







RESEARCH ARTICLE

Open Access



The use of PanDrugs to prioritize anticancer drug treatments in a case of T-ALL based on individual genomic data

Pablo Fernández-Navarro^{1,2†} , Pilar López-Nieva^{3,4,5†} , Elena Piñeiro-Yañez⁶ , Gonzalo Carreño-Tarragona⁷ , Joaquín Martínez-López⁷, Raúl Sánchez Pérez⁸, Ángel Aroca⁸, Fátima Al-Shahrour⁶ , María Ángeles Cobos-Fernández^{3,4} and José Fernández-Piqueras^{3,4,5*} 

Abstract

Background: Acute T-cell lymphoblastic leukaemia (T-ALL) is an aggressive disorder derived from immature thymocytes. The variability observed in clinical responses on this type of tumours to treatments, the high toxicity of current protocols and the poor prognosis of patients with relapse or refractory make it urgent to find less toxic and more effective therapies in the context of a personalized medicine of precision.

Methods: Whole exome sequencing and RNAseq were performed on DNA and RNA respectively, extracted of a bone marrow sample from a patient diagnosed with tumour primary T-ALL and double negative thymocytes from thymus control samples. We used PanDrugs, a computational resource to propose pharmacological therapies based on our experimental results, including lists of variants and genes. We extend the possible therapeutic options for the patient by taking into account multiple genomic events potentially sensitive to a treatment, the context of the pathway and the pharmacological evidence already known by large-scale experiments.

Results: As a proof-of-principle we used next-generation-sequencing technologies (Whole Exome Sequencing and RNA-Sequencing) in a case of diagnosed Pro-T acute lymphoblastic leukaemia. We identified 689 disease-causing mutations involving 308 genes, as well as multiple fusion transcript variants, alternative splicing, and 6652 genes with at least one principal isoform significantly deregulated. Only 12 genes, with 27 pathogenic gene variants, were among the most frequently mutated ones in this type of lymphoproliferative disorder. Among them, 5 variants detected in *CTCF*, *FBXW7*, *JAK1*, *NOTCH1* and *WT1* genes have not yet been reported in T-ALL pathogenesis.

Conclusions: Personalized genomic medicine is a therapeutic approach involving the use of an individual's information data to tailor drug therapy. Implementing bioinformatics platform PanDrugs enables us to propose a prioritized list of anticancer drugs as the best theoretical therapeutic candidates to treat this patient has been the goal of this article. Of note, most of the proposed drugs are not being yet considered in the clinical practice of this type of cancer opening up the approach of new treatment possibilities.

Keywords: T-ALL, Next-generation sequencing technologies, PanDrugs, Precision oncology, Personalized precision medicine, Translational bioinformatics, Cancer genomics, In silico prescription, Targeted therapy, Druggable genome

* Correspondence: jfpiqueras@cbm.csic.es

[†]Pablo Fernández-Navarro and Pilar López-Nieva are co-first authors

³Department of Cellular Biology and Immunology, Severo Ochoa Molecular Biology Center (CBMSO), CSIC-Madrid Autonomous University, Madrid 28049, Spain

⁴Institute of Health Research Jiménez Díaz Foundation, Madrid 28040, Spain

Full list of author information is available at the end of the article



Background

Acute leukaemia of the lymphoid lineage (ALL) is the most common form of childhood leukaemia. Based on the immunophenotype of the leukaemia cells we are able to classify ALL into T-cell acute lymphoblastic (T-ALL) and B-cell precursor (B-ALL) leukaemia. In particular, T-ALL is biologically and genetically heterogeneous with gene expression signatures that identify different biological and clinical subgroups associated with T cell arrest at different stages of thymocyte development [1], most often manifests with extensive diffuse infiltration of the bone marrow and blood involvement [2].

T-ALL results from a multistep transformation process in which accumulating genetic alterations co-ordinately disrupt key oncogenic, tumour suppressor and developmental pathways responsible for the normal control of cell growth, proliferation, survival and differentiation during thymocyte development [1]. Despite undoubted successes, the toxicity of intensified chemotherapies treatments, chemotherapy resistance and the outcomes of patients with relapsed or refractory ALL remain poor [1, 3]. It is therefore still necessary develop appropriate strategies to enable us to identify more effective, therefore, less toxic treatments taking into account the patient genetic profile. The application of Next-Generation Sequencing (NGS) techniques has produced an unprecedented body of knowledge concerning the molecular pathogenesis of these haematological disorders allowing the discovery of multiple genetic and epigenetic alterations underpinning tumour development.

Personalized medicine is gaining recognition due to limitations with standard diagnosis and treatment [4]; due to the high rates of variability observed in clinical responses to treatments, which probably reflects underlying molecular heterogeneity. Furthermore, new classes of molecularly targeted drugs have been developed [5] although its potential could still be better utilized. Identifying which genetic variants may be targetable by current therapies presents a difficult challenge in personalized cancer medicine [6]. The question raised in this work is whether the availability of molecular data provided by whole exome and transcriptome sequencing could serve to guide the selection of site-specific treatments in a patient with T-ALL as a proof of principle. We have used the bioinformatics platform PanDrugs [7] as a feasible method to address the gap between raw genomic data and clinical usefulness, identifying genetic abnormalities that can be matched to drug therapies that may not have otherwise been considered. This could be a challenge to the implementation and uptake of genomics-based screening and diagnosis to map the appropriate actions.

Methods

Primary tumour and control samples

The University Hospital 12 Octubre (Madrid, Spain) provided us a tumour primary T-ALL sample (bone marrow).

Tumour blasts were isolated from primary sample by flow cytometry sorting as CD7⁺ CD45⁺ cells. Sample was diagnosed as Pro-T acute lymphoblastic leukaemia according to World Health Organization Classification of Haematological Malignancies and recommendations from the European childhood lymphoma pathology panel.

Normalization next generation sequencing data is necessary to eliminate cell-specific biases prior to downstream analyses. Thymus control samples, were provide by La Paz University Hospital (Madrid, Spain). Due to Double Negative thymocytes (DN) are the less common fraction of cells multiplex these DN fractions by performing a single experiment on a pool of all DN cells, also pooling donors reduces variability. To create the initial pool of DN cells, isolation of thymocyte subpopulations were performed in five human paediatric thymuses of patients with only heart diseases aged 1 month to 4 years, removed during corrective cardiac surgery, using autoMACS Pro (Miltenyi Biotec) with appropriate MicroBeads. Immature thymocytes were enriched from thymocyte suspensions using the sheep red blood cell (SRBC) rosetting technique. Early progenitors (DN) were isolated as CD34⁺ cells. Purity was determined by flow cytometry using the following antibody: CD34-PE (MACS Miltenyi Biotec).

Whole exome sequencing (WES)

DNA extraction was performed using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. All isolated DNA samples were quantified by spectrophotometry, using NanoDrop (ThermoFisher Scientific, Waltham, MA, USA), and fluorimetry, using the Qubit[®] dsDNA HS and/or BR assay kits (ThermoFisher Scientific Inc.). WES analyses were performed with an Illumina HiSeq2000 sequencing platform using a paired end 2 X 100 read strategy and an Agilent's SureSelect Target Enrichment System for 71 Mb. Sequencing will be done with a 100x of coverage. Processing of the raw data was done using RubioSeq pipeline [8] where the reads were aligned against the last version of human genome reference (GRCh38/hg38 assembly) using the BWA-Mem algorithm [9]. Alignment was then processed to (i) realign known indel regions, (ii) remove duplicate reads, and (iii) recalibrate quality scores. The variant calling process for SNVs and Indels identification was done using the combined results from GATK [10] and MuTect2 [11]. Python scripts were developed to combine variants.

Variant annotations

Variants were annotated following the logic in PanDrugs, which integrates information from the Variant Effect Predictor of Ensembl [12] and additional databases. We used the versions 90 of Ensembl, 85 of COSMIC [13], and the releases 87.0 of KEGG [14], 1.53 of

ClinVar [15], 31.0 of Pfam [16], 2018_07 of UniProt (UniProt Consortium 2018) and 69.0 of InterPro [17]. Genes included in a list with the most frequently altered genes in T-cell lymphoblastic neoplasia were also indicated.

Massive mRNA sequencing

Total RNA was obtained using TriPure Reagent (Roche Applied Science, Indianapolis, IN, USA), following manufacturer's instructions. RNA integrity Numbers (RIN) were in the range of 7.2–9.8. Sequencing of tumour-derived mRNA (RNA-Seq) was analysed after filtering total RNA by removal of Ribosomal RNA. Libraries were sequenced using an Illumina HiSeq2500 instrument (Illumina Inc., San Diego, CA, USA). Estimation of RNA abundance was calculated with Cufflinks2.2.1 software using the Ensembl GRCh37/hg19p5 annotation for human genome. All these molecular analyses were performed by the Sequencing and Bioinformatics services of Sistemas Genómicos S.L. (Valencia, Spain; <https://www.sistemasgenomicos.com/en/>) in two replicates.

Identification of fusion transcripts and alternative splicing variants (ATEs)

Interpretation of RNA-Seq data using the predictive algorithm EricScript, a computational framework for the discovery of gene fusions in paired-end RNA-Seq data developed in R, perl and bash scripts. This software uses the BWA51 aligner to perform the mapping on the transcriptome reference and BLAT for the recalibration of the exon junction reference. In this study, we have used EricScript 0.5.5b and Ensembl GRCh37.73 as a transcriptome reference [18]. RNA-Sequencing data were also used to identify ATEs using CUFFLINKs [19].

PCR, sanger sequencing

Polymerase-Chain-Reaction (PCR) and Sanger sequencing were used to validate novel mutations. Sanger DNA sequencing of PCR-amplified fusion sequences were performed with the specific primers indicated in Additional file 1: Table S1.

PanDrugs

PanDrugs (<http://www.pandrugs.org>) provides a bioinformatics platform to prioritize anticancer drug treatments. The current version integrates data from 24 primary sources and supports 56,297 drug-target associations obtained from 4804 genes and 9092 unique compounds. Selected target genes can be divided into direct targets, biomarkers and pathway members [7].

During the processing PanDrugs computes a Gene Score and a Drug Score. The Gene Score (GScore, in the range of 0 to 1) measures the biological relevance of the gene and is estimated through the (i) cancer essentiality and vulnerability (by studying RNAi cell lines), (ii)

relevance in cancer (using Cancer genes Census, Tumor-Portal, Driver Gene, OncoScope, and inclusion in a list with the most frequently altered genes in T-cell lymphoblastic neoplasia), (iii) biological Impact (using Functional impact predictors such as Variant Effect predictor from ENSEMBL 16 and different predictive algorithms, VEP relevant consequence, Essentiality score, Domains and Zygosity), (iv) frequency (GMAF 1000 genomes, COSMIC and gnomAD), and (v) clinical implications (ClinVar). The Drug Score (DScore, in the range of -1 to 1) measures the suitability of the drug and considers (i) drug-cancer type indication, (ii) the drug clinical status, (iii) the gene-drug relationship, (iv) the number of curated databases supporting that relationship, and (v) collective gene impact.

To obtain the therapeutic options for this patient case, PanDrugs was queried 3 times with different types of molecular evidences: filtered variants, top 500 up-regulated genes and top 500 down-regulated genes. Filtered variants were provided as input for the Genomic Variants query option using a VCF file with converted GRCh37/hg19 assembly coordinates. The deregulated genes were selected using as criteria the log₂ based fold-change combined with an adjusted *p*-value < 0.05 and provided as input for the Genes query option.

In the three strategies we selected the most relevant therapies dividing them into 2 tiers: (i) tier 1 with the Best Therapeutic Candidates (therapies with DScore > 0.7 and GScore > 0.6), and (ii) tier 2 with the therapies with DScore > 0.7 and GScore > 0.5. For the filtered variants, we considered the drug-gene associations where the causal alteration matched the input variant and those without specification of causal alteration. For deregulated genes, we selected the therapeutic candidates where the alteration in the drug-gene association is an expression change or a copy number alteration (that can be translated into changes in the expression) in the same direction observed in the deregulated genes. The selected treatments in the three approaches were combined. Resistances arisen in some approach were used to exclude therapies suggested by the others.

Results

Clinical data evidenced a case of pro-T acute lymphoblastic leukaemia

Sixteen years old patient presented with a six weeks progressive cough, asthenia, hyporexia and lose of weight. The blood tests showed hyperleukocytosis ($152 \times 10^9/L$), anaemia (99 g/L) and thrombocytopenia ($83 \times 10^9/L$) with an increase of uric acid and lactate dehydrogenase (LDH). Chest X-ray presented mediastinum widening. A bone marrow biopsy was done showing 97% of blast cells with an immunophenotype compatible with a Pro-T acute lymphoblastic leukaemia. Cytogenetic analysis

revealed 47, XY, + 16 [20] and 48, XY, + 9, + 16 [3] karyotypes, negative by FISH for deletion of *MYB* [6q23] and a translocation/inversion of the T cell receptor locus (TCR) (14q11).

Molecular data revealed multiple candidate genes, fusion transcripts and alternative splicing variants

Whole Exome Sequencing (WES) and Massive transcriptome sequencing (RNA-Seq) were used to identify relevant genetic alterations including gene variants, gene expression levels, fusion transcripts and alternative splicing variants.

Whole exome sequencing

WES analysis and annotation process was performed as described in methods. We filtered gene variants using two main criteria: (i) population frequency, to select only somatic variants occurring in the tumour cells (GMAF or gnomAD < 0.01); (ii) functional impact of mutations, picking out those variants with high or moderate impact predicted to be pathogenic by at least two predictive algorithms. Additionally, we used the APPRIS Database to discard mutations affecting non-functional transcript-isoforms. A total of 689 gene variants, involving 308 genes, met those criteria. These genes were then categorized by GAD-Disease using the Functional Annotation tools from the Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources 6.8 (<https://david.ncifcrf.gov/>) [21]; Additional file 2: Table S2).

Scientific data available hitherto indicate that each T-ALL case only accumulates 10 to 20 biologically relevant genomic lesions, on average, as necessary events that cooperate during the development and progression of this type of leukaemia [22]. According to the information in Tumour Portal, Role Driver and Genetic Association Database (GAD_Disease data) 183 out of the 689 variants are in 77 genes previously involved in cancer. Only 12 genes with 27 presumably pathogenic gene variants were among the most frequently mutated ones in this type of leukaemia [1, 20, 23, 24]: *ARID1A*, *CTCF*, *DNM2*, *FAT1*, *FBXW7*, *H3F3A*, *JAK1*, *JAK3*, *KMT2D*, *NOTCH1*, *PHF6*, and *WT1*. Interestingly, the affectation of 4 of these genes (*DNM2*, *JAK1*, *JAK3* and *CTCF*) has been described in Early T-cell Precursor Acute lymphoblastic leukaemia (ETP T-ALL) [1, 25–27]. The T > C substitution found in the *NF1* gene is an existing variant (re2525574), which causes a stop lost effect in two defective non-functional transcripts that in addition are subjected to Non-sense Mediated Decay (NMD) (Fig. 1a).

To our knowledge 5 gene variants detected in, *CTCF*, *FBXW7*, *JAK1*, *NOTCH1* and *WT1* genes have not yet been demonstrated in T-ALL pathogenesis. Sanger sequencing (Fig. 1b) verified novel mutations in these genes. First, a homozygous insertion of an A after C (C

to CA) in *WT1*, which generates a high-impact frameshift variant that ends in a termination codon 18 amino acids after resulting in truncation of the C-terminal zinc finger domains of this transcription factor (c.1100dupR; p.Val371CysfsTer14). Similar mutations are frequently associated with oncogenic expression of the *TLX1*, *TLX3* and *HOXA* oncogenes [28]. Second, a heterozygous presumably activating missense-variant at the pseudo kinase domain of the JAK1 protein (c.2413 T > G; p.Phe805Va). Third, a heterozygous inactivating missense variant in the *FBXW7* gene (c.1634A > T; p.Tyr545Phe), which overlaps with the three main isoforms (α , β and γ). Fourth, a presumably activating heterozygous missense variant at the HD-N domain of the NOTCH protein /c.4775 T > C; p.Phe1592Ser). Fifth, an inactivating high-impact frameshift mutation at the *CTCF* gene, which generates a premature stop codon (c.950_951delCA; p.Thr317ArgfsTer91).

Massive transcriptome sequencing (RNA-Seq)

RNA-Seq analysis and annotation process was performed as indicated in the methods section. Significant deregulation was established calculating the log₂ Fold Change (log₂FC) by comparing patient sample expression data with the expression data of normal paediatric DN thymocytes (CD34+ mix), in two replicates. Absolute fold change values equal or greater than 1.5 were considered as thresholds of significance. With this stringency filtering criterium there were 6652 genes with at least one principal isoform significantly deregulated. Of these, 3575 have at least one principal isoform up regulated; 3436 exhibited at least one down regulated main isoform and, surprisingly, we detected 359 genes with at least one major isoform up and another down (Additional file 3: Table S3).

Cross-talk between exome and transcriptome data revealed 94 genes that exhibited pathogenic mutations and significant deregulation (52 up and 42 down) (Additional file 4: Table S4). Of them, five genes are in the list of most frequently altered ones in T-ALL (*FBXW7*, *FAT1*, *FAT2*, *FAT3* and *PHF6*) (Additional file 5: Table S5). Notably, 6558 genes without pathogenic mutations were significantly deregulated (3523 with some isoform up and 3393 with some isoform down) (Additional file 6: Table S6) and some of them (25 genes) are included in the list of most frequently altered genes in T-ALL (13 up and 12 down) (Additional file 7: Table S7). Up-regulated genes included *MYC*, *NOTCH2*, *FLT3*, *TLX3*, *TET1*, *TYK2*, *LMO2*, *AKT1*, *DNMT3B*, *HDAC5*, *HDAC8*, *KDM7A*, and *SMARCA1*. Down regulated genes included *CDKN2A*, *CDKN2B*, *NSD2*, *TP53* (*TP53-008*; $\Delta 133p53$ isoform), *HDAC6*, *IDH1*, *PHF6*, *CDH1*, *EPHA7*, *FAS* and *NSD2* (Fig. 2).

Fusion transcripts

Many pediatric cancers are characterized by gene fusion events that result in aberrant activity of the encoded

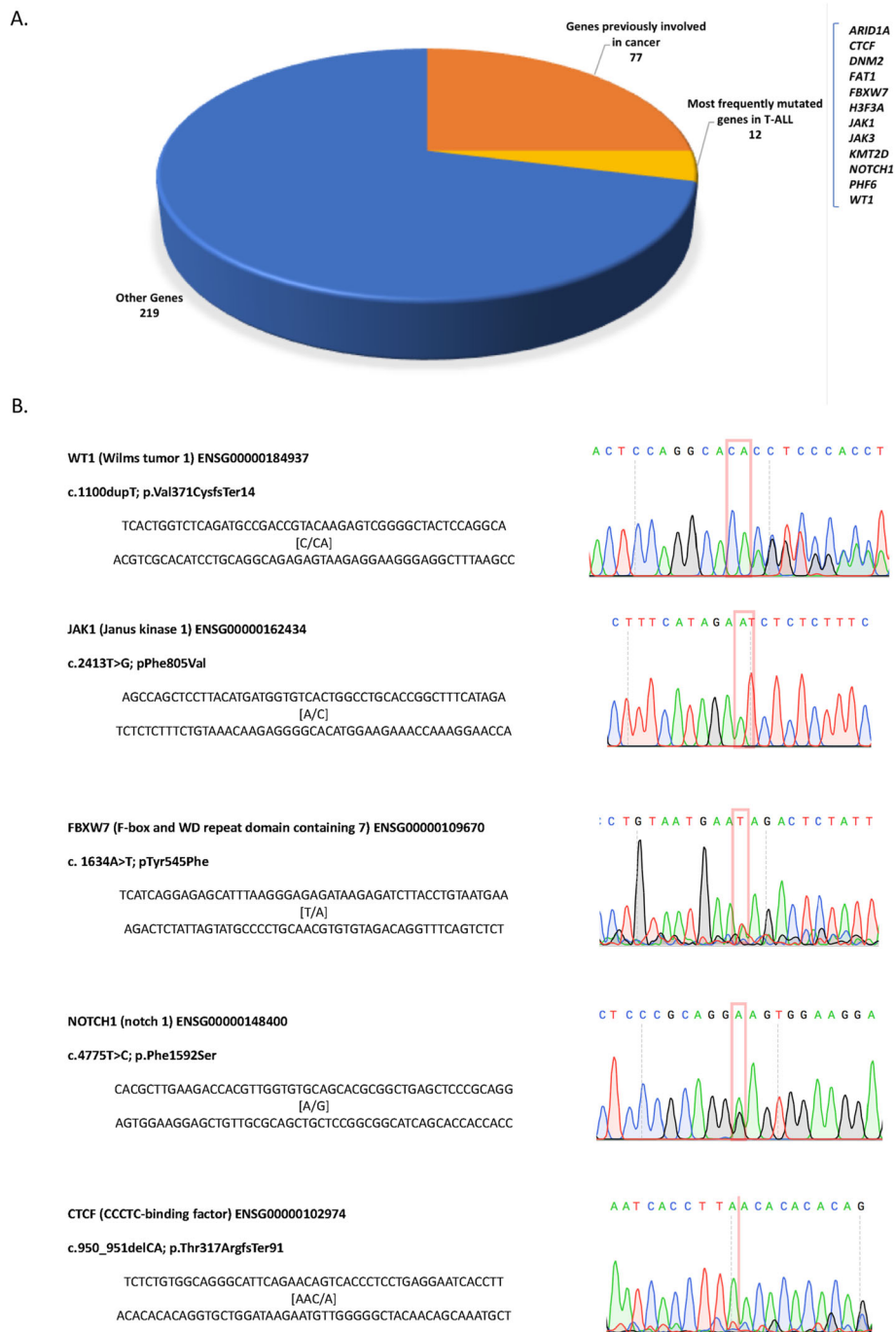
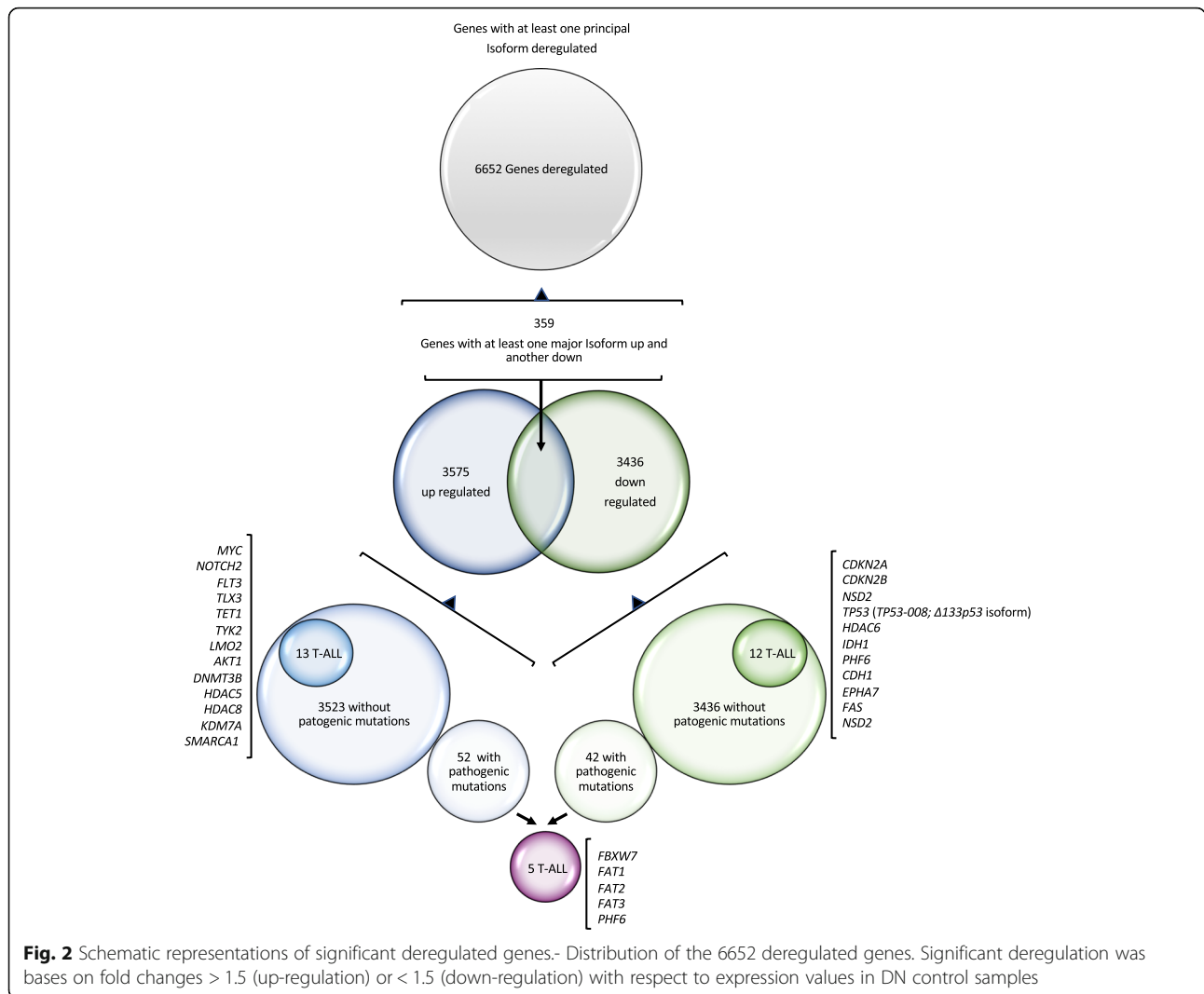


Fig. 1 Schematic representations of the Whole Exome variants predicted to be pathogenic. **a.-** Distribution of 689 gene variants involving functional transcripts-isoforms of 308 genes, which met filtering criteria to be considered pathogenic. **b.-** Mutation validation, of fifth new gene variants detected in the patient

proteins. Interpretation of RNA-Seq data using the predictive algorithm EricScript (EricScore ≥ 0.5) allow us to detect 126 fusion transcripts not previously described in T-ALL [20] (Additional file 8: Table S8). These fusion events identified by RNA-Seq may have unique biologic and diagnostic relevance.

Alternative splicing variants

Relative few significant ATEs have been reported in previous studies with T-ALL patients [20]. In our case, we detected novels junctions in *FTL3* and *KMT2D* with a known acceptor and a novel donor site that may be of functional consequences in the case of *KMT2D* gene.



ATEs in *KMT2D*, *TCF7* and *CNOT6* might also have negative implications due to the loss of critical domains (Additional file 9:Table S9).

Proposal of personalized and prioritized drug treatments

Identifying which genetic variants may be targetable by current therapies in this patient has been accomplished by using PanDrugs, a new computational methodology that provides a catalogue of candidate drugs and targetable genes estimated from a list of gene variants and deregulated genes provided by genomic analyses. This tool considers multiple targetable mutations, deregulations and the protein pathway-specific activity to prioritize a list of druggable genes categorized as direct targets, biomarkers and pathway members [7].

In order to evaluate the relevance of driver mutations, gene variant annotations of this patient were filtered by (i) population frequency (GMAF and gnomAD < 0.01),

(ii) consequences of high and moderate impact according to Ensembl classification and (iii) affection of canonical or unknown isoforms (Additional file 10: Table S10). An approach using the combination of the two general strategies based on gene mutations and significant gene deregulation suggested, as the best candidate selection, a total of 20 prioritized drugs supported by scores nearest to 1 in both GScore and D-Score values and should therefore be seen as the most effective approaches. All these drugs have the approval to be used in the treatment of different types of cancer (including blood cancer). Most of them would function as targeted therapy. Genes with GScore above the Tier’s threshold include mutated marker genes such as *MAP 2K3*, *ARID1A*, *MAP4K5*, *PKHD1* and *JAK3*, which have a genetic status associated with the drug response but the protein product is not the drug target itself. Other deregulated genes, such as *NF1*, *FGFR1*, *FLT3* and *KIT*, encode proteins that can be

improve cure rates and minimize toxicities in childhood ALL.

As indicated the PanDrugs analysis of the tumour sample for this patient identified druggable genetic alterations showing a list of 20 prioritized drugs as the best candidate selection. Since genes with GScore above the Tier's threshold include mutated marker genes such as MAP2K3 it is not surprising that Trametinib dimethyl sulfoxide (DScore 0.95), a highly selective inhibitor of MEK1 and MEK2 activity that controls the Mitogene Activated Protein Kinase (MAPK) signalling pathway, is the first recommended option to treat this patient. This drug has proved to improve overall survival in adult patients with unresectable or metastatic melanoma with a BRAF V600 mutation [31] and could be useful for the treatment of specific T-ALL subsets [23].

Lenalidome (DScore 0.932), Thalidomide (DScore 0.923) and Pomalidomide (DScore 0.901) are immunomodulatory drugs that have shown activity against the activation of tumor necrosis factor (TNF) pathway probably through the mutation of *MAP2K3* in our patient. This means that control and effectively blocks the development of abnormal cells, prevents the growth of blood vessels within tumors and also stimulates specialized cells of the immune system to attack the abnormal cells. These drugs have been used in multiple myeloma treatment but Lenalidomide also for some myelodysplastic syndromes and mantle cell lymphoma [32].

Other antineoplastics molecular target inhibitors as Dasatinib (DScore 0.933), which inhibits STAT5B signalling [33], Bosutinib (DScore 0.921), Ponatinib (DScore 0.976) and Nilotinib (DScore 0.927) tyrosine-kinase inhibitors designed for the treatment of BCR_ABL positive neoplasms, mainly in chronic myeloid leukaemia but also acute lymphoblastic leukaemia, have also off-target effects on other tyrosine-kinases. However, Dasatinib could be discarded on the basis of criteria of resistance (shaded in red in Table 1).

In addition drugs as Ibrutinib [23] (DScore 0.822) and Acalabrutinib (DScore 0.812) Burton's tyrosine-kinase inhibitors used in chronic lymphoid leukemia and mantle-cell lymphoma shows activity against JAK3 [34], which is mutated in our patient. Also *FLT3* [35], a gene that is upregulated in our case is inhibited by Sorafenib a kinase inhibitor drug approved for the treatment of primary kidney cancer (advanced renal cell carcinoma), advanced primary liver cancer (hepatocellular carcinoma) FLT3-ITD positive AML and radioactive iodine resistant advanced thyroid carcinoma.

Other drugs already used for T-ALL chemotherapy as Vinblastine (DScore 0.852) what causes M phase specific cell cycle arrest by disrupting microtubule assembly and proper formation of the mitotic spindle and the kinetochore or Etoposide (DScore 0.892) witch forms a ternary

complex with DNA and the topoisomerase II enzyme (which aids in DNA unwinding), prevents re-ligation of the DNA strands, and by doing so causes DNA strands to break [3, 36] are also suggested by PanDrugs thus supporting the reliability of this bioinformatics application (see Additional file 11: Table S11 for further details).

Conclusions

It is well known that complex diseases as cancer should not be considered as a single entity. Personalized medicine is a therapeutic approach involving the use of individual's information (genetic and epigenetic) to tailor drug therapy instead of one-size-fits-all medicine. The current approach to drug development assumes that all patients with a particular condition respond similarly to a given drug. This paper provided a framework for T-ALL patients based on the use of PanDrugs to integrate whole exome sequencing and RNA-Sequencing data into the proposal of a prioritized list of drugs, which could be clinically actionable in the context of a personalized medicine of precision. This approach is toward truly precision cancer care. Furthermore drugs directed to the activity of the surrounding interactors in the biological pathway of a mutated gene could be used in combination to avoid possible compensatory mechanisms of resistance to drugs. It means that patients with different types of cancer could receive similar treatments on the basis of the genomic diagnosis. Of note, most of the proposed drugs in this T-ALL case are not being yet considered in the clinical practice of this type of cancer, opening up the approach of new treatment possibilities. At present, many of the proposed drugs are approved on the basis of clinical trials on large populations in tumours other than T-ALL so the risk of failure is lower, because the drugs have already been found to be safe, the time frame for drug reprofiling can be reduced, because most of the preclinical testing, safety assessment and formulation development will be completed. However regulatory considerations, organizational hurdles and patent considerations must be taken into account. Repurposing of these drugs for T-ALL would require validation of the results of treatments in in vitro models that have the same genetic characteristics as the samples of the patients to be treated as well as in vivo patient-derived xenografts and eventually in trials that allow repositioning of the proposed drugs.

The speed, accuracy and accessibility of next-generation sequencing (NGS) have driven the arrival of precision medicine, its mandatory to assume that this revolution must be transferred to its applicability to patients. Bioinformatics tools such as Pandrugs will allow, using the information obtained by the sequencing platforms, to improve the effectiveness of the treatments, reducing unwanted side effects and favoring survival rates.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12885-019-6209-9>.

Additional file 1: Table S1. Primer list. Description of primers required for Sanger sequencing.

Additional file 2: Table S2. WES annotation process.

Additional file 3: Table S3. Deregulated genes after RNA-Sequencing.

Additional file 4: Table S4. Results of crossing exome and transcriptome data.

Additional file 5: Table S5. Genes included in the most frequently altered ones in T-ALL.

Additional file 6: Table S6. Genes without pathogenic mutations but significantly deregulated.

Additional file 7: Table S7. Genes without pathogenic mutations but significantly deregulated included in the list of most frequent altered genes.

Additional file 8: Table S8. Fusion transcripts.

Additional file 9: Table S9. Alternative Splicing Variants.

Additional file 10: Table S10. Gene variants selected for PanDrugs.

Additional file 11: Table S 11. Therapies. Mutations. Deregulation UP. Deregulation DOWN.

Abbreviations

ALL: Acute leukaemia of the lymphoid lineage; ATEs: Alternative Splicing Variants; B-ALL: B-cell precursor leukemia; ClinVar: Clinical implications; DAVID: Visualization and Integrated Discovery Bioinformatics Resources; DN: Double Negative; DNA: Deoxyribonucleic acid; DScore: Drug Score; ETP T-ALL: Early T-cell Precursor Acute lymphoblastic leukaemia; GAD: Genetic Association Database; GScore: Gene Score; INDELS: Insertion-deletions; LDH: Lactate dehydrogenase; log2FC: log2 Fold Change; MAPK: Mitogen Activated Protein Kinase; NGS: Next-Generation Sequencing; NMD: Non-sense Mediated Decay; PCR: Polymerase-Chain-Reaction; RIN: RNA integrity Numbers; RNA: Ribonucleic acid; RNA-Seq: Massive transcriptome sequencing; SNV: Single Nucleotide Variations; SRBC: Sheep Red Blood Cell; T-ALL: Acute T-cell lymphoblastic leukaemia; TCR: T cell receptor; TNF: Tumor necrosis factor; WES: Whole Exome sequencing

Acknowledgements

We thank all patients who were willing to donate their samples without their support the research work would not be possible.

Authors' contributions

PFN and PLN are co-first authors. PFN, PLN, developed the concepts, designed the experiments and contributed to the writing of the manuscript; PFN, EPY, FA conducted all the bioinformatics analyses; PLN performed experiments and analysis; GCT, JML, provide tumour sample and clinical data; RSP, AA, provide paediatric thymuses; MACF contributes with technical support; JFP directed the study, analysed the results and wrote the manuscript. All authors have read and approved the final manuscript.

Funding

This research was made possible through funding by the Spanish *Ministry of Science, Innovation and Universities* (RTI2018–093330-B_100); Spanish Ministry of Economy and Competitiveness (SAF2015–70561-R); MINECO/FEDER, EU; BES-2013-065740); Ramón Areces Foundation (CIVP19S7917); the Autonomous Community of Madrid, Spain (B2017/BMD-3778; LINFOMAS-CM); the Spanish Association Against Cancer (AECC, 2018; PROYE18054PIRI); and the *Institute of Health Carlos III, ISCIII (ACCI-CIBERER-17)*. Institutional grants from the Ramón Areces Foundation and the Santander Bank to the Severo Ochoa Molecular Biology Center (CBMSO) are also acknowledged. These projects only provide financial support for our experiments.

Availability of data and materials

The webtool is freely accessible at <http://www.pandrug.org> and through its programmatic API or docker image.

Ethics approval and consent to participate

Patients provided written informed consent before study enrollment. The informed consent form was reviewed and approved by Institutional review board from the Research Ethics Committee of Autonomous University of Madrid (previous references are CEI 31–773 and CEI-70-1260). The participant and/or their parents had already provided written informed consent with the guiding principles of the Declaration of Helsinki.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Cancer and Environmental Epidemiology Unit, National Center for Epidemiology, Carlos III Institute of Health, Madrid 28029, Spain. ²Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid 28029, Spain. ³Department of Cellular Biology and Immunology, Severo Ochoa Molecular Biology Center (CBMSO), CSIC-Madrid Autonomous University, Madrid 28049, Spain. ⁴Institute of Health Research Jiménez Díaz Foundation, Madrid 28040, Spain. ⁵Consortium for Biomedical Research in Rare Diseases (CIBERER), Carlos III Institute of Health, Madrid 28029, Spain. ⁶Bioinformatics Unit, Structural Biology and Biocomputing Programme, Spanish National Cancer Research Center (CNIO), Madrid 28029, Spain. ⁷Hematology Department, Hospital Universitario 12 de Octubre, Madrid 28041, Spain. ⁸Department of Congenital Cardiac Surgery, Hospital Universitario La Paz, Madrid 28046, Spain.

Received: 13 June 2019 Accepted: 25 September 2019

Published online: 26 October 2019

References

- Belver L, Ferrando A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat Rev Cancer*. 2016;16(8):494–507.
- de Leval L, Bisig B, Thielen C, Boniver J, Gaulard P. Molecular classification of T-cell lymphomas. *Crit Rev Oncol Hematol*. 2009;72(2):125–43.
- Litzow MR, Ferrando AA. How I treat T-cell acute lymphoblastic leukemia in adults. *Blood*. 2015;126(7):833–41.
- Le Tourneau C, Borcoman E, Kamal M. Molecular profiling in precision medicine oncology. *Nat Med*. 2019;25(5):711–2.
- Intlekofer AM, Younes A. Precision therapy for lymphoma—current state and future directions. *Nat Rev Clin Oncol*. 2014;11(10):585–96.
- Gomez-Lopez G, Dopazo J, Cigudosa JC, Valencia A, Al-Shahrour F. Precision medicine needs pioneering clinical bioinformaticians. *Brief Bioinform*. 2019; 20(3):752–66.
- Piñero-Yáñez E, Reboiro-Jato M, Gómez-López G, Perales-Patón J, Troulé K, Rodríguez JM, Tejero H, Shimamura T, López-Casas PP, Carretero J, et al. PanDrugs: a novel method to prioritize anticancer drug treatments according to individual genomic data. *Genome Medicine*. 2018;10(1):41.
- Rubio-Camarillo M, Gomez-Lopez G, Fernandez JM, Valencia A, Pisanò DG. RUBioSeq: a suite of parallelized pipelines to automate exome variation and bisulfite-seq analyses. *Bioinformatics*. 2013;29(13):1687–9.
- Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM; 2013.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013;43:11 10 11–33.
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213–9.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl variant effect predictor. *Genome Biol*. 2016;17:122.
- Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, Cole CG, Ward S, Dawson E, Ponting L, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017;45(D1):D777–83.

14. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353–61.
15. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46(D1):D1062–7.
16. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2016;44(D1):D279–85.
17. Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, Chang HY, Dosztanyi Z, El-Gebali S, Fraser M, et al. InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* 2017;45(D1):D190–9.
18. Benelli M, Pescucci C, Marseglia G, Severgnini M, Torricelli F, Magi A. Discovering chimeric transcripts in paired-end RNA-seq data by using EricScript. *Bioinformatics.* 2012;28(24):3232–9.
19. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics.* 2011;27(17):2325–9.
20. Atak ZK, Gianfelici V, Hulsemans G, Keersmaecker KD, Devasia AG, Geerdens E, Mentens N, Chiaretti S, Durinck K, Uyttebroeck A, et al. Comprehensive analysis of Transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. *PLoS Genet.* 2013;9(12):e1003997.
21. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44–57.
22. Becker KG, Barnes KC, Bright TJ, Wang SA. The genetic association database. *Nat Genet.* 2004;36(5):431–2.
23. Girardi T, Vicente C, Cools J, De Keersmaecker K. The genetics and molecular biology of T-ALL. *Blood.* 2017;129(9):1113–23.
24. Peirs S, Van der Meulen J, Van de Walle I, Taghon T, Speleman F, Poppe B, Van Vlierberghe P. Epigenetics in T-cell acute lymphoblastic leukemia. *Immunol Rev.* 2015;263(1):50–67.
25. Coustan-Smith E, Mullighan CG, Onciu M, Behm FG, Raimondi SC, Pei D, Cheng C, Su X, Rubnitz JE, Basso G, et al. Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol.* 2009;10(2):147–56.
26. Van Vlierberghe P, Ambesi-Impiombato A, Perez-Garcia A, Haydu JE, Rigo I, Hadler M, Tosello V, Della Gatta G, Paietta E, Racevskis J, et al. ETV6 mutations in early immature human T cell leukemias. *J Exp Med.* 2011;208(13):2571–9.
27. Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature.* 2012;481(7380):157–63.
28. Cameselle-Teijeiro J, Pardal F, Eloy C, Ruiz-Ponte C, Celestino R, Castro P, Soares P, Sobrinho-Simões M. Follicular thyroid carcinoma with an unusual glomeruloid pattern of growth. *Hum Pathol.* 2008;39(10):1540–7.
29. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. Integrative clinical genomics of advanced prostate Cancer. *Cell.* 2015;162(2):454.
30. Horak P, Frohling S, Glimm H. Integrating next-generation sequencing into clinical oncology: strategies, promises and pitfalls. *ESMO Open.* 2016;1(5):e000094.
31. Salama AK, Kim KB. MEK inhibition in the treatment of advanced melanoma. *Curr Oncol Rep.* 2013;15(5):473–82.
32. Mercurio A, Adriani G, Catalano A, Carocci A, Rao L, Lentini G, Cavalluzzi MM, Franchini C, Vacca A, Corbo F. A mini-review on thalidomide: chemistry, mechanisms of action, therapeutic potential and anti-Angiogenic properties in multiple myeloma. *Curr Med Chem.* 2017;24(25):2736–44.
33. Nam S, Williams A, Vultur A, List A, Bhalla K, Smith D, Lee FY, Jove R. Dasatinib (BMS-354825) inhibits Stat5 signaling associated with apoptosis in chronic myelogenous leukemia cells. *Mol Cancer Ther.* 2007;6(4):1400–5.
34. Davids MS, Brown JR. Ibrutinib: a first in class covalent inhibitor of Bruton's tyrosine kinase. *Future Oncol.* 2014;10(6):957–67.
35. Lierman E, Lahortiga I, Van Mieghroet H, Mentens N, Marynen P, Cools J. The ability of sorafenib to inhibit oncogenic PDGFRbeta and FLT3 mutants and overcome resistance to other small molecule inhibitors. *Haematologica.* 2007;92(1):27–34.
36. Pui CH, Mullighan CG, Evans WE, Relling MV. Pediatric acute lymphoblastic leukemia: where are we going and how do we get there? *Blood.* 2012;120(6):1165–74.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

