

Novel Candidate Cancer Genes Identified by a Large-Scale Cross-Species Comparative Oncogenomics Approach

Jenny Mattison¹, Jaap Kool², Anthony G. Uren², Jeroen de Ridder^{2,3}, Lodewyk Wessels², Jos Jonkers², Graham R. Bignell¹, Adam Butler¹, Alistair G. Rust¹, Markus Brosch¹, Catherine H. Wilson¹, Louise van der Weyden¹, David A. Largaespada⁴, Michael R. Stratton¹, P. Andy Futreal¹, Maarten van Lohuizen², Anton Berns², Lara S. Collier⁵, Tim Hubbard¹, and David J. Adams¹

Abstract

Comparative genomic hybridization (CGH) can reveal important disease genes but the large regions identified could sometimes contain hundreds of genes. Here we combine high-resolution CGH analysis of 598 human cancer cell lines with insertion sites isolated from 1,005 mouse tumors induced with the murine leukemia virus (MuLV). This cross-species oncogenomic analysis revealed candidate tumor suppressor genes and oncogenes mutated in both human and mouse tumors, making them strong candidates for novel cancer genes. A significant number of these genes contained binding sites for the stem cell transcription factors Oct4 and Nanog. Notably, mice carrying tumors with insertions in or near stem cell module genes, which are thought to participate in cell self-renewal, died significantly faster than mice without these insertions. A comparison of the profile we identified to that induced with the *Sleeping Beauty* (SB) transposon system revealed significant differences in the profile of recurrently mutated genes. Collectively, this work provides a rich catalogue of new candidate cancer genes for functional analysis. *Cancer Res*; 70(3); 883–95. ©2010 AACR.

Introduction

Tumors form in humans when a cell gains a selective advantage over other cells and manages to evade the checkpoints that would normally suppress its growth or result in apoptosis. The acquisition of this behavior is thought to occur as a result of the development of somatic mutations that deregulate gene function (1). These somatic mutations sometimes interact with predisposing germline mutations to promote tumor formation, and it is the profile of somatic and germ line mutations found in a tumor that ultimately dictate its presentation and clinical course (2). Somatic mutations in human tumors may result from a multitude of genetic insults

generating different types of lesions in the genome (3). With the exception of point mutations, these lesions are rarely focal and often encompass many genes. Profiling allelic imbalances found in human tumors is a powerful tool for identifying cancer gene-containing loci, the most commonly used approach being array comparative genomic hybridization (CGH; ref. 4). Although the resolution of this technique has improved dramatically, copy number gains and losses in human tumors are usually large, and rearrangements often encompass many genes that do not contribute to tumorigenesis. Therefore, differentiating “driver” cancer genes from “passenger” genes requires validation in other systems.

Tumors in mice can be generated using insertional mutagens such as viruses (5, 6) and transposons (7, 8) and because these elements deregulate gene function either by integrating in or near a cancer gene, they “tag” cancer loci, facilitating their identification. Viruses such as murine leukemia virus (MuLV) and the mouse mammary tumor virus have been used extensively for cancer gene identification. Screens using these viruses have been proven to identify relevant cancer genes because the genes *Myb*, *Pim1*, and *Bmi1* were identified using these mutagens (5, 9), and were subsequently shown to be genes relevant to cancer formation in humans (9). Similarly, transposons such as *Sleeping Beauty* (SB) have been shown to be potent insertional mutagens in mice (7, 8, 10). Importantly, both viruses and transposons are particularly powerful tools for identifying cooperating mutations between genes, as was shown previously for *Myc* and *Bim1* (11), and more recently, for *p19* and *Braf* (7), and for *Notch1*, *Rasgrp1*, and *Sox8* (8).

Cross-species cancer gene analysis, which integrates genome-wide cancer data sets from human and other species,

Authors' Affiliations: ¹The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; ²The Cancer Genomics Centre, Centre of Biomedical Genetics and The Academic Medical Centre, the Netherlands Cancer Institute, Plesmanlaan, Amsterdam, the Netherlands; ³Delft University of Technology, Delft, the Netherlands; ⁴Department of Genetics, Cell Biology and Development, Center for Genome Engineering, Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota; and ⁵School of Pharmacy, University of Wisconsin-Madison, Madison, Wisconsin

Note: Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

J. Kool, A.G. Uren, and J. de Ridder contributed equally to this work.

M. van Lohuizen, A. Berns, L.S. Collier, T. Hubbard, and D.J. Adams are co-senior authors.

Corresponding Author: David J. Adams, Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, United Kingdom. Phone: 44-1223-834-244, ext. 8684; Fax: 44-1223-494919; E-mail: da1@sanger.ac.uk.

doi: 10.1158/0008-5472.CAN-09-1737

©2010 American Association for Cancer Research.

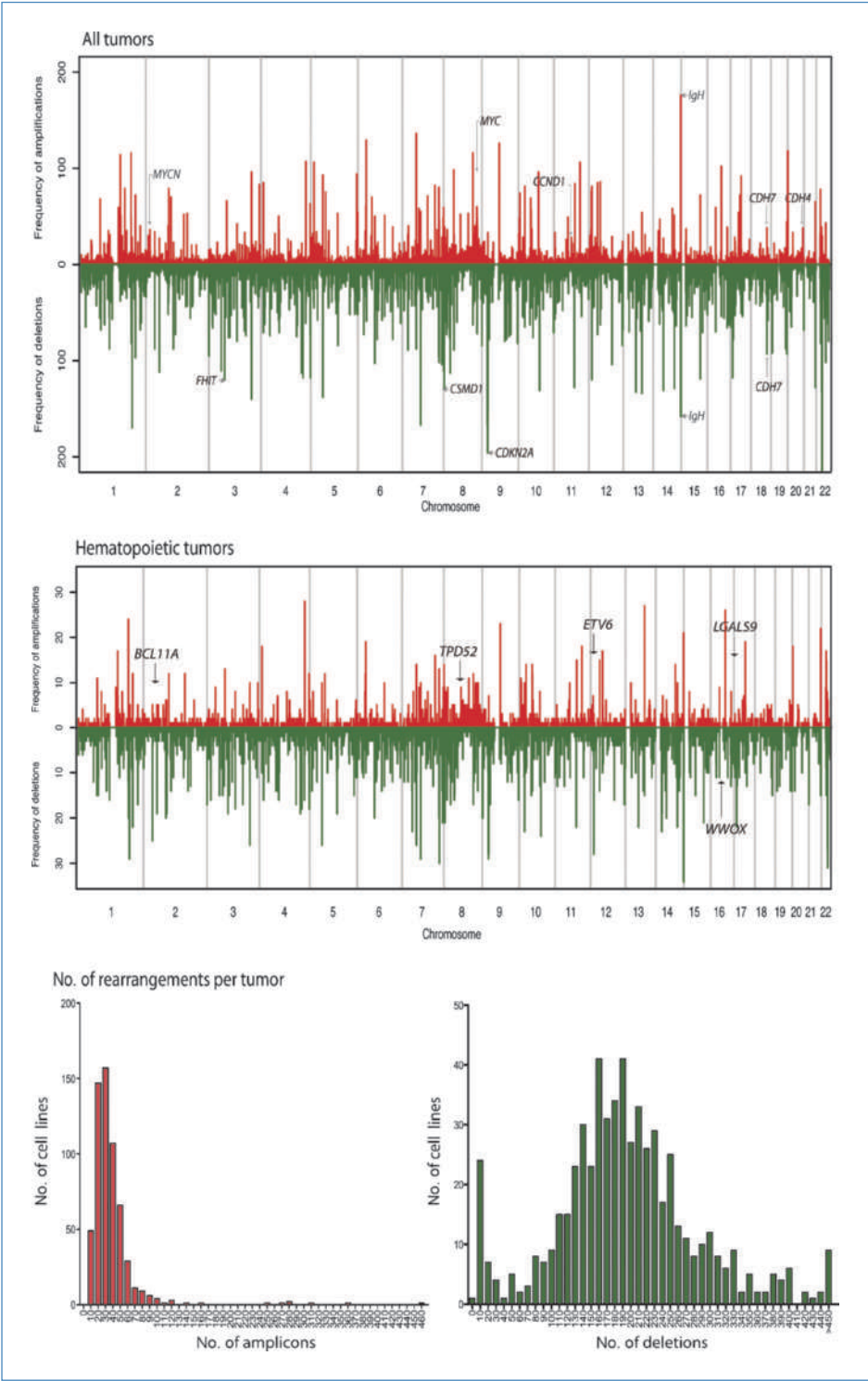


Figure 1. Global overview of genome-wide high-resolution CGH of 598 cancer cell lines. Cell lines from 29 different tissues were subjected to CGH on the Affy SNP6 platform. Significant copy number alterations were defined as those with a copy number gain of >1.7 or loss of <0.3 . The frequency of significant gains (red) and losses (green) across the autosomes in 1 Mb bins is shown. The same analysis is also done for the hematopoietic subset of the tumors. The frequency of amplifications and deletions per cell line for the entire panel of tumors is also shown.

represents a potentially powerful approach for identifying and validating genes involved in tumorigenesis. This approach has been used successfully in several instances, most recently, to identify the cancer genes *NEDD9* (12) in melanoma, and *YAPI* (13) and *DLC1* (14) in liver cancer. Here,

we present a high-resolution comparative oncogenomic analysis performed using CGH data from 598 human cancer cell lines and $>1,000$ murine lymphomas. Using insertional mutagenesis data sets generated with both MuLV and the SB transposon system, we identify candidate cancer genes mutated in

both mouse and human tumors and predict that some common insertion site (CIS) genes may play a role in driving a program of tumor self-renewal. This work significantly extends our previous study (6) in which we performed cross-species analysis on low-resolution CGH data against <500 MuLV-induced tumors, and provides a comprehensive genome-wide profile of candidate cancer genes at high resolution.

Materials and Methods

CGH. Five hundred and ninety-eight human cancer cell lines derived from 29 different tissues (Supplementary

Table S1) were analyzed using the Affymetrix Genome-Wide Human SNP 6.0 array. Several analytic approaches were tested with DNA copy and merge levels showing optimal results and use of compute resources (Supplementary Table S2). Analysis was performed as described in the Supplementary Methods. The CGH data is available for download in MIAME format (15).

Insertional mutagenesis and mouse tumor panels.

MuLV was used to induce tumors on a pure FVB background as described previously (6). SB tumors were induced on an F1 C57BL/6J-FVB background by breeding together an allele of the SB transposase knocked into the *Rosa26* locus (8) and a low copy transposon line, LC76 (chromosome 1) or LC68

Figure 2. MuLV and SB insertions across the mouse genome. A, plots show the density of insertions calculated using the kernel convolution method (18). X-axis, chromosomes (CISs are shown in green below each plot); red line, the threshold above which insertions cluster into significant CIS ($P < 0.05$). Gene names in red, genes that contain significant CISs in both screens; gene names in black, genes that contain significant CISs that are unique to one screen. *, artifactual CIS in *En2*. Inset, insertions in *Pten* which are exclusively found in SB-derived tumors. Independent insertions from the same tumor were given the same color. B, Western blot for Myc protein levels in SB-derived thymic lymphomas. Control thymus (WT) represents protein from a normal thymus from an age-matched mouse. As a positive control, a thymic lymphomas induced with MuLV, and known to carry *Myc* insertions, were also analyzed by Western blotting (MuLV).

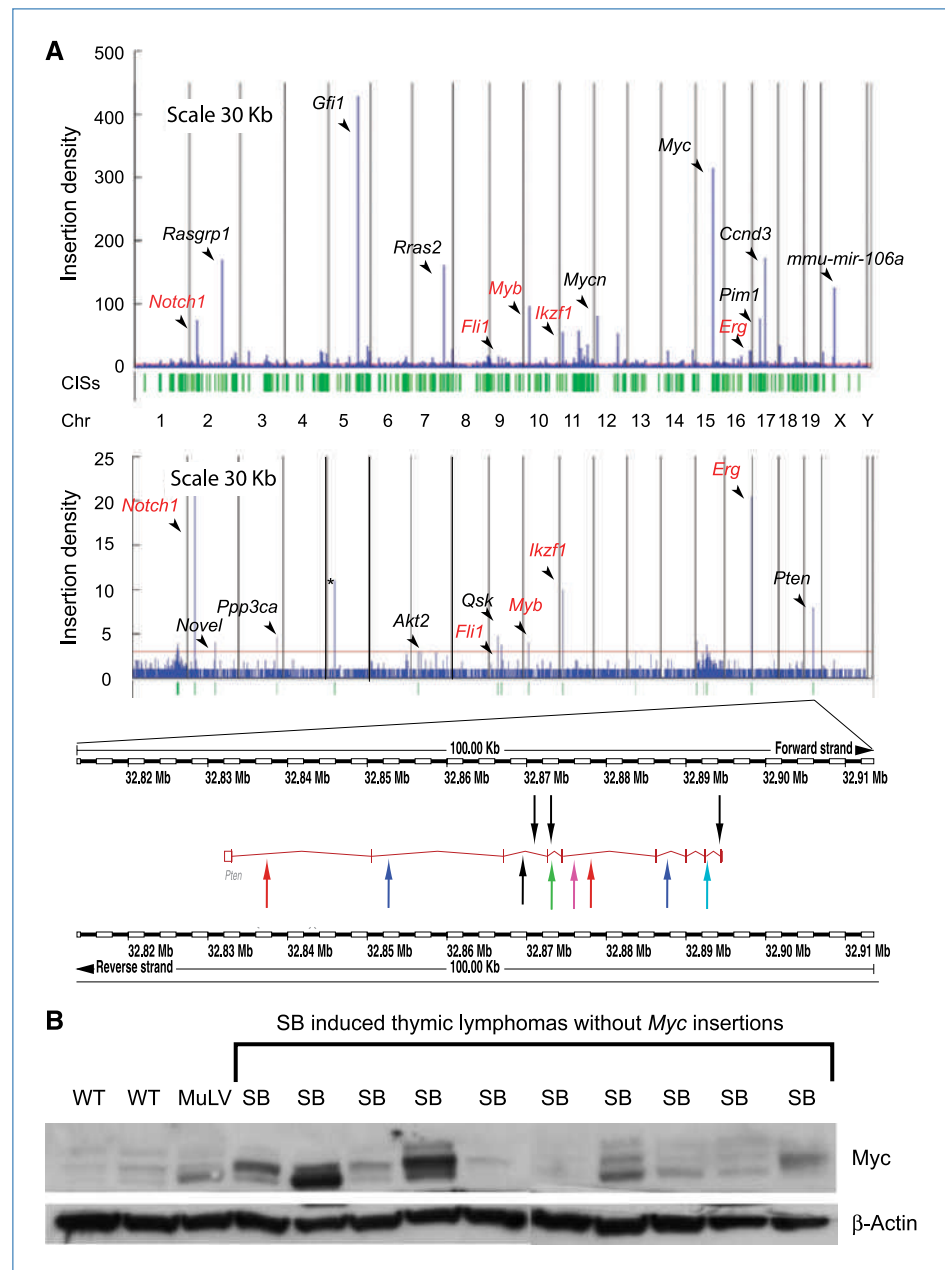


Table 1. List of mouse MuLV CIS genes that are in recurrent amplicons and deletions in human tumors**CIS genes that are recurrently amplified**

CIS gene	Mouse Ensembl ID	Human Ensembl ID	Position of CIS relative to gene	<i>Sleeping Beauty</i> CIS gene	Cancer Census gene description	COSMIC	All CGH
<i>Wwox</i>	ENSMUSG00000004637	ENSG00000186153	Inside	—	—	—	—
<i>Etv6</i>	ENSMUSG000000030199	ENSG00000139083	Inside	—	Dominant	—	—
<i>Myc</i>	ENSMUSG000000022346	ENSG00000136997	Upstream	—	Dominant	—	—
<i>Ccnd2</i>	ENSMUSG00000000184	ENSG00000118971	Upstream	—	Dominant	—	—
<i>Mycn</i>	ENSMUSG000000037169	ENSG00000134323	Inside	—	Dominant	COSMIC	Amplified
<i>Ccnd1</i>	ENSMUSG000000070348	ENSG00000110092	Upstream	—	Dominant	—	—
<i>Ikzf3</i>	ENSMUSG000000018168	ENSG00000161405	Inside	—	—	COSMIC	—
<i>Sla</i>	ENSMUSG000000022372	ENSG00000155926	Inside	—	—	—	—
<i>Lgals9</i>	ENSMUSG000000001123	ENSG00000171916	Upstream	—	—	—	—
<i>Pml</i>	ENSMUSG000000036986	ENSG00000140464	Inside	—	Dominant	COSMIC	—
<i>Itpr2</i>	ENSMUSG000000030287	ENSG00000123104	Upstream	—	—	COSMIC	—
<i>D12Ert553e</i>	ENSMUSG000000020589	ENSG00000197872	Downstream	—	—	—	—
<i>Fyb</i>	ENSMUSG000000022148	ENSG00000082074	Inside	—	—	—	—
<i>Slc1a3</i>	ENSMUSG000000005360	ENSG00000079215	Downstream	—	—	—	—
<i>Cugbp2</i>	ENSMUSG000000002107	ENSG00000048740	Inside	—	—	—	Amplified
<i>Capsl</i>	ENSMUSG000000039676	ENSG00000152611	Downstream	—	—	—	—
<i>Sdk1</i>	ENSMUSG000000039683	ENSG00000146555	Downstream	—	—	—	—
<i>Trp53inp1</i>	ENSMUSG000000028211	ENSG00000164938	Upstream	—	—	—	—
<i>Erg</i>	ENSMUSG000000040732	ENSG00000157554	Upstream	SB	Dominant	—	—
<i>Ptp4a3</i>	ENSMUSG000000059895	ENSG00000184489	Inside	—	—	—	—
<i>Mcl1</i>	ENSMUSG000000038612	ENSG00000143384	Downstream	—	—	—	Amplified
<i>Evi1</i>	ENSMUSG000000027684	ENSG00000085276	Inside	—	Dominant	COSMIC	—
<i>Flt3</i>	ENSMUSG000000042817	ENSG00000122025	Inside	—	Dominant	COSMIC	—
<i>1600014C10Rik</i>	ENSMUSG000000054676	ENSG00000131943	Inside	—	—	—	—
<i>Pag1</i>	ENSMUSG000000027508	ENSG00000076641	Upstream	—	—	—	—
<i>Anp32e</i>	ENSMUSG000000015749	ENSG00000143401	Upstream	—	—	—	Amplified
<i>Tpd52</i>	ENSMUSG000000027506	ENSG00000076554	Upstream	—	—	—	—

CIS genes that are recurrently deleted

CIS gene	Mouse Ensembl ID	Human Ensembl ID	Position of CIS relative to gene	<i>Sleeping Beauty</i> CIS gene	Cancer Census gene description	COSMIC	All CGH
<i>Wwox</i>	ENSMUSG000000004637	ENSG00000186153	Inside	—	—	—	—
<i>Etv6</i>	ENSMUSG000000030199	ENSG00000139083	Inside	—	Dominant	—	Deleted
<i>Zfp438</i>	ENSMUSG0000000050945	ENSG00000183621	Upstream	—	—	—	—
<i>Lgals9</i>	ENSMUSG000000001123	ENSG00000171916	Upstream	—	—	—	Deleted
<i>Dym</i>	ENSMUSG000000035765	ENSG00000141627	Inside	—	—	—	—
<i>Sdk1</i>	ENSMUSG000000039683	ENSG00000146555	Downstream	—	—	—	Deleted
<i>Acot11</i>	ENSMUSG000000034853	ENSG00000162390	Upstream	—	—	—	—
<i>Notch1</i>	ENSMUSG000000026923	ENSG00000148400	Inside	SB	Dominant	COSMIC	—
<i>Dock8</i>	ENSMUSG000000052085	ENSG00000107099	Upstream	—	—	—	Deleted
<i>Ntn1</i>	ENSMUSG000000020902	ENSG00000065320	Downstream	—	—	—	Deleted
<i>Vps13d</i>	ENSMUSG000000020220	ENSG00000048707	Inside	—	—	—	—
<i>Pik3cd</i>	ENSMUSG000000039936	ENSG00000171608	Upstream	—	—	—	—
<i>Ccnd2</i>	ENSMUSG000000000184	ENSG00000118971	Upstream	—	Dominant	—	—

(Continued on the following page)

Table 1. List of mouse MuLV CIS genes that are in recurrent amplicons and deletions in human tumors (Cont'd)

CNV	Nanog BS	Oct4 BS	ES cell module	No. of amplicons	P	No. of other genes in minimal amplified region
CNV	Nanog	Oct4	ES	69	0.0041	0
—	—	—	—	54	0.0043	0
—	Nanog	—	ES	55	0.0043	0
—	—	—	ES	25	0.0101	0
—	Nanog	Oct4	ES	25	0.0101	0
—	—	—	ES	23	0.0111	0
—	—	Oct4	—	22	0.0121	0
—	—	—	—	20	0.0132	1
CNV	—	—	—	17	0.0154	0
CNV	—	Oct4	—	14	0.0212	0
—	—	—	—	14	0.0212	0
—	Nanog	—	—	14	0.0212	0
—	Nanog	Oct4	—	14	0.0212	6
—	—	—	—	13	0.0248	9
—	—	Oct4	—	13	0.0248	0
—	—	—	—	13	0.0248	18
CNV	Nanog	Oct4	—	13	0.0248	0
—	—	—	—	12	0.0295	6
—	—	—	—	10	0.0365	0
—	—	—	ES	10	0.0365	30
—	—	—	—	10	0.0365	20
—	—	—	—	9	0.0447	0
—	—	—	—	9	0.0447	0
—	—	—	—	9	0.0447	0
—	—	—	—	9	0.0447	15
—	—	—	ES	9	0.0447	54
—	Nanog	—	—	9	0.0447	15
CNV	Nanog BS	Oct4 BS	ES cell module	No. of deletions	P	No. of other genes in minimal deleted region
CNV	Nanog	Oct4	ES	87	0.0011	0
—	—	—	—	60	0.0018	0
—	—	—	—	33	0.0077	0
CNV	—	—	—	31	0.0083	0
—	—	—	—	19	0.0138	0
CNV	Nanog	Oct4	—	19	0.0138	0
—	—	—	—	15	0.0176	0
CNV	—	—	ES	12	0.0237	0
CNV	—	—	—	11	0.0265	0
—	Nanog	—	—	10	0.0282	0
—	—	—	—	10	0.0282	0
—	—	Oct4	—	8	0.0382	1
—	—	—	ES	7	0.0422	0

(Continued on the following page)

Table 1. List of mouse MuLV CIS genes that are in recurrent amplicons and deletions in human tumors (Cont'd)

CIS genes that are recurrently amplified

CIS gene	Mouse Ensembl ID	Human Ensembl ID	Position of CIS relative to gene	<i>Sleeping Beauty</i> CIS gene	Cancer Census gene description	COSMIC	All CGH
<i>Jazf1</i>	ENSMUSG00000063568	ENSG00000153814	Inside	—	Dominant	—	Deleted
<i>Rcbtb2</i>	ENSMUSG00000022106	ENSG00000136161	Upstream	—	—	COSMIC	Deleted
<i>Zfp608</i>	ENSMUSG000000052713	ENSG00000168916	Upstream	—	—	COSMIC	—
<i>Zmiz1</i>	ENSMUSG000000007817	ENSG00000108175	Upstream/Inside	—	—	—	—
<i>Ets1</i>	ENSMUSG000000032035	ENSG00000134954	Upstream	—	—	COSMIC	—

NOTE: Location of the CIS relative to the gene is indicated, the presence of SB CIS, the designation of the gene in the Cancer Gene Census, the presence of mutations in the gene in the COSMIC database and whether the amplicon or deletion was also identified in the survey of copy number changes in analysis of human ALL (17) is indicated. Many of the recurrently rearranged regions of the human genome overlap with germline CNV regions as indicated in the CNV column. The presence of Oct4 and Nanog binding sites (BS) in the gene and whether the gene is a component of the ES cell module gene set (26) is also indicated. The *P* value was calculated for each CIS gene by counting the number of non-CIS genes within a higher number of amplicons/deletions and dividing it by the total number of non-CIS genes in the genome. Finally, the number of genes in the minimal amplified or deleted region was calculated.

(chromosome 15), described previously (7). Tumors were collected when mice became moribund. The SB tumor panel we used in our analysis is described in detail in Collier and colleagues (16), with the exception of 10 tumors, which were on a Bloom heterozygous background.⁶ Immunophenotyping of SB tumors revealed that the majority are of T-cell origin (16). Immunophenotyping of the MuLV tumors (6) indicated that they are either of T-cell or B-cell origin.

Insertion site isolation, analysis, and assigning insertions to genes. These methods are provided in the Supplementary Methods.

Cross-species analysis. Human orthologues of the mouse candidate genes and their genomic coordinates on NCBI 36 were extracted from Ensembl v45_36f. We chose a threshold copy number ratio of ≥ 1.7 for amplicons because this was the lowest threshold at which we observed an overrepresentation of orthologues from mouse candidates, compared with orthologues from other mouse genes, in amplified regions ($P = 8.46 \times 10^{-3}$ using the two-tailed Fisher exact test for genes amplified in two or more cell lines). Likewise, for deletions, we chose a threshold copy number of ≤ 0.3 , which was the highest threshold at which we observed an overrepresentation of orthologues in deleted regions in one or more cell lines ($P = 4.67 \times 10^{-3}$). Copy number variation (CNV) data was obtained and processed as described in the Supplementary Methods. Shared regions of deletion and amplification in pediatric acute lymphoblastic leukemias (ALL) were obtained from Mullighan and colleagues (17).

Cancer mutation data sets. Catalogue of Somatic Mutations in Cancer (COSMIC; ref. 18), Cancer Gene Census (3), and exon resequencing data sets were analyzed as described in the Supplementary Methods. The orthologues of all mouse

genes were extracted from Ensembl v45 and the number of amplicons/deletions containing each gene was calculated. Non-CIS genes were ranked according to the number of amplicons/deletions in which they resided, and a *P* value was calculated for each CIS gene by counting the number of non-CIS genes with a higher number of amplicons/deletions and dividing it by the total number of non-CIS genes. *P* values for the overrepresentation of CIS genes in COSMIC (18), Cancer Gene Census, and Sjoblom and colleagues (19) data sets were calculated using the one-tailed Fisher exact test. Only genes with mouse orthologues were included in the analysis.

Analysis of Oct4 and Nanog transcription factor binding sites and embryonic stem cell module genes. Ensembl identifiers and human orthologues were extracted from Ensembl BioMart. *P* values for the over-representation of genes with Oct4 and Nanog binding sites among CIS genes were calculated using the one-tailed Fisher exact test. To perform this analysis, we used chromatin immunoprecipitation-paired end ditag of 3,006 Nanog binding sites, 2,408 of which were found in 1,923 Ensembl mouse genes (20). Likewise, Oct4 binding sites in 817 mouse Ensembl genes, including 797 encoding proteins or miRNAs, was derived from 1,083 Oct4 binding sites (20). The embryonic stem (ES) cell module gene list was obtained from Wong and colleagues (21).

Western blotting for Myc expression. Western blotting for Myc expression was performed using standard procedures. The antibody used for these experiments was anti-Myc (SC-42/C-33) from Santa Cruz.

Pathway analysis. We analyzed MuLV CIS genes to determine if there were overrepresented Pfam (22) domains, KEGG (23), or GO pathways using the DAVID Tool (24).

Results

CGH. Across the 598 cell lines, the average number of statistically significant gains of copy number per cell line ≥ 1.7

⁶ L. Collier, unpublished data.

Table 1. List of mouse MuLV CIS genes that are in recurrent amplicons and deletions in human tumors (Cont'd)

CNV	Nanog BS	Oct4 BS	ES cell module	No. of amplicons	P	No. of other genes in minimal amplified region
—	—	—	—	7	0.0422	0
—	Nanog	—	—	7	0.0422	0
—	—	Oct4	—	6	0.0500	0
—	—	—	—	6	0.0500	0

was 34.03 (± 36.57). The average size of these amplicons was 299.10 ($\pm 1,667.93$) kb and an average of 2.99 (± 14.50) genes were found in each amplicon. The average number of statistically significant losses per cell line was 204.10 (± 194.36). These losses were on average 196.87 ($\pm 3,058.58$) kb in size, encompassing 2.61 (± 32.98) genes. Figure 1 shows the global overview of the distribution of the amplifications and deletions in this collection of cell lines, and in the hematopoietic subset. In total, we identified 2,424 amplifications and 14,010 deletions across the entire cell line panel.

Analysis of lymphomas induced using MuLV. We generated 1,005 murine lymphomas by infecting newborn mice with MuLV as described previously (6). The majority of the tumors were from mice on a wild-type [231], *p19* knockout [228], or *p53* knockout [126] background (Supplementary Table S3). Collectively, we generated 134,985 DNA sequencing reads from 1,734 splinkerette reactions. The insertion site sequences of a subset of these tumors have been published previously (6). We mapped 86,187 reads to the mouse genome assembly NCBI m36, identifying 22,579 insertion sites with an average of 22.47 (± 11.30) insertions per tumor. These data were analyzed using a kernel convolution-based algorithm (25), identifying 447 statistically significant CIS at a kernel width of 30 kb. The vast majority of these insertion sites were in genic regions of the genome. Candidate genes (Supplementary Table S4) were assigned to CIS using the criteria described in the Supplementary Methods.

Analysis of SB tumor panel. We performed splinkerette reactions from both ends of 73 SB-induced tumors, generating 10,791 DNA insertion site reads. Among these reads, 6,281 could be mapped to the mouse genome, identifying 2,643 insertion sites, 35.72 (± 18.77) per tumor. Seventy of the tumors analyzed were lymphomas. Two tumors were retrospectively classified as high-grade gliomas, and one a skin tumor. Using the kernel convolution framework (25), we identified 21 statistically significant CIS at a kernel width of 30 kb (Supplementary Table S5). Again, the majority of these CIS were in genic regions of the genome. Eighteen candidate genes were identified in the vicinity of these CIS using the criteria outlined in the Supplementary Methods. These CIS were filtered as described in the Supplementary Methods to remove CIS associated with local hopping, as well as other artifacts,

which resulted in nine CIS that were used for downstream analysis.

The genome-wide distribution of insertion sites in MuLV- and SB-induced lymphomas. Having identified insertion sites and CIS from 1,005 MuLV-induced lymphomas and 70 SB-induced lymphomas, we compared their genome-wide distributions (Fig. 2). The most frequently mutated genes in MuLV-induced lymphomas were *Gfi1/Evi5*, *c-Myc/Pvt1*, and *Ccnd3*. These genes had insertion densities of 427.28, 314.19, and 172.09, respectively, using the kernel convolution method of CIS detection (25) at a kernel width of 30 kb, which was determined to yield optimal sensitivity with this data set. Remarkably, in the SB data set, we found no insertions in or around these genes ($P < 0.0001$). This might reflect the bias of retroviruses to insert themselves into particular sites in the genome. Similarly, we identified a CIS in the tumor suppressor gene *Pten* (six tumors, $P < 0.05$; Fig. 2) in the SB panel, several tumors were found to contain multiple insertions in *Pten* which are presumably biallelic or insertions derived from tumor subclones, but we did not detect a single *Pten* insertion in any of the 1,005 tumors from the MuLV data set ($P < 0.0001$). This strongly suggests that the SB transposon (T2/Onc) used for these studies and MuLV are unique mutagens with complementary mutagenic profiles. Intriguingly, we found that despite carrying no insertions in or near the oncogene *Myc* many SB tumors showed a significant upregulation in Myc protein levels (Fig. 2). Although there were distinct differences in the profile of genes mutated using MuLV and the SB transposon system, several genes were frequently mutated by both mutagens. These included *Notch1*, *Myb*, *Ikzf1*, and *Fli1*.

Cross-species comparative analysis of human cancer data sets with the MuLV and SB data sets. Of the 439 CIS genes identified in MuLV-induced tumors, we were able to identify 384 orthologous genes within the human genome. Similarly, we were able to identify human orthologues for the nine SB CIS genes. Sixty-nine human orthologues of mouse genes predicted to be mutated by MuLV were genes with mutations in the COSMIC database (ref. 19; $P = 1.36 \times 10^{-9}$). Similarly, 36 of the human orthologues of mouse genes predicted to be mutated by MuLV were oncogenes described in the Cancer Gene Census ($P = 7.88 \times 10^{-18}$). In contrast, only three orthologues were found to be mutated in the data set

Table 2. List of mouse MuLV CIS genes that are in recurrent amplicons and deletions in human hematopoietic tumors

CIS genes that are recurrently amplified in human hematopoietic tumors

CIS gene	Mouse Ensembl ID	Human Ensembl ID	Position of CIS relative to gene	<i>Sleeping Beauty</i> CIS gene	Cancer Census gene description	COSMIC	All CGH
<i>Wwox</i>	ENSMUSG00000004637	ENSG00000186153	Inside	—	—	—	—
<i>Myc</i>	ENSMUSG000000022346	ENSG00000136997	Upstream	—	Dominant	—	—
<i>Ccnd2</i>	ENSMUSG00000000184	ENSG00000118971	Upstream	—	Dominant	—	—
<i>Cugbp2</i>	ENSMUSG000000002107	ENSG00000048740	Inside	—	—	—	Amplified
<i>Tpd52</i>	ENSMUSG000000027506	ENSG00000076554	Upstream	—	—	—	—
<i>Pag1</i>	ENSMUSG000000027508	ENSG00000076641	Upstream	—	—	—	—
<i>Trp53inp1</i>	ENSMUSG000000028211	ENSG00000164938	Upstream	—	—	—	—
<i>Sla</i>	ENSMUSG000000022372	ENSG00000155926	Inside	—	—	—	—
<i>Nedd4l</i>	ENSMUSG000000024589	ENSG00000049759	Upstream	—	—	—	—
<i>Pml</i>	ENSMUSG000000036986	ENSG00000140464	Upstream	—	Dominant	COSMIC	—
<i>Mcl1</i>	ENSMUSG000000038612	ENSG00000143384	Downstream	—	—	—	Amplified
<i>Bcl11b</i>	ENSMUSG000000048251	ENSG00000127152	Inside	—	Dominant	COSMIC	—
<i>Ptp4a3</i>	ENSMUSG000000059895	ENSG00000184489	Inside	—	—	—	—
<i>Anp32e</i>	ENSMUSG000000015749	ENSG00000143401	Downstream	—	—	—	Amplified
<i>Ikzf3</i>	ENSMUSG000000018168	ENSG00000161405	Inside	—	—	COSMIC	Deleted
<i>Mbd2</i>	ENSMUSG000000024513	ENSG00000134046	Upstream	—	—	—	—
<i>Evi1</i>	ENSMUSG000000027684	ENSG00000085276	Inside	—	Dominant	COSMIC	—
<i>Rorc</i>	ENSMUSG000000028150	ENSG00000143365	Inside	—	—	COSMIC	Amplified
<i>Vpreb2</i>	ENSMUSG000000059280	ENSG00000169575	Upstream	—	—	—	Amplified

CIS genes that are recurrently deleted in human hematopoietic tumors

CIS gene	Mouse Ensembl ID	Human Ensembl ID	Position of CIS relative to gene	<i>Sleeping Beauty</i> CIS gene	Cancer Census gene description	COSMIC	All CGH
<i>Wwox</i>	ENSMUSG000000004637	ENSG00000186153	Inside	—	—	—	—
<i>Etv6</i>	ENSMUSG000000030199	ENSG00000139083	Inside	—	Dominant	—	Deleted
<i>Lgals9</i>	ENSMUSG000000001123	ENSG00000171916	Upstream	—	—	—	Deleted
<i>Rcbtb2</i>	ENSMUSG000000022106	ENSG00000136161	Upstream	—	—	COSMIC	Deleted
<i>Ccnd2</i>	ENSMUSG000000000184	ENSG00000118971	Upstream	—	Dominant	—	—
<i>Notch1</i>	ENSMUSG000000026923	ENSG00000148400	Inside	SB	Dominant	COSMIC	Amplified
<i>Zfp438</i>	ENSMUSG000000050945	ENSG00000183621	Upstream	—	—	—	Amplified
<i>Zmiz1</i>	ENSMUSG000000007817	ENSG00000108175	Inside	—	—	—	—
<i>Vps13d</i>	ENSMUSG000000020220	ENSG00000048707	Inside	—	—	—	—
<i>Ntn1</i>	ENSMUSG000000020902	ENSG00000065320	Downstream	—	—	—	Deleted
<i>Fut8</i>	ENSMUSG000000021065	ENSG00000033170	Upstream	—	—	—	—
<i>Tbc1d1</i>	ENSMUSG000000029174	ENSG00000065882	Downstream	—	—	—	—
<i>Foxp1</i>	ENSMUSG000000030067	ENSG00000114861	Upstream	—	—	COSMIC	—
<i>Dym</i>	ENSMUSG000000035765	ENSG00000141627	Inside	—	—	—	—
<i>Sdk1</i>	ENSMUSG000000039683	ENSG00000146555	Downstream	—	—	—	Deleted
<i>Vpreb2</i>	ENSMUSG000000059280	ENSG00000169575	Upstream	—	—	—	Amplified

NOTE: The column designations are the same as described for Table 1.

from Sjöblom and colleagues (ref. 19; $P = 0.74$). This might reflect the fact that the Sjöblom data set was an exon resequencing study of breast and bowel tumors exclusively, and therefore, it might be biased against those genes mutated

in tumors of the hematopoietic system, and genes disrupted by large-scale rearrangements. Similarly, five genes from the SB data set were also genes within the COSMIC database ($P = 4.04 \times 10^{-4}$), and six were within the Cancer Gene Census

Table 2. List of mouse MuLV CIS genes that are in recurrent amplicons and deletions in human hematopoietic tumors

CNV	Nanog BS	Oct4 BS	ES cell module	No. of amplicons	<i>P</i>	No. of other genes in minimal amplified region
CNV	Nanog	Oct4	ES	9	0.0024	0
—	Nanog	—	ES	7	0.0025	0
—	—	—	ES	5	0.0052	0
—	—	Oct4	—	5	0.0052	0
—	Nanog	—	—	4	0.0196	25
—	—	—	—	4	0.0196	25
—	—	—	—	4	0.0196	20
—	—	—	—	3	0.0314	5
CNV	Nanog	—	—	3	0.0314	9
CNV	—	Oct4	—	3	0.0314	0
—	—	—	—	3	0.0314	20
—	—	—	—	3	0.0314	0
—	—	—	ES	3	0.0314	30
—	—	—	ES	2	0.0478	63
—	—	Oct4	—	2	0.0478	0
—	—	—	—	2	0.0478	34
—	—	—	—	2	0.0478	0
—	—	—	—	2	0.0478	8
—	—	—	—	2	0.0478	3
CNV	Nanog BS	Oct4 BS	ES cell module	No. of deletions	<i>P</i>	No. of other genes in minimal deleted region
CNV	Nanog	Oct4	ES	16	0.0014	0
—	—	—	—	11	0.0067	0
CNV	—	—	—	5	0.0194	0
—	—	—	—	4	0.0229	1
—	—	—	ES	3	0.0280	0
CNV	—	—	ES	3	0.0280	0
—	—	—	—	3	0.0280	0
—	—	Oct4	—	2	0.0379	0
—	—	—	—	2	0.0379	0
—	Nanog	—	—	2	0.0379	0
—	—	—	—	2	0.0379	0
—	Nanog	—	—	2	0.0379	0
—	Nanog	—	—	2	0.0379	0
—	—	—	—	2	0.0379	0
CNV	Nanog	Oct4	—	2	0.0379	0
—	—	—	—	2	0.0379	0

($P = 4.26 \times 10^{-6}$). This analysis reveals that using MuLV or the SB transposon system for cancer gene discovery has significant predictive power for those genes relevant to tumor formation in humans.

Cross-species comparative analysis of the human CGH and mouse MuLV data sets. There were 9,681 human genes with orthologues in the mouse genome found within

amplicons of human tumors. Two hundred and thirty-two of these genes were retroviral CIS genes, which is greater than the number expected by chance ($P = 4.47 \times 10^{-3}$). Twenty-seven CIS genes showed significant recurrent amplification in humans compared with non-CIS genes ($P < 0.05$; Table 1). Nine of these genes were designated dominant cancer genes in the Cancer Gene Census, a significantly higher number

than expected by chance ($P = 2.85 \times 10^{-4}$). Eighteen retroviral CIS genes showed recurrent deletion ($P \leq 0.05$; Table 1). Seven of these genes contained intragenic CIS, which is not significantly different from the number of other CIS genes with intragenic CIS ($P = 0.990$). This probably reflects the fact that MuLV is primarily a dominantly acting mutagen. Five genes (*CCND2*, *ETV6*, *LGALS9*, *SDK1*, and *WWOX*) were both significantly amplified and significantly deleted. This is a larger overlap than expected by chance ($P = 1.12 \times 10^{-4}$) and might suggest that some of these genes reside in unstable regions of the genome. Indeed, several of the recurrently amplified or deleted genes overlap with regions of germline CNV identified previously (ref. 21; Table 1). We also observed significant overlap of the copy number signatures in our survey of copy number alterations with those from a large CGH analysis of ALLs, which provides cross-platform validation (17). In addition, we performed the same analysis focusing just on the hematopoietic cell lines. The orthologues of 71 retroviral CIS genes were found within amplicons in human tumors of hematopoietic or lymphoid origin (Table 2). Nineteen CIS gene orthologues were shown to be recurrently amplified across the hematopoietic and lymphoid subset of the tumor panel ($P < 0.05$). Fourteen of these genes were also significantly amplified across all cell lines. Sixteen retroviral CIS genes were found in a significant number of deletions in tumors of hematopoietic and lymphoid origin ($P < 0.05$), and 11 of these genes were also found to be mutated in the analysis

using the entire collection of tumor cell lines (Table 2). Six of these genes contained intragenic CIS (Table 2).

Identification of Nanog and Oct4 binding sites in MuLV CIS genes and the effect of mutations in ES cell module genes on tumor latency.

In an attempt to ascribe putative functions for the genes we identified in our analysis, we next set out to determine if they contained binding sites for the transcription factors Oct4 (26) and Nanog (27), which play an important role in ES cell self-renewal. Many genes implicated in the regulation of embryonic “stemness” have been shown to play a role in tumor self-renewal and aggressiveness (21, 28). Remarkably, there was a highly significant enrichment of genes containing Oct4 and Nanog binding sites among those genes linked to CIS in MuLV-induced mouse tumors ($P = 1.64 \times 10^{-5}$ and $P = 5.86 \times 10^{-4}$, for Oct4 and Nanog, respectively). None of the genes linked to SB CIS had Nanog or Oct4 binding sites ($P = 1$ for both tests); however, this might reflect the small size of the data set. Mutations in ES cell module genes, of which the presence of Oct4 or Nanog binding sites is a common feature, have been proposed to be predictive of tumor aggressiveness (21, 28). We found that mice that carried tumors with MuLV insertions in or near ES cell module genes (21) became moribund at a significantly accelerated rate compared with mice that carried tumors without mutations linked to ES cell module genes (Fig. 3; $P < 0.0001$). The most frequently mutated ES cell module genes were *Myc*, *Myb*, and *Notch1* for MuLV,

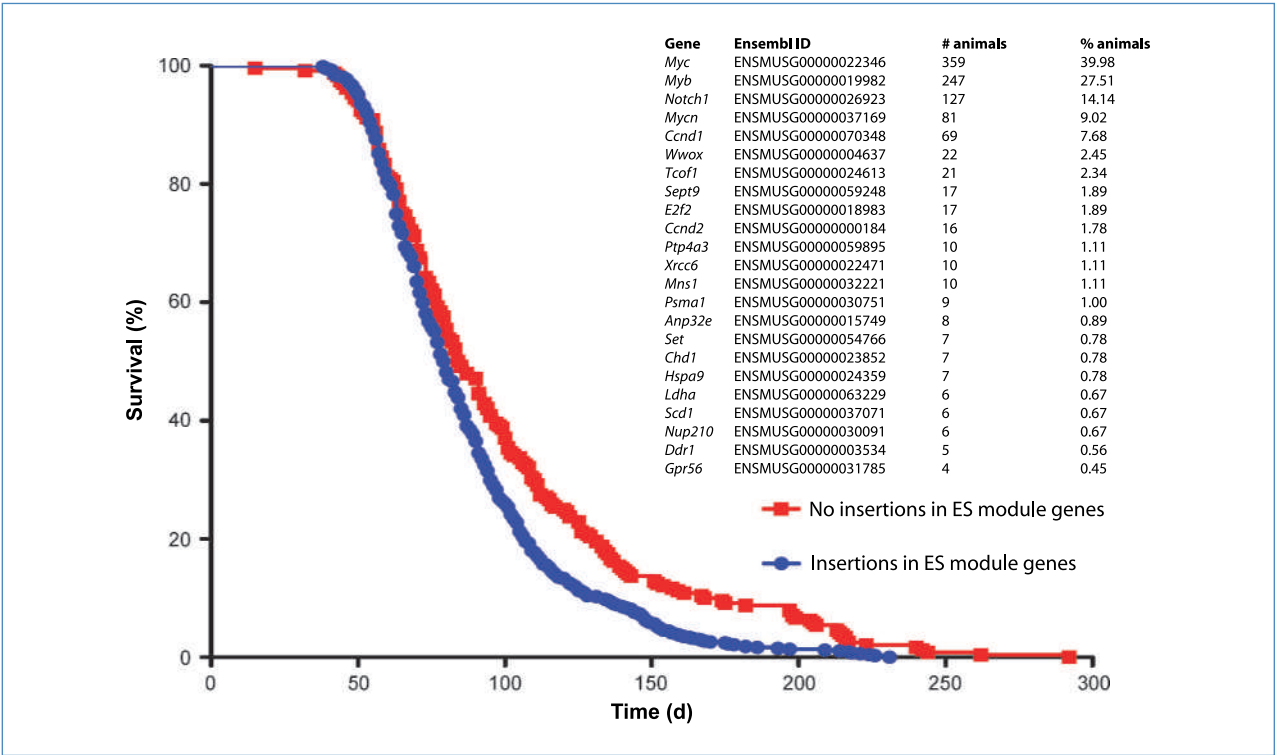


Figure 3. Mice carrying MuLV-induced tumors with CIS linked to ES module genes have significantly reduced survival compared with mice without mutations affecting these genes. A list of the most commonly mutated genes and their mutation frequencies are shown. Log-rank (Mantel-Cox) test ($P < 0.0001$).

whereas *Notch1* and *Myb* were the only ES cell module genes mutated by SB (Tables 1 and 2).

Pathways analysis. KEGG, GO, and DAVID analysis revealed an overrepresentation of MuLV (Supplementary Table S6) and SB (Supplementary Table S7) CIS genes in pathways known to participate in cancer formation and hematopoiesis. Kinase domains were also overrepresented in MuLV CIS genes.

Discussion

New high-throughput genomic analysis techniques such as massively parallel sequencing and ultra high-resolution CGH are identifying remarkable heterogeneity in cancer genomes (29), implicating a multitude of genes and pathways in oncogenesis and cancer progression. Determining which of these rearrangements have actually driven tumor initiation and progression will be a significant undertaking. Ideally, validation of genetic rearrangements should involve systematic experimental evaluation. However, few of the experimental approaches that may be used for validating cancer genes are high-throughput and, with the exception of animal models, most are unable to faithfully recapitulate the genetic and cellular context in which cancers form. Forward genetic screens in mice are a powerful tool for cancer gene discovery because tumors are formed via somatic mutation and, like human tumors, undergo a process of evolution resulting in the emergence of a malignant clone (9). When used in combination as part of a comparative oncogenomics approach, high-resolution analysis of human cancer genomes by CGH and insertion sites derived from mouse tumors represents a powerful way of identifying new genes relevant to oncogenesis.

In this study, we identified 27 genes that were recurrently amplified in human tumors in which the orthologous mouse gene was a site of clonal retroviral insertions in murine lymphomas induced using MuLV (Table 1). Similarly, we identified 18 genes that were recurrently deleted in human tumors and were also CIS genes in the MuLV data set (Table 1). Using the same approach, we identified 19 recurrently amplified and 16 recurrently deleted CIS genes by comparison to the CGH data for the hematopoietic subset of the tumor panel (Table 2). Reassuringly, we identified known dominantly active oncogenes from the Cancer Gene Census (3), as well as genes from the COSMIC (18) database which were somatically mutated in human cancers. Many of the genes that we predict to be potential cancer genes were, however, novel. Importantly, several of the genes we identified in our analysis were found in regions of the genome either recurrently amplified or deleted in a large survey of human ALLs (17), which provides cross-platform validation. Several genes, such as *WWOX*, were both recurrently amplified and deleted (Tables 1 and 2). This might reflect the fact that these genes are located in unstable or fragile regions of the genome (30). Indeed, many of the genes that we identified in our analysis were also found to be located in CNV regions of the human genome (31). This does not exclude them from being cancer genes but may indicate something of the underlying genomic architecture in which they reside. Intriguingly, we observed

several deletions that removed the entire *NOTCH1* locus, and other deletions that removed internal exons of *NOTCH1* and potentially result in the formation of oncogenic NOTCH-IC protein (Supplementary Fig. S1). Similarly, we observed a recurrent exon-specific deletion within *ETSI* that potentially generates a neomorphic allele (Supplementary Fig. S2). Importantly, there were several CIS genes identified in our analysis (Tables 1 and 2) that were designated as dominantly active in the cancer gene census (3), but which we found to be deleted in our panel of human tumors. These include *Etv6* (17) and *Bcl11b* (32). It is possible that these genes function in both gain and loss of function roles in tumorigenesis. One of the most compelling genes we identified in our analysis was the protein tyrosine phosphatase type IVA, member 3 gene (*Ptp4a3*), which was amplified in 10 tumors and contained multiple intragenic insertions (Table 1). PTP genes are a small class of prenylated protein tyrosine phosphatases implicated in many cellular processes including growth.

Just as a cross-species oncogenomics approach is a powerful method of identifying genes that may be of importance in human cancer formation, performing forward genetic screens in mice with multiple mutagens is a potentially powerful way of identifying functionally important cancer genes. In this study, we isolated insertion site sequences from tumors generated using both MuLV and the SB transposon system. Analysis revealed that these mutagens have a remarkably different mutagenic profile. Although *Myc/Pvt1*, *Gfi1/Evi5*, and *Ccnd3* were frequently mutated in MuLV tumors, this was not the case in SB tumors (Fig. 2). χ^2 analysis of the mutation profiles revealed a statistically significant difference in each case ($P < 0.0001$). Similarly, *Pten* was mutated in 6 of the 73 SB tumors but not in any of the 1,005 MuLV tumors ($P < 0.0001$). The fact that we did not detect SB insertions in or around *Myc* is striking because activation of *MYC* is a critical event in the development of many forms of human lymphoma and because one of the transposon donors was located on chromosome 15, the same chromosome as *Myc*, which should have favored insertions into *Myc* by local hopping. To investigate this further, we took 10 SB-induced thymic lymphomas and performed Western blotting to compare the level of *Myc* protein expression with wild-type thymus (Fig. 2). In at least five cases, we observed elevated *Myc* protein levels. The fact that there are no insertions in or near *Myc* in these SB tumors raises the question of whether the T2/Onc transposon is capable of inserting near this gene and activating expression. Possibly, the MSCV promoter in T2/Onc is in an unfavorable context to activate *Myc*, or that the *Myc* locus is in an unfavorable context for SB transposition, or that the *Myc* locus is amplified, which would make insertions into *Myc* redundant. Similarly, we observed no SB insertions in or near *Gfi1*, which was frequently mutated by MuLV. In the experiments described in this article, the mice treated with MuLV were on a pure FVB background, whereas the SB tumors were collected from mice that were on a hybrid C57BL/6J-FVB background. It is possible that some of the differences in the insertion profiles we describe are due to different preferences for viral or transposon integration on these different genetic backgrounds. However, insertions of

MuLV into *Myc* and *Gfi1* have been shown to occur on most genetic backgrounds including in C57BL/6J hybrids (33, 34). The observation that SB tumors contain insertions in *Pten*, which were not found in MuLV-induced tumors, is in keeping with the suggestion that *Pten* plays an important role in T-cell lymphomagenesis (35). As we have shown previously (16), immunophenotyping of the SB tumors we used in our analysis revealed that the majority were CD4/CD8 double-positive T-cell tumors, or were B220⁺ and therefore B cell-derived. The occasional SB tumor seemed to have two malignant cell clones. MuLV-induced tumors are either CD3⁺ or B220⁺, i.e., either T cell- or B cell-derived (6). It remains possible that some of the differences between the insertion profiles observed between the SB and MuLV tumors may be due to different subtypes of disease.

In this study, we also illustrate that genes with Oct4 and Nanog binding sites are enriched in genes found to be at MuLV CIS and that MuLV insertions in or near stem cell module genes is predictive of decreased survival (Fig. 3). Importantly, the most frequent stem cell module genes mutated were *Myc*, *Myb*, and *Notch 1*. Using immunophenotyping data for 349 of the MuLV tumors (6), we determined that there was not a significant difference in the CD3 (T cell) or B220 (B cell) marker status between tumors with or without insertions linked to stem cell module genes, although the subclassification of these lymphomas with additional markers may be revealing. Finally, we also identified overrepresented KEGG and GO pathways, and Pfam domains in our analysis. Not surprisingly, these pathways and genes included those implicated in hematopoiesis, development, and in important cellular processes such as cell division and transcription.

In conclusion, we have performed extensive cross-species comparative analysis, identifying a large number of candidate cancer genes that now represent worthy targets for further functional validation in model systems. We also illustrate that cross-species oncogenomics is a powerful tool for cancer gene identification.

Disclosure of Potential Conflicts of Interest

The University of Minnesota has a pending patent on the process of using transposons such as SB for cancer gene discovery. D.A. Largaespada and L.S. Collier are named among the inventors. The other authors have declared no conflict of interest.

Acknowledgments

We thank L. Bendzick, V. Maklakova, and M. Derezinski for technical assistance.

Grant Support

Cancer Research-UK and Wellcome Trust (D.J. Adams); NWO Genomics program and the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (J. Kool, A.G. Uren, L. Wessels, J. Jonkers, A. Butler, and M. van Lohuizen); the BioRange program of the Netherlands Bioinformatics Centre, which is supported by a BSIK grant through the Netherlands Genomics Initiative (J. de Ridder); the Cancer Genomics Centre through a Netherlands Genomics Initiative (J. Kool and A.G. Uren); Wellcome Trust (the Cancer Genome Project); the Kay Kendall Leukemia Fund (L. van der Weyden); and the National Cancer Institute (K01CA122183) and an American Cancer Society pre-doctoral fellowship (L.S. Collier).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 5/13/09; revised 10/26/09; accepted 11/11/09; published OnlineFirst 1/26/10.

References

- Goymer P. Natural selection: the evolution of cancer. *Nature* 2008; 454:1046–8.
- Gabriel S. Variation in the human genome and the inherited basis of common disease. *Semin Oncol* 2006;33:S46–9.
- Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. *Nat Rev Cancer* 2004;4:177–83.
- Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998;20:207–11.
- Mikkers H, Berns A. Retroviral insertional mutagenesis: tagging cancer pathways. *Adv Cancer Res* 2003;88:53–99.
- Uren AG, Kool J, Matentzoglou K, et al. Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. *Cell* 2008;133:727–41.
- Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* 2005;436:272–6.
- Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* 2005;436:221–6.
- Uren AG, Kool J, Berns A, van Lohuizen M. Retroviral insertional mutagenesis: past, present and future. *Oncogene* 2005;24:7656–72.
- Collier LS, Largaespada DA. Transforming science: cancer gene identification. *Curr Opin Genet Dev* 2006;16:23–9.
- van Lohuizen M, Verbeek S, Scheijen B, Wientjens E, van der Gulden H, Berns A. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* 1991;65:737–52.
- Kim M, Gans JD, Nogueira C, et al. Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene. *Cell* 2006;125: 1269–81.
- Zender L, Spector MS, Xue W, et al. Identification and validation of oncogenes in liver cancer using an integrative oncogenomic approach. *Cell* 2006;125:1253–67.
- Xue W, Krasnitz A, Lucito R, et al. DLC1 is a chromosome 8p tumor suppressor whose loss promotes hepatocellular carcinoma. *Genes Dev* 2008;22:1439–44.
- Available from: <http://www.sanger.ac.uk/genetics/CGP/Archive/>.
- Collier LS, Adams DJ, Hackett CS, et al. Whole-body sleeping beauty mutagenesis can cause penetrant leukemia/lymphoma and rare high-grade glioma without associated embryonic lethality. *Cancer Res* 2009;69:8429–37.
- Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007; 446:758–64.
- Forbes SA, Bhamra G, Bamford S, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* 2008. Chapter 10:Unit 10 1.
- Sjoberg T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science* 2006; 314:268–74.
- Loh YH, Wu Q, Chew JL, et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* 2006;38:431–40.
- Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module

- map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* 2008;2:333–44.
22. Finn RD, Tate J, Mistry J, et al. The Pfam protein families database. *Nucleic Acids Res* 2008;36:D281–8.
 23. Kanehisa M, Araki M, Goto S, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 2008;36:D480–4.
 24. Dennis G, Jr., Sherman BT, Hosack DA, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 2003;4:P3.
 25. de Ridder J, Uren A, Kool J, Reinders M, Wessels L. Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. *PLoS Comput Biol* 2006;2:e166.
 26. Nichols J, Zevnik B, Anastasiadis K, et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 1998;95:379–91.
 27. Chambers I, Colby D, Robertson M, et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* 2003;113:643–55.
 28. Ben-Porath I, Thomson MW, Carey VJ, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* 2008;40:499–507.
 29. Campbell PJ, Stephens PJ, Pleasance ED, et al. Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat Genet* 2008;40:722–9.
 30. Durkin SG, Glover TW. Chromosome fragile sites. *Annu Rev Genet* 2007;41:169–92.
 31. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature* 2006;444:444–54.
 32. Kamimura K, Mishima Y, Obata M, Endo T, Aoyagi Y, Kominami R. Lack of Bcl11b tumor suppressor results in vulnerability to DNA replication stress and damages. *Oncogene* 2007;26:5840–50.
 33. Broussard DR, Mertz JA, Lozano M, Dudley JP. Selection for c-myc integration sites in polyclonal T-cell lymphomas. *J Virol* 2002;76:2087–99.
 34. Dudley JP, Mertz JA, Rajan L, Lozano M, Broussard DR. What retroviruses teach us about the involvement of c-Myc in leukemias and lymphomas. *Leukemia* 2002;16:1086–98.
 35. Lenz G, Wright GW, Emre NC, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci U S A* 2008;105:13520–5.