Supporting Information

Mullighan et al. 10.1073/pnas.0811761106

SI Methods

 $\tilde{\mathbf{A}}$

Patients and Treatment. Patients were enrolled in the Children's Oncology Group P9906 trial and treated with an augmented reinduction/reconsolidation strategy (Berlin–Frankfurt– Münster regimen) (1). The cohort has been described previously in detail (2). All patients were high risk based on the presence of central nervous system or testicular disease, *MLL* rearrangement, or based on age, sex, and presentation leukocyte count (3). *BCR*-*ABL1* and hypodiploid ALLs and cases of primary induction failure were excluded. High-hyperdiploid (as defined by trisomy of chromosomes 4 and 10 on cytogenetic analysis) and *ETV6*-*RUNX1* cases were excluded unless central nervous system or testicular involvement was present at diagnosis. A total of 221 enrolled cases had suitable material for single-nucleotide polymorphism (SNP) array analysis (2), and JAK sequencing was performed for 187 patients with available DNA, SNP array, and gene expression profiling data. Twenty-two cases (11.8%) were *TCF3*-*PBX1*-positive, 18 (9.6%) harbored *MLL* rearrangements, 1 (0.5%) was hyperdiploid, and 1 was *ETV6*-*RUNX1*-positive. A total of 145 cases (77.5%) lacked a recurring chromosomal abnormality. The clinical protocol was approved by the National Cancer Institute and by the Institutional Review Board at each of the Children's Oncology Group institutions. Patients and/or a parent/guardian provided informed consent to participate in the clinical trial and for future research using clinical specimens.

Genomic Resequencing. Resequencing of the coding exons of *JAK1, JAK2, JAK3,* and *TYK2* was performed by Agencourt Biosciences. PCR parameters are available upon request. Sequencing traces and primer information have been deposited

- 1. Nachman JB, et al. (1998) Augmented post-induction therapy for children with highrisk acute lymphoblastic leukemia and a slow response to initial therapy. *N Engl J Med* 338:1663–1671.
- 2. Mullighan CG, et al. (2009) Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med* 360:470–480.
- 3. Shuster JJ, et al. (1999) Identification of newly diagnosed children with acute lymphocytic leukemia at high risk for relapse. *Cancer Res Ther Control* 9:101–107.
- 4. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8:175–185.
- 5. Ewing B, Green P (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 8:186–194.
- 6. Zhang J, et al. (2005) SNPdetector: A software tool for sensitive and accurate SNP detection. *PLoS Comput Biol* 1:e53.
- 7. Zhang J, et al. (2007) Systematic analysis of genetic alterations in tumors using Cancer Genome WorkBench (CGWB). *Genome Res* 17:1111–1117.

with the National Center for Biotechnology Information (NCBI) trace archive (www.ncbi.nlm.nih.gov/Traces/trace.cgi?). Base calls and quality scores were determined by using PHRED (4, 5), and sequence variations were analyzed and annotated by using the SNPdetector (6) and IndelDetector (7) software in an annotation pipeline (8). All putative sequence mutations were confirmed by repeat genomic PCR and sequencing of both tumor and remission DNA.

Structural Modeling of JAK2 Mutations. Mutant molecular models were generated by using the Modweb server (9). For the pseudokinase domain, residues 533–821 were modeled against the epidermal growth factor receptor kinase domain residues 4–292, which share 25% sequence identity (PDB code 2ITQ, Chain A) (10). Model scores were 1.00 for all mutants. For the kinase domain, models were generated by fitting the mutant sequences against the structure of the JAK2 kinase domain complexed with an inhibitor (PDB code 2B7A) (11). Figures of structural models were generated with Pymol (12).

JAK Homology Alignment. Protein sequences for JAK1 and JAK2 homologs, as well as the human JAK3 and TYK2 genes, were obtained from the NCBI Entrez web site. Protein sequences were aligned with ClustalX version 2.0.10 (13). The degree of amino acid conservation within the alignment was calculated by using the Consurf server (14). Inputs were the ClustalX alignment file and the JAK2 structural file 2b7a.pdb from PDB (www.rcsb.org/pdb/home/home.do). We edited 2b7a.pdb by replacing occurrences of the unconventional residue ''PTR'' with ''TYR.''

- 8. Zhang J, Rowe WL, Struewing JP, Buetow KH (2002) HapScope: A software system for automated and visual analysis of functionally annotated haplotypes. *Nucleic Acids Res* 30:5213–5221.
- 9. Eswar N, et al. (2003) Tools for comparative protein structure modeling and analysis. *Nucleic Acids Res* 31:3375–3380.
- 10. Yun CH, et al. (2007) Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* 11:217–227.
- 11. Lucet IS, et al. (2006) The structural basis of Janus kinase 2 inhibition by a potent and specific pan-Janus kinase inhibitor. *Blood* 107:176–183.
- 12. DeLano WL (2002) The PyMOL Molecular Graphics System. Available at http:// www.pymol.org.
- 13. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
- 14. Landau M, et al. (2005) ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* 33:W299–W302.

Fig. S1. Sequencing traces of representative *JAK2* mutations showing corresponding tumor and normal sequences.

Fig. S2. Alignment of JAK orthologs. The alignment of the JAK1 and JAK2 orthologs in multiple species in the regions where the 3 somatic JAK2 kinase domain and 3 JAK1 mutations were found. The mismatched residues are shown in red text, whereas the background corresponds to the ConSurf conservation color code. The JAK1 region shown (residues 640–693) corresponds to residues 599–622 in JAK2. The JAK1 V658F mutation corresponds to the JAK2 V617F (underlined) mutation.

PNAS PNAS

 $\,6$

8

7

Fig. S3. Modeling of the structural effects of JAK2 mutations. (*A*) Cartoon diagrams showing structural superposition of models for pseudokinase domain for wild-type (yellow), I682F (green), R683G (cyan), R683S (blue), and QCinsR683 (magenta). Close-up views for models showing residue substitutions as sticks are on the left, with nitrogens shown in blue and oxygens in red. The JAK2 residues I682 and R683 map to the junction between the N and C lobes of the pseudokinase domain. All 4 pseudokinase domain mutations identified affect these residues, and they are predicted to influence the structure, and likely the dynamics, of the loops that pack together at the interlobe interface. Substitution of I682 to the bulkier phenylalanine, or insertion of glycine-cysteine at this position, is predicted to displace the adjacent side-chain of R683. Because R683 is predicted to directly contact the N lobe of the pseudokinase domain, either displacement by the I682F mutation or the QC insertion or substitution to a smaller serine or glycine residue, should alter N–C lobe interactions and may result in a loss of the inhibitory activity of the pseudokinase domain. (*B*) Cartoon diagrams showing structural superposition of models for kinase domain, with sites of mutations shown as magenta sticks, nitrogens in blue, and oxygens in red. The location of T875, mutated in the acute megakaryocytic cell line CHRF-288-11 [Mercher T, et al. (2006) Blood 108:2770-2779], is shown in yellow sticks for comparison. R867Q and D873N map to the β 2- β 3 loop of the kinase domain and are predicted to alter surface electrostatic properties of this region. The D873N mutation would alter hydrogen-bonding contacts to the T875 residue. The R867Q mutation would alter the adjacent surface because R867 normally extends in the direction of D873 and T875, making contacts to the intervening residue, N873. The P933R mutation lies in the JAK2 kinase hinge region, adjacent to the ATP-binding site [Lucet IS, et al. (2006) *Blood* 107:176–183]. This proline residue is conserved among JAK kinases and is thought to impart rigidity to this hinge that may be important for catalytic activity.

Table S1. Clinical features of the P9906 cases studied

PNAS PNAS

Table S2. Gentic alterations and karyotypic abnormalities in patients with JAK mutations

PNAS PNAS

Copy number alterations are deletions unless otherwise indicated. *IL3RA/CSF2RA* alterations include focal deletions at this locus at the pseudoautosomal region of Xp22.3 and Yp11.3, or larger deletions/gains adjacent to this locus. deln, deletion; iAmp21, intrachromosomal amplification of chromosome 21. *There was no evidence of trisomy 21 on analysis of germ-line single-nucleotide polymorphism array data for these cases. †Down syndrome-associated ALL.

Table S3. Clinical characteristics of cases harboring JAK mutations

e, exon; WBC, presentation leukocyte count.

*Down syndrome-associated ALL.

PNAS PNAS

Table S4. Multivariable analysis of associations between genetic and clinical variables and outcome (Cox regression model)

MRD, minimal residual disease; WBC, peripheral blood leukocyte count at diagnosis (\times 10⁹/L).

PNAS PNAS

WBC, peripheral blood leukocyte count at diagnosis, \times 10⁹/L.

PNAS PNAS

*Analyses performed by using the EFS-PHREG procedure in SAS.

†Analyses performed by using the Fine and Gray method in S-Plus.