Supporting Information

Jiang et al. 10.1073/pnas.1213021110

SI Materials and Methods

Patient Samples and Processing. Primary samples from patients with chronic myeloid leukemia (CML) were obtained from consenting patients at the University of California at San Diego, Stanford University, and the University of Toronto Health Network, according to Institutional Review Board-approved protocols. Primary patient samples or normal blood samples (AllCells or Lonza) were CD34-selected or FACS-sorted for RNA-Seq, quantitative RT-PCR (qRT-PCR), and hematopoietic progenitor assays. For primary patient samples, peripheral blood mononuclear cells (PBMC) were extracted from peripheral blood following Ficoll density centrifugation. $CD34^+$ cells then were purified by magnetic bead separation (MACS; Miltenyi). Hematopoietic progenitors or GMP were purified by FACS using human-specific CD34 and CD38 antibodies and a lineage mixture as previously described (1, 2). For RNA-seq and qRT-PCR analyses, FACS-purified progenitor cells from normal cord blood, CML chronic phase (CP), or CML blast crisis (BC) samples were sorted directly into RLT buffer (Qiagen) as described previously (3). Lysates were processed for RNA extraction using RNeasy Micro kits (Qiagen).

Transcriptome Profiling and Analysis by RNA-Seq. RNA extracted from 50,000 normal cord blood, CML CP, or CML BC progenitor cells (20 ng) was analyzed for purity, and samples with RNA integrity numbers ≥7 were prepared for RNA-seq analysis. RNA was processed using SMART cDNA synthesis (Clontech) and NEBNext paired-end DNA Sample Prep Kit (New England BioLabs) to prepare libraries. Samples were sequenced on Illumina HiSeq2000 instruments with an average of ∼161 million Chastity-passed paired reads, 50 bp in length, per sample.

RNA-Seq Whole-Gene Analysis. As described previously (4), the paired-end Chastity-passed reads were aligned to the human reference genome (version hg19/GRch37) plus exon junction sequences constructed from all known transcript models in Refseq, EnsEMBL and known genes from the University of California, Santa Cruz (UCSC) database using Burrows–Wheeler Aligner software (version 0.5.7) (5) with default parameters except that the Smith–Waterman alignment was disabled. After the alignment, the reads that spanned exon–exon junctions and did not map to the human reference genome were repositioned as large-gapped alignments in the genome based on the coordinates of the exons used to construct exon junction sequences. For gene-expression quantification, we developed gene-coverage analysis software algorithms to calculate the per-base coverage across all exons annotated by EnsEMBL (version 59) and to sum up the coverage across all nonredundant exonic bases. From the coverage, we computed read counts and normalized the read counts by gene lengths and library depths, i.e., RPKM (6). Fold changes in gene-expression levels were calculated by transforming all RPKM values using the equation 1 + X.

RNA-Seq Isoform Analysis. For isoform analysis, paired reads were aligned to the hg19 reference genome with Alexa-seq (7), and isoform expression levels were computed with the Cufflinks software packages (v.1.1.0) (8) using default parameters plus upper-quartile normalization. A filtered version of the UCSC KnownGenes (9) was used as the input reference genome annotation; isoforms for which RNA-seq reads did not support (i.e., align to) every exon and splice junction were filtered out of the KnownGenes collection before input into Cufflinks.

Differential Gene Expression, Network, and Pathway Analyses. Unsupervised hierarchical clustering was performed using log2-transformed RPKM values that were subtracted by the median across each row (genes) for centering. The 2,228 genes expressed differentially in BC and CP CML were used to identify significantly enriched pathways and networks of differentially expressed genes using Ingenuity Pathway Analysis (IPA) (Ingenuity Systems, www.ingenuity.com), and the Reactome Functional Interaction (FI) plugin (10) on Cytoscape (version 2.8.1) essentially as previously described (11). The significantly enriched pathways from the Reactome FI plugin were filtered at false-discovery rate (FDR) <0.01, and the overlapping relationship between the significant pathways was visualized using EnrichmentMap (version 1.2) (12) with the Reactome FI plugin. The genes were ranked by the significance of differential expression represented in FDR.

Adenosine-to-Inosine RNA Editing Analysis of RNA-Seq Data. The number of RNA-seq reads specifying an A(T) or G(C) was determined at each of the putative RNA editing sites identified in previous work performed by others (13) for each CP ($n = 8$) and BC ($n = 8$) sample. To be considered as a site, that site must have at least five overlapping RNA-Seq reads in each of the eight CP samples and in each of eight BC samples. Then, for each site, the percentage of reads that contained a G was calculated for each sample. The percentages were averaged in eight CP and eight BC samples, and then the differences between these two percentages were computed (BC − CP). A two-sided t test was performed using the BC percentages and the CP percentages for each site. For sites with associated P values < 0.05, we computed the difference in average editing percentages (Fig. 1*D*). RNA editing fold changes were calculated by transforming all editing percentages using the equation 1 + X.

qRT-PCR. Five thousand to fifty thousand progenitor cells were CD34-selected or FACS-purified into RLT buffer, total RNA was isolated, and cDNA was synthesized as described previously (2). Primers were designed to detect human *ADAR1* (total and isoform-specific), *PU.1*, *GATA1*, *BCR-ABL*, *GSK3β*, and *HPRT* (Table S2). qRT-PCR was performed in duplicate on an iCycler using SYBR GreenER Super Mix (Invitrogen), 5 ng of template mRNA, and 0.4 mM of each forward and reverse primer. *HPRT* mRNA transcript levels were used to normalize each gene.

Cell Treatments, Lentivirus Preparation, and Transduction of Hematopoietic Progenitors. We have developed a lentiviral human BCR-ABL1 vector in the pCDH-EF1-T2A (skip)-GFP lentiviral vector (System Biosciences) (Fig. S3*A*). Vector control consists of the pCDH-GFP backbone. Lentiviral vector plasmids expressing ADAR1 p150 (no. OHS5898-101190065), or GFP-expressing backbone vector or shRNA specific to human ADAR1 (no. RHS4531), or backbone vector containing a nontargeting control shRNA (shControl), were purchased from Open Biosystems (Thermo Fisher Scientific).

For in vitro treatment of normal cord blood progenitors with inflammatory molecules, $CD34^+$ cord blood cells were incubated in microwell plates containing StemPro (Invitrogen) medium supplemented with human IL-3, GM-CSF, and stem cell factor (SCF) in the presence of IFN-γ, TNF-α (both from Cell Signaling Technology), or vehicle control for 48 h at 10-fold increasing concentrations. The concentrations used for TNF-α were 5 ng/mL, 50 ng/mL, 500 ng/mL, and 5 μg/mL, and for IFN-γ

concentrations were 50 ng/mL, 500 ng/mL, 5 μg/mL, and 50 μg/mL. RNA was extracted, qRT-PCR analysis was performed for ADAR1 isoforms as described above, and data were normalized to HPRT expression levels.

For lentivirus production, all lentiviruses were generated as described previously (2), and transduction efficiencies were validated by transduction of 293T cells and normal cord blood. Transductions of human hematopoietic stem cell and CML progenitors or GMP were performed at a multiplicity of infection (MOI) of 50–200. For transduction of human cord blood with lenti-BCR-ABL or vector control for RNA-seq, qRT-PCR, and nanoproteomics analyses, $CD34^+$ cells were selected using magnetic beads (MACS; Miltenyi Biotec), and cells were transduced and cultured for 2–5 d in microwell plates containing StemPro (Invitrogen) medium supplemented with human IL-3, GM-CSF, and SCF. RNA was harvested in RLT buffer as described above for qRT-PCR and RNA-seq analyses, and a subset of wells was harvested in lysis buffer for nanoproteomics analyses.

For transduction of leukemia stem cells (LSC) or human cord blood for in vitro colony assays, cells were stained and sorted by FACSAria (Becton Dickinson) for the $CD34^+CD38^+Lin^-PI^-$ cells, which then were transduced with lentiviruses in 96-well plate (10,000–20,000 cells per well) for 48–72 h and then transferred to methylcellulose for colony assays as described previously (2, 3). Transduction of normal cord blood progenitors with lentiviral vectors driving human ADAR1 (p150) overexpression (lenti-ADAR1) or ADAR1-targeting shRNA (lenti-shADAR1) compared with control backbone lentiviral vectors showed that lenti-ADAR1 conferred approximately two- to threefold higher expression than endogenous human ADAR1 p150 levels (Fig. S3*B*), whereas lenti-shADAR1 transduction resulted in ∼50% knockdown of endogenous human ADAR1 as confirmed by qRT-PCR (Fig. S3*C*). For in vivo experiments using lenti-shADAR1 knockdown, qRT-PCR analysis of lenti-shADAR1–transduced CML BC progenitors before primary transplant showed a 50% reduction in ADAR1 p150 mRNA levels compared with control vector (shControl) (Fig. 4*A*). Similarly, qRT-PCR analysis of pooled human progenitor cells harvested from the bone marrow of primary transplant recipients of CML BC progenitors transduced with lenti-shADAR1 maintained a 50% reduction in ADAR1 p150 mRNA levels compared with control vector (shControl) (Fig. 4*A*).

For lenti-ADAR1 transduction and qRT-PCR in individual colonies (Fig. S3*G*), cord blood ($n = 3$) and CML CP ($n = 3$) progenitors were transduced with lenti-ADAR1 p150 or vector (ORF) control (48 h), followed by coculture (48 h) with mouse bone marrow stromal cells as previously described (14) and then were transferred to methylcellulose (2 wk). Individual ADAR1 $p150^+$ colonies were collected and processed for qRT-PCR with splice isoform-specific primers for lentivirus-driven ADAR1 p150, misspliced GSK3β (exons 8 and 9 deleted), and GSK3β (exon 9 deleted) and were normalized to HPRT.

Hematopoietic Progenitor Colony Assays. Following lentivirus transduction of FACSAria-sorted cord blood or CML patient progenitors, GFP^+ cells were sorted into MethoCult medium (50–100 cells per well of a 12-well plate). After 2 wk, total colonies were counted, and individual colonies for each condition were picked and replated in fresh MethoCult medium as previously described (3). Secondary colonies were counted after an additional 2 wk of culture.

Xenograft Models, Serial Transplantation, and Analysis of Human Cell Engraftment. Normal or CML $CD34^+$ cells were selected using magnetic beads (MACS; Miltenyi Biotec). Equal numbers of cells were transduced with shADAR1 or shControl lentiviruses for 3–5 d and then were and transplanted intrahepatically into immunocompromised $RAG2^{-/-}\gamma_c^{-/-}$ mice (10,000–20,000 cells per pup) according to established methods (1, 2) and as described above. All animal experiments were approved by the Animal Experimental Committee of the University of California at San Diego and were performed according to National Institutes of Health recommendations for animal use. Mice transplanted with CML LSC or no-transplant controls were screened for tumor formation or human engraftment in peripheral blood by FACS 8–10 wk after transplantation. At 10–12 wk, mice were killed, and single-cell suspensions of hematopoietic organs and tumors were analyzed via FACS for human cell engraftment. Bone marrow cells were $CD34^+$ selected to serial transplantation into $RAG2^{-/-}\gamma_c^{-/-}$ mice (10,000–20,000 cells per pup), and mice were killed after 10–12 wk.

Nanofluidic Phospho-Proteomic Immunoassay. Nanofluidic phosphoproteomic immunoassay experiments were performed with the Nanopro 1000 instrument (Cell Biosciences), and samples were analyzed in triplicate. Samples were prepared using 4 nL of 10 mg/mL lysate diluted to 0.2 mg/mL in 200 nL HNG [20 mM Hepes (pH 7.5), 25 μM NaCl, 100 μL/mL glycerol, Sigma Phosphatase Inhibitor Mixture 1 diluted 1:100, and Calbiochem Protease Inhibitor diluted 1:100]. Two hundred nanoliters of sample mix containing internal pI standards was added. The Firefly system first performed a charge-based separation (isoelectric focusing) in a 5-cm-long, 100-μm i.d. capillary. Predicted pI values were calculated with Scansite. Each sample was run on a panel of different pH gradients (pH 3–10 and pH 2–11) to optimize the resolution of different peak patterns. After separation and photoactivated in-capillary immobilization, CRKL was detected using CRKL-specific antibody (Cell Signaling Technology). B2-microglobulin (B2M) (Upstate) antibody was used to normalize the amount of loaded protein. The peaks were quantified by manually selecting the start and end of each peak and a flat baseline and calculating the area under the curve. The nanoproteomics data were normalized to B2M by dividing the measured peak area for the protein of interest by the measured peak area for B2M.

Statistical Analysis. ***qRT-PCR.*** ADAR1 isoforms, *BCR-ABL*, *PU.1*, *GATA1,* and *GSK3β* expression data were assessed for normal distribution. The Student *t* test or one-way ANOVA was used to calculate changes in ADAR1 expression levels based on whether two or more groups were compared. Results were expressed as means ± SEM. Correlation analysis of HPRT-adjusted expression levels between groups (ADAR1, BCR-ABL, PU.1, GATA1, and GSK3β) was performed by Pearson correlation and calculation of r^2 values.

RNA-seq. For differential gene-expression analysis, we used DESeq[36] (version 1.6.1) in R (version 2.14.1) and identified differentially expressed genes at an FDR of 10%. For hierarchical clustering of RNA-Seq–based gene-expression analyses, we used the complete linkage and Spearman's rank correlation coefficient as distance metrics for both sample and gene clusterings. The clustering results were visualized using Treeview (Java Cluster software, available at http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm). For the presentation in Fig. 1, we rotated the subtree of BC-01, CB-02, CB-03, and BC-25 to align BC-25 with other BC samples for comparison. Volcano plots of comparisons of RNA editing and isoform expression were performed by plotting the absolute value of the base 10 logarithm of the *t* test *P* values and the base 2 logarithm of the fold change in expression, and reference lines were plotted where $P < 0.05$ and at fold changes of ±2.5 fold.

Hematopoietic progenitor assays. The Student *t* test was used to compare colony counts in CML BC versus CML CP and normal human progenitors. Statistical analyses were performed with the aid of FlowJo, GraphPad Prism, Microsoft Excel, R, and SAS software. Results were expressed as means ± SEM.

1. Jamieson CH, et al. (2004) Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N Engl J Med* 351(7):657–667.
2. Abrahamsson AE, et al. (2009) Glycogen synthase kinase 3beta missplicing contributes to leukemia stem cell generation. *Proc Natl Acad Sci USA* 106(10):3925–3929.
3. Geron I, et al. (2008) Selective inhibition of JAK2-driven erythroid differentiation of polycythemia vera progenitors. *Cancer Cell* 13(4):321–330.
4. Morin R, et al. (2008) Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques* 45(1):81–94.
5. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
6. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621–628.
7. Griffith M, et al. (2010) Alternative expression analysis by RNA sequencing. *Nat Methods* 7(10):843–847.
8. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–515.
9. Hsu F, et al. (2006) The UCSC Known Genes. *Bioinformatics* 22(9):1036–1046.
10. Wu G, Feng X, Stein L (2010) A human functional protein interaction network and its application to cancer data analysis. *Genome Biol* 11(5):R53.
11. Shah SP, et al. (2012) The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486(7403):395–399.
12. Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010) Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* 5 (11):e13984.
13. Bahn JH, et al. (2012) Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res* 22(1):142–150.
14. Hogge DE, Lansdorp PM, Reid D, Gerhard B, Eaves CJ (1996) Enhanced detection, maintenance, and differentiation of primitive human hematopoietic cells in cultures containing murine fibroblasts engineered to produce human steel factor, interleukin-3, and granulocyte colony-stimulating factor. *Blood* 88(10): 3765–3773.

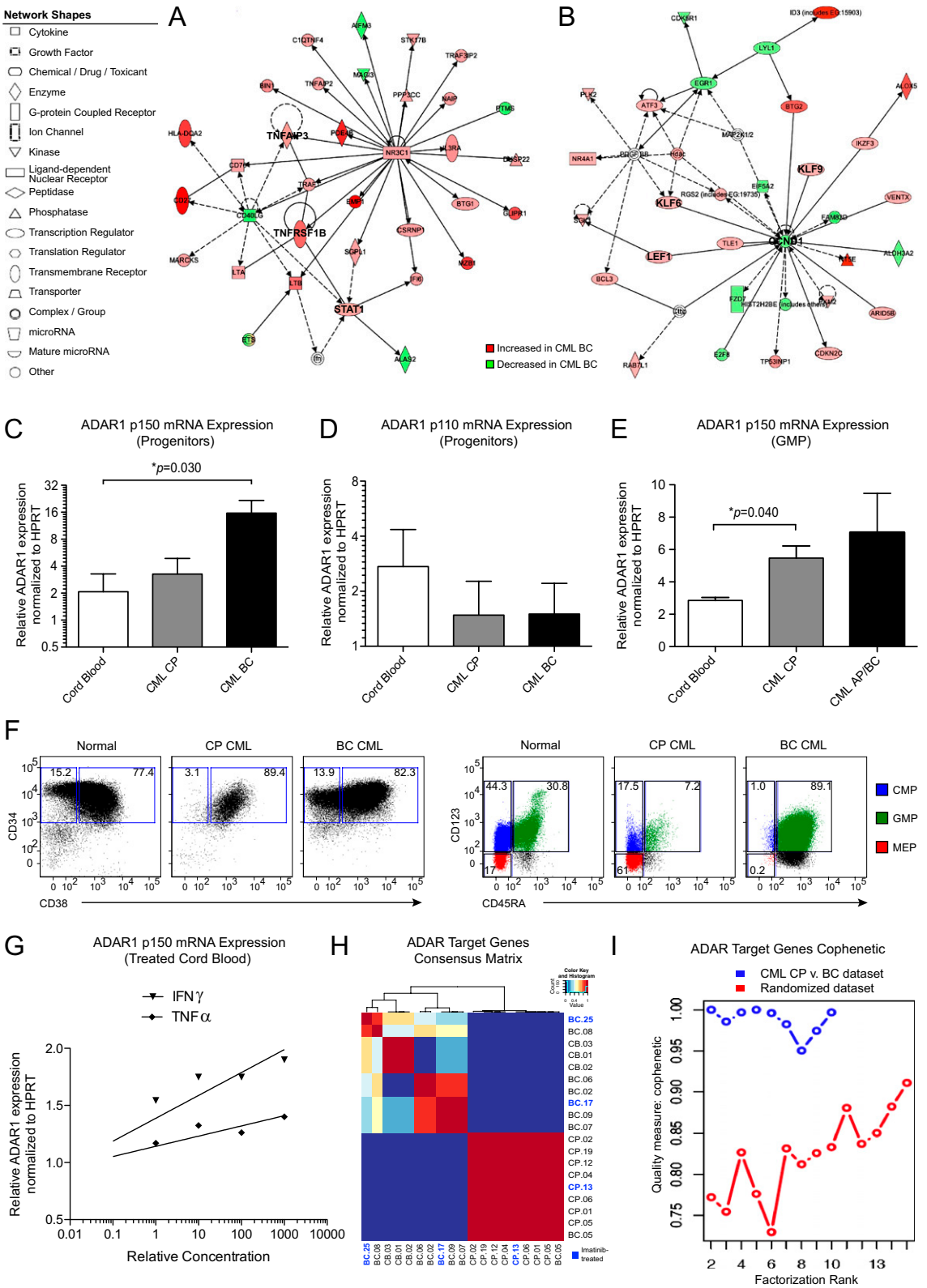

Fig. S1. Activation of inflammatory genes in BC CML and cophenetic analysis corresponding to consensus matrices. (*A* and *B*) RNA-seq–based network analysis showing activation of inflammatory pathway genes (*A*) and cell growth and reprogramming factors (*B*) in $CD34^{+}38^{+}Lin^{-}$ progenitor cells isolated from BC CML ($n = 8$) compared with CP ($n = 8$) (red, up-regulated genes; green, down-regulated genes). Networks were generated through the use of IPA (Ingenuity Systems, www.ingenuity.com). (*C* and *D*) RNA isolated from $CD34^{+}38^{+}Lin^{-}$ cells from normal cord blood ($n = 8$), CML CP ($n = 6$), and CML BC ($n = 7$) samples was analyzed by qRT-PCR with isoform-specific primers detecting human ADAR1 p150 (*C*) and p110 (*D*), and cycle threshold (Ct) values were normalized to HPRT. $P < 0.05$ compared with normal cord blood by Student *t* test. (*E*) RNA from granulocyte-macrophage progenitors (GMP) that were FACS-sorted from normal cord blood ($n = 3$), CML CP ($n = 5$), and CML accelerated phase (AP, $n = 1$)/BC ($n = 2$) samples was analyzed by qRT-PCR with isoform-specific primers detecting

Legend continued on following page

human ADAR1 p150, and Ct values were normalized to HPRT. $P < 0.05$ compared with normal cord blood by Student *t* test. (*F*) Representative FACS plots of normal, CML CP, and CML BC peripheral blood samples. CMP, common myeloid progenitors; GMP, granulocyte-macrophage progenitors; MEP, megakaryocyte-erythroid progenitors. (*G*) Relative ADAR1 p150 expression levels were measured by qRT-PCR in normal cord blood ($n = 1$) incubated with IFN-γ or TNF-α for 48 h, and linear regression analysis was performed comparing p150 mRNA expression with IFN-γ ($r^2 = 0.8098$) and TNF-α ($r^2 = 0.8349$) relative concentration. (*H*) Consensus matrix from 200 random starts, using nonnegative matrix factorization clustering of the differentially expressed genes in BC versus CP CML, computed at $k = 3$ for the above samples. Dark blue indicates that samples were never clustered in the same cluster, and red indicates that samples always were in the same cluster; the major clusters found previously are recapitulated. Patients treated with imatinib are denoted in blue. (*I*) Cophenetic correlation coefficients corresponding to consensus matrices computed at factorization ranks between 2 and 15. The blue plot indicates the cophenetic correlation coefficients calculated using our dataset, and the red plot indicates the cophenetic correlation coefficients calculated using a randomized dataset, supporting three major groupings in the data.

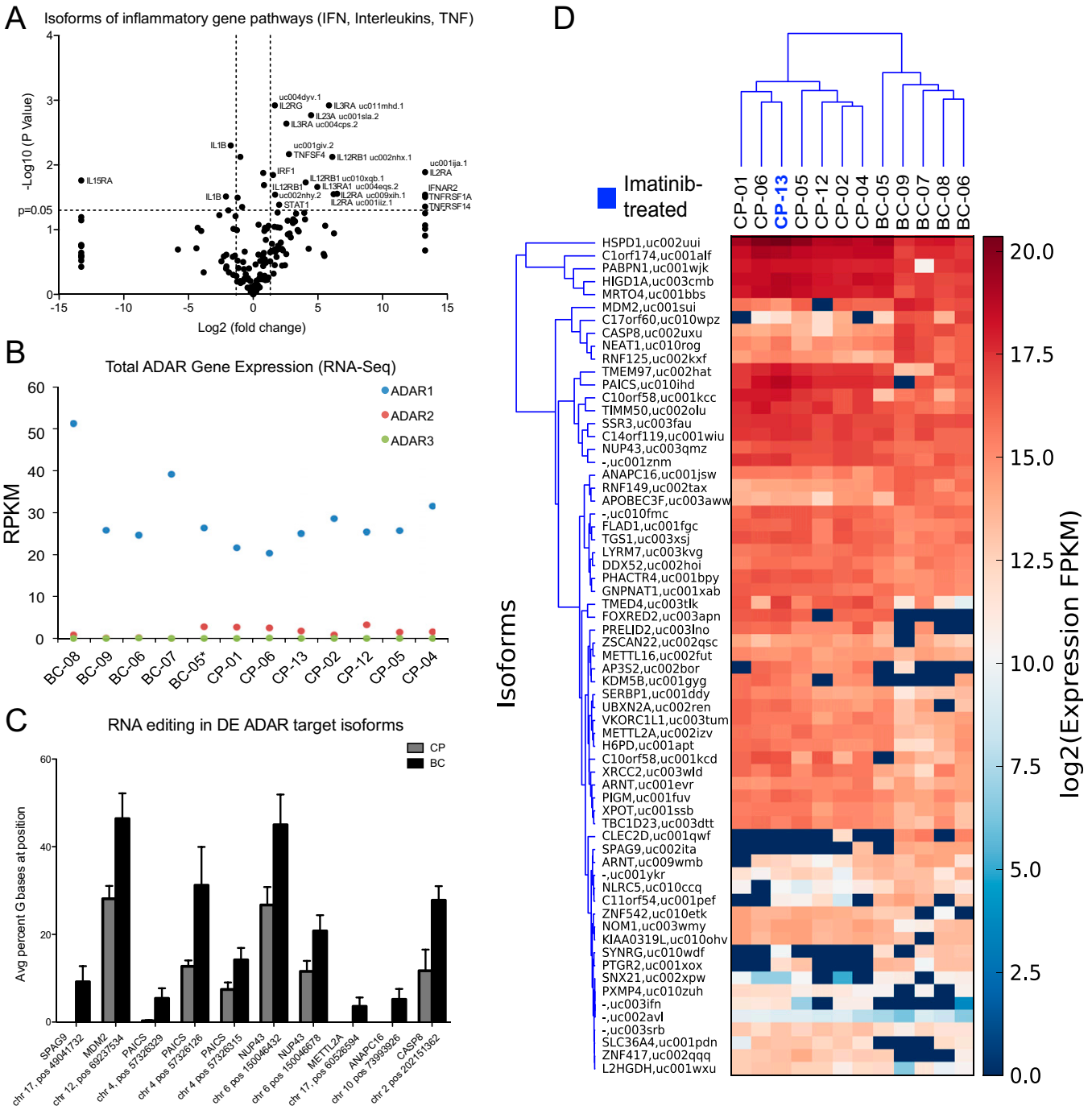

Fig. S2. RNA-seq analyses of inflammatory pathways, profiling of ADAR family genes across CML patient samples, and analysis of RNA editing and ADAR target gene-expression patterns by RNA-seq. (*A*) Isoform-level RNA-seq analysis showing differential expression and preferential up-regulation of inflammatory receptor molecule isoforms in CML BC ($n = 5$) versus CP ($n = 7$). (*B*) Total gene-expression analysis of ADAR1 (ADAR), ADAR2 (ADARB1), and ADAR3 (ADARB2) over individual CP ($n = 7$) and BC ($n = 5$) CML samples. (*C*) Editing percentages in differentially edited ADAR target sites located in transcripts that were differentially expressed in BC ($n = 5$) and CP ($n = 7$) CML. All sites shown were differentially edited at a significance of $P < 0.05$ by Student *t* test. (*D*) The top 65 isoforms of ADAR targets differentially expressed in BC CML progenitors and CP samples were identified and presented as a heatmap for each transcript in each sample ($n = 7$ CP, $n = 5$ BC). The clustergram was generated using (average) standardized Euclidean distance between (clusters of) expression vectors.

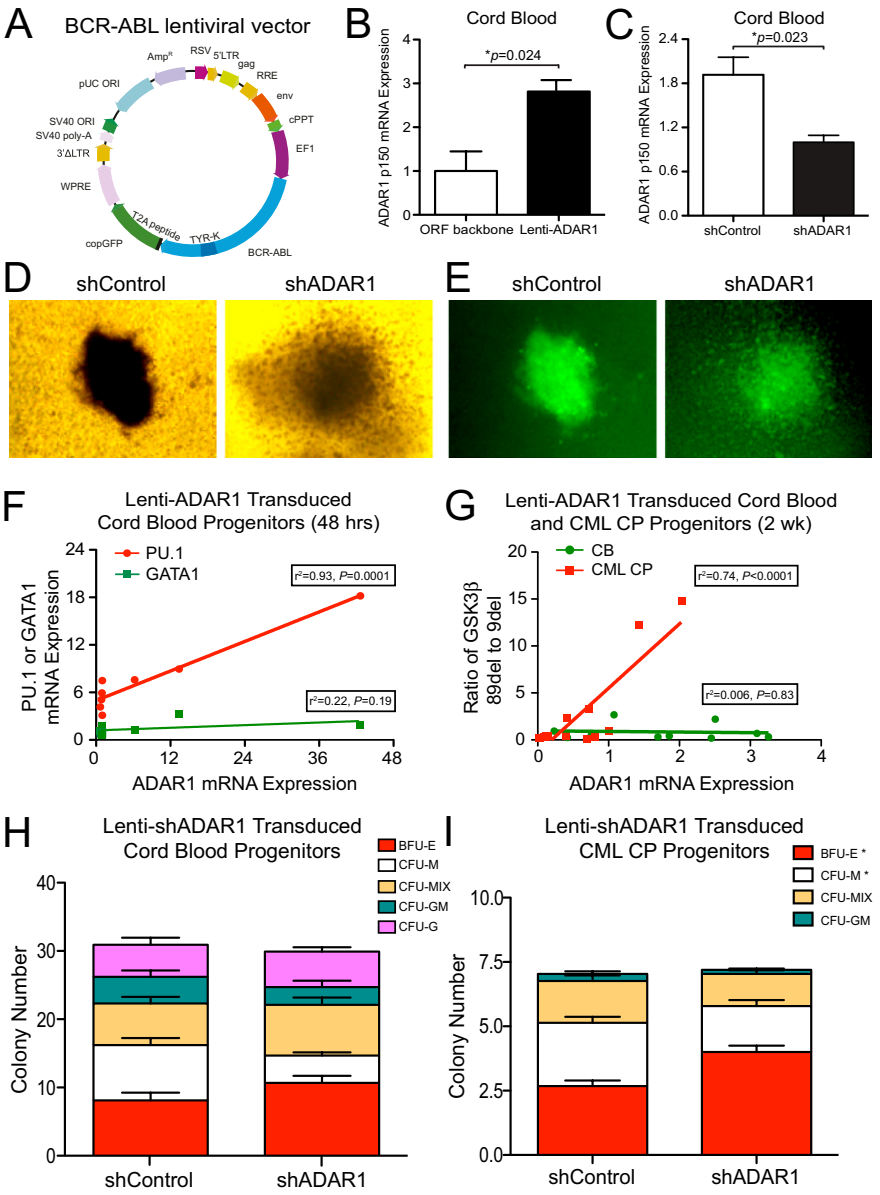

Fig. S3. Lentivirus transduction in normal cord blood and primary CML samples. FACSAria-sorted $CD34^+CD38^+Lin^-$ progenitor cells were transduced with lentiviral vectors and analyzed by qRT-PCR. (*A*) Diagram of the BCRABL1 p210 lentivirus vector construction. (*B*) qRT-PCR analysis of lenti-ADAR–transduced cord blood ($n = 1$) progenitors showing a two- to threefold increase in total ADAR1 mRNA compared with vector-transduced (ORF backbone) controls. (*C*) qRT-PCR analysis of lenti-shADAR–transduced cord blood progenitors showing a 50% reduction in total ADAR1 mRNA compared with control vector (shControl). (*D*) Bright-field microscopy showing normal cord blood colonies transduced with shRNA vectors. (*E*) Fluorescent imaging showing GFP^+ normal cord blood colonies transduced with shRNA vectors. (*F*) Pearson correlation analysis of HPRT-normalized ADAR1 mRNA levels and PU.1 ($r^2 = 0.9252$) or GATA1 ($r^2 = 0.2210$) in cord blood ($n = 2$) progenitors transduced (48 h) with ADAR1 p150 lentivirus at increasing MOI. (*G*) qRT-PCR of GSK3β splice variants in individual colonies derived from lenti-ADAR1–transduced cord blood ($n = 3$) or CML CP ($n = 3$) progenitors compared with vector-transduced controls. (*H*) Lentiviral shADAR1 transduction and hematopoietic progenitor colony assays in normal cord blood progenitors compared with vector control. (*I*) Colony formation in lentiviral shADAR1-transduced CML CP progenitors compared with shControl vector-transduced controls. $*P < 0.05$ compared with vector-transduced controls by Student *t* test.

Table S1. Summary of CML chronic-phase and blast-crisis patient samples

Patient ID	Sex/ age	Date	Sample type	WBC count (K/mm^3)	% blast (PB)	Treatment	% blast (BM)	Cytogenetics	Immunophenotyping
CP-01	M/60	Nov.13, 2008	CP	189	<5	None	<5	*t* (9;22)(q34;q11)	N/A
CP-02	F/63	May 23, 2008	CP	326	5	None	<5	46,XX,*t* (9;22)(q34;q11.2)[20], nuc ish(ABL1x3), (BCRx3), (ABL1 con BCRx2)[194/200]	N/A
CP-04	M/44	Oct.14, 2008	CP	306	5.8	None	N/A	46,XY,*t* (9;22)(q34;q11.2)[20]	N/A
CP-05	M/26	Sept. 21, 2009	CP	231	<1	None	<5	46,XY,*t* (9;22)(q34;q11.2)[20]	N/A
CP-06	F/62	Sept. 25, 2009	CP	87.7	<5	None	<5	*t* (9;22)(q34;q11.2)	N/A
CP-07	F/33	March 25, 2003	CP	134	0	N/A	N/A	*t* (9;22)(q34;q11.2)	N/A
CP-08	M/56	Jan, 27, 2099	CP	381	<5	None	N/A	*t* (9;22)(q34;q11.2)	N/A
CP-12	N/A	Aug. 26, 2009	CP	390	<5	None	N/A	N/A	N/A
CP-13	F/46	Sept. 22, 2011	CP	320	13	Imatinib	N/A	46,XX, *t* (9;22)(q34;q11.2), add(17)(p11.2~13) & T315I	CD34+ 6% of CD45+
CP-19	M/40	20-Oct-10	CP	221	13	None	<5	*t* (9;22)(q34;q11.2)	N/A
BC-02	M/34	Aug. 26, 2004	BC	241	92	None	90	*t* (9;22)(q34;q11.2)	CD11b 90%; CD13 99%; CD33 99%; CD34 88%; CD56 95%; HLA-DR 95%; CD10 21%; CD19 15%
BC-04	M/20	July 29, 2008	BC	622	68	None	10	*t* (9;22)(q34;q11.2)	N/A
BC-05	M/43	Dec. 8, 2003	BC	82.4	32	None	5	*t* (9;22)(q34;q11.2)	CD11b 95%; CD13 95%; CD33 95%; CD4 95%; CD117 48%; HLA-DR 90%
BC-06	M/30	Oct. 26, 1993	BC	170	94	Hydroxyurea	N/A	46,XY, *t* (9;22)(q34;q11.2)	CD13 70%; CD14 0.28; CD33 74%; CD19 48%; HLA-DR 49%
BC-07	M/48	Oct. 26, 1993	BC	209	86	Hydroxyurea	N/A	*t* (9;22)(q34;q11.2) abn 7	CD13 76%; CD33 17%; CD19 90%; HLA-DR 88%; CD34 82%; CD10 84%
BC-08	M/53	July 27, 2000	BC	98	83	Hydroxyurea	90	46,XY,*t* (9;22)(q34;q11), add(18)(q21).nuc ish 9q34(ABLx3), 22q11(BCRx2)	CD11 49%; CD13 100%; CD33 95%; CD56 91%; HLA-DR 92%; CD34 85%
BC-09	M/65	Oct, 17, 1991	BC	72	42	None	90	45,XY,-7, *t* (6;17;18) (p21.3;q23;p11.3), *t* (9;22)(q34;q11)	CD13 45%; CD14 68%; CD33 67%; HLA-DR 58%
BC-10	M/40	Sept. 21, 2093	BC	133	82	None	N/A	*t* (9;22)(q34;q11.2)	N/A
BC-12	F/47	July 26, 2009	BC	265	45	Hydroxyurea	N/A	*t* (9;22)(q34;q11.2)	N/A
BC-17	M/29	July 2, 2010	Lymphoid -BC	24	18	Imatinib	N/A	*t* (9;22)(q34;q11.2)	CD34 46%; HLA-DR 83%; CD10 99%; CD19 99%; TdT 99%
BC-19	M/46	Nov. 23, 2007	BC	127	30	Imatinib followed by dasatinib	40–50	*t* (9;22)-T315I	CD34+, CD13+, HLA-DR+, CD117+, MPO+, CD33+, weak CD79a, aberrant CD7+
BC-25	M/58	June 21, 2009	BC	9	32	Imatinib and hydroxyurea	N/A	*t* (9;22) and inv3(q21q26)	CD117$^-$, HLA-DR$^-$, CD33dim, CD38dim

Samples were collected before treatment except for samples noted in the table that received treatment with hydroxyurea and/or tyrosine kinase inhibitor therapy with imatinib. Patient identifiers denoted in bold and underlined were used in the RNA-Seq studies. BM, bone marrow; NA, data not available; PB, peripheral blood.

Table S2. qRT-PCR primers

Gene name	Forward primer (5′-3′)	Reverse primer (5′-3′)
ADAR1 (p150)	aacgaaagcgaaattgaacc	gggtgtagtatccgctgagg
ADAR1 (p110)	gactgaaggtagagaaggctacg	tgcacttcctcgggacac
ADAR1 (total)	tgctgctgaattcaagttgg	tcgttctccccaatcaagac
PU.1	aggcgtgcaaaatggaagg	tggcgttggtatagatccgtg
GATA1	actcgaaaccgcaaggcat	accaccataaagccaccagct
BCR-ABL	ctccagactgtccacagcat	ccctgaggctcaaagtcaga
mGSK3β (exon 8,9 del)	ctgtgtgttggctgagctgt	gggtcggaagaccttgatta
GSK3β (exon 9 del)	ctgtgtgttggctgagctgt	gtcggaagaccttagtccaa
HPRT	cgtcttgctcgagatgtgatg	tttatagccccccttgagcac

Dataset S1. Whole-transcriptome sequencing-based RNA editing analysis of significantly differentially edited ADAR target sites in CP ($n = 8$) versus BC ($n = 8$) CML $CD34^{+}CD38^{+}Lin^{-}$ progenitor cells.

Dataset S1