Evolutionary trajectories and cell-of-origin of *TCF3-ZNF384* and *PTPN11* mutations revealed by genome-wide multi-layer OMICs in a pair of infant monozygotic twins with concordant B-cell ALL

Clara Bueno[1,2,*], J Ramón Tejedor[3,*], Rachael Bashford-Rogers[4,*] Laura González-Silva[5,*], Rafael Valdés-Mas[6], Antonio Agraz-Doblás[1,2,5], Rafael Díaz de la Guardia[1], Chrystele Bilhou-Nabera[8], Nassera Abermil[8], Hélène Guermouche[8], Mario F Fraga[9], Agustín F Fernández[3], Paola Ballerini[7,£], Ignacio Varela[5,£], Pablo Menendez[1,2,10,£].

[1]Josep Carreras Leukemia Research Institute and Department of Biomedicine, School of Medicine. University of Barcelona. Barcelona. Spain.
[2]Centro de Investigación Biomédica en Red de Cáncer (CIBER-ONC), ISCIII, Barcelona, Spain.
[3]Institute of Oncology of Asturias (IUOPA), ISPA-HUCA.
[4]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.
[5]Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria-CSIC, Santander, Spain.
[6]Dreamgenics S.L. Oviedo. Spain.
[7]Pediatric Hematology, A. Trousseau Hospital, Paris, France.
[8]Sorbonne Université, Inserm, Centre de Recherche Saint-Antoine, CRSA, AP-HP, Hôpital Saint-Antoine, Hématologie Biologique, F-75012 Paris, France
[9]Nanomaterials & Nanotechnology Research Center (CINN-CSIC), Universidad de Oviedo, Asturias, Spain.
[10]Instituciò Catalana de Recerca i Estudis Avançats (ICREA). Barcelona. Spain.

Running Title: Natural history of TCF3-ZNF384 and PTPN11 mutations in B-cell ALL

Key words: B-cell ALL, monozygotic twins, TCF3-ZNF384, PTPN11, BCR, prenatal origin, genome-wide sequencing

*These authors equally contributed and should be considered first co-authors.
£These authors equally contributed and should be considered senior co-authors.

Correspondence should be addressed to:

Paola Ballerini MD & Pablo Menendez PhD
Josep Carreras Leukemia Research Institute
School of Medicine. University of Barcelona.
Casanova 143, 08036. Barcelona. Spain.
paola.ballerini@aphp.fr ; ignacio.varela@unican.es; pmenendez@carrerasresearch.org

ABSTRACT

B-cell acute lymphoblastic leukemia (B-cell ALL) is the most common cancer in childhood. Concordance of B-cell ALL in monozygotic twins results from intraplacental spread of an initiated pre-leukemic clone. Studying identical twins with B-cell ALL provides a unique and tractable model for deciphering the developmental timing of pre- and post-natal mutations contributing to clonal evolution. To date, this has mainly focused on major cytogenetic subgroups of childhood B-cell ALL, including MLL fusions, ETV6-RUNX1, hyperdiploidy and BCR-ABL1. However, formal demonstration of the prenatal origin and "backtracking" the natural history of the leukemia remains understudied in "B-other"/Normal Karyotype (NK) B-cell ALL. Here, we performed whole genome DNA-, B-cell receptor (BCR)-, and DNA bisulphite-sequencing on a pair of 8-month-old monozygotic twins diagnosed with concordant "B-other"/NK B-cell ALL. We show that t(12;19)/*TCF3-ZNF384* is an initiating prenatal lesion shared by both leukemic twins, and different *PTPN11* mutations represent postnatal secondary hits for overt B-cell ALL. Moreover, both leukemic twins displayed a strikingly similar genome-wide methylome landscape. Deep-sequencing analysis of BCR repertoires suggests a primitive fetal hematopoietic progenitor with a germline or an incompletely rearranged (pre-VDJ) IgH locus as the cell-of-origin for *TCF3-ZNF384* and *PTPN11* mutations. Our data support a parallel and convergent genetic and epigenetic evolution of this concordant B-cell ALL.

INTRODUCTION

B-cell acute lymphoblastic leukemia (B-cell ALL) is the most common cancer in childhood, and is molecularly very heterogeneous [1]. Identical twins with concordant B-cell ALL have provided a unique and tractable model to delineate the natural history and clonal evolution of the leukemia [1, 2, 3]. These studies led to the "two hit" model for pediatric B-cell ALL elucidated by Greaves and colleagues [4], which provides unambiguous evidence that an initiating alteration, often occurring *in utero*, generates a preleukemic clone, which eventually gives rise to an overt leukemia upon acquisition of secondary cooperating mutations [5]. Importantly, genome-wide sequencing has revealed a strikingly silent mutational landscape in infant and pediatric B-cell ALL, further indicative of a developmental cancer of prenatal origin with a short window of time to accumulate molecular alterations [6, 7, 8]. Furthermore, a complex branching, rather than lineal, model of mutation acquisition in B-cell ALL is now well-accepted [7, 9, 10].

The stepwise molecular pathogenesis of infant/pediatric B-cell ALL following prenatal initiation from a common clone *in utero* has been demonstrated for major cytogenetic subgroups of infant/pediatric B-cell ALL including MLL fusions [11, 12, 13], ETV6-RUNX1+ [14, 15, 16], hyperdiploid [5], BCR-ABL1+ and Ikaros-mutated [17]. However, formal demonstration of the prenatal origin and "backtracking" the natural history of the leukemia remains understudied in the major subgroup, so-called "B-other" ALL. This subgroup accounts for ~30% of *de novo* diagnosis, and is molecularly and cytogenetically characterized by a normal karyotype (NK) and the absence of fusion oncogene-encoding chromosomal translocations, aneuploidy and copy

number alterations and/or mutations recurrently observed in B-cell ALL, thus challenging the use of a patient-specific molecular tag to backtrack the disease pathogenesis/origin [1, 18, 19, 20, 21].

Predicated on the hypothesis that shared identical mutations are prenatal in origin and twin-specific mutations are likely to be postnatal and secondary, we performed genome-wide multi-layer OMICs analysis, including whole genome DNA (WG-seq), B-cell receptor (BCR-seq), and whole genome DNA bisulphite (WGB-seq) sequencing on a pair of 8-month-old monozygotic twins diagnosed with concordant B-other/NK B-cell ALL. We show that t(12;19)/*TCF3-ZNF384* is an initiating prenatal lesion shared by both leukemic twins, while different *PTPN11* mutations represent secondary postnatal hits necessary for overt B-cell ALL. Genome-wide DNA methylation analysis of leukemic cells and fetal liver (FL)-derived healthy B-cell progenitors (FL-BCPs) of both twins revealed a highly similar methylome landscape. Deep sequencing analysis of BCR repertoires suggested a primitive fetal progenitor with a germline or an incompletely rearranged (pre-VDJ) IgH locus as the cell-of-origin for *TCF3-ZNF384* and *PTPN11* mutations in these leukemic monozygotic twins. Our data support a parallel and convergent genetic and epigenetic evolution of this concordant B-cell ALL.

PATIENTS AND METHODS

Clinico-biological features of twin patients

A pair of 8-month-old male monozygotic twins (referred hereafter as twin A and twin B) were diagnosed in July 2017 at the Armand Trousseau Hospital, Paris, with concordant B-cell ALL. Both twins showed a diploid karyotype with an uncharacterized addition of genomic material in chromosome 12p13 (**Fig S1A**, **Table 1**). Cytogenetic, fluorescence *in situ* hybridization (FISH) and molecular analysis ruled out recurrent chromosomal translocations/fusion oncogenes (aneuploidy, 11q23/MLL, t(9;22)/BCR-ABL1, t(12;21)TEL-AML1, t(1;19)/E2A-PBX1) and gene mutations (*RAS, DUX4, PAX5, IKZF, CDKN, ERG*, among others) associated with B-cell ALL (**Table 1**). Flow cytometry revealed a pro-B immunophenotype (CD10-CD34+CD19+, **Fig S1B**). Peripheral blood (PB) diagnostic blasts were FACS-purified (purity >98%, data not shown), and genomic DNA was extracted from 2×10^6 leukemic blasts using the Maxwell RSC Blood DNA kit (Promega), and quantified with a NanoDrop Spectrophotometer. To confirm the somatic nature of the mutations, both PB mononuclear cells at complete remission (CR) and BM-derived mesenchymal stem/stromal cells were used as controls. **Table 1** shows the clinic-biological features of both twins. For genome-wide methylome profiling, CD34+CD19+ FL-BCPs and leukemic blasts from two unrelated infants diagnosed with NK B-cell ALL were used [6, 22, 23]. The parents of the leukemic twins signed an informed consent and the study was approved by the Institute Review Board of the Clinic Hospital of Barcelona (HCB/2013/8539) for using leftover material for this study. Methods on WG-seq, BCR-seq, WGB-seq, and bioinformatic analyses are detailed in **Supplemental Methods** and **Table S1**.

RESULTS

TCF3-ZNF384 is an initiating prenatal lesion in infant monozygotic twins with concordant B-cell ALL

Routine cytogenetic and molecular diagnostics of both monozygotic twins with concordant B-cell ALL revealed the absence of recurrent chromosomal translocations/fusion oncogenes and gene mutations associated with B-cell ALL (**Table 1**, and data not shown), thus classifying the leukemia as B-other/NK B-cell ALL [8, 20]. We thus decided to perform high-coverage WG-seq in purified diagnostic leukemic blasts, and PB cells at CR from both monozygotic twins to comprehensively characterize the developmental timing of acquired somatic mutations in B-cell ALL. We obtained a minimum of 360 Gb of sequence/sample (>100× depth); more than 90% of sequence showed a quality score >30 and >99% of reads mapping to the human reference genome (**Table S2**). We found a total of 227 and 261 somatic mutations in twin A and twin B, respectively (**Fig 1A**, **Table S3**), which represents a very low mutational load (~0.07 mut/Mb) and is in line with the silent mutational landscape reported for pediatric B-cell ALL [6, 7]. Notably, as many as 75 somatic mutations were shared between both twins, suggesting a common prenatal origin of the leukemia (**Fig 1A, Fig 2**). Of these 75 somatic mutations, only three affected coding regions: a t(12;19)(p13;p13) encoding the *TCF3-ZNF384* fusion gene (**Fig 1A,B**) [22], and a point mutation in both *SDAD1* and *UQCR11*, predicted to produce deleterious effects on the protein (**Fig 1A, Fig 2**). These prenatal coding mutations showed a mutant allele frequency (MAF) near 50%, thus representing initial events in the original leukemic clone. A rearrangement of the ZNF384 gene was shown by FISH analysis, with its partner located on chromosome 19, presumably the TCF3 gene. The *TCF3-ZNF384* fusion gene was confirmed by ~~FISH and~~

reverse transcriptase-polymerase chain reaction (RT-PCR) (**Fig 1C-E**), and the other coding mutations were validated using orthogonal sequencing strategies (**Table S4**).

Parallel and convergent evolution of B-cell ALL in both twins

In addition to the 75 shared mutations, we found 152 and 186 somatic mutations exclusive for twin A and twin B, respectively, suggesting that, after birth, the leukemia evolved in parallel and twin-specific mutations accumulated postnatally (**Fig 1, Fig 2, Table S3**). Regarding coding mutations, twin B specifically acquired a non-synonymous mutation in *WDR87*. A non-synonymous mutation in CAPN6 was also originally identified in twin A but failed to be further validated. Of note, we found that each twin acquired a distinct nonsynonymous mutation in *PTPN11*, a known tumor suppressor gene in pediatric B-cell ALL [24] (**Fig 1,2, Fig S2**). Both coding mutations in *PTPN11* were subclonal and were predicted by SHIFT and POLYPHEN algorithms to have a deleterious impact on protein function, likely representing postnatal secondary oncogenic hits cooperating with the *TCF3-ZNF384* fusion gene for overt leukemia development. Thus, our data suggest a convergent evolution in both leukemias as a result of a strong selection to inactivate *PTPN11*.

Genome-wide DNA methylome profiling reveals a highly similar epigenetic landscape in infant monozygotic twins with concordant B-cell ALL.

To further explore the evolutionary traits of the concordant B-cell ALL in both twins, we used WGB-seq to survey the DNA methylation status of concordant B-cell ALLs. An average of 18.5 million CpG dinucleotides per sample were measured (>10× sequencing coverage), and >2.5 million of the CpG sites were shared across all interrogated samples (**Table S5**). B-cell ALL cells from both monozygotic twins were

significantly hypomethylated as compared with their healthy FL-BCP counterparts (**Fig 3A**). In fact, among the ~190,000 and ~176,000 differentially-methylated CpG (dmCpG) sites found for twin A and B, respectively, the number of hypomethylated CpG sites largely outperformed the number of hypermethylated ones (**Fig 3B**). However, we observed very limited DNA methylation changes between the twins' leukemic cells (**Fig 3B**, green bars). Indeed, both twins displayed a very similar enrichment of dmCpG sites, both hyper- and hypomethylated, with regard to their CpG context (**Fig 3C**) or CpG genomic location (**Fig 3C-E**). The relatively few dmCpG sites observed between twins might reflect interindividual variability rather than a divergent epigenetic evolution of the leukemia, reinforcing a parallel and convergent genetic but also epigenetic evolution of the B-cell ALL in this pair of monozygotic twins.

Whether pathogenic chromosomal rearrangements and/or protein-coding DNA mutations in cancer are associated with chromatin and/or DNA methylation changes remains elusive. We therefore wanted to gain insights into the methylation status of the loci specifically found rearranged or mutated in both twins (*TCF3*, *ZNF384*, *SAD1*, *WRD87*, and *PTPN11*). Strikingly, in contrast to healthy FL-BCPs and B-cell ALL cells from unrelated infant NK B-cell ALL, we found that B-cell ALL cells from both twins displayed concordant and specific significant DNA hypomethylation in both *TCF3* and *ZNF384* loci, in the vicinity of the translocation genomic breakpoints (**Fig 3F,G**). DNA hypomethylation occurred in both CpG-rich areas surrounding the promoter regions and also towards the 3' end of the *TCF3* gene (**Figure 3F,G**). Similarly, B-cell ALL cells from both twins also showed significant DNA hypomethylation within the promoter region of *SDAD1*, *WDR87* and *PTPN11*, coinciding with CpG islands and CpG shore regions (**Fig S3A-D**). *PTPN11* also showed specific hypomethylation on its own gene body (**Fig S3D**). Of note, DNA methylation changes were not observed in *MLL* or *TP53* loci (**Fig S3E,F**), further supporting the specificity of the DNA hypomethylation observed in the loci rearranged/mutated in both twins with concordant leukemia. These

data suggest that the pathogenic mutations may trigger a cellular response towards compensating/enhancing gene expression of affected genes.

Deep-sequencing analysis of BCR repertoires suggests an early pre-VDJ common progenitor as the cell-of-origin for t(12;19)/*TCF3-ZNF384* and *PTPN11* mutations in B-cell ALL monozygotic twins.

We next analyzed BCR repertoires to gain insights into the clonal composition of immunoglobulin heavy chain (IgH) rearrangements in diagnostic samples from both monozygotic twins. BCRs are generated during B-cell differentiation and represent unique markers for each B-cell clone. Because each BCR sequence provides a molecular tag for a different B-cell clone, BCR-seq of the IgH locus provides a detailed view of the B-cell population dynamics and clone tracking [6, 25]. We therefore performed BCR-seq on *TCF3-ZNF384*+ B-cell ALL PB blasts from the monozygotic twins to address whether they expressed fully rearranged BCRs from which major B-cell clonal expansion may be observed. We employed a PCR-based method with additional incorporation of unique molecular barcodes, allowing for accurate quantitation of relative B-cell clone frequency [6, 25]. After BCR sequence filtering, each sample yielded 9562 and 22790 BCRs (6362 and 11832 unique BCRs) for twin A and B, respectively.

We first delineated the relative clonality of both twins and found that, similar to what has been reported for infant MLLr pro-B ALL, the BCR repertoires from infant twins with *TCF3-ZNF384*+ B-cell ALL did not exhibit significantly expanded VDJ rearranged B-cell clones (**Fig 4A,B**). This is in contrast to non-MLL, non-infant B-cell ALL including t(1;19)/TCF3-PBX1+, t(12;21)/ETV6-RUNX1+ and t(9;22)/BCR-ABL1+, which were all found significantly clonal, with large B-cell clones comprising 3–40% of total BCRs (**Fig 4A,B**) [6, 25]. Furthermore, we found that the BCR clonal overlap between the monozygotic twins was 3-fold higher (60 *vs* 20%) than between twins and healthy donors (**Fig 4C**). Because the *TCF3-ZNF384* rearrangement, and *SDAD1* and *UQCR11* mutations are clonal, and *PTPN11* mutations are found in major clones, the lack of

B-cell clonal expansion together with the leukemia phenotype developmentally stalled at the pro-B stage supports a model in which the genomic drivers in this pair of infant monozygotic twins with concordant B-cell ALL originate in a primitive fetal progenitor with a germline or an incompletely rearranged (pre-VDJ) IgH locus [6].

DISCUSSION

Genomic studies on twins with concordant B-cell ALL, which is relatively rare, have been very informative with respect to the identification of early and shared clonal oncogenic hits and secondary, postnatal events [2, 3]. The prenatal origin and the natural history of the leukemia has been elegantly backtracked in major cytogenetic subgroups of infant/pediatric B-cell ALL including MLL fusions [11, 12, 13], ETV6-RUNX1+ [14, 15, 16], hyperdiploid [5], BCR-ABL1+ and Ikaros-mutated [17]. However, it remains understudied in patients diagnosed with B-other/NK B-cell ALL, which is likely due, at least in part, to the absence of a prospective and tractable patient-specific genetic marker. Here, we applied WG-seq to delineate the natural history and evolutionary trajectory of concordant B-other/NK B-cell ALL in a pair of infant monozygotic twins. To the best of knowledge, this is the first study reporting that t(12;19)/*TCF3-ZNF384* is a clonal lesion shared by both leukemic twins, therefore arising *in utero* (**Fig 2**). *TCF3-ZNF384* is very likely to be an early prenatal initiating mutation in B-cell ALL. This view is supported by a recent study demonstrating that hematopoietic lineage determination is altered by the expression of TCF3-ZNF384 in human stem and progenitor cell populations, resulting in acute leukemia [26].*TCF3-ZNF384* fusion was reported to be present in 2.4% of childhood B-cell ALL (out of 291 B-cell ALL patients enrolled), and linked to a $CD10^{-/dim}$ pro-B phenotype [22], in line with the twins here reported. In addition to *TCF3-ZNF348*, we also found that mutations (predicted to produce deleterious effects on the protein) in *SDAD1* and *UQCR11*, two genes rarely found mutated in B-

cell ALL, are also shared by both leukemic twins with a MAF near 50%, thus representing initial mutations. Whether *SDAD1* and *UQCR11* represent leukemia-initiating events or are just passenger mutations present in the initial leukemia close needs further research.

A total of 152 and 186 twin-specific mutations were found, suggesting that once the twins separated after birth, the leukemia evolved in parallel and non-shared twin-specific mutations accumulated postnatally (**Figure 2**). Importantly, among these mutations, we found that each twin acquired a distinct non-synonymous coding mutation in *PTPN11*, suggesting a convergent evolution in both leukemias as a result of a strong selection to inactivate *PTPN11* (**Fig 2**). Both coding mutations in *PTPN11* were subclonal and predicted to generate a deleterious protein, likely representing postnatal secondary oncogenic hits cooperating with *TCF3-ZNF384* for leukemia development. Our work adds to the accumulating evidence that *PTPN11* mutations represent a secondary oncogenic event in childhood B-cell ALL, as previously reported for hyperdiploid B-cell ALL [24]. Somatically acquired *PTPN11* mutations occur in near 10% of childhood B-cell ALL patients [27], and are largely mutually exclusive with *NRAS* and *KRAS* mutations suggesting a signaling impairment through the RAS pathway [24]. A striking observation of our study was the same observed latency, since both twins were diagnosed the same day. The "identical latency" may be explained by the fact that both twins share the same coding mutations (*SDAD1, UQCR11*, and *PTPN11*) and *TCF3-ZNF384* fusion with a very high MAF (~50% for prenatal and 24–42% for postnatal mutations).

As far as we know, whole genome methylome studies have not previously been performed in monozygotic twins with concordant B-cell ALL. To further explore the evolutionary traits of the concordant B-cell ALL, we employed WGB-seq to profile DNA methylation, and found very limited changes between the twins' leukemic cells. The strikingly similar methylome landscape might reflect interindividual variability rather than

a divergent epigenetic evolution of the leukemia, reinforcing a parallel and convergent genetic but also epigenetic evolution of the B-cell ALL. Whether pathogenic chromosomal rearrangements and/or protein-coding DNA mutations in cancer are linked to chromatin and/or DNA methylation changes remains poorly understood. B-cell ALL cells from both twins displayed concordant and specific significant DNA hypomethylation in both *TCF3* and *ZNF384* loci, in the vicinity of the translocation genomic breakpoints, as well as in mutated *SDAD1* and *PTPN11* genes. These data are surprising and suggest that the initiating chromosomal rearrangements or secondary pathogenic DNA mutations may trigger a cellular response towards compensating/enhancing gene expression of affected genes. What to do with the hypomethylation observed in mutated genes and genes involved in the translocation? Should we keep (in results) and discuss (here) what these data would mean? Is this data strong and conclusive enough to be reported?

The leukemic twins showed a very low mutational load, which is in accord with the silent mutational landscape for pediatric B-cell ALL [6, 7]. Although *TCF3-ZNF384* and *PTPN11* by themselves may be sufficient to induce B-cell ALL, the contribution of such mutations in leukemogenesis should consider the nature of both the fetal target cell for transformation and the leukemia initiating cell, according to the increasingly accepted stochastic stem cell model of B-ALL [28, 29]. Using high-throughput BCR-seq of the IgH locus we delineated the dynamics-clonality of B-cell populations and found that, similar to infant MLL-rearranged pro-B-cell ALL, the twin's leukemic cells exhibited a BCR repertoire composed of thousands of minor, non-expanded VDJ rearranged IgH B-cell clones. Because *TCF3-ZNF384* is clonal and *PTPN11* mutations are found in clones of relatively large size, they are likely to originate in primitive fetal progenitors that have a germline or an incompletely rearranged (pre-VDJ) IgH locus [6, 25]. In sum, this is the first report on the natural history, evolutionary trajectories and cell-of-origin of *TCF3-ZNF384* and *PTPN11* mutations in a pair of infant monozygotic twins with concordant B-cell ALL.

AUTHOR CONTRIBUTION: C.B, J.R.T, R.B-R, L.G-S, R.V-M, A.A-D and R.D.dlG: performed experiments and analyzed data. C.B, A.F.F, M.F.F, C.B-N, N.A, P.B, I.V, and P.M conceived the study, designed experiments, and analyzed data. C.B, J.R.T, R.B-R, R.V-M, I.V and P.M contributed to artwork and wrote the manuscript.

ACKNOWLEDGMENTS. Financial support for this work was obtained from the European Research Council (CoG-2014-646903 and PoC-2018-811220 to PM, and StG-2014-637904 to IV), the Generalitat de Catalunya (SGR330 and PERIS 2017-2019) to PM, the Spanish Ministry of Economy and Competitiveness (SAF2016-80481-R and SAF2016-76758-R) to PM and IV, the Spanish Association against cancer (AECC-CI-2015) to CB, the Health Institute Carlos III (ISCIII/FEDER, PI17/01028) to CB, PM also acknowledges financial support from the Obra Social La Caixa-Fundaciò Josep Carreras. JRT is supported by a Juan de la Cierva Formación fellowship by the Spanish Ministry of Science and Innovation (FJCI-2015-26965) and IUOPA. RV-M is supported by a Torres Quevedo fellowship by Spanish Ministry of Science and Innovation (PTQ-16-08623). P.M is an investigator of the Spanish Cell Therapy cooperative network (TERCEL).

CONFLICT OF INTEREST DISCLOSURE: The authors have nothing to disclose.

REFERENCES

1. Hunger, S.P. & Mullighan, C.G. Redefining ALL classification: toward detecting high-risk ALL and implementing precision medicine. *Blood* 2015 **125**, 3977-3987.

2. Greaves, M.F., Maia, A.T., Wiemels, J.L. & Ford, A.M. Leukemia in twins: lessons in natural history. *Blood* 2003 **102**, 2321-2333.

3. Maia, A.T. *et al.* Protracted postnatal natural histories in childhood leukemia. *Genes Chromosomes Cancer* 2004 **39**, 335-340.

4. Greaves, M. Molecular genetics, natural history and the demise of childhood leukaemia. *Eur J Cancer* 1999 **35**, 1941-1953.

5. Bateman, C.M. *et al.* Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia. *Blood* 2010 **115**, 3553-3558.

6. Agraz-Doblas, A. *et al.* Unravelling the cellular origin and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis. *Haematologica* 2019.

7. Anderson, K. *et al.* Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* 2011 **469**, 356-361.

8. Mullighan, C.G. The genomic landscape of acute lymphoblastic leukemia in children and young adults. *Hematology Am Soc Hematol Educ Program* 2014 **2014**, 174-180.

9. Greaves, M. & Maley, C.C. Clonal evolution in cancer. *Nature* 2012 **481**, 306-313.

10. Notta, F. *et al.* Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature* 2011 **469**, 362-367.

11. Ford, A.M. *et al.* Fetal origins of the TEL-AML1 fusion gene in identical twins with leukemia. *Proc Natl Acad Sci U S A* 1998 **95**, 4584-4588.

12. Sanjuan-Pla, A. *et al.* Revisiting the biology of infant t(4;11)/MLL-AF4+ B-cell acute lymphoblastic leukemia. *Blood* 2015 **126**, 2676-2685.

13. Urtishak, K.A. *et al.* Unique Familial MLL(KMT2A)-Rearranged Precursor B-Cell Infant Acute Lymphoblastic Leukemia in Non-twin Siblings. *Pediatr Blood Cancer* 2016 **63**, 1175-1180.

14. Alpar, D. *et al.* Clonal origins of ETV6-RUNX1(+) acute lymphoblastic leukemia: studies in monozygotic twins. *Leukemia* 2015 **29**, 839-846.

15. Bungaro, S. *et al.* Genomic analysis of different clonal evolution in a twin pair with t(12;21) positive acute lymphoblastic leukemia sharing the same prenatal clone. *Leukemia* 2008 **22**, 208-211.

16. Wiemels, J.L., Ford, A.M., Van Wering, E.R., Postma, A. & Greaves, M. Protracted and variable latency of acute lymphoblastic leukemia after TEL-AML1 gene fusion in utero. *Blood* 1999 **94**, 1057-1062.

17. Cazzaniga, G. *et al.* Developmental origins and impact of BCR-ABL1 fusion and IKZF1 deletions in monozygotic twins with Ph+ acute lymphoblastic leukemia. *Blood* 2011 **118**, 5559-5564.

18. Galbiati, M. *et al.* Natural history of acute lymphoblastic leukemia in neurofibromatosis type 1 monozygotic twins. *Leukemia* 2013 **27**, 1778-1781.

19. Ma, Y. *et al.* Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* 2013 **110**, 7429-7433.

20. Moorman, A.V. New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia. *Haematologica* 2016 **101**, 407-416.

21. Zaliova, M. *et al.* Genomic landscape of pediatric B-other acute lymphoblastic leukemia in a consecutive European cohort. *Haematologica* 2019.

22. Hirabayashi, S. *et al.* ZNF384-related fusion genes define a subgroup of childhood B-cell precursor acute lymphoblastic leukemia with a characteristic immunotype. *Haematologica* 2017 **102**, 118-129.

23. Munoz-Lopez, A. *et al.* Cellular Ontogeny and Hierarchy Influence the Reprogramming Efficiency of Human B Cells into Induced Pluripotent Stem Cells. *Stem Cells* 2016 **34**, 581-587.

24. Molteni, C.G. *et al.* PTPN11 mutations in childhood acute lymphoblastic leukemia occur as a secondary event associated with high hyperdiploidy. *Leukemia* 2010 **24**, 232-235.

25. Bashford-Rogers, R.J. *et al.* Eye on the B-ALL: B-cell receptor repertoires reveal persistence of numerous B-lymphoblastic leukemia subclones from diagnosis to relapse. *Leukemia* 2016 **30**, 2312-2321.

26. Dickerson, K., Yoshihara, H., Janke, L. & Mullighan, C. Expression of TCF3-ZNF384 in Human Hematopoietic Cells Induces Lineage Disruption in Vitro and Acute Leukemia In Vivo. ASH Annual Meeting; 2018; San Diego, California. USA; 2018.

27. Tartaglia, M. *et al.* Genetic evidence for lineage-related and differentiation stage-related contribution of somatic PTPN11 mutations to leukemogenesis in childhood acute leukemia. *Blood* 2004 **104**, 307-313.

28. Elder, A. *et al.* Abundant and equipotent founder cells establish and maintain acute lymphoblastic leukaemia. *Leukemia* 2017 **31**, 2577-2586.

29. Rehe, K. *et al.* Acute B lymphoblastic leukaemia-propagating cells are present at high frequency in diverse lymphoblast populations. *EMBO Mol Med* 2013 **5**, 38-51.

FIGURE LEGENDS

Figure 1. Mutational landscape of B-cell ALL monozygotic twins. (A) Circos plot (http://circos.ca/) representing the total number of mutations identified at diagnosis for each twin. Twin A is shown in pink (inner ring). Twin B is shown in blue (outer ring). Chromosome idiograms are indicated inside the graph. Copy number alterations are represented as lines within the grey rings. Venn diagrams inside the graph show the total number of exclusive (pink, twin A; blue, twin B) and shared (yellow) somatic mutations. Coding mutations are annotated outside the graph according to genomic gene location. Substitution mutant allele frequency is shown in brackets. Genomic chromosomal rearrangements are represented with black lines connecting chromosomal breakpoints. **(B)** Left, schematic representation depicting the genomic breakpoints of the t(12;19)(p13;p13) determined at the base pair level. Right, Sanger sequencing electropherogram confirming the t(12;19)(p13;p13)/*TCF3-ZNF384* breakpoint. **(C-E)** Interphase FISH **(C)** and RT-PCR **(D,E)**, validating the *TCF3-ZNF384* fusion in the twin's leukemic blasts.

Figure 2. B-cell ALL natural history reconstruction in the monozygotic twins. Graphical reconstruction of the number of mutations accumulated during tumor history. Approximately one-third of the somatic mutations (shown in yellow), including the translocation responsible for the expression of the TCF3-ZNF384

fusion gene, occurs before birth when both twins shared the same circulatory system. Coding mutations are indicated. Leukemia in both twins evolved independently after birth with twin-specific postnatal mutations.

Figure 3. **Genome-wide DNA methylation analysis of B-cell ALL in monozygotic twins**. **(A)** Violin plots depicting the global DNA methylation levels (β-value distributions) of CpG sites identified by whole genome DNA bisulphite sequencing. For healthy B-cell progenitors (BCPs), the average DNA methylation value of two fetal liver-derived pools of BCP is represented. **(B)** Total number of differentially hyper- and hypomethylated CpGs observed for the indicated comparisons between twins and healthy BCPs. The inner graph represents the total number of dmCpG sites. **(C)** Stacked barplots indicating the relative proportion of significant hyper- and hypomethylated CpGs between twins, or between each twin and healthy BCPs, regarding their CpG context (*left panel*) or the CpG genomic location (*right panel*), respectively, as indicated in the colored legend. **(D,E)** Ideograms representing the genomic location of differentially DNA hypermethylated (**D**) and hypomethylated (**E**) CpG sites obtained for twin A (pink) and twin B (blue) as compared with healthy BCPs. **(F,G)** Colored line plots illustrate the DNA methylation status for *TCF3* and *ZNF384* genes in leukemic blasts from both monozygotic twins with B-cell ALL (pink, blue), two unrelated infants with Normal Karyotype B-cell ALL (yellow) and from healthy BCPs (green). CpG site location and CpG context are mapped to their corresponding genomic coordinates as represented in the lower panels. Areas of differential methylation between leukemic twins and other samples are orange-shaded. Areas where translocation genomic breakpoints in *TCF3* and *ZNF384* occurred are blue-shaded.

Figure 4. Analysis of BCR repertoires suggests a pre-VDJ common progenitor as the cell-of-origin for t(12;19)/*TCF3-ZNF384* in B-cell ALL monozygotic twins. (A) *Left panels*, cloud-plots of BCR repertoires of a representative healthy donor (HD) and an infant MLL-rearranged pro-B-ALL depicting the existence of many minor diagnostic non-expanded B-cell clones. *Middle panels*, cloud-plots of BCR

repertoires of a representative t(12;21)/TEL-AML1+ and t(1;19)/E2A-PBX1+ pediatric B-cell ALL showing high clonality of B-cell clones. *Right panels*, cloud-plots of BCR repertoires of TCF3-ZNF384-rearranged pro-B ALL in both monozygotic twins. Each vertex represents a unique BCR sequence, and relative vertex size is proportional to the number of identical reads. For the HD and B-cell ALL twin samples, each vertex is represented by a pie chart indicating the percentage of each isotype, where blue = IgD/M, red = IgA1/2, yellow = IgG1/2, green = IgG3, and grey = IgE. **(B)** Boxplots of the largest and second largest BCR clone size for HD, non-t(4;11)+ B-cell ALL, and twins with concordant B-cell ALL. **(C)** Histogram representing the number of BCR clones that are shared between twin A and twin B (blue) or between unrelated HD and the twins (red).

Figure S1. Karyotype and immunophenotype of diagnostic bone marrow cells from the monozygotic twins. (A) Representative apparently normal karyotypes of both twins. **(B).** Representative FACS dot plots confirming a CD10- pro-B immunophenotype of both twins.

Figure S2. Validation of postnatal secondary coding mutations by Sanger sequencing.

Figure S3. DNA methylation status of the indicated genes in leukemic blasts from both monozygotic twins with B-cell ALL, two unrelated infants with NK B-cell ALL and from healthy FL-derived B-cell progenitors. Colored line plots illustrate the CpG methylation landscape of *SDAD1* (**A**), *UQCR11* (**B**), *WDR87* (**C**), *PTPN11* (**D**), *MLL* (**E**) and *TP53* (**F**) genes. Healthy B-cell progenitors are shown in green. Twin A is shown in pink, twin B is shown in blue, and NK B-cell ALL is shown in yellow. CpG site location and CpG context are mapped to their corresponding genomic coordinates as represented in the lower panels. Areas of differential methylation between leukemic twins and other samples are orange-shaded.

Coding (exonic) mutations identified by WG-seq are represented with yellow triangle marks (when shared by both twins), pink triangle marks (specific for twin A), and blue triangle marks (specific for twin B).

3rt