

Published in final edited form as:

*Nat Genet.* 2013 October ; 45(10): 1226–1231. doi:10.1038/ng.2754.

## Identification of a recurrent germline *PAX5* mutation and susceptibility to pre-B cell acute lymphoblastic leukemia

A full list of authors and affiliations appears at the end of the article.

# These authors contributed equally to this work.

Somatic alterations of the lymphoid transcription factor gene *PAX5* are a hallmark of B-progenitor acute lymphoblastic leukemia (B-ALL)<sup>1-3</sup>, but inherited mutations of *PAX5* have not previously been described. Here, we report a novel heterozygous germline variant, c.547G>A (p.Gly183Ser), in the octapeptide domain of *PAX5* that was found to segregate with disease in two unrelated kindreds with autosomal dominant B-ALL. Leukemic cells from all patients in both families exhibited 9p deletion, with loss-of-heterozygosity and retention of the mutant *PAX5* allele at 9p13. Two additional sporadic ALL cases with 9p loss harbored somatic *PAX5* Gly183 substitutions. Functional and gene expression analysis of the *PAX5* mutation demonstrated relatively reduced transcriptional activity. These data extend the role of *PAX5* alterations in the pathogenesis of pre-B ALL, and implicate *PAX5* in a novel syndrome of susceptibility to pre-B cell neoplasia.

B-cell precursor acute lymphoblastic leukemia (ALL) is the most common pediatric malignancy. There is a 2-4 fold increased risk of developing the disease in children of affected siblings<sup>4</sup>, and in occasional cases ALL is inherited as a Mendelian disorder<sup>5</sup>. *PAX5*, encoding the B-cell lineage transcription factor paired box 5, is somatically deleted, rearranged or otherwise mutated in approximately 30% of sporadic B-progenitor ALL cases<sup>1,3,6-9</sup>. In *PAX5*-deficient mice, B-cell development is arrested at the pro-B-cell stage,

Correspondence should be addressed to M.S.H. (horwitz@uw.edu), C.G.M. (charles.mullighan@stjude.org) and K.O. (offitk@mskcc.org).

### AUTHOR CONTRIBUTIONS

K.O., C.G.M., M.S.H., J.T.S., S.S., K.A.S., E.W., A.T., J.V., C.M., S.L., R.J.K., M.D., D.A., and S.W.L. conceived and designed the experiment. S.S., K.A.S., E.W., A.T., J.V., C.M., J.W., J.Y., X.G., C.M., R.J.K., A.V., N.D.S., D.A., M.S.H., C.G.M., K.O., M.K., T.W., K.L., T.K., D.B., J.L., L.O., S.C.R., P.M., M.F., G.R., and J.C. performed the experiments. S.S., K.A.S., E.W., A.T., J.V., C.M., J.W., J.Y., R.J.K., N.D.S., M.S.H., C.G.M., K.O., L.W., J.Z., G.W., M.R., P.N., J.M., S.C., G.S., and J.C. performed statistical analysis. S.S., K.A.S., E.W., A.T., J.V., C.M., J.W., J.Y., C.M., R.R., M.C., R.M., M.H., S.L., R.J.K., A.V., N.D.S., D.A., C.H., H.S.S., S.W.L., M.S.H., C.G.M., K.O., L.W., J.H., J.Z., G.W., M.R., P.N., J.M., S.C., G.S., and J.C. analyzed the data. E.W., C.M., J.T.S., S.J., M.H., J.S., V.M., S.P., D.G.H., D.Z., G.C., S.L., S.M.L., R.L., A.V., K.L.N., M.D., D.A., C.H., H.S.S., S.W.L., M.S.H., C.G.M., K.O., R.P.K., A.S., J.J.L., K.T., R.S., S.H., J.S., D.W., D.R., P.M., J.Z., G.W., J.M., S.C., J.D., and K.O. contributed reagents, materials, and analysis tools. K.O., C.G.M., M.S.H., S.S., K.A.S., E.W., A.T., J.V., C.M., J.Y., R.J.K., and S.W.L. wrote the paper. K.O., C.G.M., M.S.H., J.T.S., R.J.K., M.D., D.A., and S.W.L. jointly supervised the research.

**URLS.** dbSNP 137, <http://www.ncbi.nlm.nih.gov/projects/SNP/>; NHLBI exome sequencing project, <http://evs.gs.washington.edu/EVS/>; European Genome-phenome Archive, <http://www.ebi.ac.uk/ega/>. Public data portal for results from the St. Jude–Washington University Pediatric Cancer Genome Project, <http://explore.pediatriccancergenomeproject.org/>.

### Accession codes.

Transcriptome and whole exome sequencing data and SNP microarray data have been deposited in the European Genome-phenome Archive (EGA) which is hosted by the European Bioinformatics Institute (EBI) under EGAS00001000447. Mouse Affymetrix gene expression data are deposited in the Gene Expression Omnibus (GEO) under accession code GSE45260 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE45260>)

Supplementary information is available in the online version of the paper.

Supplementary information contains Supplementary Note, Supplementary Figures 1-8, Supplementary Tables 1-15 and Supplementary files 1-6.

### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests

and *in vitro* these cells can differentiate into other lymphoid and myeloid lineages<sup>10</sup>. *PAX5* is also essential for maintaining the identity and function of mature B-cells<sup>11</sup>, and its deletion in mature B cells results in dedifferentiation to pro-B cells and aggressive lymphomagenesis<sup>12</sup>.

A heterozygous germline *PAX5* variant, c.547G>A (NM\_016734), p.Gly183Ser (NP\_057953), was identified by exome sequencing in two families, one of Puerto Rican ancestry (family 1; Figure 1a) and the other of African-American ancestry (family 2; Figure 1b, Supplementary Note). This variant had not been previously described in public databases (Exome Variant Server, 1000 Genomes and dbSNP 137), or previous sequencing analyses of ALL and cancer genomes<sup>1,2,9</sup>. All affected family members had B-cell precursor ALL and all available diagnostic and relapse leukemic samples from both families demonstrated loss of 9p through i(9)(q10) or dicentric chromosomes involving 9q, both of which resulted in loss of the wild-type *PAX5* allele and retention of *PAX5* p.Gly183Ser (Figure 1c, Supplementary Figure 1 and Supplementary Table 1).

The germline *PAX5* p.Gly183Ser mutation segregated with the leukemia in both kindreds, however, several unaffected obligate carriers (Family 1: II3, III2, III3 and Family 2: II, I2, II2, II3) were also observed, suggesting incomplete penetrance. Unaffected mutation carriers and patients at the time of ALL diagnosis had normal immunoglobulin levels and no laboratory or clinical evidence of impaired B-cell function. Sanger sequencing of cDNA from peripheral blood of unaffected carriers indicated biallelic transcription of *PAX5* (data not shown). The only mutated gene common to both families was *PAX5* and no germline copy number aberrations were found to be shared between patients (Supplementary tables 2 and 3).

To determine whether p.Gly183Ser arose independently in each kindred or instead reflects common ancestry, the risk haplotypes of each family were compared. The families share a 4.7kb haplotype, spanning five SNPs (Figure 1d, Supplementary Note). The relatively small size of the shared haplotype and principal component analysis of genome-wide single nucleotide polymorphism genotype data (Supplementary Figure 2) together imply that the two families are not recently related and differ in ethnicity. Moreover, given the reduced fitness due to increased susceptibility for childhood ALL, it is unlikely that such a lethal mutation could be propagated over time. Because this haplotype is relatively frequent worldwide (Supplementary Table 4), it is likely that each family's mutation arose independently.

Genomic profiling of tumor samples demonstrated expression of the p.Gly183Ser mutant *PAX5* in diagnostic and relapse tumor specimens from affected members of family 2, an average of 1 chimeric fusion and 9 non-silent sequence variants per case, and homozygous deletion of *CDKN2A/CDKN2B* in all cases due to loss of 9p and focal deletion of the second allele. Apart from loss of 9p, no other somatic sequence mutations or structural rearrangements were shared by the affected family members (Supplementary Tables 1, 5-12)

As somatic i(9)(q10)/dic(9;v) abnormalities were seen in all of the familial leukemias, we sequenced *PAX5* in 44 additional sporadic pre-B ALL cases with i(9)(q10)/dic(9;v) aberrations to assess whether *PAX5* mutations frequently co-occur with loss of 9p. Two leukemic samples revealed octapeptide mutations p.Gly183Ser, and p.Gly183Val, and in others previously reported variants including p.Pro80Arg and p.Val26Gly<sup>1</sup> were observed (Table 1, Figure 2a). We examined the frequency of non-silent *PAX5* somatic sequence mutations in a cohort of B-ALL cases with i(9)(q10)/dic(9;v) (n=28) and two cohorts of B-ALL without i(9)(q10)/dic(9;v) (n=183) and n=221.<sup>1,2</sup> We observed a significantly higher frequency of *PAX5* mutations in the i(9)/dic(9) cohort (p=0.0001). No novel germline *PAX5*

mutations were detected in 39 families with a history of two or more cases of cancer including at least one hematological cancer, although one familial case of ALL harbored a dic(9;20)(p11;q11.1) and a somatic p.Pro80Arg variant (Table 1 and Supplementary Materials).

Previously identified *PAX5* somatic mutations commonly result in significant reduction in transcriptional activation mediated by *PAX5*. Downstream targets of *PAX5* include *CD19* and *CD79A* (Ig $\alpha$  or mb-1)<sup>13</sup>. We examined the transactivating activity of wild type and mutant *PAX5* alleles using a *PAX5*-dependent reporter gene assay, containing copies of a high-affinity *PAX5*-binding site derived from the *CD19* promoter<sup>14</sup>. Both p.Gly183Ser and Val mutations resulted in partial but significant reduction in transcriptional activation ( $P < 0.0001$  for both alleles, Figure 3a). Additionally, there was no detectable differences in subcellular localization of wild type and p.Gly183Ser *PAX5* (Supplementary Figure 3). To study the effect of this mutation on *CD79A* expression, we expressed mutant and wild type *PAX5* in J558L $\mu$ M, a mouse plasmacytoma cell line that does not express *PAX5* or *CD79A*. Enforced expression of *PAX5* results in expression of *CD79A* and assembly of the surface immunoglobulin (sIgM) complex. The amount of sIgM expression may be used to assess the transcriptional activity of *PAX5* alleles on the *CD79A* promoter<sup>1</sup>. Both p.Gly183 alleles resulted in a significant reduction in sIgM expression compared to wild-type *PAX5* ( $P < 0.0001$ , Figure 3b). These results suggest that the *PAX5* G183 mutations result in partial loss of *PAX5* activity.

The identified missense variant, p.Gly183Ser, is located at a conserved residue in the octapeptide (OP) domain of *PAX5* that mediates interaction with Groucho transcriptional corepressors<sup>15</sup> (Figure 2b). Previous studies have shown that GRG4 (also known as TLE4 in humans) represses *PAX5*-dependent luciferase activity in cells expressing wild type *PAX5* but not in cells expressing *PAX5* octapeptide domain mutants<sup>15</sup>. We observed GRG4-mediated repression of the transcriptional activity of *PAX5* wild-type and p.Gly183Ser (Figure 3c), suggesting the effect of this mutation is not mediated by an altered interaction with GRG4.

To further explore the effect of the variant on downstream targets, we performed genome-wide transcriptional profiling of J558L $\mu$ M cells transduced with empty vector, *PAX5* wild-type or mutant alleles (either all transduced cells marked by RFP expression, or the subset of cells expressing sIgM), and examined the expression of *PAX5* activated and repressed genes previously defined in *Pax5*<sup>-/-</sup> murine pro- and mature B cells<sup>16-19</sup> and in human *ETV6-RUNX1* B-ALL<sup>1</sup>. Examining all *PAX5* expressing cells, we observed profound deregulation of *PAX5* activated and repressed genes in J558L $\mu$ M cells expressing known loss of function (e.g. the common exon 2-6 deletion that results in a truncating frameshift *PAX5* allele) or strongly hypomorphic alleles (*PAX5* p.Pro80Arg) and less marked deregulation in p.Gly183Ser or Val expressing cells ( $P$  values for each mutant allele versus wild-type were all  $P < 0.001$  (Supplementary Figures 4 and 5). Comparing sorted sIgM positive *PAX5* p.Gly183Ser to *PAX5* wild-type cells, we observed reduced expression of genes activated by *PAX5* in pro-B and mature B-cells ( $P = 1.4 \times 10^{-4}$  and  $P = 3.8 \times 10^{-4}$  respectively, Supplementary Tables 13-15)

We next examined the transcriptional consequences of *PAX5* p.Gly183Ser by performing transcriptome sequencing (mRNA-seq) of diagnosis and relapse samples obtained from two patients in kindred 2, and of 139 sporadic childhood B-progenitor ALL samples. We performed gene set enrichment analysis incorporating gene sets of *PAX5*-mutated *ETV6-RUNX1* ALL (one third of which harbor focal *PAX5* deletions)<sup>1</sup>, *PAX5* regulated genes in *Pax5*<sup>-/-</sup> mice<sup>16-19</sup> and genes regulated during murine B lymphoid development<sup>20</sup>. As a limited set of genes are known to be regulated in both murine pro- and mature B cells, and

as the overlap between mouse and human *PAX5* regulated genes are unknown, we used all previously published *PAX5* regulated genes, and genes regulated during murine B cell development<sup>16-20</sup> in an unbiased approach to explore the effects of the *PAX5* p.Gly183 mutations on direct and indirect transcriptional targets of *PAX5*. This showed striking enrichment of genes deregulated in *ETV6-RUNX1/PAX5*-mutated ALL, *PAX5* activated and repressed genes (including *CD19*, *CD72* and *CD79a*), and genes regulated during murine B lymphoid development in the signature of familial *PAX5* p.Gly183Ser versus sporadic B-ALL (Figure 3d and Supplementary Figures 6 and 7. We also analyzed the overlap of previously published data and the expression differences between the familial ALL tumor samples and other B-ALL cases stratified by *PAX5* mutation status (Supplementary file 6; Supplementary Figure 8). Together, these results suggest that *PAX5* p.Gly183Ser results in attenuation of *PAX5* function and deregulation of *PAX5* target genes that is less severe than the previously reported p.Pro80Arg and D2-6 alleles that result in marked or complete loss of *PAX5* activity.

The *PAX5* deletions, translocations and sequence mutations identified as somatic events in B-ALL commonly affect the DNA-binding and transactivation domains and result in complete loss or marked attenuation of *PAX5* transcriptional activity, but are rarely homozygous and not observed as inherited variants. Moreover, *Pax5* loss promotes the development of B-ALL in experimental models which are commonly accompanied by the acquisition of second hits in *Pax5*<sup>21</sup>, indicating that profound loss of *PAX5* activity is commonly a central event in leukemogenesis. In contrast, the inherited *PAX5* p.Gly183Ser mutation results in modest attenuation of *PAX5* activity, and is accompanied by somatic loss of the wild-type *PAX5* allele due to 9p alterations during leukemogenesis. This model is also consistent with the finding of a significant association of somatic *PAX5* hypomorphic mutations coincident with complete loss of the normal *PAX5* allele in leukemic cells absent 9p. These observations suggest that a severe reduction in *PAX5* activity is incompatible with normal B lymphoid development and is deleterious in carriers, but by contrast, the partial hypomorphic p.Gly183Ser allele is tolerated as a germline allele but additional genetic events further reducing *PAX5* activity are required to establish the leukemic clone. The universal finding of deletion of wild-type *PAX5* in all familial ALL cases, rather than the acquisition of additional hypomorphic *PAX5* mutations, suggests that a complete loss of wild-type *PAX5* activity is required for developmental arrest and loss of maturation. This is supported by our transcriptional profiling of J558L $\mu$ M p.Gly183Ser cells and familial leukemias showing deregulation of *PAX5* target gene expression that is significant but less marked than known loss-of-function mutations. The differences in transcriptional profiles of some target gene panels were not as robustly observed as in mouse model systems, presumably due to inherent germline and somatic genetic and epigenetic variability in human leukemias. In addition, ongoing studies will be of interest to fully characterize the functional consequences of *PAX5* octapeptide domain mutations.

Our findings have clinical implications with regard to options for pre-implantation genetic diagnosis, and the possible significance of somatic 9p alterations as a harbinger of a germline *PAX5* mutation. The recent identification of germline *TP53* mutations in familial ALL<sup>20,22</sup> and the data presented here strongly implicating *PAX5* mutations in a novel syndrome of inherited susceptibility to pre-B cell ALL, indicate that further sequencing of affected kindreds is required to define the full spectrum of germline variations contributing to ALL pathogenesis.

## METHODS

Methods and any associated references are available in the online version of the paper

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Authors

Sohela Shah<sup>#1,2</sup>, Kasmintan A. Schrader<sup>#1,2</sup>, Esmé Waanders<sup>#3,4</sup>, Andrew E. Timms<sup>#5</sup>, Joseph Vijai<sup>#1,2</sup>, Cornelius Miething<sup>#1</sup>, Jeremy Wechsler<sup>5</sup>, Jun Yang<sup>6</sup>, James Hayes<sup>1</sup>, Robert J. Klein<sup>1,2</sup>, Jinghui Zhang<sup>7</sup>, Lei Wei<sup>7,3</sup>, Gang Wu<sup>7</sup>, Michael Rusch<sup>7</sup>, Panduka Nagahawatte<sup>7</sup>, Jing Ma<sup>8</sup>, Shann-Ching Chen<sup>3</sup>, Guangchun Song<sup>3</sup>, Jinjun Cheng<sup>3,8</sup>, Paul Meyers<sup>9</sup>, Deepa Bhojwani<sup>10</sup>, Suresh Jhanwar<sup>11</sup>, Peter Maslak<sup>12</sup>, Martin Fleisher<sup>13</sup>, Jason Littman<sup>2</sup>, Lily Offit<sup>2</sup>, Rohini Rau-Murthy<sup>2</sup>, Megan Harlan Fleischut<sup>2</sup>, Marina Corines<sup>2</sup>, Rajmohan Murali<sup>11</sup>, Xiaoni Gao<sup>1</sup>, Christopher Manschreck<sup>2</sup>, Thomas Kitzing<sup>1</sup>, Vundavalli V. Murty<sup>14</sup>, Susana Raimondi<sup>3</sup>, Roland P. Kuiper<sup>4</sup>, Annet Simons<sup>4</sup>, Joshua D. Schiffman<sup>15</sup>, Kenan Onel<sup>16</sup>, Sharon E. Plon<sup>17</sup>, David Wheeler<sup>17</sup>, Deborah Ritter<sup>17</sup>, David S. Ziegler<sup>18</sup>, Kathy Tucker<sup>19</sup>, Rosemary Sutton<sup>20</sup>, Georgia Chenevix-Trench<sup>21</sup>, Jun Li<sup>21</sup>, David G. Huntsman<sup>22</sup>, Samantha Hansford<sup>22</sup>, Janine Senz<sup>22</sup>, Thomas Walsh<sup>23</sup>, Ming Lee<sup>23</sup>, Christopher N. Hahn<sup>24</sup>, Kathryn Roberts<sup>3</sup>, Mary-Claire King<sup>23</sup>, Sarah M. Lo<sup>25</sup>, Ross L. Levine<sup>26</sup>, Agnes Viale<sup>27</sup>, Nicholas D. Socci<sup>28</sup>, Katherine L. Nathanson<sup>29</sup>, Hamish S. Scott<sup>24</sup>, Mark Daly<sup>30</sup>, Steven M. Lipkin<sup>31</sup>, Scott W. Lowe<sup>1</sup>, James R. Downing<sup>3</sup>, David Altshuler<sup>30</sup>, John T. Sandlund<sup>#10</sup>, Marshall S. Horwitz<sup>#5</sup>, Charles G. Mullighan<sup>#3</sup>, and Kenneth Offit<sup>#1,2</sup>

## Affiliations

<sup>1</sup>Cancer Biology and Genetics Program, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. <sup>2</sup>Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. <sup>3</sup>Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. <sup>4</sup>Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences and Radboud Institute for Oncology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. <sup>5</sup>Department of Pathology, University of Washington, Seattle, Washington, USA. <sup>6</sup>Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. <sup>7</sup>Department of Computational Biology and Bioinformatics, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. <sup>8</sup>Pediatric Cancer Genome Project Laboratory, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. <sup>9</sup>Department of Pediatrics, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. <sup>10</sup>Department of Oncology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. <sup>11</sup>Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. <sup>12</sup>Hematology Laboratory Service, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. <sup>13</sup>Clinical Chemistry Service, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. <sup>14</sup>Department of Pathology and Cell Biology, Columbia University, New York, New York, USA. <sup>15</sup>High Risk Pediatric Cancer Clinic, Huntsman Cancer Institute/Primary Children's Medical Center, University of Utah, Salt Lake City, Utah, USA. <sup>16</sup>Pediatrics, University of Chicago, Chicago, IL, USA. <sup>17</sup>Texas Children's Cancer Center and Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA. <sup>18</sup>Kids Cancer Centre, Sydney Children's Hospital, Sydney, New South Wales, Australia, Children's Cancer Institute Australia for Medical Research, Lowy Cancer Research Centre, University of New South Wales, Sydney, New South Wales, Australia. <sup>19</sup>Hereditary Cancer Clinic Prince of Wales Hospital, Randwick, Australia. <sup>20</sup>Children's Cancer Institute Australia for Medical Research,

University of New South Wales, Randwick, Australia. <sup>21</sup>Cancer Genetics Laboratory, The Queensland Institute of Medical Research, Herston, Australia. <sup>22</sup>Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada. <sup>23</sup>Departments of Medicine and Genome Sciences, University of Washington, Seattle, Washington, USA. <sup>24</sup>Department of Molecular Pathology, SA Pathology and Centre for Cancer Biology, Adelaide, Australia. <sup>25</sup>Department of Pediatrics, Weill Cornell College of Medicine, New York, New York, USA. <sup>26</sup>Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. <sup>27</sup>Genomics Core Laboratory, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. <sup>28</sup>Bioinformatics Core, Memorial Sloan-Kettering Cancer Center, New York, New York, USA. <sup>29</sup>Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. <sup>30</sup>Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA <sup>31</sup>Department of Medicine, Weill Cornell College of Medicine, New York, New York, USA.

## Acknowledgments

We thank M.A.S. Moore and S. Jae-Hung for their contributions to ongoing tumor studies; G. Dressler for the murine Grg4 construct; M. Busslinger for the *luc*-CD19 reporter construct; J. Hagman for providing a *PAX5* vector and the J558L $\mu$ M cell line; D. Payne-Turner for technical assistance; Dr. Wenjian Yang and Dr. Colton Smith (Pharmaceutical Sciences) for their assistance in the haplotype analyses and the Tissue Resources Core Facility, Pediatric Cancer Genome Project Core Facility and Flow Cytometry and Cell Sorting Core Facility of St Jude Children's Research Hospital. We thank the families for their generous participation in these studies. This project was supported by grant I5-A523 from the Starr Cancer Consortium, the Robert and Kate Niehaus Clinical Cancer Genetics Initiative, the Sabin Family Research Fund, the Lymphoma Foundation, Geoffrey Beene Cancer Research Center grant 78730, the Cancer Prevention and Research Institute of Texas Grant RP101089, NSW Priority of the Knights of the Order of Saint John, Matthew Bell Foundation, US National Cancer Institute of the National Institutes of Health (NIH) Comprehensive Cancer Center Core Grant CA21765, the American Lebanese Syrian Associated Charities of St Jude Children's Research Hospital, and grant R01DK58161. R.P.K. is funded by a grant from the Dutch Cancer Society (KUN2009-4298). T.K. is supported by the German Research Foundation Postdoctoral Fellowship (KI1605/1-1). C.G.M. is a Pew Scholar in the Biomedical Sciences, and a St Baldrick's Scholar. K.G.R. is supported by a National Health and Medical Research Council (NHMRC, Australia) CJ Martin Postdoctoral Fellowship. K.A.S. is funded by the Canadian Institutes of Health Research. A.T. is supported by T32GM007454. G.C.-T. is a Senior Principal Research Fellow of the NHMRC. E.W. is funded by the Dutch Cancer Society, project number KUN2012-5366.

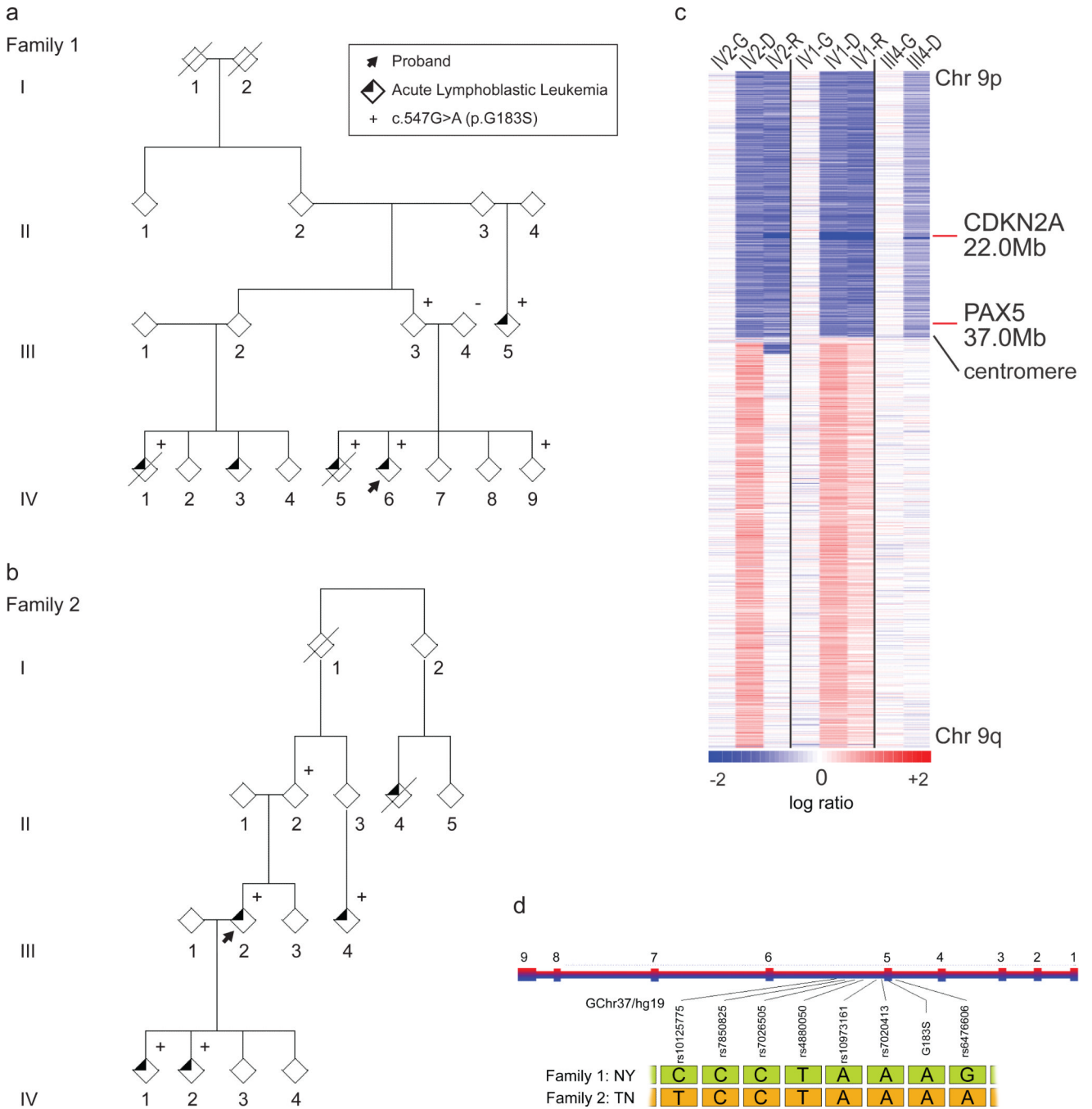
## REFERENCES

1. Mullighan CG, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*. 2007; 446:758–64. [PubMed: 17344859]
2. Mullighan CG, et al. Deletion of *IKZF1* and prognosis in acute lymphoblastic leukemia. *N Engl J Med*. 2009; 360:470–80. [PubMed: 19129520]
3. Kuiper RP, et al. High-resolution genomic profiling of childhood ALL reveals novel recurrent genetic lesions affecting pathways involved in lymphocyte differentiation and cell cycle progression. *Leukemia*. 2007; 21:1258–66. [PubMed: 17443227]
4. Hemminki K, Jiang Y. Risks among siblings and twins for childhood acute lymphoid leukaemia: results from the Swedish Family-Cancer Database. *Leukemia*. 2002; 16:297–8. [PubMed: 11840299]
5. Pui CH, Robison LL, Look AT. Acute lymphoblastic leukaemia. *Lancet*. 2008; 371:1030–43. [PubMed: 18358930]
6. Mullighan CG, Downing JR. Global genomic characterization of acute lymphoblastic leukemia. *Semin Hematol*. 2009; 46:3–15. [PubMed: 19100363]
7. Mullighan CG, et al. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of *Ikaros*. *Nature*. 2008; 453:110–4. [PubMed: 18408710]

8. Nebral K, et al. Incidence and diversity of PAX5 fusion genes in childhood acute lymphoblastic leukemia. *Leukemia*. 2009; 23:134–43. [PubMed: 19020546]
9. Zhang J, et al. Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. *Blood*. 2011; 118:3080–7. [PubMed: 21680795]
10. Nutt SL, Heavey B, Rolink AG, Busslinger M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature*. 1999; 401:556–62. [PubMed: 10524622]
11. Horcher M, Souabni A, Busslinger M. Pax5/BSAP maintains the identity of B cells in late B lymphopoiesis. *Immunity*. 2001; 14:779–90. [PubMed: 11420047]
12. Cobaleda C, Jochum W, Busslinger M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature*. 2007; 449:473–7. [PubMed: 17851532]
13. Maier H, Colbert J, Fitzsimmons D, Clark DR, Hagman J. Activation of the early B-cell-specific mb-1 (Ig-alpha) gene by Pax-5 is dependent on an unmethylated Ets binding site. *Mol Cell Biol*. 2003; 23:1946–60. [PubMed: 12612069]
14. Czerny T, Busslinger M. DNA-binding and transactivation properties of Pax-6: three amino acids in the paired domain are responsible for the different sequence recognition of Pax-6 and BSAP (Pax-5). *Mol Cell Biol*. 1995; 15:2858–71. [PubMed: 7739566]
15. Eberhard D, Jimenez G, Heavey B, Busslinger M. Transcriptional repression by Pax5 (BSAP) through interaction with corepressors of the Groucho family. *EMBO J*. 2000; 19:2292–303. [PubMed: 10811620]
16. Pridans C, et al. Identification of Pax5 target genes in early B cell differentiation. *J Immunol*. 2008; 180:1719–28. [PubMed: 18209069]
17. Revilla IDR, et al. The B-cell identity factor Pax5 regulates distinct transcriptional programmes in early and late B lymphopoiesis. *EMBO J*. 2012; 31:3130–46. [PubMed: 22669466]
18. Delogu A, et al. Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity*. 2006; 24:269–81. [PubMed: 16546096]
19. Schebesta A, et al. Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function. *Immunity*. 2007; 27:49–63. [PubMed: 17658281]
20. Holmfeldt L, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nat Genet*. 2013; 45:242–52. [PubMed: 23334668]
21. Dang J, Mullighan CG, Phillips LA, Mehta P, Downing JR. Retroviral and Chemical Mutagenesis Identifies Pax5 as a Tumor Suppressor in B-Progenitor Acute Lymphoblastic Leukemia. *Blood (ASH Annual Meeting Abstracts)*. 2008; 112:1789.
22. Powell BC, et al. Identification of TP53 as an acute lymphocytic leukemia susceptibility gene through exome sequencing. *Pediatr Blood Cancer*. 2012
23. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics*. 2011; 43:491–8. [PubMed: 21478889]
24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–60. [PubMed: 19451168]
25. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20:1297–303. [PubMed: 20644199]
26. Pounds S, et al. Reference alignment of SNP microarray signals for copy number analysis of tumors. *Bioinformatics*. 2009; 25:315–21. [PubMed: 19052058]
27. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*. 2004; 5:557–72. [PubMed: 15475419]
28. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*. 2007; 23:657–63. [PubMed: 17234643]
29. Mullighan CG. Single nucleotide polymorphism microarray analysis of genetic alterations in cancer. *Methods in molecular biology*. 2011; 730:235–58. [PubMed: 21431646]
30. Lin M, et al. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics*. 2004; 20:1233–40. [PubMed: 14871870]

31. Cai Y, Brophy PD, Levitan I, Stifani S, Dressler GR. Groucho suppresses Pax2 transactivation by inhibition of JNK-mediated phosphorylation. *EMBO J.* 2003; 22:5522–9. [PubMed: 14532124]
32. Lundblad A, et al. Immunochemical studies on mouse myeloma proteins with specificity for dextran or for levan. *Immunochemistry.* 1972; 9:535–44. [PubMed: 5036150]
33. Sitia R, Neuberger MS, Milstein C. Regulation of membrane IgM expression in secretory B cells: translational and post-translational events. *EMBO J.* 1987; 6:3969–77. [PubMed: 3127200]
34. Maier H, et al. Requirements for selective recruitment of Ets proteins and activation of mb-1/Ig-alpha gene transcription by Pax-5 (BSAP). *Nucleic Acids Res.* 2003; 31:5483–9. [PubMed: 14500810]
35. Hombach J, Tsubata T, Leclercq L, Stappert H, Reth M. Molecular components of the B-cell antigen receptor complex of the IgM class. *Nature.* 1990; 343:760–2. [PubMed: 2304550]
36. Team, RDC. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2009.
37. Gentleman RC, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* 2004; 5:R80. [PubMed: 15461798]
38. Irizarry RA, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics.* 2003; 4:249–64. [PubMed: 12925520]
39. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007; 8:118–27. [PubMed: 16632515]
40. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3. 2004 Article3.
41. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B.* 1995; 57:289–300.
42. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005; 102:15545–50. [PubMed: 16199517]
43. Zhang J, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature.* 2012; 481:157–63. [PubMed: 22237106]
44. Trapnell C, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010; 28:511–5. [PubMed: 20436464]

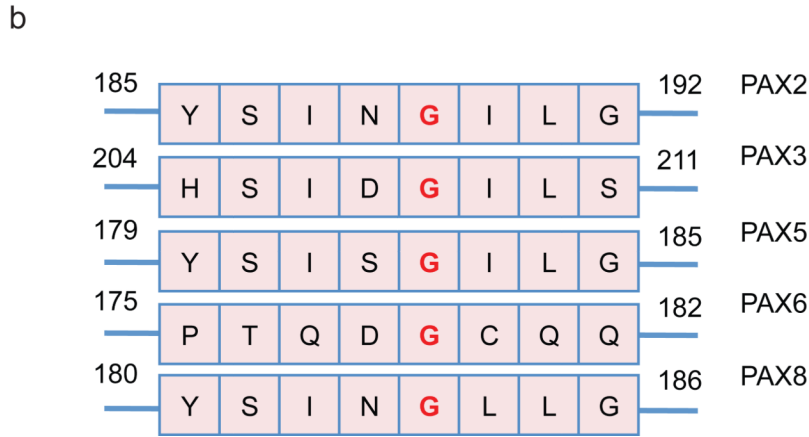
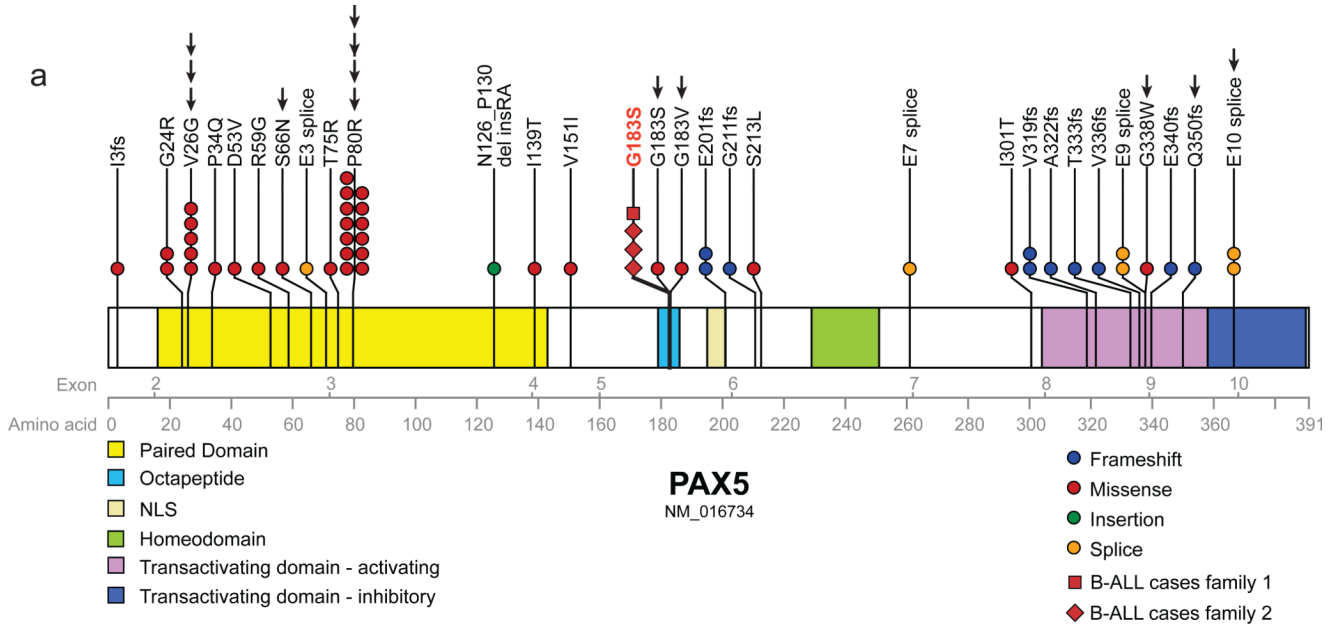




**Figure 1. Familial Pre-B cell ALL associated with i(9)(q10) and dic(9;v) in two families harboring a novel, recurrent germline p.Gly183Ser variant**

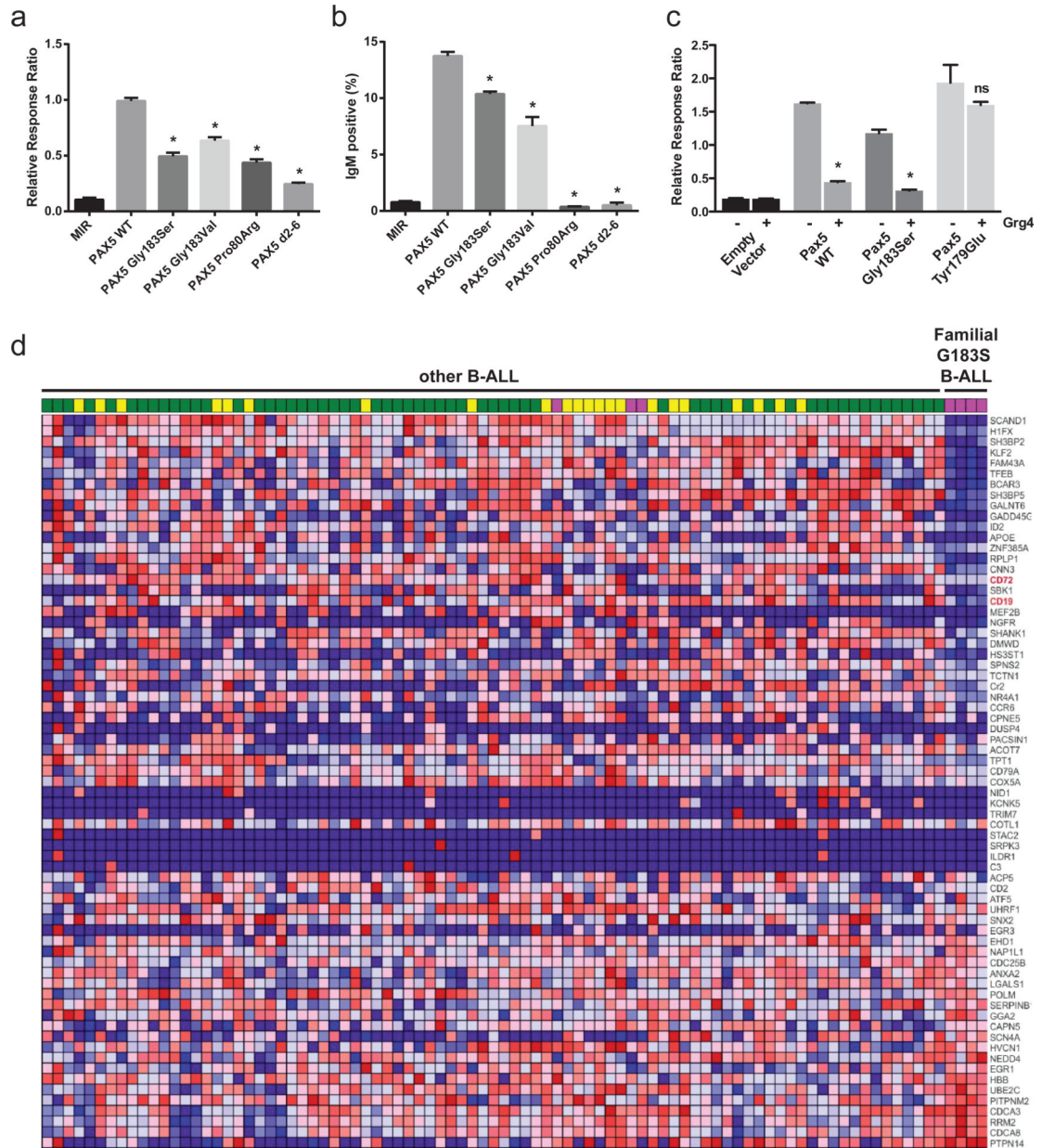
**a.** Family 1 of Puerto Rican ancestry. The proband is denoted by an arrow. Exome sequencing was undertaken in germline DNA from all available affected (IV1, IV5, IV6, III5) and unaffected (IV9, III3, III4) individuals as well as the diagnostic leukemic sample from IV6. p.Gly183Ser variant status denoted by (+/-). **b.** Family 2 of African-American ancestry. The proband is denoted by an arrow. Exome sequencing was undertaken in diagnostic, remission and relapse leukemic samples from individuals III4, IV1, and IV2. p.Gly183Ser variant status denoted by (+/-). **c.** Chromosome 9 copy number heat map for SNP6.0 microarray data of germline and tumor samples from three members of Family 2.

These data demonstrate the common feature of loss of 9p in the tumor specimens. Note the focal dark blue band denoting homozygous loss of CDKN2A/B in all samples. Blue indicates deletions and red indicates gains. G, germline; D, diagnosis; R, relapse sample. **d.** The haplotype flanking the p.Gly183Ser mutation. A five SNP haplotype rs7850825 to rs7020413 (Chr9:36.997-37.002 Mb) proximal to the mutation was concordant in both family 1 and family 2. However, the distal end flanking the mutation rs6476606 was discordant.



**Figure 2. Recurrent PAX5 mutations in ALL**

**a.** Gene schematic of *PAX5* showing the exons (upper grey numbers), amino acid residues (lower grey numbers), protein domains (as denoted by colored legend) and position of the germline p.Gly183Ser variant (in red) in relation to the somatic *PAX5* mutations described in this study (n=13, arrows) and somatic mutations described previously in B-ALL<sup>1,2,20</sup>. Primary leukemic samples with confirmed retention of the germline p.Gly183Ser variant denoted by the square shape (Family 1) and diamond shape (Family 2). In one case of i(9)/dic(9) ALL, we found both a heterozygous Val26Gly and a heterozygous Gln350fs mutation, indicating polyclonality of the tumor. **b.** Conservation of the octapeptide domain in selected PAX family members.



**Figure 3. Attenuated transcriptional activity of PAX5 p.Gly183Ser**  
**a.** Transcriptional activity of PAX5 variants compared to wild-type using a Pax5-dependent reporter gene assay in 293T cells. Bars show mean ( $\pm$  s.e.m.) luciferase activity of six individual experiments with triplicate measurements (PAX5 p.Gly183Val and PAX5 d2-6 four experiments with triplicate measurements). Asterisks indicate significant difference calculated by Dunnett's test ( $P < 0.0001$ ). MIR, MSCV-IRES-mRFP empty vector. **b.** Transcriptional activity of PAX5 variants using CD79A-dependent sIgM expression in the murine J558L $\mu$ M plasmocytoma cell line. Percentages indicate proportion of mRFP positive cells that show sIgM expression. Bars show mean ( $\pm$  s.e.m.) sIgM expression in two individual experiments with triplicate replicates each. Asterisks indicate significant difference calculated by Dunnett's test ( $P < 0.0001$ ). MIR, MSCV-IRES-mRFP empty vector. **c.** PAX5-dependent reporter gene assay of PAX5 wild-type and PAX5 p.Gly183Ser run in triplicate as above, with or without co-transfection of 0.05 $\mu$ g of Grg4 as indicated. A PAX5

p.Tyr179Glu (Y179E) mutant that is deficient in binding to Grg4 and empty vector were used as controls. Asterisks indicate significant differences by two-tailed t-test ( $p < 0.0001$ ). **d.** Heatmap of *PAX5* activated genes in mature B cells. Four samples from family 2 (diagnosis and relapse samples from individuals IV1 and IV2) show differential expression of *PAX5* activated genes when compared to a group of 139 sporadic B-ALL cases. This indicates an effect of the pGly183Ser mutation on *PAX5* function. Red indicates high expression, blue represents low expression. *PAX5* mutation status is indicated by the colors above the samples. Green indicates wild type *PAX5*, yellow indicates heterozygosity for a *PAX5* mutation, and pink indicates biallelic *PAX5* mutations.

**Table 1***PAX5* mutations found in familial and sporadic i(9)(q10)/dic(9;v) pre-B cell ALL samples

Inheritance	Patient	Mutation	Tumor status	Germline status
Family 1	IV6	c.547G>A = p.Gly183Ser	Homozygous	Heterozygous
Family 2	III4	c.547G>A = p.Gly183Ser	Homozygous	Heterozygous
Family 2	IV1 D	c.547G>A = p.Gly183Ser	Homozygous	Heterozygous
Family 2	IV1 R	c.547G>A = p.Gly183Ser	Homozygous	
Family 2	IV2 D	c.547G>A = p.Gly183Ser	Homozygous	Heterozygous
Family 2	IV2 R	c.547G>A = p.Gly183Ser	Homozygous	
Familial <sup>1</sup>		c.239C>G = p.Pro80Arg (tumor shows dic(9;20)(p11;q11.1))	Homozygous	Wildtype
Sporadic		c.77T>G = p.Val26Gly	Heterozygous	Wildtype
Sporadic		c.77T>G = p.Val26Gly	Heterozygous	Wildtype
Sporadic		c.77T>G = p.Val26Gly	Heterozygous	Wildtype
Sporadic		c.197G>A = p.Ser66Asn	Homozygous	ND
Sporadic		c.239C>G = p.Pro80Arg	Homozygous	Wildtype
Sporadic		c.239C>G = p.Pro80Arg	Homozygous	Wildtype
Sporadic		c.239C>G = p.Pro80Arg	Homozygous	Wildtype
Sporadic		c.547G>A = p.Gly183Ser	Homozygous	ND
Sporadic		c.548G>T = p.Gly183Val	Heterozygous	Wildtype
Sporadic		c.1012G>T = p.Gly338Trp	Heterozygous	Wildtype
Sporadic		c.1049-1051delAGTinsGTCCG = p.Gln350fs	Heterozygous	Wildtype
Sporadic		c.1100_1100+15 del16bp (IVS9 splice)	heterozygous	ND

<sup>1</sup>Familial case as reported in main text. ND = Not Determinable. Germline DNA either not tested or not available. *PAX5* accession No: NM\_016734