SUPPLEMENTARY

Molecular ontogeny of donor-derived follicular lymphomas occurring after hematopoietic cell transplantation

SUPPLEMENTARY CLINICAL INFORMATION

For initial cytoreduction of chronic phase CML the patient received hydroxyurea. The myeloablative conditioning regimen consisted of cyclophosphamide (1800 mg/m^2 × 2 days) and fractionated total body irradiation (1400 cGy in 7 fractions over 4 days) followed by methotrexate and tacrolimus for graft-versus-host disease prophylaxis. Interferon alpha-2b dose was 3 M Units/day SC, imatinib 400 mg/day PO.

SUPPLEMENTARY METHODS

Isolation of genomic DNA and total RNA

DNA from FFPE tissue was isolated using QIAamp DNA FFPE Tissue kit (Qiagen, USA). DNA from fresh frozen cells was isolated using Qiagen Genomic-tip 20/G and Buffer set (Qiagen, USA). Total RNA from FFPE tissue was isolated using RNeasy FFPE kit (Qiagen, USA) and first strand cDNA synthesis was performed using random hexamer primers and Super Script III First-Stand Synthesis System for RT-PCR (Invitrogen, Life Technologies, USA).

Immunohistochemical (IHC) analysis

Immunohistochemistry was performed using 5-µm-thick formalin fixed paraffin embedded (FFPE) tissue sections. Slides were soaked in xylene, passed through graded alcohols, and put in distilled water. Slides were pretreated DAKO high pH target retrieval solution (for anti-CD20, anti-BCL6, anti-ARID1A/BAF250A (Santa Cruz); DAKO USA, Carpinteria, CA) with 1mM ethylenediamine tetraacedic acid buffer pH 8.0 (for anti-BCL2, Zymed Laboratories/ Invitrogen, Calsbad, CA) or with 10mM citrate buffer pH 6.0 (for anti-CD10 and anti-ARID1A/BAF250 (Sigma-Aldrich); Invitrogen) in a steam pressure cooker (Biocare Decloaking Chamber CD2008US, Biocare Biomedical, Concord, CA) at manufacturers recommended settings. All further steps were performed at room temperature in a hydrate chamber. The slides were blocked for endogenous peroxidase activity with peroxidase block (DAKO), washed 5 minutes in buffer, and incubated with serum free protein block (DAKO) for 20 minutes. The antibodies and dilutions used for immunohistochemistry are outlined in the main manuscript. After washing, antibodies were detected using species appropriate (rabbit or mouse) Envision kit (DAKO) and DAB and counterstained with Harris hematoxylin. Of note, the anti-ARID1A/BAF250A antibody recognizes amino acids 600-1018, *i.e.* N-terminal of codon R1276.

Real-time quantitative TaqMan PCR

Real-time quantitative PCR (Applied Biosystems TaqMan) was performed in 50uL reactions (Applied Biosystems Taqman buffer) with 1 nmol of each primer and 250 pmol of a translocation specific probe (quencher dye 6-carboxy-tetramethylrhodamin (TAMRA), reporter dye 6-carboxy-fluorescin (FAM)) hybridizing to the junction region. The *BCL2*/*IGH* translocation was amplified from genomic DNA

of the DLI sample, TA cloned into PCR2.1-TOPO (Invitrogen, USA) and confirmed by Sanger sequencing. Standard curves were generated using serial tenfold dilutions of clone copy numbers (range 10^6 to 10^1) for the BCL2/JH6 translocation and GAPDH. 750 ng of the DLI sample were tested in triplicates for BCL2/JH6 and GAPDH. RQ-PCR cycling conditions consisted of 50°C for 2 minutes (allows for elimination of contamination by uracil-N-glycosylase), initial denaturation with 95°C for 10 minutes, and 45 cycles of 15 seconds at 95°C and a 1-minute combined annealing/extension step at 60°C (ABI 7300 thermocycler).

Exome sequencing and computational analysis

Genomic DNA from each follicular lymphoma and the DLI was fragmented and library construction was carried out according to the SureSelect Target Enrichment System for the Applied Biosystems SOLiD system (Agilent Technologies). Briefly, 3 ug of genomic DNA was sheared to an average size of 150 base pairs, ligated to adapters and hybridized to biotinylated capture oligonucleotides (SureSelect Human All Exon Kit - 38Mb); the captured portion of the library was amplified by PCR. Approximately 50 bases from one end and 35 bases from the other end of each library fragment were sequenced using the Applied Biosystems SOLiD4 instrument (Life Technologies) and reagents (Applied Biosystems).

Sequence data was mapped to the human reference genome, hg18, using mapreads, and variants were detected using diBayes (Bioscope v1.2, Applied Biosystems). Mapping statistics are available in Table S1. Variants were annotated to RefSeq release 35 using custom python scripts (see below, *Annotation of single nucleotide substitution and small insertions and deletions for exome sequencing*). The pipeline for establishing a validation set of sequence variants is outlined in Figure S2. Details on selection criteria and a summary of Sanger sequencing validation are provided in Tables S2-3 and *Validation of exome sequencing by Sanger sequencing* (see below).

Annotation of single nucleotide substitution and small insertions and deletions for exome sequencing

A custom, in-house pipeline was developed using Python to annotate the output of diBayes with information including gene context using RefSeq, characterized polymorphisms (dbSNP release 130), amino acid consequence, ratio of detection, average quality and supporting reads' orientation distribution and unique starting sites. Additionally, this pipeline automated the process of intersecting the data between samples to elucidate candidates for somatic mutations seen only in the donor and recipient libraries. Small insertions and deletions (InDels) were detected by remapping the sequence data using BFAST v.0.6.2a, with the standard parameters for SOLiD data as recommended in the BFAST manual. The mapped reads were processed using the pileup subroutine of Samtools v.0.1.12a. The results from the pileup were filtered against the following criteria: consensus quality >5, SNP quality >5, mapping quality >15, coverage >5, %evidence of variant > 15%. The remaining variants were annotated with information including gene context, using RefSeq, and known InDels using dbSNP130. Variants outside of coding regions were ignored. The remaining variants were intersected and keyed by chromosome : coordinate. This list of candidates was sorted by a composite score of SNP quality, consensus quality, %evidence of variants, and coverage: Score = (Consensus Quality + SNP Quality) * Variant Coverage / (Total Coverage)^2. Indels for validation were selected from the top of this score-sorted list.

Venn diagrams outlining the intersection of SNV and InDel candidates before and after subtracting known polymorphisms (dbSNP release 130) are depicted in Figure S3.

Validation of exome sequencing by Sanger sequencing

After subtraction of known polymorphisms (db SNP release 130), non-coding and synonymous variants, intersection with DLI and subtraction of probable private germline polymorphisms (>2 reads in DLI, Phred score>20, %variants>10%), annotated sequencing variants were filtered (Phred Score >30, number of different start sites >2, reads' orientation diversely distributed, coverage $\geq$5 reads), sorted and ranked by quality metrics, visually inspected by Integrated Genome Browser (IGV 2.0, Broad Institute (1)) and selected for validation by PCR and Sanger sequencing from both lymphomas, the DLI and appropriate germline DNA (Figures S2-3. Tables S2-S3).

Validation rate of coding single nucleotide variants (SNVs) and insertions/deletions InDels is summarized in Table S3. Validation rate for SNVs was 51% (49 out of 96; 34 different SNVs), and for insertions/deletions (InDels) 17% (10 out of 60; 6 different InDels). Validated somatic mutations (19 SNVs and 2 InDels) were analyzed using software tools to predict functional impact of missense mutations (CanPredict (2), http://www.cgl.ucsf.edu/Research/genentech/canpredict/index.html) (Table S4) and selected for further analysis and experiments.

References

1. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24-6.

2. Kaminker JS, Zhang Y, Waugh A, Haverty PM, Peters B, Sebisanovic D, et al. Distinguishing cancer-associated missense mutations from common polymorphisms. Cancer research. 2007;67:465-73.

Supplementary Figure Legends

Figure S1. Sorting was performed on the DLI sample based on CD19 expression. Standard curve for quantitative PCR demonstrates that detection of the BCL2/IGH product from the CD19-negative DLI population is below the detection threshold and in the range of control human DNA (HDNA).

Figure S2. Exome sequencing data analysis workflow for the donor's and recipient's FL and the DLI sample.

Figure S3. Venn diagrams of (**A**) all SNVs and (**B**) all InDels from exome sequencing of the donor's FL, the recipient's FL and the DLI. (**C**) and (**D**) are the corresponding Venn diagrams displaying only coding, non-synonymous SNVs that are not reported in dbSNP release 130.

Figure S4. A. Immunohistochemistry for CD10 at low power demonstrating lymphoma cell infiltration of 40-45% and 60-65% in the donor's and the recipient's lymphoma, respectively. **B.** Immunohistochemistry for ARID1A/BAF250 is shown in both lymphomas and a tonsil from a healthy individual with a different anti-BAF250 antibody (Sigma-Aldrich) than used in Figure 3E.