High-throughput retroviral tagging to identify components of specific signaling pathways in cancer

Harald Mikkers[1], John Allen[1], Puck Knipscheer[1], Lieke Romeyn[1], Augustinus Hart[2], Edwin Vink[1] & Anton Berns[1]

Published online: 19 August 2002, doi:10.1038/ng950

Genetic screens carried out in lower organisms such as yeast[1], *Drosophila melanogaster*[2] and *Caenorhabditis elegans*[3] have revealed many signaling pathways. For example, components of the RAS signaling cascade were identified using a mutant eye phenotype in *D. melanogaster* as a readout[2]. Screening is usually based on enhancing or suppressing a phenotype by way of a known mutation in a particular signaling pathway. Such *in vivo* screens have been difficult to carry out in mammals, however, owing to their relatively long generation times and the limited number of animals that can be screened. Here we describe an *in vivo* mammalian genetic screen used to identify components of pathways contributing to oncogenic transformation. We applied retroviral insertional mutagenesis in *Myc* transgenic (Eμ*Myc*) mice lacking expression of *Pim1* and *Pim2* to search for genes that can substitute for *Pim1* and *Pim2* in lymphomagenesis. We determined the chromosomal positions of 477 retroviral insertion sites (RISs) derived from 38 tumors from Eμ*Myc Pim1*$^{-/-}$ *Pim2*$^{-/-}$ mice and 27 tumors from Eμ*Myc* control mice using the Ensembl and Celera annotated mouse genome databases. There were 52 sites occupied by proviruses in more than one tumor. These common insertion sites (CISs) are likely to contain genes contributing to tumorigenesis. Comparison of the RISs in tumors of *Pim*-null mice with the RISs in tumors of Eμ*Myc* control mice indicated that 10 of the 52 CISs belong to the *Pim* complementation group. In addition, we found that *Pim3* is selectively activated in *Pim*-null tumor cells, which supports the validity of our approach.

Retroviral insertions in the genome can transform host cells by activating proto-oncogenes or inactivating tumor-suppressor genes[4]. Multiple rounds of retroviral insertional mutagenesis yield a full-blown tumor in which proviral insertions mark the genes collaborating in stepwise tumor development. Thus, retroviral insertions are instrumental in the clonal outgrowth of the incipient tumor cell. In accordance with this notion, two or three CISs within a single tumor are often occupied by proviruses. We have previously shown co-activation of the Pim family of serine/threonine kinases and either Myc or Nmyc1 in retrovirus-induced tumors[5,6]. The cooperation between the *Myc* and *Pim* proto-oncogenes was proven using transgenic experiments in which Eμ*Myc*Eμ*Pim1* and Eμ*Myc*Eμ*Pim2* double-transgenic mice succumbed around birth to pre-B cell leukemia[7,8]. Although the frequent retroviral activation of *Pim1* established the role of the *Pim* genes in retrovirus-induced lymphomagenesis, the crucial downstream targets of the Pim kinases are elusive. Candidate Pim substrates such as P100 (ref. 9), CDC25A (ref. 10), HP1γ (ref. 11), TFAF2/SNX6 (ref. 12), SOCS1 (ref. 13) and NFATC (ref. 14) have been described, but it is still unclear to what extent they contribute to Pim-mediated transformation.

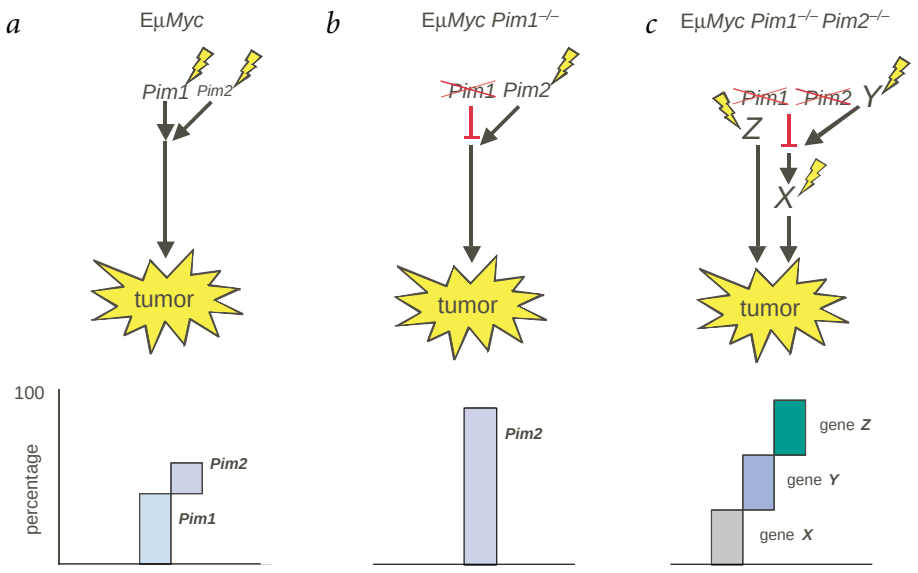

Fig. 1 Retroviral tagging in lymphoma-prone Eμ*Myc* mice that are sensitized to activation of the Pim pathway. ***a***, MoMuLV infection of Eμ*Myc* mice yields lymphomas, of which 40% and 15% have retrovirus-activated *Pim1* and *Pim2* alleles, respectively. ***b***, In 90% of the lymphomas generated in Eμ*Myc* mice lacking expression of *Pim1*, the Pim pathway has been activated through proviral insertions near *Pim2*. ***c***, Retroviral insertional mutagenesis in Eμ*Myc Pim1*$^{-/-}$ *Pim2*$^{-/-}$ mice is expected to yield lymphomas with activated oncogenic Pim signaling, either by mutation of a gene in a parallel pathway (*Y*), a gene downstream of Pim (*X*) or a *Pim*-related gene (*Z*).

[1]*Division of Molecular Genetics and Centre of Biomedical Genetics,* [2]*Division of Radiotherapy, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. Correspondence should be addressed to A.B. (e-mail: a.berns@nki.nl).*

Table 1 • Predicted frequencies of random proviral insertions in the mouse genome

	Expected number of random CISs[a]			Two insertions			Three insertions			Four insertions		
Number of tags	Efr = 0.001	Efr = 0.005	Efr = 0.01	Efr = 0.001	Efr = 0.005	Efr = 0.01	Efr = 0.001	Efr = 0.005	Efr = 0.01	Efr = 0.001	Efr = 0.005	Efr = 0.01
10,000	10	50	100	0.26 kb	1.3 kb	2.6 kb	12 kb	27 kb	39 kb	50 kb	88 kb	113 kb
5,000	5	25	50	0.5 kb	2.6 kb	5.2 kb	24 kb	54 kb	77 kb	99 kb	176 kb	227 kb
2,500	2.5	12.5	25	1.04 kb	5.2 kb	10.4 kb	47 kb	108 kb	155 kb	198 kb	351 kb	454 kb
2,000	2	10	20	1.3 kb	6.5 kb	13 kb	59 kb	135 kb	193 kb	248 kb	439 kb	567 kb
1,000	1	5	10	2.6 kb	13 kb	26 kb	118 kb	269 kb	386 kb	495 kb	878 kb	1,134 kb
500	0.5	2.5	5	5.2 kb	26 kb	52 kb	236 kb	538 kb	772 kb	991 kb	1,757 kb	2,267 kb

Ignoring end-of-chromosome effects, random proviral insertions into the mouse genome follow a Poisson distribution. The expected fraction (Efr) indicates the fraction of the total number of proviral insertion sites expected to be random CIS clusters within the specified distance. For example, 2,500 tags will contain 2.5 CISs consisting of 2 random insertions within 1.04 kb, 2.5 CISs of 3 random insertions within 47 kb, and so on. The calculated distances are based on the available mouse genome sequence at Celera (2.6×10^6 kb). [a]The expected number of CISs is the mean number of clusters for $n = \infty$ experiments.

To gain more insight into the oncogenic signaling network in which the Pim proteins act and to identify crucial downstream Pim targets, we established a mammalian *in vivo* enhancer screen similar to the genetic screens carried out in lower organisms. The synergism between *Myc* and *Pim* in lymphomagenesis probably sensitizes those mice that are deficient for Pim but express high levels of Myc to developing lymphomas in which genes acting either downstream of or parallel to *Pim* have been mutated. In fact, the percentage of retroviral activations of *Pim* observed in Eμ*Myc Pim1*$^{-/-}$ mice was almost twice that observed in Eμ*Myc* mice (Fig. 1*a*,*b*). We observed proviral activations of *Pim2*, which encodes a protein that is 57% identical to Pim1 (ref. 15), in 90% of the lymphomas in *Pim1*-null mice. These observations underscore the selective advantage of *Pim* activation in the presence of high Myc levels and suggest a strategy for identifying genes that rescue loss of Pim function in lymphomas containing activated *Myc*. In the current study we implement this strategy by infecting Eμ*Myc* newborns lacking expression of both *Pim1* and *Pim2* with Moloney murine leukemia virus (MoMuLV) (Fig. 1*c*), carrying out high-throughput sequence analyses of the proviral insertion sites, mapping the insertions and nearby candidate target genes (using the annotated mouse

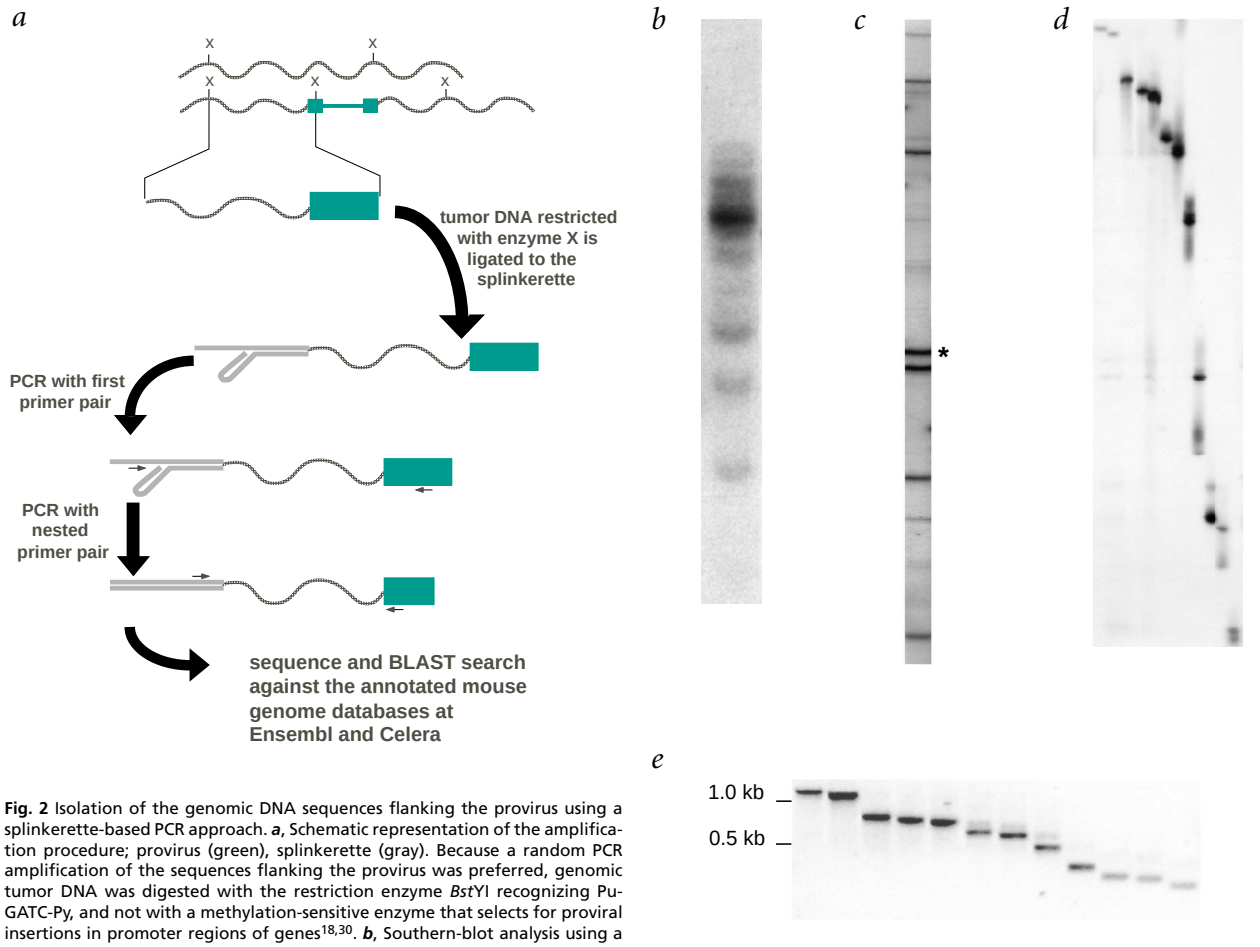

Fig. 2 Isolation of the genomic DNA sequences flanking the provirus using a splinkerette-based PCR approach. ***a***, Schematic representation of the amplification procedure; provirus (green), splinkerette (gray). Because a random PCR amplification of the sequences flanking the provirus was preferred, genomic tumor DNA was digested with the restriction enzyme *Bst*YI recognizing Pu-GATC-Py, and not with a methylation-sensitive enzyme that selects for proviral insertions in promoter regions of genes[18,30]. ***b***, Southern-blot analysis using a viral LTR probe shows the number of proviral insertions in a tumor. The lymphoma analyzed carries a large number of retroviral insertions, indicating that this tumor is of oligoclonal origin. ***c***, First radioactive splinkerette-based PCR on the same tumor. The asterisk marks the internal MoMuLV fragment amplified. ***d***, Nested radioactive splinkerette-based PCR on the excised fragments. ***e***, Final PCR amplification of the provirus flanking sequences yields ready-to-sequence DNA fragments.

genomic sequence database at Celera Genomics and Ensembl[16]) and assigning CISs to complementation groups.

The concept of insertional mutagenesis has so far been based on the assumption that the existence of CISs is due to a selective advantage associated with insertion in that site. Identifying a large number of insertion sites (as in this study) may, however, increase the likelihood of incorrectly labeling a site as a CIS. It is therefore necessary to attach parameters of significance to the definition of a CIS. We propose to assign to each CIS identified an estimate of the non-randomness of its occurrence. In a set of 500 CISs, random insertion events may account for approximately 2.5 clusters, of 2 insertions each, within 26 kb (Table 1). Random occurrence of larger CIS clusters (consisting of three or more retroviral insertions) in a data set of 500 is much less likely. It should be noted that the predicted occurrences of CISs do not objectively equal the likelihood of insertions contributing to tumorigenesis. This is because the measure of non-randomness does not take into account preferential insertion due to chromatin structure or sequence context. 'Cold' and 'hot' spots for transposon insertions have been reported for a variety of genomes.

Table 2 • Common retroviral insertion sites in Eμ*Myc* tumors

CIS name	Candidate gene	Accession ID	Candidate protein family	Mouse chr.	Human chr.	No. isolated tags	No. insertions
Dkmi1	*Dst*	ENSMUSG00000026131	actin cross-linking protein	1	6p11–p12	2	2
Dkmi2	*Ly108*	ENSMUSG00000015314	carcinoembryonic antigen	1	1	2	1
Cis1	*Ptma*	ENSMUSG00000026238	nuclear protein	1	2q35–q34	1	ND
Nki3[a]	*Zfhx1b*	ENSMUSG00000026872	zinc-finger homeobox protein	2	2q22	1	ND
Dkmi3	*Ptpn1*	ENSMUSG00000027540/	TYR phosphatase	2	20q13.1–13.2	5	ND
Dkmi4	*Set*/ ND/ 1190004A01Rik	NM 023871 ENSMUSG00000026785 ENSMUSG00000015335	nucleosome assembly protein/ protein kinase/ ND	2	9q34	3	ND
Gfi1b	*Gfi1b*	ENSMUSG00000026815	transcription factor	2	9q34.13	2	2
Notch1	*Notch1*	ENSMUSG00000026923	receptor	2	9q34.3	2	ND
Bmi1	*Bmi1*	ENSMUSG00000026739	polycomb protein	2	10p13	17	17
Evi18	*RasGrp1*	ENSMUSG00000027347	RAS exchange factor	2	15q15	8	3
Dkmi5	*ND*	ND	ND	2	20	2	1
Dkmi6	*Pkig*	ENSMUSG00000035268	protein kinase inhibitor	2	20q12–q13.1	2	ND
Dkmi7	*Mcl1*	ENSMUSG00000038612	BCL-2–related	3	1q21	3	1
Dkmi8	*Cla3*	ENSMUSG00000015749	ND	3	1q21.1	2	ND
Lef1	*Lef1*	ENSMUSG00000027985	transcription factor	3	4q23–q25	3	2[c]
Evi55[a]	*Camk2d*	ENSMUSG00000027970	SER/THR kinase	3	4	2[b]	ND
Cis2	ND	ND	ND	4	9	1	ND
Nki11[a]	*Runx3*	ENSMUSG00000028814	transcription factor	4	1p36	2[b]	ND
Evi62[a]	*E2f2*/ *Idb3*	ENSMUSG00000007872/ ENSMUSG00000018983	HLH factor/ transcription factor	4	1p36	2[b]	ND
Evi143[a]	*Ak4*/ *Lepr*	NM 009647/ ENSMUSG00000028529	adenylate kinase/ leptin receptor	4	9p24–p13	1	ND
Evi58[a]	5830400A04Rik	ENSMUSG00000029204	RAS-related	5	4p13	1	ND
Gfi1*/ *Evi5	*Gfi1*/ *Evi5*	ENSMUSG00000029275/ ENSMUSG00000011831	transcription repressor/ cell-cycle protein	5	1p22	15	15
Dkmi9	*Kdr*	ENSMUSG00000029232	TYR kinase receptor	5	4q12	3	3
Kit	*Kit*	ENSMUSG00000005672	TYR kinase receptor	5	4q12	1	3
Dkmi10	ND	ENSMUSG00000035273	heparanase	5	4q21.3	3	2[c]
Nki16	ND	ENSMUSG00000029471	SER/THR kinase	5	12q24.31	1	ND
Evi65[a]	*Coro1c*/ *Selp1*	ENSMUSG00000004530/ NM 009151	actin binding protein/ selectin	5	12q24.1	1	ND
Evi78[a]	*Calm2*/ ND	NM 007589/ ENSMUSG00000030349	calcium binding protein/ ribosomal protein	6	2p21	1	ND
Ccnd2	*Ccnd2*	ENSMUSG00000000184	cell-cycle regulator	6	12p13	3	5
Cis3	ND/*Wnt5b*	mCG49753/ ENSMUSG00000030170	F-box protein/ growth factor	6	12p13.3	1	ND
Evi167[a]	*Sema4b*	ENSMUSG00000030539	receptor	7	15q26.1	2	ND
Cis4	PD	LOC243990	ND	7	ND	1	ND
Dkmi11	PD	mCG60113	ND	7	10q25	2	2
Dkmi12	*Rras2*/ *Copb1*	ENSMUSG00000038142/ ENSMUSG00000030754	RAS-related/ beta coat protein	7	11pter–p15	3	3
Evi83[a]	*Swap70*	ENSMUSG00000031015	coiled-coil BCR binding protein	7	11p15	1	ND
Dkmi13	*Nttp1*	ENSMUSG00000037887	TYR/THR phosphatase	7	11p15.5	3	ND
Dkmi14	PD	mCG57816	*Drosophila* protein CG5765	8	13q34	2	2
Evi86[a]	*Irs2*	ENSMUSG00000038894	docking protein	8	13q34	1	ND
Evi97[a]	ORF23 like/ ND	mCG10088/mCG57228	KIAA1865/ ND	8	14q24.3	1	ND
Dkmi15	ND	ND	ND	8	16p12	3	4
Evi92[a]	*Gab1*/ *Apm1*	ENSMUSG00000003033/ ENSMUSG00000031714	clathrin coat protein/ growth factor receptor associated	8	4	1	ND
Cis5	1100001J13Rik/ *Mshra*	ENSMUSG00000001472/ ENSMUSG00000041188	ND/ receptor	8	16q24.3	1	ND
Cis6	*Lyl1*	NM008535	transcription factor	8/10	19p13.2/ 10q22	1	ND

Table 2 • (continued)

Fli1	*Fli1*	ENSMUSG00000016087	transcription factor	9	11q24.1–24.3	1	ND
Ets1	*Ets1*	ENSMUSG00000032035	transcription factor	9	11q23.3	2	ND
Dkmi16	*Madh3*	ENSMUSG00000032402	transcription factor	9	15q14–q15	4	ND
Dkmi17	*Tcf12*	ENSMUSG00000032228	transcription factor	9	15q21	2	2
Cis7	*Kif9*/ PD	ENSMUSG00000032489/ ENSMUSG00000032483	kinesin-related/ KELCH-like	9	3p21	1	ND
Evi100[a]	2700018N07	ENSMUSG00000041012	ND	9	ND	1	ND
Cis8	*Mknk2*	ENSMUSG00000020190	MAPK interacting protein	10	19p13.3	1	ND
Dkmi18[d]	PD	ENSMUSG00000020258	ND	9/11	3p21/ 5q33.1	7	ND
Myb	*Myb*	ENSMUSG00000019982	transcription factor	10	6q23.3–q24	7	ND
Dkmi19	*Hbs1l/Myb*	NM019702/ ENSMUSG00000019982	elongation factor/ transcription factor	10	6q23–q24	5	10
Nki28[a]	PD/ *Galgt1*	ENSMUSG00000040462/ ENSMUSG00000006731	ND/ transferase	10	12q13	1	ND
Tcfe2a	*Tcfe2a*	ENSMUSG00000020167	transcription factor	10	19p13.3	2	2
Evi158[a]	*Nfic*	ENSMUSG00000020237	transcription factor	10	19p13.3	1	ND
Evi106[a]	2810013G11Rik	ENSMUSG00000020280	ND	11	2p16	1	ND
Evi9	*Bcl11a*	ENSMUSG00000000861	transcription factor	11	2	4	ND
Il9r	*Il9r*	ENSMUSG00000020279	interleukin receptor	11	5q35	1[b]	ND
Evi159[a]	*Supt4h*	ENSMUSG00000020485	transcription suppressor	11	17q21–q23	1	ND
Cis9	*Grb7/ Znf1a*	ENSMUSG00000019312/ ENSMUSG00000018168	docking protein/ transcription factor	11	17q21.2	1	ND
Cis10	ND/ *Cdc6*	ENSMUSG00000038013/ ENSMUSG00000017499	WASP interacting/ cell-cycle regulator	11	17q21.3	2	ND
Cis11	*Stat5a/ Stat5b/ Stat3*	ENSMUSG00000004043/ ENSMUSG00000020919/ ENSMUSG00000004040	transcription factors	11	17q11.2	1	ND
Cis12	PD	ENSMUSG00000034168	transcription factor	12	14q24.3	1	ND
Cis13	*Irf4*	ENSMUSG00000021356	transcription factor	13	6p25–p23	1	ND
Evi112[a]	*Trim25/ Txnrd1*	AL022677/ ENSMUSG00000020250	transcription factor/ thioredoxin reductase	11/ 10	17q11.2/ 12q23.3	1	ND
Dkmi20	*Cryabp1*	ENSMUSG00000021366	transcription factor	13	6p24–p22.3	2	3
Nki33	PD	ENSMUSG00000021755	ND	13	5p13.2	1	ND
Dkmi21	*Ptp4a*	ENSMUSG00000022606	TYR phosphatase	15	8q24	2	ND
Pim3	*Pim3*	AF086624	SER/THRkinase	15	22q13.3	5	9
Evi163[a]	PD	ENSMUSG00000022462	amino acid transporter	15	12q13.11	1	ND
Dkmi22	*Kcnh3*/ PD	ENSMUSG00000037579/ ENSMUSG00000037570	potassium channel/ transcription factor	15	12q13.12	2	ND
Cis14	PD	LOC239926	ND	15	ND	1	ND
Dkmi23	PD/ *Runx1*	mCG60609/ ENSMUSG00000022952	ND/transcription factor	16	21q22	3	4
Evi13	*Runx1*	ENSMUSG00000022952	transcription factor	16	21q22.12	4	3[c]
Cis15	1810055P05Rik	ENSMUSG00000023883	transcription factor	17	6q27	1	ND
Pim1[e]	*Pim1*	ENSMUSG00000024014	SER/THR kinase	17	6p21	9	–
Evi14	*Ccnd3/ Tbn* pending	ENSMUSG00000034165/ ENSMUSG00000023980	cell-cycle regulator/ chromatin associated protein	17	6p21	6	ND
Dkmi24	PD/ PD	mCG55784/ENSMUSG00000041683	ND	17	6p21	2	1
Dkmi25	*TsgA2*	ENSMUSG00000024034	phosphatidyl inositol kinase	17	21q22.3	2	ND
Dkmi26	*Fsrg1*	ENSMUSG00000024335	bromodomain containing protein	17	4p16.3	2	ND
Tpl2	*Tpl2*	ENSMUSG00000024235	SER/THR kinase	18	10p11	7	7
Evi136[a]	*Egr1*	ENSMUSG00000038418	transcription factor	18	5q31.1	1	ND
Evi153[a]	*Hmg1*/ 1810041M12Rik	mCG9361/ ENSMUSG00000035765	HMG box protein/ ND	18	13q12	1	ND
Dkmi27	*Fbxw4*	ENSMUSG00000040913	F-box/WD40-repeat	19	10q24–q25	4	1
Evi17[a]	*Rasgrp2*	ENSMUSG00000032946	RAS exchange factor	19	11q13	1	ND
Dkmi28	*Vegfb*	ENSMUSG00000024962	growth factor	19	11q13	2	4
Dkmi29	*Cd6*	ENSMUSG00000024670	scavanger receptor	19	11q13	2	1
Pim2[e]	*Pim2*	ENSMUSG00000031155	SER/THR kinase	X	Xp11.23	2	–
Nki37[a]	*Elf4*	ENSMUSG00000031103	transcription factor	X	Xq26	1	ND
Dkmi30	1200013B08Rik	ENSMUSG00000031101	TYR kinase	X	Xq25-26.3	2	–[f]

Where gene names are specified, RNA/protein expression altered by proviruses has been demonstrated. Candidate genes are genes adjacent to the provirus. Gene accession number at Ensembl (ENSMUSG/LOC), Celera (mCG) or NCBI. The number of insertions is the number of retroviral insertions observed in 38 Eμ*Myc Pim1*$^{-/-}$ *Pim2*$^{-/-}$ tumors as determined by Southern-blot analysis. 'ND' indicates that the number of insertions was not determined for the CIS. CISs in bold have been described previously or the affected genes have been identified by altered mRNA or protein expression. Underlined gene or chromosome names are those used by Celera. 'PD' indicates a predicted gene according to the Ensembl analysis pipeline, and '*PD*' indicates a predicted gene according to the Celera Discovery System. '*Dkmi*' indicates a double knockout *Myc* insertion. [a]RISs overlapping with CISs identified by Suzuki *et al.* (*Evi*) or Lund *et al.* (*Nki*) based on 3 or more insertions within 100 kb. [b]Two independent retroviral insertions (distance > 26 kb). [c]CIS consists of two clusters, of which only one has been checked by Southern-blot analysis. [d]Gene-rich region of 50 kb harboring, according to Celera, four genes encoding the candidate proteins RAN-related (mCG50456), PP2C-like phosphatase (mCG19525), ACTIN depolymerization factor (*Ptk9l*; mCG19506), WD-repeat-containing protein (mCG19514). [e]*Pim1* and *Pim2* insertions were isolated only from Eμ*Myc* lymphomas. [f]No rearrangments were observed, indicating the subclonal nature of the retroviral insertions.

Table 3 • Common retroviral insertion sites substituting for *Pim1* and *Pim2* in lymphomagenesis

CIS name	Gene	Protein family	No. insertions[a]	Insertion type	*P* value[b]
Pim3	*Pim3*	SER/THR kinase	9	5′ promoter, 5′ or 3′ enhancer	0.000
Kit	*Kit*	TYR kinase receptor	3	5′ enhancer	0.025
Tpl2	*Tpl2*	SER/THR kinase	7	activating truncation	0.000
Ccnd2	*Ccnd2*	cell-cycle regulator	5	5′ promoter and 5′ enhancer	0.002
Dkmi1	ND	*actin cross-linking*	2		0.088
Dkmi9	ND	*TYR kinase receptor*	3		0.025
Dkmi11	ND	ND	2		0.088
Dkmi15[c]	ND	ND	4		0.007
Dkmi20	ND	*transcription factor*	3		0.025
Dkmi28	ND	*growth factor*	4		0.007

[a]Number of proviral insertions in 38 Eμ*Myc Pim1*$^{-/-}$ *Pim2*$^{-/-}$ tumors detected by Southern-blot analysis. [b]*P* value of 38 Eμ*Myc Pim1*$^{-/-}$ *Pim2*$^{-/-}$ tumors compared with 89 control tumors using Fisher's exact test. [c]CIS consists of two loci 40 kb or 125 kb apart according to Celera and Ensembl, respectively. Protein descriptions in italics represent the proteins encoded by candidate genes 'ND' means that the genes affected by the retroviral insertions have not been determined.

Confirmation of the contribution of CISs to tumorigenesis relies on the identification of the affected gene and evidence that aberrant expression of that gene reproduces specific aspects of the tumor phenotype.

To identify the genomic sequences flanking most proviruses, we designed an efficient, PCR-based splinkerette amplification procedure (Fig. 2). A splinkerette is an adaptor molecule containing a hairpin loop that prevents nonspecific PCR amplification[17]. The applied splinkerette amplification is preferred over the previously described inverse PCR (IPCR) method[18] for two reasons. First, this technique is not based on a long-range PCR amplification, which may increase the size of the amplified fragments and limit the recovery of provirus flanking sequences. Because the complete annotated mouse genomic sequence is available at Ensembl/Celera, the size of the amplified sequences flanking proviruses no longer has an advantage for identifying CISs. Second, this method does not require cloning in bacterial hosts, which increases the speed of the isolation of proviral-flanking sequences. We used this splinkerette procedure to analyze the sequences of 477 RISs from 38 lymphomas from Eμ*MycPim1*$^{-/-}$*Pim2*$^{-/-}$ mice and 27 lymphomas from control Eμ*Myc* mice. This group represents approximately 60% of all retroviral insertions present in the tumors, a substantial fraction of which were of oligoclonal origin. The sequence-analyzed fraction of RISs corresponded to an average of approximately seven insertions per tumor (Table 4). We then compared the 477 RIS sequences against the Celera annotated mouse genomic database and found that 176 of the RISs represented 52 CISs (Table 2; for a complete overview, see Web Table A online). Comparison of the RISs with previously identified CISs, and the RISs and CISs characterized by Suzuki *et al.* (ref. 19; this issue) and Lund *et al.* (ref. 25; this issue), yielded a total of 91 independent CISs in this tumor panel (Table 2 and Table 4).

In Eμ*Myc* mice lacking expression of *Pim1*, the pressure to activate the Pim pathway by means of proviral insertions in *Pim2* is very high. In Eμ*Myc* mice nullizygous with respect to both *Pim1* and *Pim2*, the selective advantage conferred by retroviral activation of the Pim pathway is likely to remain unchanged. Therefore, genes that can substitute for *Pim1* and *Pim2* in lymphomagenesis can be expected to fall into one of the following categories (Fig. 1*c*): genes encoding proteins that directly substitute for the function of Pim1 or Pim2; genes that encode targets or other downstream components of the *Pim1*/*Pim2* signaling pathway; or genes whose proteins function in pathways parallel to *Pim1* or *Pim2* that independently activate a similar crucial oncogene target. One common retroviral insertion that was identified in five independent lymphomas from Eμ*Myc Pim1*$^{-/-}$ *Pim2*$^{-/-}$ mice affected *Pim3*, the third member of the *Pim* family and thus a prime candidate for the first category of genes that might substitute for *Pim1* or *Pim2*. At the amino-acid level, Pim3 is 71% and 61% identical to Pim1 and Pim2, respectively. Knockout experiments have shown a high degree of redundancy between *Pim1* and *Pim3*, suggesting a similar function for the encoded proteins (H.M., unpublished data). Southern-blot analysis showed insertions near *Pim3* in 9 of 38 lymphomas from Eμ*Myc Pim1*$^{-/-}$ *Pim2*$^{-/-}$ mice. The discovery that *Pim3* is preferentially activated in tumors lacking expression of *Pim1* and *Pim2* underscores the pathway-specificity of this screen.

To assign a gene to the *Pim* complementation group (which consists of genes encoding proteins that can either fully or partially substitute for Pim in lymphomagenesis), retroviral insertions near the corresponding genes should be preferentially absent in tumors of mice that express *Pim1* and/or *Pim2* or, if present, should be mutually exclusive with insertions near one of the *Pim* genes. To test the validity of this hypothesis, we carried out Southern-blot analysis for insertions near *Pim3* in 89 lymphomas from Eμ*Myc*, Eμ*Myc Pim1*$^{-/-}$ and Eμ*Myc Pim2*$^{-/-}$ control mice, of which 61% showed retroviral activation of either *Pim1* or *Pim2*. Retroviral insertions near *Pim3* were observed in only one tumor, and this tumor did not carry insertions near *Pim1* or *Pim2*. We subsequently analyzed the whole tumor panel of 89 lymphomas from Eμ*Myc* control and 38 Eμ*Myc Pim1*$^{-/-}$ *Pim2*$^{-/-}$ mice by Southern blotting, using CISs depicted in Table 2 as probes. Nine CISs, identified as *Kit*, *Ccnd2, Tpl2*, *Dkmi1*, *Dkmi9*, *Dkmi11*, *Dkmi15*, *Dkmi20* and *Dkmi28*, were found to be mutually exclusive with *Pim1*, *Pim2* and *Pim3* (Table 3). Within this group of

Table 4 • Cancer loci are efficiently identified by retroviral tagging

No. tumors	No. tags	No. tags per tumor	No. tags CISs[a]	% tags CISs	CISs per tumor
65	477	7.40	230	48	3.53
	known CISs		86	18	1.32
	new CISs		90	19	1.38
	RIS/CISs[b]		39	8	0.6
	new CISs (RIS/RIS)[c]		15	3	0.23
	PIM-substituting CISs		25[d]		0.68[e]

[a]Number of proviral tags identifying a CIS. [b]Single RISs from this study belong to CISs identified by Suzuki *et al.*[19] or Lund *et al.*[20]. [c]Comparison of the RISs from this study with the RISs isolated by Suzuki *et al.*[19] and Lund *et al.*[20] revealed additional CISs. [d]Southern-blot analysis showed 34 RISs substituting for Pim. [e]Calculations are based on 38 Eμ*Myc Pim1*$^{-/-}$*Pim2*$^{-/-}$ tumors.

Pim-complementing loci, four of the affected genes (*Pim3*, *Kit*, *Ccnd2* and *Tpl2*) were identified by altered expression (data not shown). The observation that these loci belong to the Pim complementation group suggests that the proteins encoded by the affected genes act either downstream of or parallel to Pim.

Despite the unknown position of the proteins encoded by the *Pim*-complementing genes relative to Pim signaling, the varying nature of these proteins argues that Pim proteins, like members of the Myc family, have a central role in a complex signaling network. In a pathway model that fits this hypothesis, Pim acts as a modulator of cross-talk between stem-cell factor–induced Kit signaling and cytokine signaling pathways (see Web Fig. A online). To induce a maximum proliferative effect, cytokines require the synergistic action of stem-cell factor[21–23]. Genes induced by interleukins but not stem-cell factor, such as *Pim*, are prime candidates for involvement in the cross-talk mediating the synergistic proliferative effect[8,24]. The modulating role for Pim is supported by the observations that enforced *Pim1* expression reconstitutes the number of lymphocytes in *Rag*-deficient and common-γ–deficient mice[25], and that Pim1 is recruited to the receptor complexes where it associates with the suppressor of cytokine signaling[13].

The use of genetically modified mice in combination with high-throughput analyses of retroviral insertions and the availability of the complete mouse genomic sequence have permitted us to focus on specific oncogenic signaling pathways by means of an *in vivo* mammalian genetic screen. The strategy described here can indicate whether the protein encoded by a candidate gene belongs to a particular signaling network, and permits a more focused approach in subsequent biochemical analyses. Thus, it represents a mammalian equivalent of the powerful *D. melanogaster* and *C. elegans* genetic screens. In addition, the methodology we used allows combination of data from independent panels, as illustrated by the additional CISs that were identified upon comparison of the RISs from different panels.

Methods

Mice and MoMuLV infection. The Eμ*Myc* mice[26] were bred with *Pim1*$^{-/-}$ *Pim 1neo59* mice[24] and *Pim2*$^{-/-}$ *Pim2 K180* mice (J.A., unpublished data) to generate Eμ*Myc Pim1*$^{-/-}$, Eμ*Myc Pim2*$^{-/-}$ and Eμ*Myc Pim1*$^{-/-}$ *Pim2*$^{-/-}$ mice. We infected newborns with 1×10^5 infectious units of MoMuLV. We killed moribund mice and isolated lymphomas. All animal experiments were approved by the Dutch Animal Research Committee.

Statistical analysis. Because exact calculations and simulations show that end-of-chromosome effects can be ignored for a set of several hundred proviral insertions, random insertions in the genome follow a Poisson distribution in which the distance between two adjacent insertions is exponentially distributed. This means that the probability that the distance is at most x equals $1 - e^{-x/\mu}$, where μ equals the mean distance $G/(b + 1)$, where G is the sequenced genome size (2.6×10^9) and b is the total number of insertions. For small values of x/μ (<0.05), the probability can be approximated by x/μ.

The number of insertions to the right of a selected insertion in a fixed window W also follows a Poisson distribution. This means that the probability of at least m such extra insertions equals $1 - \exp(-\lambda)(1 + \lambda + \lambda^2/2 + \ldots + \lambda^{m-1}/(m-1)!)$, where λ equals the mean number of insertions in window W: $W \times b/G$ and $(m-1)! = 1 \times 2 \times 3 \times \ldots \times (m-1)$. For $m = 1$ (a cluster of two insertions), this equals $1 - \exp(-\lambda)$, or approximately λ. For $m = 2$ (a cluster of three insertions), this equals $1 - \exp(-\lambda)(1 + \lambda)$, or approximately $1 - (1 - \lambda + \lambda^2/2) \times (1 + \lambda) = \lambda^2/2$. For a cluster of 4 insertions, this equals $1 - \exp(-\lambda)(1 + \lambda + \lambda^2/2)$, or approximately $1 - (1 - \lambda + \lambda^2/2 - \lambda^3/6 + \lambda^4/24) \times (1 + \lambda + \lambda^2/2) = \lambda^3/6 - \lambda^4/8$. These approximations only hold for a small mean number of insertion clusters in the window (<0.05).

Southern-blot analysis of CISs. Genomic tumor DNA (10 μg) was digested with the appropriate restriction enzyme, separated on a 0.7% agarose gel and transferred to Hybond-N membranes (Amersham). We analyzed the number of proviral insertions and the number of insertions into the known CISs *Pim1*, *Pim2*, *Bmi1* and *Gfi1* using the probes and restriction enzymes as described previously[5,15,28,29]. Genomic fragments, free of repetitive sequences, that flanked the proviruses and hybridized to a CIS were used as probes to analyze the frequency at which a provirus inserted into these loci.

Isolation of the proviral insertion sites. Tumor DNA (3 μg) was digested with *Bst*YI (New England Biolabs) and the enzyme was subsequently inactivated. We generated the splinkerette adaptor by annealing the splinkerette oligonucleotides HMSpAA and HMSpBB (primer sequences are available on request). Both oligonucleotides contain modifications of a splinkerette described previously[17]. The oligonucleotides (150 pmol each) were denatured at 95 °C for 3 min and subsequently cooled to room temperature at a rate of 1 °C per 15 s using a thermocycler (PTC100, Perkin Elmer). We ligated 600 ng of genomic tumor DNA digested with *Bst*YI to the splinkerette oligonucleotide (molar ratio 1:10) with 4 U T4 DNA ligase (Roche Diagnostics) in a final volume of 40 μl. To avoid amplification of the internal 3′ MoMuLV fragment, we digested the ligated fragments with 10 U of *Eco*RV in a total volume of 100 μl. Ligation mixtures were desalted in a Microcon YM-30 (Amicon BioSeparations).

We amplified MoMuLV-flanking sequences with a radioactive long terminal repeat (LTR)–specific primer, AB949, and a splinkerette primer, HMSp1 (primer sequences are available upon request). Primer AB949 (10 pmol) was radioactively labeled with [γ-^{32}P]ATP (3 μCu) using T4 polynucleotide kinase (PNK) (0.2 U; Roche Diagnostics).

The 50 μl PCR mixture contained 150 ng ligated tumor DNA, 10 pmol each primer, 300 nmol dNTPs, 1 U PfuITurbo and 1 × PfuITurbo buffer (Stratagene). The hot-start PCR conditions were 3 min at 94 °C (2 cycles); 15 s at 94 °C, 30 s at 68 °C, 3.5 min at 72 °C (27 cycles); 15 s at 94 °C, 30 s at 66 °C, 3.5 min at 72 °C, and 5 min at 72 °C. We concentrated radioactive PCR fragments using a Microcon-YM30 (Amicon BioSeparations) and then separated them on a 3.5% denaturing polyacrylamide gel. The gels were dried onto 3-mm Wattman paper and exposed overnight to X-Omat AR films (Kodak). We excised amplified fragments from the gel and boiled them for 30 min in 100 μl TE. We used 1 μl of the DNA solution for a nested amplification with a ^{32}P-labeled virus-specific primer, HM001, and a non-radioactive splinkerette-specific primer, HMSp2 (primer sequences are available upon request). We carried out nested PCR with 5 pmol of each primer, 200 nM of each dNTP, 1.75 mM Mg, 1 U *Taq* polymerase (Gibco BRL) and 1 × PCR buffer (Gibco BRL) in a final volume of 20 μl. The PCR conditions were 15 s at 94 °C, 30 s at 60 °C, 3 min at 72 °C for either 25 (for fragments < 400 bp) or 28 cycles (for fragments > 400 bp). We separated the re-amplified fragments on a 3.5% denaturing polyacrylamide gel and isolated them as described above. We then re-amplified 1 μl of the amplified fragments in a non-radioactive PCR of 25 cycles under the conditions as described for the radioactive nested PCR.

We treated the nested PCR mixture with 0.5 U exonuclease and 0.5 U shrimp alkaline phoshatase, according to the manufacturer's (Amersham) instructions. We used about 25 ng of the PCR product in the sequence reaction containing BigDye terminator mix (Perkin Elmer) and primer HM001. In addition, we used HMSp2 as primer for sequencing of amplified fragments larger than 500 bp. We carried out automated sequence analysis on an ABI 377 (Perkin Elmer). The sequences were processed with Sequencher 3.1.1 and subjected to BLAST analysis against the annotated mouse genome databases at Celera (release 1.2) and Ensembl (version 6.3a.1).

GenBank accession numbers. The accession numbers for the 477 flanking sequences of the retroviral insertions in the Eμ*Myc* tumors (*Dkm*) are AY127080 through AY127557). Further information is available at http://protagdb.nki.nl.

Note: Supplementary information is available on the Nature Genetics website.

Acknowledgments

We wish to thank R. Regnerus for assistance in genotyping the mice; N. Bosnie, L. Rijswijk, A. Zwerver, T. Maidment, C. Spaans and F. van der Ahé for animal care; and J. Jonkers and R. van Amerongen for critical reading of the manuscript. This work was supported by the Dutch Cancer Society (H.M.) and the Leukemia Society of America (J.A.).

Competing interests statement

The authors declare that they have no competing financial interests.

Received 23 April; accepted 8 July 2002.

1. Nasmyth, K. At the heart of the budding yeast cell cycle. *Trends Genet.* **12**, 405–412 (1996).
2. Wassarman, D.A., Therrien, M. & Rubin, G.M. The Ras signaling pathway in *Drosophila*. *Curr. Opin. Genet. Dev.* **5**, 44–50 (1995).
3. Sternberg, P.W. & Han, M. Genetics of RAS signaling in *C. elegans*. *Trends Genet.* **14**, 466–472 (1998).
4. Jonkers, J. & Berns, A. Retroviral insertional mutagenesis as a strategy to identify cancer genes. *Biochim. Biophys. Acta* **1287**, 29–57 (1996).
5. van Lohuizen, M. *et al*. Predisposition to lymphomagenesis in *Pim1* transgenic mice: cooperation with c-*Myc* and N-*Myc* in murine leukemia virus-induced tumors. *Cell* **56**, 673–682 (1989).
6. Selten, G., Cuypers, H.T., Zijlstra, M., Melief, C. & Berns, A. Involvement of c-*Myc* in MuLV-induced T cell lymphomas in mice: frequency and mechanisms of activation. *EMBO J.* **3**, 3215–3222 (1984).
7. Verbeek, S. *et al*. Mice bearing the *EμMyc* and E*μPim1* transgenes develop pre-B-cell leukemia prenatally. *Mol. Cell. Biol.* **11**, 1176–1179 (1991).
8. Allen, J.D., Verhoeven, E., Domen, J., van der Valk, M. & Berns, A. *Pim2* transgene induces lymphoid tumors, exhibiting potent synergy with c-*Myc*. *Oncogene* **15**, 1133–1141 (1997).
9. Leverson, J.D. *et al*. PIM-1 kinase and P100 cooperate to enhance c-*Myb* activity. *Mol. Cell* **2**, 417–425 (1998).
10. Mochizuki, T. *et al*. Physical and functional interactions between PIM-1 kinase and CDC25A phosphatase. Implications for the PIM-1-mediated activation of the c-MYC signaling pathway. *J. Biol. Chem.* **274**, 18659–18666 (1999).
11. Koike, N., Maita, H., Taira, T., Ariga, H. & Iguchi-Ariga, S.M. Identification of heterochromatin protein 1 (HP1) as a phosphorylation target by PIM-1 kinase and the effect of phosphorylation on the transcriptional repression function of HP1(γ). *FEBS Lett.* **467**, 17–21 (2000).
12. Ishibashi, Y. *et al*. PIM-1 translocates sorting NEXIN6/TRAF4-associated factor 2 from cytoplasm to nucleus. *FEBS Lett.* **506**, 33–38 (2001).
13 Rainio, E.M., Sandholm, J. & Koskinen P.J. Cutting edge: transcriptional activity of NFATc1 is enhanced by the PIM-1 kinase. *J. Immunol.* **168**, 1524–1527 (2002).
14. van der Lugt, N.M. *et al*. Proviral tagging in *EμMyc* transgenic mice lacking the *Pim1* proto-oncogene leads to compensatory activation of *Pim2*. *EMBO J.* **14**, 2536–2544 (1995).
15. Allen, J.D. & Berns, A. Complementation tagging of cooperating oncogenes in knockout mice. *Semin. Cancer Biol.* **7**, 299–306 (1996).
16. Craig, N.L. Target site selection in transposition. *Annu. Rev. Biochem.* **66**, 437–474 (1997).
17. Devon, R.S., Porteous, D.J. & Brookes, A.J. Splinkerettes—improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res.* **23**, 1644–1645 (1995).
18. Li, J. *et al*. Leukaemia disease genes: large-scale cloning and pathway predictions. *Nature Genet.* **23**, 348–353 (1999).
19. Suzuki, T. *et al*. Retroviral tagging in the post-genome era identifies new genes involved in cancer. *Nature Genet.* **32**, 86–94 (2002).
20. Lund, A.H. *et al*. Genome-wide retroviral insertional tagging of cancer genes in Cdkn2a-deficient mice. *Nature Genet.* **32**, 80–85 (2002).
21. Miyazawa, K. *et al*. Recombinant human interleukin-9 induces protein tyrosine phosphorylation and synergizes with steel factor to stimulate proliferation of the human factor-dependent cell line, M07e. *Blood* **80**, 1685–1692 (1992).
22. Tsuji, K., Lyman, S.D., Sudo, T., Clark, S.C. & Ogawa, M. Enhancement of murine hematopoiesis by synergistic interactions between steel factor (ligand for c-kit), interleukin-11, and other early acting factors in culture. *Blood* **79**, 2855–2860 (1992).
23. Laird, P.W. *et al*. *In vivo* analysis of Pim-1 deficiency. *Nucleic Acids Res*. **21**, 4750–4755 (1993).
24. Jacobs, H. *et al*. PIM1 reconstitutes thymus cellularity in interleukin 7- and common γ chain-mutant mice and permits thymocyte maturation in Rag- but not CD3γ-deficient mice. *J. Exp. Med.* **190**, 1059–1068 (1999).
25. Losman, J. *et al*. IL-4 signaling is regulated through the recruitment of phosphatases, kinases, and SOCS proteins to the receptor complex. *Cold Spring Harb. Symp. Quant. Biol.* **64**, 405–416 (1999).
26. van Lohuizen, M. *et al*. Identification of cooperating oncogenes in *EμMyc* transgenic mice by provirus tagging. *Cell* **65**, 737–752 (1991).
27. te Riele, H., Maandag, E.R., Clarke, A., Hooper, M. & Berns, A. Consecutive inactivation of both alleles of the *Pim1* proto-oncogene by homologous recombination in embryonic stem cells. *Nature* **348**, 649–651 (1990).
28. Cuypers, H.T. *et al*. Murine leukemia virus-induced T-cell lymphomagenesis: integration of proviruses in a distinct chromosomal region. *Cell* **37**, 141–150 (1984).
29. Scheijen, B., Jonkers, J., Acton, D. & Berns, A. Characterization of *Pal1*, a common proviral insertion site in murine leukemia virus-induced lymphomas of c-*Myc* and *Pim1* transgenic mice. *J. Virol.* **71**, 9–16 (1997).
30. Nakamura, T., Largaespada, D.A., Shaughnessy, J.D. Jr, Jenkins, N.A. & Copeland, N.G. Cooperative activation of Hoxa and Pbx1-related genes in murine myeloid leukaemias. *Nature Genet.* **12**, 149–153 (1996).

High-throughput retroviral tagging to identify components of specific signaling pathways in cancer

H Mikkers, J Allen, P Knipscheer, L Romeyn, A Hart, E Vink & A Berns

Nature Genet. **32**, 153–159 (2002).

By error, several corrections were not made to proofs while preparing the manuscript for the press.
In the reference list, reference 25 (Losman *et al.*) should be inserted as reference 13. As a consequence, references 13–24 should be renumbered as 14–25.
In the text, the following changes should be made:
On page 154, in the second column, reference 16 should be placed at the end of the sentence "These observations underscore the selective advantage…" Reference 16 should also be removed from the first line on page 155.
On page 155, in the second column, reference 17 should be placed at the end of the sentence "'Cold' and 'hot' spots for transposon insertions…"
On page 157, in the first full paragraph, reference 17 should be reference 18, reference 18 should be reference 19, reference 19 should be reference 20, and reference 25 should be reference 21.
On page 158, in the first full paragraph, references 21–23 should be references 22,23.

Genetics, cytokines and human infectious disease: lessons from weakly pathogenic mycobacteria and salmonellae

T H M Ottenhoff, F A W Verreck, E G R Lichtenauer-Kaligis, M A Hoeve, O Sanal & J T van Dissel

Nature Genet. **32**, 97–105 (2002).

On page 97, paragraph 2, line 13, 'IL-29' should be 'IL-27'.
On page 98, line 11, 'seemed to be' should be 'was', as this has been demonstrated to be the case.
On page 98, Fig. 1, a mistake in the color coding occurred, and the affected genes that appear in purple should be colored red.
On page 100, line 1, 'IL12Rb1' should read 'IL12Rβ1'.

Distal ureter morphogenesis depends on epithelial cell remodeling mediated by vitamin A and Ret

E Batourina, C Choi, N Paragas, N Bello, T Hensle, F D Constantini, A Schuchardt, R L Bacallao & C L Mendelsohn

Nature Genet. **32**, 109–115 (2002).

doi:10.1038/ng952

The article was missing a reference to Web Movie A, which should have appeared on the last line of page 110 together with the reference to Web Fig. A.

Targeted mutation of *Cyln2* in the Williams syndrome critical region links CLIP-115 haploinsufficiency to neurodevelopmental abnormalities in mice

C C Hoogenraad, B Koekkoek, A Akhmanova, H Krugers, B Dortland, M Miedema, A van Alphen, W M Kistler, M Jaegle, M Koutsourakis, N Van Camp, M Verhoye, A van der Linden, I Kaverina, F Grosveld, C I De Zeeuw & N Galjart

Nature Genet. **32**, 116–127 (2002).

doi:10.1038/ng954

The article mistakenly contained a note stating that supplementary information was available on the *Nature Genetics* website. Instead, it should have contained a brief paragraph in the Methods section listing a separate website where additional information can be found.

High-throughput retroviral tagging to identify components of specific signaling pathways in cancer

H Mikkers, J Allen, P Knipscheer, L Romeyn, A Hart, E Vink & A Berns

Nature Genet. **32**, 153–159 (2002).

doi:10.1038/ng950

By error, several corrections were not made to proofs while preparing the manuscript for the press.

In the reference list, reference 25 (Losman *et al.*) should be inserted as reference 13. As a consequence, references 13–24 should be renumbered as 14–25.
In the text, the following changes should be made:
On page 154, in the second column, reference 16 should be placed at the end of the sentence "These observations underscore the selective advantage…". Reference 16 should also be removed from the first line on page 155.
On page 155, in the second column, reference 17 should be placed at the end of the sentence "'Cold' and 'hot' spots for transposon insertions…".
On page 157, in the first full paragraph, reference 17 should be reference 18, reference 18 should be reference 19, reference 19 should be reference 20, and reference 25 should be reference 21.
On page 158, in the first full paragraph, references 21–23 should be references 22,23.

New genes involved in cancer identified by retroviral tagging

T Suzuki, H Shen, K Akagi, H C Morse III, J D Malley, D Q Naiman, N A Jenkins & N G Copeland

Nature Genet. **32**, 166–174 (2002).

doi:10.1038/ng949

By error, the subpanel labels for Fig. 1 on page 172 were offset to the right of the corresponding subpanels.