcot, UK).

2. CRISPR/Cas9 editing

Sequences for human *MX1* (accession number NC_000007.14) and *ETV6* (accession numbers NC_000012.12) were retrieved for gRNA design by the Invitrogen TrueDesign Genome Editor tool (threshold at 1 nucleotides of mismatch tolerance). gRNA sequences and the adjacent PAM sequence used in this study are reported in **Supplementary Table 1**. GeneArt Precision gRNA Synthesis Kit (Invitrogen, Inchinnan, UK) was used to synthesise customised gRNAs via in vitro transcription (IVT). The transfection of the gRNA and Cas9 endonuclease was achieved by assembly and delivery of an RNP complex via electroporation using the Neon Transfection System Kit (Invitrogen). The RNP complex was assembled at 1:2 molar ratio of TrueCut™ Cas9 Protein v2 (Invitrogen) to synthesised gRNA, and incubated at 37°C for 15 minutes prior to electroporation. Clones harbouring the translocation were isolated by single cell cloning by limiting dilution and PCR screening.

3. Amplification of t(7;12) fusion junctions by polymerase chain reaction (PCR)

PCR on genomic DNA was performed using High-Fidelity Phusion polymerase (Invitrogen) using manufacturer's instructions. PCR primers (**Supplementary Table 2**) were designed using Primer-BLAST (NCBI). The complete genomic sequences of human *MX1* and *ETV6* were obtained from the NCBI Gene and Nucleotide Database with accession numbers NC_000007.14 for *MX1* and NC_000012.12 for *ETV6*. The amplified fragments of the correct estimated size for K562 Mut3 were excised, purified from the gels, and sequenced (**Supplementary Figure 1**).

4. Fluorescence in situ hybridisation (FISH)

Preparation of metaphase chromosomes was performed by colcemid treatment (0.05 µg/mL) as described (Federico *et al.*, 2019) and fixed in methanol:acetic acid. FISH was performed using commercially available (MetaSystems, Altlußheim, Germany) and PAC-derived probes (**Supplementary Table 3**). PAC-derived probes were extracted from bacterial cultures and fluorescently labelled by nick translation (Roche, Mannheim, Germany). Following hybridisation and washes according to published protocols (Federico *et al.*, 2019), images were captured using a Leica DM4000 fluorescence microscope (Leica Microsystems, Wetzlar, Germany). A minimum of 200 nuclei were captured for analysis. Representative

images of K562 control and K562-t(7;12) nuclei hybridised with the probe combinations and the interpretation of signal patterns are shown in **Supplementary Figure 2**.

5. RNL analysis

Radial nuclear location (RNL) was calculated using 2D analysis of nuclei hybridised by FISH as previously described (4). The RNL is defined as a numerical value corresponding to the positioning of a FISH signal as the ratio of the nuclear radius. RNL ranges from 0 to 1, with 0 marking the interior and 1 marking the outer extreme. A minimum of 200 nuclei per condition are analysed; RNL is expressed taking into consideration the median value of all signals \pm confidence interval, with a 0.650 and lower defining the nuclear interior (4).

6. Real-time quantitative PCR (qPCR)

Total RNA was extracted using the Monarch Total RNA Miniprep Kit (New England Biolabs, Hitchin, UK), from which complementary DNA (cDNA) was synthesised by reverse transcription using High-Capacity RNA-to-cDNA Kit (Applied Biosystems, Waltham, US). qPCR was performed using FastGene 2x IC Green Universal qPCR Mix (fluorescein) (Nippon Genetics, Düren, Germany) for each gene with primers listed in **Supplementary Table 4**. Differential gene expression was calculated using the $\Delta\Delta C_t$ method. *HPRT-1* was used as endogenous reference gene.

7. Colony forming assay

Cells were plated onto semi-solid methylcellulose for colony-forming assays (CFC), in Methocult H4434 Classic (StemCell Technologies, Cambridge, UK). Cells were first suspended in IMDM (Gibco) + 20% FBS (Gibco) at the desired concentration (1000 – 10000 cells / plate). Plates were incubated at 37°C and 5% CO₂ for 10-12 days, when colonies were scored.

8. Long-Term Culture Initiating Cell Assay (LTC-IC)

The MS5 stroma cells were plated in a 96-well plate at a density of 30 000 cells per plate. 45 000 CD34+ HSCPs were added to each well in 100 μ l of H5100 Myelocult (StemCell Technologies), 1 μ M hydrocortisone (StemCell Technologies), 20 ng/ml of IL-3, TPO, G-CSF, and SCF (PeproTech). The medium was replaced weekly and the assay was conducted for 4 weeks. At the end of the 4th week, 100 μ l of Methocult H4434 (StemCell Technologies) was added to each well, and colonies were scored after 10 days. If colonies were present, the methylcellulose-based culture was dissociated by extensive resuspension and washes in PBS and cells were replated onto MS5 stroma for a further 5 week period, followed by methylcellulose addition.

9. Erythroid differentiation assay

K562 cells were subjected to erythroid differentiation by induction by DMSO (Fisher Scientific, Paisley, UK) over a period of 6 days, as described (2). 100 000 cells were seeded. At E0, 1.5% DMSO was added onto the culture medium (RPMI + 10% FBS + 1% P/S). RNA was collected at E0, E2, and E6 for subsequent qPCR analysis of erythroid gene markers *HBB*, *EPOR*, and *TAL1*.

10. RNA sequencing analysis

RNA sequencing was performed by MacroGen Europe BV (Amsterdam, Netherlands). A minimum of 1 µg (20 ng/µl) of high-quality total RNA (extracted using Monarch Total RNA Miniprep Kit, NEB) was supplied for sequencing. MacroGen Europe BV constructed libraries using Illumina Truseq Stranded Total RNA library preparation with Ribozero rRNA depletion, and performed sequencing was performed on a Novaseq 6000 platform, at 50M paired-end reads per sample. RNA-seq results in the form of fastq raw reads were analysed with the open source software package of the Tuxedo Suite. Tophat2 with Bowtie2 were used to map paired-end reads to the reference Homo sapiens genome build GRCh38 (5,6). GENCODE38 (7) was used as the reference human genome annotation. Aligned reads were filtered by quality using samtools (8) with a minimum selection threshold set at 30. Transcript assembly and quantification was achieved using Cufflinks (9). Differential expression between sample and control was performed by collapsing technical replicates for each condition and the use of the Cuffdiff tool (9). The differential expression was expressed in the form of log2 fold change between sample and control, and deemed statistically significant by a lower p value of 0.05 and false discovery rate (FDR) of 0.1.

11. Gene Ontology analysis

Gene ontology (GO) analysis was performed in ExpressAnalyst (available at www.expressanalyst.ca) using PANTHER Biological Process (BP) and Molecular Function (MF) repositories. GO terms and pathways were filtered by p value and false discovery rate (FDR) with a cut-off of ≤ 0.05 for meaningful association.

12. Statistical analysis and visualisation

Statistical significance was calculated by two-tailed Student's t-test on available replicates (minimum n=3). Variance is represented in barcharts by error bars of \pm standard deviation (SD). A p value ≤ 0.05 was considered statistically significant, unless otherwise stated. Statistical significance is symbolised as: p value ≤ 0.05 (*), 0.001 (**), 0.0001 (***), and 0.00001 (****). Calculations were performed in R studio version 4.1.0. Graphs were generated in Microsoft Excel 2013 and in R studio 4.1.0 using the libraries ggplot2 (v 3.3.5) and ggrepel (v 0.9.1). Multiplex Venn diagrams were generated on VennPainter (10).

13. Patient signature methods and data availability

Clinical phenotype and expression data (in counts units) were extracted from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET, <https://ocg.cancer.gov/programs/target>) available in the GDC TARGET-AML cohort last accessed on 3rd June 2022 from the University of California Santa Cruz (UCSC) Xena public repository (11). Microarray data in RMA units of expression for the TARGET AML cohort, including normal bone marrow data, was downloaded from the NCI TARGET Data portal. Microarray data in RMA units of expression for paediatric AML from Balgobind *et al.* (12) (under accession number GSE17855) was downloaded from GEO Accession Viewer. Data normalisation across arrays was achieved by Training Distribution Matching (TDM) using the R library TDM v 0.3 (13) and any remaining batch effects were removed by Limma

removeBatchEffect. Differential expression analysis was performed using Limma for R (v 3.48.3). To construct the gene signatures, all differentially expressed genes for each comparison below a p value of 0.01 were included and intersected accordingly using Venn diagrams on VennPainter (10).

14. Gene Set Enrichment Analysis (GSEA)

Custom gene signatures were used as gene sets for GSEA analysis (Subramanian *et al.*, 2005) on the GSEA software v4.2.3 on the RNA sequencing expression values in counts units of K562-t(7;12) against K562 control. GSEA was ran in 10 000 permutations on gene set using the weighted Signal2Noise metric. Enrichment plots and leading edge heatmaps were included as generated from the GSEA software.

Supplementary References

- (1) Kamel YM, Naiel A, Alshehri A, Vetter M, Saccone S, Anderson R, et al. Fluorescence in situ hybridisation assays designed for del (7q) detection uncover more complex rearrangements in myeloid leukaemia cell lines. 2014.
- (2) Arede L, Foerner E, Wind S, Kulkarni R, Domingues AF, Giotopoulos G, et al. KAT2A complexes ATAC and SAGA play unique roles in cell maintenance and identity in hematopoiesis and leukemia. *Blood advances* 2022;6(1):165-180.
- (3) Gulino GM, Bruno F, Sturiale V, Brancato D, Ragusa D, Tosi S, et al. From FISH to Hi-C: The Chromatin Architecture of the Chromosomal Region 7q36. 3, Frequently Rearranged in Leukemic Cells, Is Evolutionary Conserved. *International Journal of Molecular Sciences* 2021;22(5):2338.
- (4) Federico C, Cantarella CD, Di Mare P, Tosi S, Saccone S. The radial arrangement of the human chromosome 7 in the lymphocyte cell nucleus is associated with chromosomal band gene density. *Chromosoma* 2008;117(4):399-410.
- (5) Kim PG, Albacker CE, Lu YF, Jang IH, Lim Y, Heffner GC, et al. Signaling axis involving Hedgehog, Notch, and Scl promotes the embryonic endothelial-to-hematopoietic transition. *Proc Natl Acad Sci U S A* 2013 Jan 8;110(2):E141-50.
- (6) Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods* 2012;9(4):357-359.
- (7) Frankish A, Diekhans M, Ferreira A, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* 2019;47(D1):D766-D773.
- (8) Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25(16):2078-2079.
- (9) Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* 2012;7(3):562-578.
- (10) Lin G, Chai J, Yuan S, Mai C, Cai L, Murphy RW, et al. VennPainter: a tool for the comparison and identification of candidate genes based on Venn diagrams. *PloS one* 2016;11(4):e0154315.
- (11) Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* 2020;38(6):675-678.
- (12) Balgobind BV, Hollink IH, Arentsen-Peters ST, Zimmermann M, Harbott J, Beverloo HB, et al. Integrative analysis of type-I and type-II aberrations underscores the genetic heterogeneity of pediatric acute myeloid leukemia. *Haematologica* 2011 Oct;96(10):1478-1487.
- (13) Thompson JA, Tan J, Greene CS. Cross-platform normalization of microarray and RNA-seq data for machine learning applications. *PeerJ* 2016;4:e1621.