Supporting Information

Triviai et al. 10.1073/pnas.1401215111

SI Materials and Methods

Mice. Inbred NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice were obtained from The Jackson Laboratory in July 2009 (NSG_{UKE}) and December 2012 (NSG_{JAX}). The NSG_{UKE} mice were transferred to the specific pathogen-free facility of the Heinrich Pette Institute in April 2010 (NSG_{HPI}). The NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl} Tg(HLA-A2.1)1Enge/SzJ (NSG-HLA) mice were obtained from The Jackson Laboratory in January 2013.

Morphological and Histological Analysis of Mice. Moribund or control mice were subjected to morphologic and histologic analysis. Blood parameters were determined on a Hemavet 950 (Drew Scientific). Cytospins from bone marrow (BM) cells harvested from tibiae and femura and peripheral blood smears were stained with May-Grünwald and Giemsa stains (Sigma-Aldrich Chemie) according to the Pappenheim method. Tissue samples of liver and spleen were fixed in 4% (vol/vol) formalin and embedded in paraffin. Deparaffinated sections were stained with H&E or periodic acid Schiff (PAS) solution (Sigma-Aldrich).

FACS Analysis. Cell suspensions from bone marrow, spleen, and blood were stained with fluorescent conjugated antibodies obtained from eBiosciences, BD Biosciences, or BioLegend. Four- and six-color cytometry was performed on FACSCanto II and FACSAria (BD Biosciences) flow cytometers and analyzed with FlowJo v10.6 (Tree Star) or Diva (BD-Biosciences) analysis software.

Cell Lines, Viruses, and Virus Assays. SC1 feral mouse fibroblasts (1) and 293T human cells were maintained in DMEM supplemented with 10% (vol/vol) FCS (PAN), 2% (vol/vol) glutamine, and 1 mM sodium pyruvate. The 293T-mCat1 cells were kindly provided by Kristoffer Riecken (University Medical Center Hamburg-Eppendorf, Hamburg, Germany). Marker rescue assays, which exploit the ability of endogenous retrovirus (ERV) particles to transfer replication-defective retroviral vectors into target cells, were used to monitor ERV infectivity. Cell supernatants of spleens homogenized in PBS were filtered through 0.8-μM Millex-AA filters (Merck Millipore) and added to cultures of SC1 mouse fibroblast transduced with the SF91iGFP retroviral vector (2). After cultivation for 2 wk, serial dilutions of cell supernatant were assayed for GFP-transfer-units (GTU) on 5×10^4 cells plated into each well of a 12-well plate. Forty-eight hours after infection, the percentage of cells expressing eGFP was determined using a FACSCanto II (BD-Biosciences). Over 10^5 events were assayed to determine baseline levels. To quantitate viremia, plasma samples (20 μL and 100 μL) were added to SC1-SF91iGFP cultures (3×10^4 cells per well) in 12-well plates and spun ($640 \times g$) for 1 h at room temperature. Medium was replaced after 24 h and then in 48-h intervals. Virus titers of the supernatant were determined after 7 d of culture. Limiting dilutions of viral supernatants with known titers were tested in parallel to determine sensitivity and linear range of the assay.

Quantitative RT-PCR was performed to assess viral transcript levels. RNA was isolated by TRIzol Reagent (Invitrogen) (frozen tissue) or the RNease Micro Kit (Qiagen) (cells in culture), subjected to DNase I digestion, and converted to cDNA using High-Capacity cDNA Reverse Transcription Kits (Applied Biosystems). cDNA was used as a template for amplification of the spliced *Emv30 env* transcript or *Hprt* mRNA with Power SYBR Green PCR Master Mix (Applied Biosystems), run on a LightCycler 480 II System (Roche Applied Sciences). Primers are listed in Table S4. Data are plotted as expression of the *env* transcript relative to expression of *Hprt* in the same sample and compared with the empirically derived lowest detection limit for the *env* transcript in negative control SC1 or 293T cells (3).

Biological and Molecular Cloning of Infectious Murine Leukemia Virus. To clone infectious murine leukemia virus (MuLV), supernatant from SC1-SF91iGFP previously infected with virus secreted from the spleen of a NSG_{HPI} mouse was used to infect 293T-mCat1 cells at low multiplicities of infection (<0.3). Infected cells were then cloned by end-point dilution and 20 clones were analyzed by Southern blot analysis to confirm a single ERV copies. Marker assays confirmed that the provirus was replication competent. Overlapping proviral DNA fragments from clones #5 (ecotropic) and #11 (polytropic) were amplified using primers in Table S4, cloned into pBS, and sequenced (GenBank accession numbers KJ668270 and KJ668271, respectively). Primers spanning the U3 region of the LTR and *env* genes were used to analyze DNA isolated from 293T-mCat1 cultures infected with virus isolated from acute myeloid leukemia (AML) tumors.

NOD *Emv30* Sequence Reconstruction. The scaffolds and FM index from the de novo assembly for NOD/ShiLtJ mouse (assembly release REL-1302-Assembly) were provided by the Mouse Genomes group at the Wellcome Trust Sanger Institute and can be obtained from their FTP site (ftp://ftp-mouse.sanger.ac.uk/current_denovo) (4). We searched for the *Emv30* sequence in the scaffold sequences with BLAST+ (version 2.2.26+, blast.ncbi.nlm.nih.gov) (5) using GenBank sequence J01998.1 as query. We found hits for the query sequence on scaffold-2177464 and scaffold-404801, which we merged into a single scaffold with appropriate gaps. We filled the gap between scaffolds with SGA (6) using the FM index from the assembly. Remaining gaps in the *pol* gene were filled by PCR amplification from nonobese diabetic (NOD) DNA and sequencing. The composite Emv30 sequence is deposited in GenBank (accession number KJ668269). We mapped the insertion position with BLAT (genome.ucsc.edu) to approximately chr11:22745102 (11qA3.2) in the mm10 reference genome using unique flanking sequence as query, which was confirmed by sequence analysis of PCR fragments. Oligos used for PCR are listed in Table S4.

Southern Blot Analysis of Genomic DNA. Genomic DNA (10 μg) was digested with the indicated restriction enzyme, size separated on a 0.8% agarose Tris-Acetate-EDTA (pH 8.0) gel, and transferred onto a Biodyne nylon membrane (Pall Life Sciences). An approximately 640-bp P^{32}-radiolabeled fragment containing *Emv1 env* sequences was used to detect ecotropic provirus. Genomic DNA was isolated from spleen from the indicated mouse strains using standard methods; DNA samples from NOD/ShiLtJ (Stock#1976; P39173) and NOD.CB17-Prkdc^{scid}/J (Stock#1303; P32021) mice were purchased from The Jackson Laboratory.

Retroviral Integration Sites. Retroviral integration sites were isolated by a modified protocol of the magnetic extension primer tag selection preceding solid-phase ligation-mediated PCR (7). Briefly, genomic DNA was digested with Csp61 and subjected to primer extension using a biotinylated primer specific for the *Emv30* LTR and Taq DNA polymerase (MBI-Fermentas). The single-stranded reaction products were purified using the QIAquick PCR Purification kit (Qiagen) and coupled to Streptavidine M280Dynabeads (Invitrogen Dynal AS), followed by ligation to double-stranded linkers using T4 ligase (MBI-Fermentas). Nested PCR using primers recognizing

the linker and *Emv30* LTR sequences were used to amplify the integration site. Fragments were isolated, cloned into pUC, and sequenced. Sequences were mapped against the mm10 assembly (Genome Reference Consortium Mouse Build 38) using BLAT alignment tool (genome.ucsc.edu). PCR primers were designed to confirm integration sites in genomic DNA isolated from mouse spleens or tumor tissue (Table S4).

1. Hartley JW, Rowe WP (1975) Clonal cells lines from a feral mouse embryo which lack host-range restrictions for murine leukemia viruses. *Virology* 65(1):128–134.
2. Schwieger M, et al. (2002) AML1-ETO inhibits maturation of multiple lymphohematopoietic lineages and induces myeloblast transformation in synergy with ICSBP deficiency. *J Exp Med* 196(9):1227–1240.
3. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29(9):e45.
4. Keane TM, et al. (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477(7364):289–294.
5. Camacho C, et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.
6. Simpson JT, Durbin R (2012) Efficient de novo assembly of large genomes using compressed data structures. *Genome Res* 22(3):549–556.
7. Schmidt M, et al. (2001) Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples. *Hum Gene Ther* 12(7):743–749.

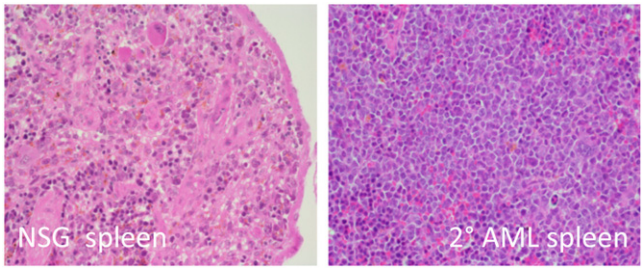

Fig. S1. AML is transplantable to secondary recipients. AML cells (5×10^6 per mouse) from bone marrow from two mice (P35-2 and P33-4) were transplanted into three NSG mice each. From each cohort, one mouse developed AML with similar characteristics of the original disease. Histological analysis of the spleen (560 mg) of the transplanted P35-2 tumor, in comparison with control NSG spleen (45 mg), demonstrates the hypercellularity and uniform predominance of myeloid blasts (H&E Staining, objective 40×).

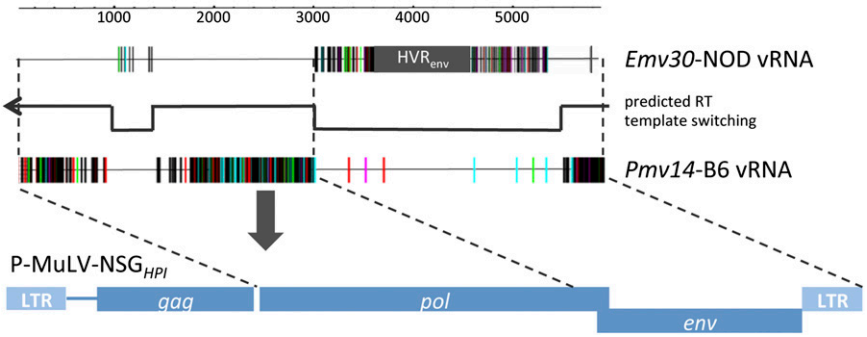

Fig. S2. Schematic representation of the recombination events leading to polytropic (P)-MuLV-NSG$_{HPI}$ isolated from AML and uninfected NSG$_{HPI}$ mice. The recombinant env sequence showed the highest identity with polytropic ERVs mapping to *Pmv19* (94.1%) and *Pmv14* (94.0%) in the C57BL/6 genome, but sequence comparison was made with *Pmv14* (Chr16: 76135043–76144023; GRCm38/mm10 Dec 2011), which has been mapped in the NOD genome (1). Alignment of Hypermut plots between P-MuLV and either Pmv14 or Emv30 sequences predict two independent recombinant events that have occurred in the generation of P-MuLV during minus-strand DNA synthesis driven by the viral reverse transcriptase (RT) using viral RNAs as templates. The Hypermut plots were generated using online tools (www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html), in which red = GG > AG, cyan = GGA > AA, green = GC > AC, magenta = GT > AT, black = not G > A transition, and yellow = gaps. The high degree of sequence variation (including deletions and insertions) in the Hypervariable Region (HVR) of P-MuLV (polytropic) and Emv30 (ecotropic) Env genes precluded generation of Hypermut plots in this region. The sequence identity between P-MuLV and Emv30 in the 5′ LTR and Gag sequences is 99%.

1. Gaskins HR, Prochazka M, Hamaguchi K, Serreze DV, Leiter EH (1992) Beta cell expression of endogenous xenotropic retrovirus distinguishes diabetes-susceptible NOD/Lt from resistant NON/Lt mice. *J Clin Invest* 90(6):2220–2227.

Table S1. Features of 10 patients from which hematopoietic stem and progenitor cells (HSPC) isolated from peripheral blood were used in xenograft studies

			Blood										
Pat Nr	Age	Sex	Hb	HCT	Leu	Plt	Cyto	JAK2 status	Fibrosis grade	Splenomegaly	IPSS	AML mice*	Splenomegaly in mice
Group I†													
P18	69	F	10.1	28.5	8.4	244	Normal	V617F	MF3	Yes	High risk	0 of 2	No (0/2)
P22	58	M	8.5	27.0	12.2	518	Normal	V617F	Unknown	Yes	Intermediate 2	0 of 2	No (0/2)
P31	57	M	12.5	36.3	3.6	99	del (4)(q25), −20,+mar	V617F	MF2	Yes	Intermediate 1	0 of 5	No (0/5)
P32	21	F	11.2	32.0	10.2	715	Normal	None	MF2	Yes	Low risk	0 of 3	No (0/3)
Group II													
P19	53	f	8.2	22.8	35.3	54	Normal	V617F	MF2	Yes	High risk	0 of 2	Yes (2/2)
P21	56	m	8.4	25.5	19.6	78	Normal	V617F	MF2	Slight	High risk	3 of 3	Yes (3/3)
P33	57	m	12.1	33.8	2.7	112	Unbalanced *t* (4;20)	V617F	MF2	Yes	Intermediate 1	5 of 5	Yes (5/5)
P35	46	m	17.7	51.8	14.6	852	Unknown	V617F	MF0	Splenectomy	Unknown	2 of 3	Yes (2/3)
P38	58	m	11.4	34.4	28.8	260	Normal	None	Accelerated	Yes	Intermediate 2	0 of 8	Yes (5/7)
P39	58	f	10	29.8	18.5	77	5q-, 13q-	V617F	MF3	Yes	High risk	3 of 5	Yes (4/5)

Cyto, cytogenetics; Fibrosis, bone marrow fibrosis grade (1); Hb, hemoglobin values (g/dL); HCT, hematocrit (%);IPSS, International Prognostic Scoring System (2); Leu, leukocyte count (x10^9/L); Pat Nr, patient number; Plt, platelet count (10^9/L).

*The increased frequency of AML induction in group II compared with group I is significant, with a double-sided *P* value (sum of small *P*s) of 0.013 determined by the Fischer's exact test (3).

†Two groups of patients were defined based on the degree of myeloproliferation (correlating with spleen size) observed in recipient mice. None of the mice receiving HSPC from patients of group 1 developed splenomegaly, as depicted in Fig. *4B*, and indicated in the last column.

1. Thiele J, et al. (2005) European consensus on grading bone marrow fibrosis and assessment of cellularity. *Haematologica* 90(8):1128–1132.
2. Cervantes F, et al. (2009) New prognostic scoring system for primary myelofibrosis based on a study of the International Working Group for Myelofibrosis Research and Treatment. *Blood* 113(13):2895–2901.
3. Uitenbroek DG (1997) SISA-Binomial. Available at www.quantitativeskills.com/sisa/distributions/binomial.htm. Accessed March 4, 2014.

Table S2. Summary of NSG_{HPI} mice that were monitored for AML induction in parallel to primary myelofibrosis (PMF)-xenografted mice

Treatment*	No. mice	Months	AML ?
PBS	6	4–12	No
CD34+_BM	7	6	No
CD34+_UCB	21	6–7	No
None	68	4–11	No
TOTAL	102	4–12	No

*Animals were intravenously injected with with PBS or $CD34^+$ HSPC isolated either from normal BM or umbilical cord blood.

Table S3. Sequence variation between *Emv30* (NOD) and *Emv2* (C57BL/6) or the proposed *Emv11* (AKR) locus

Virus Region	*Emv30* vs. *Emv2**	Amino acid substitution†	*Emv30* vs. *Emv11*‡	Amino acid substitution†
LTR-R	T 393 C		T 393 C	
LTR-U5	C 510 T			
	GT 643 del			
	C 714 T			
glyGag	C 979 T	S75L		
Gag	T 1190 C	Silent		
	C 1310 A	Silent		
	T 1525 C	L169P		
	G 1556 A	Silent	G 1556 A	Silent
	C 1790 T	Silent		
	G 1892 A	Silent		
	A 2033 G	Silent		
	A 2100 G	N361D		
Pol	A 2351 G	Silent		
			C 2756 G	Silent
	T 3024 C	Y131H		
			C 3049 T	A139V
	C 3071 T	Silent		
			G 3205 A	R191K
	G 3359 A	Silent		
	T 3797 C	C382R		
	G 3957 C	A438P§		
	T 4193 C	Silent		
			C 4252 G	T540R
	T 5364 A	S911T		
	T 5756 C	Silent		
	C 5790 T	Silent		
	T 5834 C	Silent		
	A 5956 G	Q1108R		
	A 5978 G	Silent		
	G 5996 A	Silent	G 5996 A	Silent
	G 6146 A	Silent		
Env	G 6146 A	A13T		
	A 6260 G	T50A		
	A 6296 T	T63S		
	A 6595 C	Silent		
	G 6603 A	G165E		
	A 6652 G	Silent		
	T 7096 C	Silent	T 7096 C	Silent
	T 7438 C	Silent		
	T 8145 C	I679T	T 8145 C	I679T
	A 8151 G	E681G		
LTR-U3	T 8256 C			
	G 8573 A			
			A8394 ins99bp	
LTR-R	T 8600 C		T 8600 C	

*Chr8: 123425804 123434531of the C57BL/6 genome (mm10 assembly of the Genome Reference Consortium Mouse Build 38).
†Amino acid substitution predicted from sequence change.
‡AKR-MuLV (GenBank sequence J01998.1) may deviate from the actually *Emv11* provirus from AKR mice, which has not been characterized.
§Inactivating mutation in *Emv2* pol gene.

Table S4. Oligo list for PCR reactions

Function	Fwd	Sequence	Rev	Sequence
LM-PCR				
Primer extension	cs711-btn	CTGGGGACCATCTGTTCTTGGCCTC		
First round	cs1913	CTCTCTGTACTTCCTTGTTC	cs949	GACCCGGGAGATCTGAATTC
Second round	cs1914	TGCAGTTAGCTGGCTAAGCC	cs950	AGTGGCACAGCAGTTAGG
Verification of de novo integration sites				
Chr3:30010616	cs2077	TCTGGGGAACCTTGAGACAG	cs2079	AGCCCCTAGTTCTCCACCAT
Chr7:92767833	cs2073	GCCCTAGTCCCGGTACTTTC	cs2074	TCACCTGTGCTGTGACTCCT
Chr10:37212389	cs2073	GCCCTAGTCCCGGTACTTTC	cs2075	CTGAGGTCACAAAGGGGTGT
Chr11:95246796	cs2071	ATCCTGTTTGGCCCTAGTCC	cs2072	ACCCAAAGGAGACGATTTGA
Chr14:66621123	cs2077	TCTGGGGAACCTTGAGACAG	cs2078	CCACCACTGTCTTCCTGGAG
Chr17:29788095	cs2073	GCCCTAGTCCCGGTACTTTC	cs2076	GGGTCAGGAGTTCAAGGTCA
MuLV cloning (modified from ref. 1)				
Frag1	cs1895	GCGCCAGTCCTCCGATA	cs1896	GGGGGTAAGAGCAGGGTAA
Frag2	cs1897	AGGTGTTCTCTCCTGGTCCC	cs1898	GATCCTTTCGGCTTCTCTCA
Frag9	cs2014	TTGGTCGAAAGTTAGAGCGG	cs2015	TCAGAGGTATGATTAGAGGC
Frag3	cs1899	GCTTTCCGACTTTCCCCA	cs1900	GCCAAGGTCCCAGTTTTTG
Frag4	cs1901	TCCTAACTGCCCCCGC	cs1902	GCTTGCCTTGGAAGACCC
Frag5	cs1903	ACCCCATATACGCCTGCC	cs1904	CGCACCCACACGGAGTC
Frag10	cs2018	AATTAACGCTTGCAGCTGGC	cs2019	TTCTGGACCACCACATGACC
Frag6	cs1905	AGGAGCCCCTGACTTCATATAC	cs1906	CTCGTCTTTCAAATTGGTGGTA
Frag7	cs1907	CTCCCTGTATCTCAACCACCA	cs1908	CAGAAGCGAGAAGCGAGC
Frag8	cs1909	GCATGGGAAAATACCAGAGC	cs1910	TGCAACAGCAAAAGGCTTT
Frag5-poly	cs2014	TTGGTCGAAAGTTAGAGCGG	cs2015	TCAGAGGTATGATTAGAGGC
Frag10-poly	cs250	GAACCTCGCTGGAAAGGACC	cs245	CACAACCAGCACTCTTG
Frag11-poly	cs2049	CCTTCTCAACAACCTGGGAC	cs1910	TGCAACAGCAAAAGGCTTT
Emv30 locus cloning				
Chr11-5′ flanking	cs2040	GCGTGGTCTACAGAGGGAAA	cs2041	ACGAGCCCCCAAATGAAAG
Chr11-3′ flanking	cs2042	CTGTCTCAAGGTTCCCCAGA	cs2043	GAAAATTAGGAAGGTGAAAACACA
Pol fragment	cak1	CCGCTCCAGACATTGGTCG	cak2	GTATCTGGAGGCGTCTTTATGG
Quantification of MuLV RNA				
Spliced *env* (eco)	cs2047	CCAGGGACCACCGACCCACCGT	cs2048	TAGTTGGTCCCGGTAGGCCTCG
Hprt	mkf35	GCTGGTGAAAAGGACCTCT	mkf36	CACAGGACTAGAACACCTGC
Spliced *env* (poly)	cs2053	AAAGCCTTTTGCTGTTGCAT	cs2054	CAGTGGCTTCCAGACCTCTC

1. Julien P, Thierry H, Marianne M (2006) A recombinant endogenous retrovirus amplified in a mouse neuroblastoma is involved in tumor growth in vivo. *Int J Cancer* 119(4):815–822.