Table S1. Alphabetical List of Genes deleted in the derivative chromosome of patient WHIM-09 by whole genome sequencing (related to Figure 3A)

gene_symbol	entrez_gene_id	map_location	type_of_gene	full_name
ACMSD	130013	2q21.3	protein-coding	aminocarboxymuconate semialdehyde decarboxylase
ACVR1	90	2q23-q24	protein-coding	activin A receptor, type I
ACVR1C	130399	2q24.1	protein-coding	activin A receptor, type IC
ADAM17	6868	2p25	protein-coding	ADAM metallopeptidase domain 17
ADCY3	109	2p23.3	protein-coding	adenylate cyclase 3
ANKRD30BL	554226	2q21.2	pseudo	ankyrin repeat domain 30B-like
APOB	338	2p24-p23	protein-coding	apolipoprotein B (including Ag(x) antigen)
ASXL2	55252	2p24.1	protein-coding	additional sex combs like 2 (Drosophila)
ATP6V1C2	245973	-	protein-coding	ATPase, H+ transporting, lysosomal 42kDa, V1 subunit C2
BAZ2B	29994	2q24.2	protein-coding	bromodomain adjacent to zinc finger domain, 2B
C2orf27A	29798	2q21.2	protein-coding	chromosome 2 open reading frame 27A
C2orf27B	408029	2q21	protein-coding	chromosome 2 open reading frame 27B
C2orf43	60526	2p24.1	protein-coding	chromosome 2 open reading frame 43
C2orf44	80304	2p23.3	protein-coding	chromosome 2 open reading frame 44
C2orf48	348738	2p25.1	protein-coding	chromosome 2 open reading frame 48
C2orf49	79074	2q12.1	protein-coding	chromosome 2 open reading frame 49
C2orf50	130813	2p25.1	protein-coding	chromosome 2 open reading frame 50
C2orf84	653140	2p23.3	protein-coding	chromosome 2 open reading frame 84
CCDC148	130940	2q24.1	protein-coding	coiled-coil domain containing 148
CCNT2	905	2q21.3	protein-coding	cyclin T2
CD302	9936	2q24.2	protein-coding	CD302 molecule
CENPO	79172	2p23.3	protein-coding	centromere protein O
CXCR4	7852	2q21	protein-coding	chemokine (C-X-C motif) receptor 4
CYS1	192668	2p25.1	protein-coding	cystin 1
CYTIP	9595	2q11.2	protein-coding	cytohesin 1 interacting protein

DAPL1	92196	2q24.1	protein-coding	death associated protein-like 1
DARS	1615	2q21.3	protein-coding	aspartyl-tRNA synthetase
DDX1	1653	2p24	protein-coding	DEAD (Asp-Glu-Ala-Asp) box polypeptide 1
DNAJC27	51277	2p23.3	protein-coding	DnaJ (Hsp40) homolog, subfamily C, member 27
DNAJC27-AS1	729723	2p23.3	miscRNA	DNAJC27 antisense RNA 1 (non-protein coding)
DNMT3A	1788	2p23	protein-coding	DNA (cytosine-5-)-methyltransferase 3 alpha
DPP4	1803	2q24.3	protein-coding	dipeptidyl-peptidase 4
DTNB	1838	2p24	protein-coding	dystrobrevin, beta
E2F6	1876	2p25.1	protein-coding	E2F transcription factor 6
EFR3B	22979	2p23.3	protein-coding	EFR3 homolog B (S. cerevisiae)
ERMN	57471	2q24.1	protein-coding	ermin, ERM-like protein
FAM49A	81553	2p24.2	protein-coding	family with sequence similarity 49, member A
FAM84A	151354	2p24.3	protein-coding	family with sequence similarity 84, member A
FAP	2191	2q23	protein-coding	fibroblast activation protein, alpha
FHL2	2274	2q12.2	protein-coding	four and a half LIM domains 2
FIGN	55137	2q24.3	protein-coding	fidgetin
FKBP1B	2281	2p23.3	protein-coding	FK506 binding protein 1B, 12.6 kDa
FLJ12334	400946	2p24.1	miscRNA	-
FLJ33534	285150	2p25.1	miscRNA	-
GALNT13	114805	2q24.1	protein-coding	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 13 (GalNAc-T13)
GALNT5	11227	2q24.1	protein-coding	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 5 (GalNAc-T5)
GCA	25801	2q24.2	protein-coding	grancalcin, EF-hand calcium binding protein
GCG	2641	2q36-q37	protein-coding	glucagon
GDF7	151449	2p24.1	protein-coding	growth differentiation factor 7
GEN1	348654	2p24.2	protein-coding	Gen endonuclease homolog 1 (Drosophila)
GPR39	2863	2q21-q22	protein-coding	G protein-coupled receptor 39
GPR45	11250	2q11.1-q12	protein-coding	G protein-coupled receptor 45
GREB1	9687	2p25.1	protein-coding	growth regulation by estrogen in breast cancer 1

GRHL1	29841	2p25.1	protein-coding	grainyhead-like 1 (Drosophila)
HPCAL1	3241	2p25.1	protein-coding	hippocalcin-like 1
HS1BP3	64342	2p24.1	protein-coding	HCLS1 binding protein 3
IFIH1	64135	2q24	protein-coding	interferon induced with helicase C domain 1
IL18R1	8809	2q12	protein-coding	interleukin 18 receptor 1
IL18RAP	8807	2q12	protein-coding	interleukin 18 receptor accessory protein
IL1R1	3554	2q12	protein-coding	interleukin 1 receptor, type I
IL1R2	7850	2q12	protein-coding	interleukin 1 receptor, type II
IL1RL1	9173	2q12	protein-coding	interleukin 1 receptor-like 1
IL1RL2	8808	2q12	protein-coding	interleukin 1 receptor-like 2
ITGB6	3694	2q24.2	protein-coding	integrin, beta 6
ITSN2	50618	2pter-p25.1	protein-coding	intersectin 2
KCNF1	3754	2p25	protein-coding	potassium voltage-gated channel, subfamily F, member 1
KCNH7	90134	2q24.2	protein-coding	potassium voltage-gated channel, subfamily H (eag-related), member 7
KCNJ3	3760	2q24.1	protein-coding	potassium inwardly-rectifying channel, subfamily J, member 3
KCNS3	3790	2p24	protein-coding	potassium voltage-gated channel, delayed-rectifier, subfamily S, member 3
KIF3C	3797	2p23	protein-coding	kinesin family member 3C
KLF11	8462	2p25	protein-coding	Kruppel-like factor 11
LAPTM4A	9741	2p24.1	protein-coding	lysosomal protein transmembrane 4 alpha
LCT	3938	2q21	protein-coding	lactase
LOC100129961	100129961	2q21.3	miscRNA	-
LOC100144595	100144595	-	protein-coding	-
LOC100287010	100287010	2q12.1	miscRNA	-
LOC100506421	100506421	2q12	miscRNA	-
LOC100506474	100506474	2p24	miscRNA	-
LOC100507600	100507600	-	miscRNA	-
LOC150568	150568	2q12.1	miscRNA	-

LOC284998	284998	2q12.1	miscRNA	-
LOC285000	285000	2q12.2	miscRNA	-
LOC375190	375190	2p23.3	protein-coding	-
LOC554201	554201	2q24.1	miscRNA	-
LOC645949	645949	2p24.1	miscRNA	-
LPIN1	23175	2p25.1	protein-coding	lipin 1
LY75	4065	2q24	protein-coding	lymphocyte antigen 75
LY75-CD302	100526664	2q	protein-coding	LY75-CD302 readthrough
LYPD1	116372	2q21.2	protein-coding	LY6/PLAUR domain containing 1
MAP4K4	9448	2q11.2-q12	protein-coding	mitogen-activated protein kinase kinase kinase kinase 4
MARCH7	64844	2q24.2	protein-coding	membrane-associated ring finger (C3HC4) 7
MATN3	4148	2p24-p23	protein-coding	matrilin 3
MCM6	4175	2q21	protein-coding	minichromosome maintenance complex component 6
MFSD2B	388931	2p23.3	protein-coding	major facilitator superfamily domain containing 2B
MFSD9	84804	2q12.1	protein-coding	major facilitator superfamily domain containing 9
MGAT5	4249	2q21.3	protein-coding	mannosyl (alpha-1,6-)-glycoprotein beta-1,6-N-acetyl-glucosaminyltransferase
MIR128-1	406915	2q21.3	miscRNA	microRNA 128-1
MIR1301	100302246	-	miscRNA	microRNA 1301
MIR3125	100422986	-	miscRNA	microRNA 3125
MIR3679	100500878	-	miscRNA	microRNA 3679
MIR4261	100422929	-	miscRNA	microRNA 4261
MIR4262	100422996	-	miscRNA	microRNA 4262
MIR4429	100616469	-	miscRNA	microRNA 4429
MIR4757	100616307	-	miscRNA	microRNA 4757
MIR4772	100616157	-	miscRNA	microRNA 4772
MIR4785	100616364	-	miscRNA	microRNA 4785
MIR663B	100313824	-	miscRNA	microRNA 663b
MRPS9	64965	2q12.1	protein-coding	mitochondrial ribosomal protein S9
MSGN1	343930	2p24.2	protein-coding	mesogenin 1

MYCN	4613	2p24.3	protein-coding	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)
MYCNOS	10408	2p24.1	miscRNA	MYCN opposite strand/antisense RNA (non-protein coding)
NBAS	51594	2p24	protein-coding	neuroblastoma amplified sequence
NCKAP5	344148	2q21.2	protein-coding	NCK-associated protein 5
NCOA1	8648	2p23	protein-coding	nuclear receptor coactivator 1
NOL10	79954	2p25.1	protein-coding	nucleolar protein 10
NT5C1B	93034	2p24.2	protein-coding	5'-nucleotidase, cytosolic IB
NT5C1B-RDH14	100526794	2p	protein-coding	NT5C1B-RDH14 readthrough
NTSR2	23620	2p25.1	protein-coding	neurotensin receptor 2
ODC1	4953	2p25	protein-coding	ornithine decarboxylase 1
OSR1	130497	2p24.1	protein-coding	odd-skipped related 1 (Drosophila)
PDIA6	10130	2p25.1	protein-coding	protein disulfide isomerase family A, member 6
PFN4	375189	2p23.3	protein-coding	profilin family, member 4
PKP4	8502	2q24.1	protein-coding	plakophilin 4
PLA2R1	22925	2q23-q24	protein-coding	phospholipase A2 receptor 1, 180kDa
POMC	5443	2p23.3	protein-coding	proopiomelanocortin
POTEKP	440915	2q21.1	pseudo	POTE ankyrin domain family, member K, pseudogene
POU3F3	5455	2q12.1	protein-coding	POU class 3 homeobox 3
PQLC3	130814	2p25.1	protein-coding	PQ loop repeat containing 3
PSMD14	10213	2q24.2	protein-coding	proteasome (prosome, macropain) 26S subunit, non-ATPase, 14
PTRHD1	391356	2p23.3	protein-coding	peptidyl-tRNA hydrolase domain containing 1
PUM2	23369	2p22-p21	protein-coding	pumilio homolog 2 (Drosophila)
R3HDM1	23518	2q21.3	protein-coding	R3H domain containing 1
RAB3GAP1	22930	2q21.3	protein-coding	RAB3 GTPase activating protein subunit 1 (catalytic)
RAD51AP2	729475	2p24.2	protein-coding	RAD51 associated protein 2
RBMS1	5937	2q24.2	protein-coding	RNA binding motif, single stranded interacting protein 1
RDH14	57665	2p24.2	protein-coding	retinol dehydrogenase 14 (all-trans/9-cis/11-cis)

RHOB	388	2p24	protein-coding	ras homolog gene family, member B
ROCK2	9475	2p24	protein-coding	Rho-associated, coiled-coil containing protein kinase 2
RRM2	6241	2p25-p24	protein-coding	ribonucleotide reductase M2
SDC1	6382	2p24.1	protein-coding	syndecan 1
SF3B14	51639	2pter-p25.1	protein-coding	-
SLC4A10	57282	2q23-q24	protein-coding	solute carrier family 4, sodium bicarbonate transporter, member 10
SLC9A2	6549	2q11.2	protein-coding	solute carrier family 9 (sodium/hydrogen exchanger), member 2
SLC9A4	389015	2q12.1	protein-coding	solute carrier family 9 (sodium/hydrogen exchanger), member 4
SMC6	79677	2p24.2	protein-coding	structural maintenance of chromosomes 6
SNORA80B	100302743	2p25.1	snoRNA	small nucleolar RNA, H/ACA box 80B
TAF1B	9014	2p25	protein-coding	TATA box binding protein (TBP)-associated factor, RNA polymerase I, B, 63kDa
TANC1	85461	2q24.2	protein-coding	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 1
TANK	10010	2q24-q31	protein-coding	TRAF family member-associated NFKB activator
TBR1	10716	2q24	protein-coding	T-box, brain, 1
TGFBRAP1	9392	2q12.1	protein-coding	transforming growth factor, beta receptor associated protein 1
TMEM163	81615	2q21.3	protein-coding	transmembrane protein 163
TMEM182	130827	2q12.1	protein-coding	transmembrane protein 182
TP53I3	9540	2p23.3	protein-coding	tumor protein p53 inducible protein 3
TRIB2	28951	2p24.3	protein-coding	tribbles homolog 2 (Drosophila)
TTC32	130502	2p24.1	protein-coding	tetratricopeptide repeat domain 32
UBXN4	23190	2q21.3	protein-coding	UBX domain protein 4
UPP2	151531	2q24.1	protein-coding	uridine phosphorylase 2
VSNL1	7447	2p24.3	protein-coding	visinin-like 1
WDR35	57539	2p24.1	protein-coding	WD repeat domain 35
WDSUB1	151525	2q24.2	protein-coding	WD repeat, sterile alpha motif and U-box domain containing 1

YSK4	80122	2q21.3	protein-coding	YSK4 Sps1/Ste20-related kinase homolog (S. cerevisiae)
YWHAQ	10971	2p25.1	protein-coding	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, theta polypeptide
ZRANB3	84083	2q21.3	protein-coding	zinc finger, RAN-binding domain containing 3

Table S2. List of the orientation of the 18 pieces of the derivative chromosome of patient WHIM-09 and their nucleotide site of connections derived from whole genome sequencing of neutrophil DNA (related to Figure 3B). Numbers are based on the chromosome 2 sequence of the hg19 reference version of the human genome project. Note red numbers are uncertain because repetitive sequences in the region of the connection make assigning a specific nucleotide ambiguous.

Piece 1	Piece 2	Orientation	5' End (Head)	3' End (Tail)
1	5	Head to Head	9650469	92323543
5	6	Tail to Tail	26227289	102167978
6	17	Head to Tail	102168450	156704625
17	12	Head to Head	157479872	134383040
12	3	Tail to Tail	134379440	18013669
3	15	Head to Head	18014058	137413687
15	16	Tail to Head	137402471	155186295
16	13	Tail to Tail	137413696	134383068
13	10	Head to Head	134384129	134208472
10	8	Tail to Head	133712996	132361705
8	4	Tail to Head	106220579	24238501
4	14	Tail to Tail	22647206	137344145
14	7	Head to Tail	137350425	92314297

7	**2**	**Head to Head**	**102176282**	**13364053**
2	**9**	**Tail to Tail**	**13332820**	**133689046**
9	**11**	**Head to Head**	**133712975**	**134248797**
11	**18**	**Tail to Tail**	**134208529**	**164835205**

Supplemental Figure Legends

Supplemental Figure 1. Bone marrow Testing for ALK activity (related to Figure 1D)

Two representative photomicrographs from a WHIM-09 bone marrow biopsy are shown. On the left is a typical hematoxylin and eosin stain and on the right is an immunohistochemical stain for anaplastic lymphoma receptor tyrosine kinase (*ALK*) activity. Lack of brown color indicates that although the *ALK* gene is translocated, enzymatic activity is not increased.

Supplemental Figure 2. Chromothripsis is the mechanism for loss of *CXCR4*^{R334X} in patient WHIM-09 (related to Figure 3)

(A) Microarray analysis of bone marrow cell DNA using the Affymetrix Cytoscan HD array. Analysis of copy number state and single nucleotide polymorphism variation using the Chromosome Analysis suite software revealed 7 distinct areas of chromosome 2 that were deleted (red boxes).

(B) Higher resolution microarray view of the 4th and 5th deleted regions reveals that *CXCR4* is located in the 5th deleted region.

Supplemental Figure 3. Fusion genes created by chromothripsis on chromosome 2 of patient WHIM-09 (related to Figure 3)

Gene names are indicated above the brackets delimiting gene elements. P, promoter; E, exon.

Supplemental Figure 4. Gating strategy for purification of hematopoietic stem and progenitor cells from bone marrow of patient WHIM-09 (related to Figure 4D)

Bone marrow aspirate underwent red cell lysis and washing and then was stained with marker specific monoclonal antibodies. An initial sort to enrich $CD34^+$ cells was then performed (left) prior to a second sorting procedure using the indicated gating and sorting strategy (right) to collect HSC and progenitor cells. (See also Figure 4F)

Supplemental Figure 5. Hematopoietic stem cell frequency in bone marrow from donor mice is not affected by *Cxcr4* haploinsufficiency (related to Figure 5)

Donor mice used for transplantation experiments were screened for LSK and HSC frequency in bone marrow. The following donor strains were used: $Cxcr4^{+/o}$ on a CD45.2 congenic background ($Cxcr4^{+/o}$ (CD45.2)) and its littermates ($Cxcr4^{+/+}$ (CD45.2)), and $Cxcr4^{+/S338X}$ on a CD45.1xCD45.2 background ($Cxcr4^{+/S338X}$ (CD45.1/CD45.2)) and its littermates ($Cxcr4^{+/+}$ (CD45.1/CD45.2)). A. Gating scheme (Mayle et al., 2013). HSCs are gated as $Flt3^-$ Lin^- $Sca1^+$ c-Kit^+ ($Flt3^-$LSK) which include long-term and short-term HSCs. B. Frequency of LSK cells. C. Frequency of HSCs. Data are expressed as mean $\pm$ SEM from two mice in each group, and repeated once with similar results.

Supplemental Figure 6. Lineage depletion of donor bone marrow cells does not affect the engraftment advantage of *Cxcr4* haploinsufficient cells (related to Figure 6)

(A) Engraftment advantage of $Cxcr4^{+/o}$ vs. $Cxcr4^{+/S338X}$ lineage-depleted donor bone marrow. Bone marrow cells from donors with a $Cxcr4^{+/o}$ genotype on a homozygous CD45.2 background and $Cxcr4^{+/S338X}$ on a heterozygous CD45.1/45.2 background were

mixed equally and then 5 million whole bone marrow cells (i) or 400,000 lineage depleted bone marrow cells (ii) were injected intravenously into each CD45.1 recipient mouse that had been lethally irradiated 8 hours before injection.

(B) Engraftment advantage of *Cxcr4*$^{+/o}$ vs. *Cxcr4*$^{+/+}$ lineage-depleted donor bone marrow. Bone marrow cells from donors with a *Cxcr4*$^{+/o}$ genotype on a homozygous CD45.2 background and *Cxcr4*$^{+/+}$ on a heterozygous CD45.1/CD45.2 background were mixed equally and then 5 million whole bone marrow cells or 400,000 lineage depleted bone marrow cells were injected intravenously into each CD45.1 recipient mouse that had been lethally irradiated 8 hours before injection.

All results shown in A and B are from a single experiment. Blood draws at weeks 2, 4 and 6 for flow cytometry were performed to determine which mouse bone marrow cell preparations would engraft more readily. Boxes in the FACS plots at the left of each section show the percentage of total donor cells injected that were Lin$^-$. Time course data are shown as the mean ± SEM results from whole bone marrow cell transplanted mice (n = 2) or lineage depleted bone marrow cell transplanted mice (n= 5) for the various blood leukocyte subpopulations shown.

Supplemental Figure 7. Long-term reconstitution of blood leukocytes after competitive transplantation with *Cxcr4* haploinsufficient bone marrow (related to Figure 6)

Bone marrow cells from *Cxcr4*$^{+/o}$ donors on a homozygous CD45.2 background were mixed equally with bone marrow cells from either : *Cxcr4*$^{+/S338X}$ donors on a heterozygous CD45.1/CD45.2 background (*Cxcr4*$^{+/o}$/*Cxcr4*$^{+/S338X}$) or *Cxcr4*$^{+/+}$ donors on a heterozygous CD45.1/CD45.2 background (*Cxcr4*$^{+/o}$/*Cxcr4*$^{+/+}$). Five million bone marrow cells from the mixed pairings were then

injected intravenously into each CD45.1 recipient mouse that had been lethally irradiated 8 hours before infection. Mouse blood was analyzed 39 weeks after bone marrow transplantation for total white blood cell and differential counts (5 recipient mice for each donor pair).

Extended Experimental Methods

Flow Cytometry, Cell Purification, and Culture

Patient cells were isolated from heparinized blood or bone marrow aspirate and purified by Ficoll-Hypaque centrifugation and stained as previously described (McDermott et al., 2014). Cell purification was performed either using magnetic bead selection (Miltenyi Biotec, Cambridge, MA) following manufacturer recommendations for blood samples or using a sorting cytometer (FACS Aria II, BD Biosciences, San Jose, CA) for bone marrow samples. Flow cytometric data were analyzed using FlowJo v10 software (TreeStar Inc., Ashland, OR). Peripheral blood mononuclear cells (PBMC) and purified bone marrow cells were cultured to expand $CD34^+$ hematopoietic progenitor cells (HPC), and PBMC were used to create lymphoblastoid cell lines (LCL) from Epstein-Barr virus (EBV)-transformed B cells. 100 µl of mouse blood was collected from the mandibular vein into EDTA anticoagulant tubes (Becton Dickinson, Franklin Lakes, NJ), then 2 µl of Fc block (BioLegend, San Diego, CA) was added and incubated for 10 min prior to incubation with specific antibodies on ice for 30 min. Red blood cells were then lysed with 1 ml ACK lysis buffer (Quality Biologicals, Inc., Gaithersburg, MD) for 1 min at room temperature before centrifugation. Cells were then washed twice with flow cytometry buffer, fixed with 1% paraformaldehyde, stored on ice, and analyzed using an LSRII FACS cytometer (BD Biosciences).

Mutation Detection and Sanger DNA Sequencing

Polymerase chain reaction - restriction fragment length polymorphism (PCR-RFLP) analysis was performed as previously described (Hernandez et al., 2003) using genomic DNA isolated from purified cell populations or whole blood. Amplification utilized the following two primers: 5'-ATCCTCTATGCTTTCCTTGGAGCC-3' and 5'-GTGGAAACAGATGAATGTCCACCGC-3' which

generate a 126 bp amplicon from both the wild type *CXCR4* allele and the *CXCR4*^{R334X} (1000C-->T) mutant allele that causes WHIM syndrome. The mutation destroys a naturally occurring *Bst*UI restriction enzyme site in the amplicon, such that the enzyme does not cut the mutant amplicon but generates 104 and 22 bp digested fragments from the wild type amplicon. In addition the same DNA was subjected to traditional Sanger method DNA sequencing at a Clinical Laboratory Improvement Amendment of 1988 (CLIA-88) certified core facility. We also designed a PCR assay using the following 2 primers: 5'-GGAATCTCACATCCAGAATCATGC-3' and 5'-TTGTTCTCACCTTTGGCCAGTGG-3' which selectively generate a 242 bp amplicon spanning the junction of pieces 13 and 16 in the patient's derivative chromothriptic chromosome 2. Amplicons for 3 other chromothriptic junctions yielded similar results (data not shown).

Cytogenetics

Cytogenetic analysis was performed on bone marrow aspirate samples, purified CD19^{+} B cells, and LCL cell lines using standard techniques (Quest Diagnostics, Chantilly, VA). GTW (G-banding by Trypsin treatment followed by Wright stain) banded metaphases were obtained using established harvesting and banding techniques.

Fluorescence in situ hybridization (FISH)

FISH was performed with Abbott Molecular probes (Abbott Molecular, Des Plaines, IL) following the manufacturer's recommended protocol.

Affymetrix CytoScan® HD SNP-oligo microarray

Affymetrix Cytoscan HD array (Affymetrix, Santa Clara, CA) was used to identify chromosomal breakpoints using DNA from the same cells used in cytogenetic analysis, purified neutrophils and skin fibroblasts. Microarray analysis was performed following the manufacturer's protocol. Each sample was digested with *Nsp*I, ligated to adaptors, and amplified with PCR. The PCR product was purified using a magnetic separation technique, fragmented, and labeled before hybridization to the microarray. The oligo-SNP array contains approximately 2.67 million probes, including 1.9 million copy number probes and 750,000 SNP probes. Copy number alterations were analyzed using Chromosome Analysis Suite (ChAS) software (Affymetrix).

Molecular Pathology Testing and Cytokine Analysis

A formalin-fixed, paraffin-embedded (FFPE) block was obtained from the patient's spleen, removed by surgical splenectomy in 1963. This was examined histologically and DNA was extracted using a standard protocol. The DNA was then used in the *Bst*U1 PCR-RFLP analysis. A variety of short tandem repeat (STR) loci were compared between the patient's neutrophil and buccal DNA using an STR chimerism assay developed for the tracking of bone marrow transplant engraftment at the Hematology Service, Division of Laboratory Medicine, NIH Clinical Center. This assay is capable of detecting 5-95% donor chimerism and has a standard deviation of ±5%. Purified $CD3^+$ and $CD19^+$ DNA were checked for T and B cell clonality using previously published methods (Lawnicki et al., 2003; Ramasamy et al., 1992; van Dongen et al., 2003). Finally, a specific mutation (V617F) in the *JAK2* oncogene and presence of the *BCR-ABL* gene fusion were queried using DNA sequencing and a specific real time PCR assay, respectively (further detailed below).

Whole Genome Sequencing and Analysis

DNA from fibroblasts obtained from a skin biopsy and from peripheral blood neutrophils were used to prepare paired-end libraries using the TruSeq DNA protocol following the manufacturer's instructions (Illumina Inc., San Diego, CA). For each DNA sample, two different size libraries were generated, one of 200 to 400 bp and one of 300 to 500 bp. Each library was sequenced on 2 lanes of a HiSeq 2000 next generation sequencer (Illumina) generating ~380 million reads per library of 2 x 100 bp paired-ends. Reads were trimmed for adaptor and poor quality sequence and mapped to build hg19 of the human genome using the fusion search option in publically available Tophat2 software from the McKusick-Nathans Institute of Genetic Medicine (Kim et al., 2013). The output from this software (fusion.out file) was used to identify candidate fusion junctions that were present in the neutrophil but not the fibroblast samples. All fusion junctions were verified by visual inspection using publicly available Integrative Genomics Viewer (IGV version 2.3) software from the Broad Institute (Robinson et al., 2011). Rearrangements were visualized using publicly available Circos software from the Canadian Michael Smith Genome Centre (Krzywinski et al., 2009).

Flow Cytometric Sorting, HSC selection and culture, and Colony-Forming Cell Assays

A FACS Aria II (SORP) fluorescence activated cell sorter (Becton-Dickinson Biosciences, San Jose, CA), equipped with 405 nm, 488 nm, 561 nm, and 633 nm lasers, was used to sort cells. For sorting of hematopoietic progenitors from human bone marrow, Enrich mode (Purity mask=0; Yield mask=8) was performed first using anti human CD34-Phycoerythrin (Biolegend, San Diego, CA) to select $CD34^+$ cells followed by Purity mode sorting for HSC, CLP, CMP, and GMP populations using the following monoclonal Abs:

CD135-Allophycocyanin, CD10- Fluorescein Isothiocyanate, CD45-PerCP, CD90-Alexa Fluor 700, CD45RA-Pacific Blue (all from Biolegend, San Diego, CA) and CD38-PECy7 (BD Biosciences, San Jose, CA). The following purified anti-human monoclonal antibodies labeled with fluorochromes were obtained from BD Biosciences and used for leukocyte subset sorting from whole blood: CD14-Fluorescein Isothiocyanate, CD19- Phycoerythrin, CD4-Allophycocyanin, and CD8-Pacific Blue. NK cells were purified from whole blood by positive selection using biotin-conjugated anti-human monoclonal antibody to CD56 (BD Biosciences, San Jose, CA) and anti-biotin microbeads (Miltenyi Biotec, San Diego, CA). Anti-human monoclonal antibodies used for sorting of leukocyte subsets from bone marrow included CD14-Fluorescein Isothiocyanate, CD19-Phycoerythrin, CD34-Allophycocyanin, and CD3-Alexa Fluor 700 (BD Biosciences). In other experiments, blood and bone marrow derived $CD34^+$ cells were positively selected with Miltenyi Biotec (San Diego, CA) microbeads then were cultured using the Arteriocyte (Hopkinton, MA) NANEX HSC expansion system following the manufacturer's directions. In some experiments, the $CD34^+$ cells were cultured in HSC expansion serum free media (Stem Cell Technologies, Vancouver, Canada) supplemented with the cytokine cocktail (hSCF, hTPO, hflt3-L) obtained from Peprotech (Rocky Hill, NJ) and human LDL from Athens Research and Technology (Athens, GA). Human colony-forming cell assays were done using Methocult medium purchased from Stem Cell Technologies (Vancouver, Canada) according to the manufacturer's instructions. Briefly, PBMCs were resuspended in IMDM+2% FBS medium at a concentration of $2X10^6$/ml. Then 0.3 ml of PBMCs were added to 3 ml Methocult media, mixed vigorously, and allowed to stand at room temperature for 5 minutes. Then using a syringe, 1.1 ml of the mixture was injected into a well of a 6- well-plate. Each donor was tested in

duplicate. The plate was cultured for 14 days, followed by counting the BFU-E, CFU-GM and CFU-GEMM colonies. For DNA isolation from BFU-E colonies, we collected the BFU-E colonies under light microscopic guidance and isolated genomic DNA.

T cell clonality testing

DNA was extracted from peripheral blood cells using an automated DNA extraction system and tested for quality prior to PCR amplification for detection of T cell receptor (TRG locus) gene rearrangements. A single multiplexed PCR reaction was performed as described by Lawnicki et al. using primers that interrogate TRG rearrangements involving all of the known Vg family members, and the Jg1/2, JP1/2 and JP joining segments. (Lawnicki et al., 2003) To allow for fluorescence detection, each joining region primer was covalently linked to a unique fluorescent dye. The products were analyzed by capillary electrophoresis on an ABI 3130xl Genetic Analyzer, and electropherograms were analyzed using GeneMapper software version 4 (Applied Biosystems, Life Technologies, Grand Island, NY). A polyclonal rearrangement pattern was detected and there was no evidence of a clonal T-cell process, within the detection limits of the assay. TRG PCR is capable of detecting a clonal population comprising a minimum of 2-5% of the total T-cell population, and as performed at the NIH-Clinical Center, can identify > 90-95% of all TRG gene rearrangements occurring in clonal T-cell proliferations.

B cell clonality testing

DNA from peripheral blood cells isolated as above was PCR-amplified for detection of immunoglobulin (IGH and IGk loci) gene rearrangements. For the IGH locus, two separate reactions were performed, one using consensus primers to framework region III and

the joining region of the immunoglobulin heavy chain gene (FRIII-IGH PCR) and a second using primers directed to the framework region II and the joining region of the immunoglobulin heavy chain gene (FRII-IGH PCR) as described by Ramasamy, et al. (Ramasamy et al., 1992) Primers directed to the joining region were linked to the fluorescent marker FAM (carboxyfluorescein). Two additional reactions were performed for the IGκ locus using the Biomed II primer set described by van Dongen et al and supplied by InVivoScribe Technologies (San Diego, CA) using the IGK Gene clonality assay and ABI Fluorescence Detection).(van Dongen et al., 2003) These reactions interrogate rearrangements involving the Vκ loci and Jκ, the Vκ locus and the κDE locus, and the κ intron RSS locus and the κDE locus. Products were run on an ABI 3130xl Genetic Analyzer and electropherograms were analyzed using GeneMapper software version 4 (ABI). Polyclonal rearrangement patterns were detected in all reactions and there was no evidence of clonal B cell proliferation within the limits of the assay. These assays performed at the NIH-Clinical Center are capable of detecting a clonal population comprising 2-10% of the total B cell population and can identify about 85-90% of all clonal B cell proliferations.

JAK2 and BCR-ABL Mutation Analysis

The JAK2 V617F mutation was not detected on CLIA-88 approved testing developed by Ipsogen (Qiagen, Germantown, MD). A quantitative PCR assay was used to determine the relative levels of BCR-ABL and ABL using cDNA obtained by reverse transcribing RNA extracted from peripheral blood leukocytes. The levels of BCR-ABL and ABL were quantitated by comparison to plasmid reference standards using real time PCR. The ratio (BCR-ABL/ABL) was found in WHIM-09 to be 0.12%. In patients with newly diagnosed or relapsed chronic myeloid leukemia this typically varies between 50-150%. The coefficient of variation for replicate assays at the NIH-Clinical Center is 40-50%.

Competitive Mouse Bone Marrow Repopulation Experiments

To study the selective advantage for long-term engraftment of transplanted *Cxcr4*$^{+/o}$ donor bone marrow over donor bone marrow from the *Cxcr4*$^{+/S338X}$ mouse model of WHIM syndrome (Figure 5A), bone marrow cells from *Cxcr4*$^{+/o}$ donors on a CD45.2 background and from *Cxcr4*$^{+/S338X}$ donors on a heterozygous CD45.1/CD45.2 background were mixed equally and injected intravenously into CD45.1 congenic recipient mice that had been lethally irradiated 8 hours previously. Peripheral blood was drawn every 2 weeks until day 105 after transplant and the differential was determined by flow cytometry. The monoclonal Abs used were CD45.1- PECy7 and CD45.2-eFlour450 from eBioscience, and Ly6G-APC-Cy7, CD11b-PerCP-Cy5.5, CD19-TITC, and CD3-APC from Biolegend. 5 male and 5 female recipient mice were tested with similar results and the data were combined for the Figure; the results were verified in one additional independent experiment. To study the selective advantage for long-term engraftment of transplanted *Cxcr4*$^{+/o}$ over *Cxcr4*$^{+/+}$ donor bone marrow (Figure 5B), bone marrow cells from *Cxcr4*$^{+/o}$ donors on a CD45.2 background were mixed with bone marrow cells from *Cxcr4*$^{+/+}$ donors on a CD45.1 background (42%:58%, respectively) and then injected intravenously into 5 male congenic CD45.1 and 5 male congenic CD45.2 recipient mice that had been lethally irradiated 8 hours previously. Peripheral blood was drawn every 2 weeks until day 303 after transplant and leukocyte subsets was determined by flow cytometry. The results were similar for both congenic recipients, and the data were combined in the figure shown. The results were verified in two additional independent experiments. Mice in the first experiment were sacrificed at day 303 post transplantation and bone marrow was harvested for HSC and HPC quantification by flow cytometry. The additional monoclonal Abs used for this

analysis were Lin-FITC (FITC conjugated antibodies against lineage markers including B220, CD3, CD4, CD8, CD11b, Gr1, and Ter119), Sca1-APC-Cy7, c-Kit-APC, Flt3-PE-Cy5, and IL-7Rα-Violet 605 (all from Biolegend) and biotinylated CD34-streptavidin-PE (clone RAM) from eBioscience. To study the proliferation of HSCs (Figure 6A), bone marrow cells from donors with a *Cxcr4*$^{+/o}$ genotype on a homozygous CD45.2 background were mixed (47%:53%) with bone marrow cells from donors with a *Cxcr4*$^{+/+}$ genotype on a heterozygous CD45.1/CD45.2 background and then 5 million bone marrow cells were injected intravenously into each CD45.1 recipient mouse that had been lethally irradiated 8 hours prior. Six days after bone marrow transplantation, each mouse was given 1.25 mg of BrdU I.P. Twenty hours later, the mice were euthanized for HSC proliferation analysis. The experiment was repeated once with similar results. Lineage depletion of mouse bone marrow cells was performed with a Lineage Depletion Kit from Miltenyi Biotec (San Diego, CA). Briefly, whole mouse bone marrow cells were magnetically labeled with a cocktail of biotinylated antibodies against a panel of "lineage" antigens (CD5, CD45R [B220], CD11b, Anti-Gr-1 [Ly-6G/C], 7-4, and Ter-119 antibodies) and Anti-Biotin MicroBeads. The cells were then run through magnetic column of autoMACS™ separator, and the un-labeled lineage negative cells were collected for transplantation.

Supplemental References

Hernandez, P.A., Gorlin, R.J., Lukens, J.N., Taniuchi, S., Bohinjec, J., Francois, F., Klotman, M.E., and Diaz, G.A. (2003). Mutations in the chemokine receptor gene CXCR4 are associated with WHIM syndrome, a combined immunodeficiency disease. Nat Genet *34*, 70-74.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology *14*, R36.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome research *19*, 1639-1645.

Lawnicki, L.C., Rubocki, R.J., Chan, W.C., Lytle, D.M., and Greiner, T.C. (2003). The distribution of gene segments in T-cell receptor gamma gene rearrangements demonstrates the need for multiple primer sets. The Journal of molecular diagnostics : JMD *5*, 82-87.

Mayle, A., Luo, M., Jeong, M., and Goodell, M.A. (2013). Flow cytometry analysis of murine hematopoietic stem cells. Cytometry Part A : the journal of the International Society for Analytical Cytology *83*, 27-37.

McDermott, D.H., Liu, Q., Velez, D., Lopez, L., Anaya-O'Brien, S., Ulrick, J., Kwatemaa, N., Starling, J., Fleisher, T.A., Priel, D.A.*, et al.* (2014). A phase 1 clinical trial of long-term, low-dose treatment of WHIM syndrome with the CXCR4 antagonist plerixafor. Blood *123*, 2308-2316.

Ramasamy, I., Brisco, M., and Morley, A. (1992). Improved PCR method for detecting monoclonal immunoglobulin heavy chain rearrangement in B cell neoplasms. Journal of clinical pathology *45*, 770-775.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature biotechnology *29*, 24-26.

van Dongen, J.J., Langerak, A.W., Bruggemann, M., Evans, P.A., Hummel, M., Lavender, F.L., Delabesse, E., Davi, F., Schuuring, E., Garcia-Sanz, R.*, et al.* (2003). Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia *17*, 2257-2317.