nature genetics

The genomic landscape of juvenile myelomonocytic leukemia

Elliot Stieglitz[1,38], Amaro N Taylor-Weiner[2,38], Tiffany Y Chang[1], Laura C Gelston[1], Yong-Dong Wang[3], Tali Mazor[4], Emilio Esquivel[1], Ariel Yu[1], Sara Seepo[2], Scott R Olsen[5], Mara Rosenberg[2], Sophie L Archambeault[1], Ghada Abusin[6], Kyle Beckman[1], Patrick A Brown[7], Michael Briones[8], Benjamin Carcamo[9], Todd Cooper[10], Gary V Dahl[11], Peter D Emanuel[12], Mark N Fluchel[13], Rakesh K Goyal[14], Robert J Hayashi[15], Johann Hitzler[16], Christopher Hugge[17], Y Lucy Liu[12], Yoav H Messinger[18], Donald H Mahoney Jr[19], Philip Monteleone[20], Eneida R Nemecek[21], Philip A Roehrs[22], Reuven J Schore[23], Kimo C Stine[24], Clifford M Takemoto[7], Jeffrey A Toretsky[25,26], Joseph F Costello[4], Adam B Olshen[27,28], Chip Stewart[2], Yongjin Li[3], Jing Ma[29], Robert B Gerbing[30], Todd A Alonzo[31], Gad Getz[2,32,33], Tanja A Gruber[29,34], Todd R Golub[2,35,36], Kimberly Stegmaier[2,35,36] & Mignon L Loh[1,37]

Juvenile myelomonocytic leukemia (JMML) is a myeloproliferative neoplasm (MPN) of childhood with a poor prognosis. Mutations in *NF1*, *NRAS*, *KRAS*, *PTPN11* or *CBL* occur in 85% of patients, yet there are currently no risk stratification algorithms capable of predicting which patients will be refractory to conventional treatment and could therefore be candidates for experimental therapies. In addition, few molecular pathways aside from the RAS-MAPK pathway have been identified that could serve as the basis for such novel therapeutic strategies. We therefore sought to genomically characterize serial samples from patients at diagnosis through relapse and transformation to acute myeloid leukemia to expand knowledge of the mutational spectrum in JMML. We identified recurrent mutations in genes involved in signal transduction, splicing, Polycomb repressive complex 2 (PRC2) and transcription. Notably, the number of somatic alterations present at diagnosis appears to be the major determinant of outcome.

JMML is a rare but aggressive form of childhood leukemia that exhibits both myelodysplastic and myeloproliferative properties[1]. The only curative therapy is hematopoietic stem cell transplantation (HSCT)[2]. However, some patients exhibit highly aggressive disease despite HSCT, whereas spontaneous remission is occasionally observed in others with minimal therapy[3,4]. The lack of current laboratory, genetic and clinical features to distinguish these patients[5,6] presents a clinical dilemma for physicians and parents. We hypothesized that complete genomic characterization of JMML would aid in distinguishing these cases and would further identify relevant molecular

[1]Department of Pediatrics, Benioff Children's Hospital, University of California, San Francisco, San Francisco, California, USA. [2]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [3]Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [4]Department of Neurological Surgery, University of California, San Francisco, San Francisco, California, USA. [5]Hartwell Center for Bioinformatics and Biotechnology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [6]Stead Family Department of Pediatrics, University of Iowa Carver College of Medicine, Iowa City, Iowa, USA. [7]Department of Pediatrics, Johns Hopkins Hospital, Baltimore, Maryland, USA. [8]Department of Pediatrics, Emory University School of Medicine, Aflac Cancer and Blood Disorder Center, Atlanta, Georgia, USA. [9]Department of Pediatrics, Texas Tech University, El Paso, Texas, USA. [10]Department of Pediatrics, Seattle Children's Hospital, Seattle, Washington, USA. [11]Department of Pediatrics, Stanford School of Medicine, Stanford, California, USA. [12]Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA. [13]Department of Pediatric Hematology Oncology, University of Utah, Salt Lake City, Utah, USA. [14]Department of Pediatrics, Children's Hospital of Pittsburgh of the University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA. [15]Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri, USA. [16]Division of Hematology/Oncology, The Hospital for Sick Children, Toronto, Ontario, Canada. [17]Pediatric Hematology Oncology, SSM Cardinal Glennon Children's Medical Center, St. Louis, Missouri, USA. [18]Division of Pediatric Hematology Oncology, Children's Hospitals and Clinics of Minnesota, Minneapolis, Minnesota, USA. [19]Department of Pediatrics, Texas Children's Hospital, Baylor College of Medicine, Houston, Texas, USA. [20]Pediatric Hematology Oncology, Pediatric Specialists of Lehigh Valley Hospital, Bethlehem, Pennsylvania, USA. [21]Pediatric Bone Marrow Transplant Program, Oregon Health and Science University, Portland, Oregon, USA. [22]Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [23]Division of Pediatric Oncology, Children's National Medical Center, Washington, DC, USA. [24]Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA. [25]Department of Pediatrics, Georgetown University, Washington, DC, USA. [26]Department of Oncology, Georgetown University, Washington, DC, USA. [27]Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California, USA. [28]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA. [29]Department of Pathology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [30]Department of Statistics, Children's Oncology Group, Monrovia, California, USA. [31]Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [32]Harvard Medical School, Boston, Massachusetts, USA. [33]Department of Pathology and Cancer Center, Massachusetts General Hospital, Boston, Massachusetts, USA. [34]Department of Oncology, St. Jude Children's Research Hospital, Memphis, Tennessee, USA. [35]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [36]Division of Hematology/Oncology, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [37]Department of Pediatrics, Benioff Children's Hospital, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California, USA. [38]These authors contributed equally to this work. Correspondence should be addressed to E.S. (elliot.stieglitz@ucsf.edu) or M.L.L. (mignon.loh@ucsf.edu).

Received 22 February; accepted 17 August; published online 12 October 2015; doi:10.1038/ng.3400

targets for the development of novel therapies in patients with the most aggressive disease phenotypes.

Mutations in *NF1*, *NRAS*, *KRAS*, *PTPN11* and *CBL* ('RAS pathway' mutations) currently allow for a molecular diagnosis in 85% of patients[7–11]. Recently, secondary mutations in *SETBP1* and *JAK3* were identified by whole-exome sequencing in a small number of patients with JMML at diagnosis[12]. We subsequently identified several patients who had increased allele frequency for *SETBP1* mutations at relapse. We then used droplet digital PCR (ddPCR) to show that subclonal *SETBP1* mutations were present in nearly one-third of patients with JMML at diagnosis and independently predicted relapse[13]. These findings indicated a level of genetic complexity previously unrecognized in JMML, and, given the limited numbers of patients with non-syndromic, *de novo* JMML who have had exome sequencing performed, we set out to assess the genomic landscape of JMML. We sequenced samples from patients (n = 29) with matched tumor-normal pairs. Seven of these patients also had serially acquired samples from relapse and/or transformation to acute myeloid leukemia (AML) available for sequencing. We then validated our findings in an independent cohort of 71 patients (**Supplementary Fig. 1**), of whom nine had paired diagnostic-relapse samples available. Two of the 29 patients who had exome sequencing data were suspected of having Noonan syndrome. Upon confirmation of this diagnosis, these patients were removed from all outcome analyses, which were specific to patients with somatically mutated JMML.

RESULTS

Sequencing of JMML samples using optimized algorithms

We performed whole-exome sequencing at a mean coverage of 95× (**Supplementary Table 1**) for 22 patients with paired germline-diagnosis samples and an additional seven patients with germline-diagnosis-relapse samples (**Fig. 1**). Because of the frequent contribution of germline mutations in the development of JMML[7,11], we optimized an algorithm to detect tumor in normal content (deTiN) to retrieve mutations that would otherwise have been missed using a traditional tumor-normal bioinformatics approach. Four tissue types of germline material were used as normal controls, including buccal cells, cord blood, Epstein-Barr virus (EBV)-immortalized lymphoblasts and fibroblasts. However, by comparing several within-patient germline sources that contained varying degrees of tumor content, it became evident that each tissue type had different amounts of tumor contamination in the normal sample. For example, in patient UPN2026, we first detected a heterozygous mutation in *RRAS2* in DNA from a buccal swab, but repeat sequencing of EBV-immortalized B cells showed wild-type sequence (**Supplementary Fig. 2**). We therefore implemented deTiN to both assess and correct for the purity of each germline source.

In total, we identified ten genes that were mutated beyond the five RAS pathway genes previously documented to have lesions (**Supplementary Table 2**). The mutations occur in known oncogenes and tumor suppressors that broadly fall into categories of genes affecting the RAS pathway, signal transduction, transcription factors, epigenetic regulation and the spliceosome complex (**Fig. 1** and **Table 1**).

We next carried out targeted deep sequencing (mean coverage of 1,380×) on 71 diagnostic tumors as well as paired diagnostic-relapse samples from nine patients for the 15 genes identified by exome sequencing to determine the frequency of mutations in these genes. In addition, we sequenced *JAK3*, as previous reports had identified recurrent subclonal mutations in this gene in some patients. We also included 11 patients from the discovery cohort in the resequencing cohort for validation purposes. All pathogenic mutations identified by exome sequencing were identified by targeted resequencing. Recurrent mutations were detected in 14 of the 15 genes, excluding *RRAS2*, for which mutations were only identified in the discovery cohort (**Supplementary Table 3**). The incidence of mutations in these genes is presented for a total of 100 patients, of whom 29 were in the discovery cohort and 71 were in the validation cohort (**Supplementary Fig. 3**). RNA sequencing (RNA-seq) data were available for 14 patients (15 samples) and confirmed that each of the pathogenic variants detected in exome and deep sequencing was transcribed and present in the RNA (**Supplementary Table 4**).

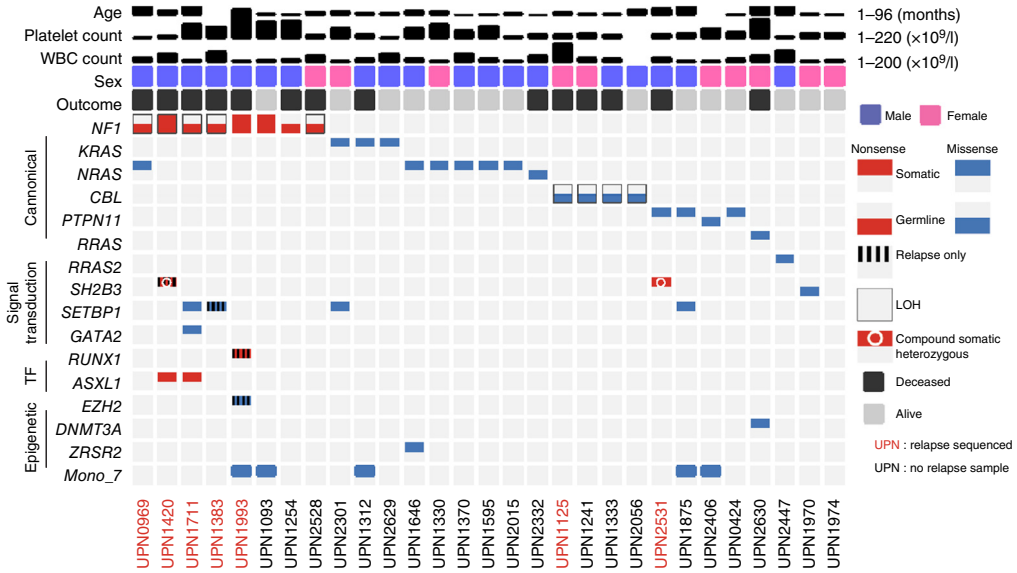

Figure 1 Mutations identified by exome sequencing. Data for 29 patients who were subjected to whole-exome sequencing are displayed. The results for each patient are presented in a single condensed column, including mutations identified in the germline and at diagnosis (black type) and relapse (red type). Germline mutations are presented by color in the bottom half of the box for any given gene, and somatic mutations are presented in the top half. Mutations only present at relapse are denoted with vertically striped bars. LOH for a single gene is annotated with a thin black rectangle outlining the mutation. Somatic compound heterozygous mutations are noted with a white circle. WBC, white blood cell; TF, transcription factor.

Table 1 Mutations detected by whole-exome sequencing excluding known RAS pathway genes

Gene	Mutation type	RefSeq accession	Amino acid change	Nucleotide change	UPN ID
ASXL1	Nonsense	NM_015338	p.Tyr591*	c.1773C>G	UPN1420
ASXL1	Frameshift	NM_015338	p.His630fs	c.1888_1910del	UPN1711
DNMT3A	Missense	NM_022552	p.Arg882Cys	c.2644C>T	UPN2630
EZH2	Missense	NM_001203247	p.Val674Leu	c.2020G>C	UPN1993
GATA2	Missense	NM_001145661	p.Asn317Ser	c.950A>G	UPN1711
RRAS	Missense	NM_006270	p.Gln87Leu	c.260A>T	UPN2630
RRAS2	Missense	NM_012250	p.Gln72Leu	c.215A>T	UPN2447
RUNX1	Frameshift	NM_001001890	p.Arg349fs	c.1047_1048insC	UPN1993
SETBP1	Missense	NM_015559	p.Gly870Ser	c.2608G>A	UPN1711
SETBP1	Missense	NM_015559	p.Asp868Asn	c.2602G>A	UPN2301
SETBP1	Missense	NM_015559	p.Ile871Thr	c.2612T>C	UPN1875
SETBP1	Missense	NM_015559	p.Gly870Ser	c.2608G>A	UPN1383
SH2B3	Frameshift	NM_005475	p.Phe390fs	c.1170delC	UPN1420
SH2B3	Nonsense	NM_005475	p.Gln258*	c.772C>T	UPN1420
SH2B3	Nonsense	NM_005475	p.Trp262*	c.785G>A	UPN2531
SH2B3	Splice site	NM_005475	p.His414_splice	c.1240_C>T splice	UPN2531
SH2B3	Missense	NM_005475	p.Glu400Lys	c.1198G>A	UPN1970
ZRSR2	Missense	NM_005089	p.Gly179Glu	c.536G>A	UPN1646

Somatic copy number alterations

The most common karyotypic abnormality identified in whole-exome sequencing was deletion of one copy of chromosome 7, which occurred in five of the exome-sequenced cases (**Supplementary Fig. 4**). One patient with t(11;17) detected by diagnostic cytogenetics had additional evidence of disruption of the *NF1* locus in FISH analysis (data not shown) and was found to have a germline mutation in this gene by whole-exome sequencing. The *NF1* allele harboring the c.2041C>T mutation had a fractional abundance of 33% in RNA transcriptome sequencing (15/45 reads). However, RNA-seq did not identify a fusion transcript in this patient, suggesting that the translocation breakpoints might be intergenic. RNA-seq in 14 additional patients did not identify any structural rearrangements resulting in chimeric transcripts. Consistent with previous reports[14,15], the only additional recurrent structural variants were copy-neutral isodisomy at 11q23.3 and 17q11.2 where *CBL* and *NF1* are located, respectively (**Supplementary Fig. 5**).

Expanding the spectrum of RAS pathway mutations

In addition to canonical *NF1*, *KRAS*, *NRAS*, *PTPN11* and *CBL* alterations, we identified a previously described *RRAS* mutation in one patient[16], as well as an *RRAS2* alteration in another patient, neither of whom harbored other driver mutations at diagnosis. Mutations in *RRAS2*, a member of the RAS GTPase superfamily encoding the TC21 oncoprotein, have previously been limited to solid malignancies, including ovarian, endometrial and squamous cell carcinomas[17], and regulate transforming growth factor (TGF)-β signaling through loss of *NF1* (ref. 18). The *RRAS2* point mutation in one of our patients results in the p.Gln72Leu amino acid substitution that was previously shown to confer transformative properties in ovarian tissue, as evidenced by increased colony formation *in vitro* and increased tumorgenicity after inoculation in mice as compared to wild-type *RRAS2* (ref. 19).

Although RAS pathway lesions have traditionally been thought to represent largely mutually exclusive events[20], coexisting mutations in *NRAS*, *KRAS*, *PTPN11*, *CBL* and *NF1* were found in 11 of 100 (11.0%) patients (**Fig. 2**). Analysis of single colonies for two patients with compound RAS pathway mutations demonstrated that both mutations occurred in the same colony (**Supplementary Table 5**). *PTPN11* and *NF1* lesions were the most frequent of these cooperative events. In addition, one patient harbored two *NRAS* lesions (encoding p.Gly13Asp and p.Gln61Lys substitutions).

NF1 mutations in patients without clinical NF1

Two patients in our discovery cohort who were not suspected of having neurofibromatosis type 1 (NF1) nevertheless harbored germline *NF1* mutations. One of these patients was heterozygous for the mutation in both the tumor and matched normal (buccal cell) samples, with no LOH noted in the tumor. Another patient had an *NF1* mutation with an allelic fraction of 10.1% in the germline tissue sample (buccal cells). The allelic fraction in the tumor increased to 45.3% after a copy-neutral duplication at 17q. Although an event with low allelic fraction in germline tissue could be explained as a somatic-only variant in tumor contamination of the normal sample, application of the deTIN algorithm estimated that no such contamination occurred (95% confidence interval (CI) for contamination = 0–3.35%), on the basis of absence of the copy number alteration in the normal tissue sample. This finding raises the possibility of somatic mosaicism, which occurs in clinical NF1 and often presents with subtle clinical features[21]. In our validation cohort, an additional six patients harbored *NF1* mutations, but only three were suspected of having clinical NF1. It is not possible to determine the somatic versus germline origin of these lesions, as germline tissue was not available for these patients. These observations are consistent with a previous study indicating that JMML can be the first manifestation of NF1 in some affected infants and young children[22].

Activated JAK-STAT corresponds with poor clinical outcome

We previously found that increased STAT5 phosphorylation in a subset of myeloid cells is a general feature of JMML[23]. The tumor suppressor LNK is encoded by *SH2B3* and negatively regulates JAK-STAT signaling[24]. Mutations in *SH2B3* were first reported in a variety of adult MPNs, including primary myelofibrosis and essential thrombocythemia, as well as in isolated erythrocytosis[25–27]. The alterations included nonsense and missense mutations affecting the pleckstrin homology (PH) domain and/or the Src homology 2 (SH2) domain. Mutations in *SH2B3* have also been identified in lymphoid malignancies, including acute lymphoblastic leukemia, as both germline and somatic events[28]. Two patients in the discovery cohort possessed compound heterozygous mutations in *SH2B3* in addition to a known JMML driver mutation. Each patient had a nonsense mutation mapping to both the PH domain and the SH2 domain (**Fig. 3a**). Whereas one patient harbored these mutations at diagnosis, the other patient only had detectable *SH2B3* mutations upon transformation to AML. The latter patient relapsed with AML 3 years after his initial diagnosis. Protein lysates collected from unsorted primary mononuclear cells from these patients showed a decrease in LNK amount by immunoblot, commensurate with the allelic fraction of the patients' nonsense mutations (**Fig. 3b**). One additional patient without any other identifiable mutations was found via exome sequencing to harbor a germline *SH2B3* mutation encoding p.Glu400Lys, corresponding to a nonsynonymous SNP that has previously been associated with idiopathic erythrocytosis[29].

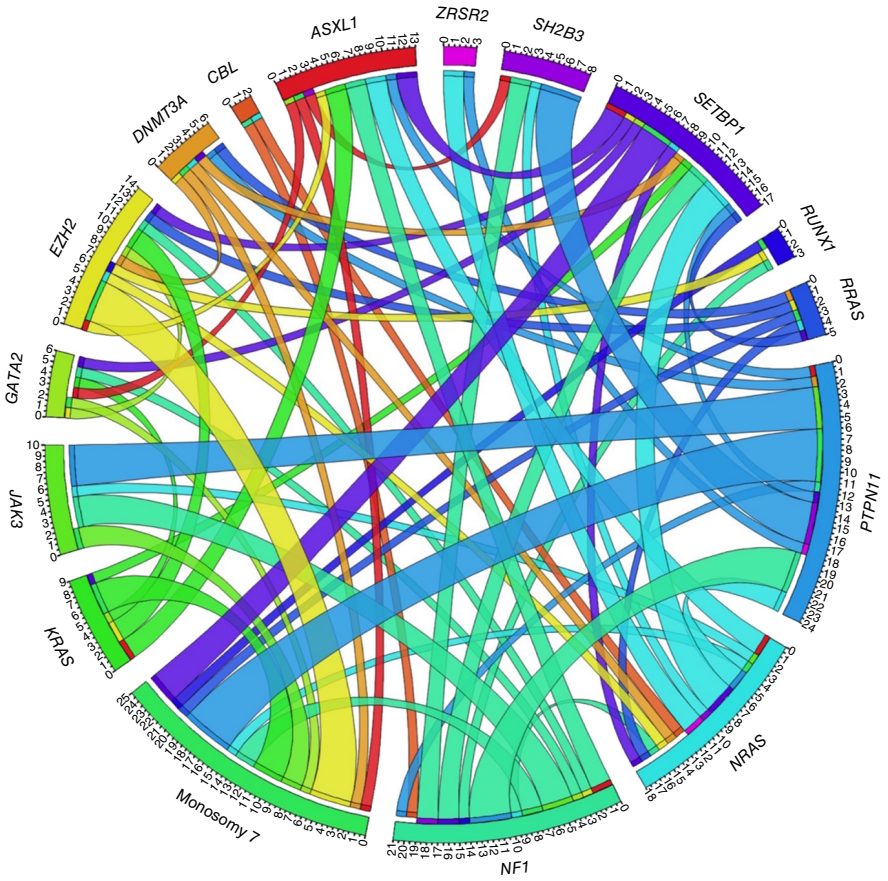

Figure 2 Circos plot of samples with at least two mutations. Using data from whole-exome sequencing and targeted resequencing, patients with at least two mutations are depicted. Associations between genomic alterations in the same patient are marked by connecting bands, with the width of each band proportional to the frequency of the association.

In the validation cohort, an additional four patients were found to harbor *SH2B3* mutations, for a cumulative incidence of 7.0% (7/100) (**Supplementary Table 3**).

We did not identify any activating *JAK3* mutations in our exome discovery cohort, but we performed resequencing with improved coverage (680×) in our validation cohort, given previous reports[12]. These studies uncovered *JAK3* mutations in six of 83 cases (7.3%). Of note, three of the four patients with *JAK3* mutations at diagnosis went on to relapse. Two other patients had lesions detected only at relapse (with allelic fractions of 8.6% and 37.4%) and presumably had subclonal mutations at diagnosis below the threshold of detection by our exome sequencing.

SETBP1 mutations are frequently subclonal at diagnosis

We detected a similar incidence of *SETBP1* mutations (7/100; 7.0%) using deep sequencing as was previously reported by Sakaguchi *et al.*[12]. An additional 17 patients were found to harbor subclonal *SETBP1* mutations using ddPCR[13]. In combination, subclonal and clonal *SETBP1* mutations occurred in 24 of the 100 patients (**Supplementary Table 6**). No mutations were found by exome sequencing in *ETNK1*, a gene that was recently reported to have mutations co-occurring with ones in *SETBP1* in atypical chronic myelogenous leukemia (aCML)[30].

Transcription factor mutations in JMML

We identified mutations in *GATA2* (ref. 31), a transcription factor broadly involved in hematopoiesis. Recent work has shown that germline *GATA2* mutations are responsible for several syndromes, leading to a predisposition to myeloid malignancies and opportunistic infections[32–34]. Two patients harbored three heterozygous mutations, all of which affected the regulatory zinc-finger (ZF1) domain. The overwhelming majority of mutations occurring in previous reports affected either the ZF1 or ZF2 domain[34–36]. One of our two patients harbored both a germline and a somatic *GATA2* mutation, as previously described[37]. The other patient was found to have a somatic variant encoding p.Asn317Ser on one allele, which has previously been detected in patients with both myelodysplastic syndrome (MDS) and AML[38,39].

A heterozygous nonsense *RUNX1* mutation was also found in one patient at diagnosis and in another patient at relapse, representing the first reports, to our knowledge, of haploinsufficient *RUNX1* mutations in JMML[40]. Haploinsufficient missense mutations in *RUNX1* have previously been reported in other malignancies such as MDS and AML and are predicted to be inactivating[41].

Wiskott-Aldrich syndrome mutations in JMML

Previous reports have documented germline *WAS* mutations in patients suspected of having JMML[42,43]. Similarly, we detected one male patient with a germline *WAS* mutation encoding p.His180Asn that has been described in several patients with Wiskott-Aldrich syndrome[44]. Interestingly, this patient also had a somatic *NRAS* mutation encoding p.Gly12Ser. To the best of our knowledge, there are no previous reports in the literature of patients with concomitant germline mutations in *WAS* and somatic mutations in RAS pathway genes.

Mutations in epigenetic modifiers are frequent in JMML

In contrast to previous reports that emphasized the rarity of genetic mutations in epigenetic modifier genes in JMML[45–47], we identified

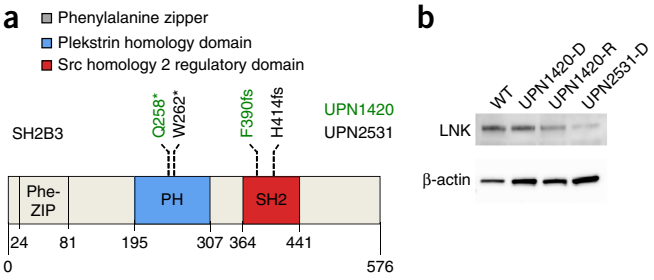

Figure 3 Mutations in *SH2B3* decrease expression of LNK. (**a**) Compound heterozygous mutations mapping to the PH and SH2 domains are presented for each patient found to harbor *SH2B3* lesions by whole-exome sequencing. (**b**) Immunoblot analysis of whole-cell lysates using antibodies to LNK and β-actin. Commensurate with the allelic fraction of each mutation (UPN1420-relapse, 31%; UPN2531-diagnosis, 37%), the expression of LNK is decreased relative to a healthy contol (WT).

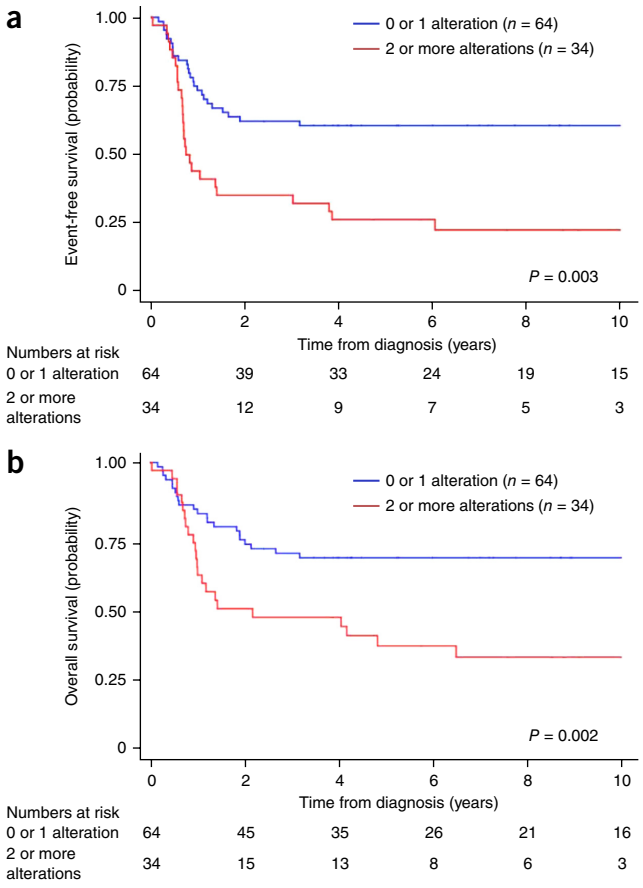

Figure 4 Event-free and overall survival of patients stratified by number of somatic alterations. Kaplan-Meier estimated event-free survival (log-rank $P = 0.003$) (**a**) and overall survival (log-rank $P = 0.002$) (**b**) rates are shown according to the number of somatic alterations at diagnosis.

such mutations in 14 of the 100 patients, for a combined incidence in our cohort of 14.0%. Components of PRC2, including *EZH2* (4/100; 4.0%) and *ASXL1* (8/100; 8.0%), were mutated at diagnosis, and tumor cells with these mutations were clonally expanded in several patients at relapse (**Supplementary Tables 2** and **3**). All four patients with mutations in *EZH2* (7q36.1) also had monosomy 7 (**Fig. 2**). We also found *DNMT3A* mutations in three patients, including one mutation that affected the previously reported Arg882 hotspot and two other frameshift mutations. Genome-wide DNA methylation analysis demonstrated that four of the five patients with *ASXL1* mutations had a globally hypermethylated profile in comparison to patients with wild-type *ASXL1* (**Supplementary Fig. 6**), whereas one patient with a *DNMT3A* mutation had a globally hypomethylated profile, consistent with previous reports[48,49].

The spliceosome complex is implicated in JMML

A member of the spliceosome complex, *ZRSR2*, was mutated in one patient in our discovery cohort as well as in two additional patients upon targeted resequencing. A pairwise analysis was carried out comparing RNA-seq data from one patient with a *ZRSR2* mutation and 14 patients wild type for *ZRSR2*. The patient with mutant *ZRSR2* had increased retention of U12-type introns in comparison to those with wild-type *ZRSR2*, consistent with previous reports in adults with MDS[50] (**Supplementary Fig. 7**). To our knowledge, these are the first reported *ZRSR2* mutations in JMML[47,51,52]. No mutations in *SF3B1*, *U2AF1* or *SRSF2* were found in our discovery phase.

Phylogenetic evolution at relapse

Focusing on patients with both diagnostic and relapse material allowed us to plot the acquisition of mutations over time (**Supplementary Fig. 8** and **Supplementary Tables 2** and **3**). All of the pathogenic mutations at diagnosis were present at relapse. However, it should be noted that, owing to the low number of mutations per sample (0.468 mutations/Mb), these analyses lacked the power to definitively track clones over time. To further investigate the clonality of the mutations at diagnosis, we analyzed individual colonies obtained from diagnostic colony-forming unit (CFU) granulocyte-macrophage assays (**Supplementary Table 7**). Patients UPN1420 and UPN2531 were particularly interesting because of the pattern of *SH2B3* mutations detected. UPN2531 harbored a *PTPN11* driver mutation along with compound heterozygous mutations in *SH2B3* at diagnosis. By contrast, UPN1420 only acquired mutations in *SH2B3* upon transformation to AML (**Supplementary Fig. 9a**). Colonies from UPN1420 were homozygous for either the p.Gln258* (19/24) or p.Phe390fs (5/24) alteration. In a model of convergent evolution, these clones were likely the result of copy number alterations occurring at 12q in individual cells that contained heterozygous *SH2B3* mutations. Copy number variation analysis of this sample at relapse showed distinct duplication and loss at 12q where *SH2B3* resides (**Supplementary Fig. 9b**). We hypothesize that the allele with the mutation encoding p.Gln258* was duplicated in one clone whereas the mutation encoding p.Phe390fs achieved homozygosity in the other.

Clinical and biological features at diagnosis

A total of 98 of the 100 children with JMML diagnosed between 2001 and 2013 at North American institutions were included in our analysis, with two patients excluded because of insufficient follow-up data (**Supplementary Table 8**). Zero or one somatic alteration (pathogenic mutation or monosomy 7) was identified in 64 (65.3%) patients at diagnosis, whereas two or more alterations were identified in 34 (34.7%) patients (mean number of alterations = 1.32, range = 0–4). Several clinical and biological characteristics showed statistically significant differences when compared between these groups (**Supplementary Table 9**). A higher proportion of patients diagnosed with two or more alterations were older ($P < 0.001$) and male ($P = 0.001$). These patients also demonstrated a higher rate of monosomy 7 ($P < 0.001$) or somatic *NF1* mutation ($P < 0.001$).

Somatic alterations at diagnosis predict outcome

Using the number of somatic events at diagnosis, we evaluated differences in event-free survival and overall survival. We initially identified a trend toward improved outcome in patients with zero or one somatic alteration in comparison to those with two or more alterations in our 27-patient exome cohort ($P = 0.12$) (**Supplementary Fig. 10a**). This trend was later validated in our independent cohort of 71 patients ($P = 0.0004$) (**Supplementary Fig. 10b**). Combining both cohorts (98 patients in total), patients with zero or one somatic alteration at diagnosis had improved outcomes in comparison to patients who had two or more alterations (event-free survival hazard ratio (HR) = 2.09, $P = 0.009$; overall survival HR = 2.02, $P = 0.018$). The 10-year event-free survival and overall survival rates according to number of somatic events at diagnosis were as follows (± standard error): zero or one alteration, 60.6 ± 6.2% and 65.1 ± 6.0% and two or more alterations, 22.7 ± 7.4% and 29.0 ± 8.3% ($P = 0.003$ and 0.002, respectively; **Fig. 4a**,**b**). Log-rank tests detected statistically significant differences in times to both outcomes on the basis of this predictor variable. Differences in event-free and overall survival for groups stratified by canonical RAS pathway driver alteration, including ones in *PTPN11*, *NF1*,

Table 2 Multivariate analysis of prognostic variables in JMML

	Event-free survival from date of diagnosis		Overall survival from date of diagnosis	
Variable	HR (95% CI)	*P*	HR (95% CI)	*P*
Somatic alterations at diagnosis				
0 or 1	1		1	
2 or more	2.65 (1.27–5.54)	**0.01**	3.13 (1.39–7.01)	**0.006**
Age at diagnosis				
<24 months	1		1	
>24 months	1.62 (0.82–3.21)	0.167	1.55 (0.73–3.28)	0.25
Platelet count at diagnosis ($\times 10^9$/l)				
≥40	1		1	
<40	1.64 (0.86–3.08)	0.128	2.19 (1.09–4.39)	**0.028**
Fetal hemoglobin at diagnosis				
Not elevated for age	1		1	
Elevated for age	1.56 (0.81–3.02)	0.184	0.93 (0.69–2.80)	0.355
NF1 status				
No	1		1	
Yes	0.85 (0.38–1.93)	0.7	0.5 (0.19–1.29)	0.15

P values <0.05 are shown in bold.

NRAS, *KRAS* and *CBL*, were not statistically significant in our cohort ($P = 0.834$) (**Supplementary Fig. 11**). Of note, event-free survival for patients with two RAS pathway mutations (14%; $n = 7$) in our cohort was significantly inferior to that for patients with a single RAS pathway mutation (62%; $n = 55$) ($P = 0.017$). This difference in outcome implies a dosage effect based on the number of RAS mutations and suggests that each mutation in a pair of compound RAS mutations has biological activity.

Previous studies have assigned poorer prognostic significance to several clinical and laboratory characteristics in patients with JMML, including older age, male sex, lower platelet count, higher fetal hemoglobin concentration, monosomy 7, *PTPN11* mutation and subclonal *SETBP1* mutation. In univariate Cox analyses (**Supplementary Table 10**), the characteristics that reached significance for event-free and overall survival were age of ≥24 months (HR = 2.09, CI = 1.20–3.62, $P = 0.009$ and HR = 2.03, CI = 1.13–3.66, $P = 0.018$) and monosomy 7 (HR = 1.24, CI = 1.01–1.52, $P = 0.040$ and HR = 1.29, CI = 1.04–1.60, $P = 0.019$). Notably, the most statistically significant covariate in this analysis was the number of somatic alterations at diagnosis (event-free survival HR = 2.65, CI = 1.27–5.54, $P = 0.01$ and overall survival HR = 3.13, CI = 1.39–7.01, $P = 0.006$). Furthermore, when a Cox multivariate regression model was applied, only the number of somatic alterations at diagnosis retained statistical significance for both event-free and overall survival. Two or more somatic events remained independently prognostic of poor outcome (HR = 2.86, CI = 1.35–6.05, $P = 0.0006$ and HR = 3.00, CI = 1.27–8.40, $P = 0.0008$) after adjusting for age, platelet count, fetal hemoglobin concentration and NF1 status at diagnosis (**Table 2**).

We assessed the relapse incidence in patients who were treated with HSCT. Of the 79 patients in our cohort who received HSCT and had follow-up data, 49 (62%) harbored zero or one alteration at diagnosis in comparison to 30 (38%) who harbored two or more alterations. As expected, the cumulative incidence of relapse from the time of HSCT was significantly higher in the cohort with two or more events at diagnosis (77.7 ± 8.56%) than in the cohort with zero or one event (38.4 ± 8.33%; log-rank $P = 0.0002$) (**Supplementary Fig. 12**).

DISCUSSION

Although previous genomic and biochemical analyses in JMML have emphasized dependence on the RAS pathway for initiation of disease[12], we hypothesized that additional pathways might be important for both disease initiation and progression. By studying patients with progressive disease and by optimizing the use of germline controls, the mutational spectrum in JMML is now expanded to include some molecules with 'off-the-shelf' therapeutic implications, whereas others will require further investigation. In addition to identifying another member in the RAS family of oncogenes mutated in JMML (*RRAS2*), we also uncovered additional genetic pathways contributing to leukemogenesis, including upregulated JAK-STAT signaling and epigenetic modification through the PRC2 and spliceosome complexes.

This report is the first to our knowledge to identify *SH2B3* mutations in JMML. In combination with a previous study that uncovered missense *JAK3* alterations[12], these data raise the possibility of using JAK inhibitor therapy in the subset of patients harboring these mutations. This is also the first report of *ZRSR2* mutations in this disease.

In contrast to chronic myelomonocytic leukemia (CMML), where epigenetic mutations are frequent[53], previous reports in JMML have focused on the rarity of these events. However, 14% of patients in our cohort harbored genetic alterations in epigenetic modifier genes. We show that *ASXL1* mutations result in globally hypermethylated profiles. Patients exhibiting aberrant methylation would be predicted to respond to DNA-hypomethylating agents, and previous case reports support the notion that these agents have clinical usefulness in JMML[54]. Currently, a phase I/II trial using 5-azacytidine in children with JMML is enrolling patients in Europe. Mutations in *DNMT3A* typically cause decreased methyltransferase activity and result in focal areas of CpG hypomethylation[55].

Obtaining germline tissue in children was of critical importance in our study because approximately 25% of patients had inherited syndromes that predispose to the development of JMML. In addition, unlike in other hematological malignancies, where one can use remission bone marrow samples for germline tissue, patients with JMML rarely respond to traditional chemotherapy before transplantation[56]. By developing innovative algorithms to identify tumor in normal (TiN) content with retrieval of minor differences in tumor-normal comparisons, we identified several variants that would have otherwise been overlooked using standard approaches. We showed that sources such as buccal swabs, cord blood and EBV-immortalized lymphocytes were all suboptimal in comparison to bone marrow fibroblasts. Although buccal samples are frequently contaminated with tumor by infiltrating leukemic cells, cord blood and EBV-immortalized samples provided biological insights into disease.

In the majority of cord blood samples, the canonical mutation was retrievable, reflecting the *in utero* nature of leukemic initiation[57,58]. Of note, the mean age at diagnosis of the five patients who had cord blood available was 6.6 months. Interestingly, no secondary mutations were identified at diagnosis in any of these children, which may explain why younger age at diagnosis has traditionally been a more favorable clinical risk factor for outcome. B cells immortalized with EBV also provide insights into the range of the hematopoietic lineages affected by JMML mutations, with many patients appearing to have clonal involvement of mutations in the B cell compartment, whereas others

only had myeloid cell involvement. There were insufficient patients available to determine how frequently such involvement occurs.

Several mutated genes new to JMML were identified in relapse samples, including *RUNX1* and *JAK3*. However, in contrast to most malignancies, where a branching model of disease evolution with mutation acquisition and dropout is common[59], JMML appears to have a unique model of progression. Specifically, mutations identified at diagnosis invariably persisted at relapse in all 16 patients analyzed over time. In addition, the cancer cell fraction for mutations identified at diagnosis by exome sequencing approached 100%, implying that all mutations occurred within a single dominant clone. Single-colony analysis in three patients with available cryopreserved cells was consistent with a linear model of acquisition for additional genetic features (**Supplementary Fig. 9** and **Supplementary Tables 5** and **7**). When developing new therapeutic strategies for these patients, it will be critical to understand how secondary mutations alter the behavior of cells in contrast to cells harboring only the primary lesion. Indeed, combinatorial therapies with agents that target the RAS pathway as well as the secondary genetic event may be necessary, as our work demonstrates that the clone harboring the secondary event frequently expands at disease progression. For example, combining a MEK inhibitor with DNA-hypomethylating agents could prove more efficacious in the correct genetic context than monotherapy with a MEK inhibitor[60,61].

Previous reports have reached different conclusions regarding the prognostic relevance of canonical RAS pathway mutations in JMML[5,6]. We show that mutations in the five RAS pathway genes are not independently prognostic of outcome, precluding their sole use in a risk stratification algorithm. Consistent with our hypothesis, we found that the number of mutations at diagnosis, rather than the type of mutations, possesses prognostic relevance.

In conclusion, we demonstrate that patients with JMML who harbor two or more somatic alterations at diagnosis have significantly worse event-free and overall survival rates than those with one or no event. In previous reports that identified older age, male sex and lower platelet count as markers of poor outcome, these clinical features were likely descriptors of patients with two or more underlying somatic alterations. Notably, in addition to identifying patients with aggressive disease, our expanded mutational spectrum will serve as a platform for studying the interactions of these secondary events in the context of hyperactive RAS–mitogen-activated protein kinase (MAPK) signaling. These studies will allow further interrogation of multiple molecular targets that can be exploited in patients predicted to have poor outcomes using conventional HSCT.

URLs. Exome Aggregation Consortium (ExAC), http://exac.broadinstitute.org/; DesignStudio, http://designstudio.illumina.com/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. High-throughput sequencing data have been deposited in the database of Genotypes and Phenotypes (dbGaP) under accession phs000973.v1.p1.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors thank the patients and their families for participating in this research, without whom this work would not have been possible. In addition, A. Ikeda, E. Raetz, N. Bunin and J. Finkelstein all assisted in this research by providing invaluable patient samples.

This work was supported by the Carlos Slim Foundation in Mexico as part of the Slim Initiative for Genomic Medicine; the St. Baldrick's Foundation (E.S.); Alex's Lemonade Stand Foundation (E.S.); the Leukemia and Lymphoma Society (grants 6059-09 and 6466-15) (M.L.L.) and a Leukemia and Lymphoma Society Scholar award (K.S.)); US National Institutes of Health, National Cancer Institute grants T32CA128583 (E.S.) and R01CA173085 (M.L.L.), National Cancer Institute Cancer Center Support grant 5P30CA082103, Children's Oncology Group (COG) Statistics and Data Center grant 1U10CA180899 (T.A.A.), COG Chair's Grant 5U10CA098543 (T.A.A.) and US National Institutes of Health grant P30CA82103 (A.B.O.); the Frank A. Campini Foundation (E.S. and M.L.L.); Hyundai Hope on Wheels (M.L.L.); and a University of California San Francisco Dean's Commitment to the Center for Advanced Technologies facility.

AUTHOR CONTRIBUTIONS

E.S., L.C.G., T.M., E.E., A.Y., K.B. and S.L.A. performed the experiments. E.S., A.N.T.-W., Y.-D.W., T.M., M.R., A.B.O., Y.L., J.M., R.B.G. and T.A.A. performed data analysis. G.A., M.B., P.A.B., B.C., T.C., G.V.D., P.D.E., M.N.F., R.K.G., R.J.H., J.H., C.H., Y.L.L., Y.H.M., D.H.M., P.M., E.R.N., P.A.R., R.J.S., K.C.S., C.M.T. and J.A.T. contributed reagents, materials and analysis tools. E.S., A.N.T.-W. and M.L.L. wrote the first draft of the manuscript. T.Y.C. performed statistical analysis. J.F.C., C.S., G.G., T.A.G., T.R.G., K.S. and M.L.L. supervised research. S.S. and S.R.O. managed the project. All coauthors contributed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Chang, T.Y., Dvorak, C.C. & Loh, M.L. Bedside to bench in juvenile myelomonocytic leukemia: insights into leukemogenesis from a rare pediatric leukemia. *Blood* **124**, 2487–2497 (2014).
2. Dvorak, C.C. & Loh, M.L. Juvenile myelomonocytic leukemia: molecular pathogenesis informs current approaches to therapy and hematopoietic cell transplantation. *Front. Pediatr.* **2**, 25 (2014).
3. Matsuda, K. *et al.* Spontaneous improvement of hematologic abnormalities in patients having juvenile myelomonocytic leukemia with specific *RAS* mutations. *Blood* **109**, 5477–5480 (2007).
4. Matsuda, K. *et al.* Long-term survival after nonintensive chemotherapy in some juvenile myelomonocytic leukemia patients with *CBL* mutations, and the possible presence of healthy persons with the mutations. *Blood* **115**, 5429–5431 (2010).
5. Bresolin, S. *et al.* Gene expression–based classification as an independent predictor of clinical outcome in juvenile myelomonocytic leukemia. *J. Clin. Oncol.* **28**, 1919–1927 (2010).
6. Flotho, C. *et al.* Genotype-phenotype correlation in cases of juvenile myelomonocytic leukemia with clonal *RAS* mutations. *Blood* **111**, 966–967, author reply 967–968 (2008).
7. Side, L. *et al.* Homozygous inactivation of the *NF1* gene in bone marrow cells from children with neurofibromatosis type 1 and malignant myeloid disorders. *N. Engl. J. Med.* **336**, 1713–1720 (1997).
8. Flotho, C. *et al. RAS* mutations and clonality analysis in children with juvenile myelomonocytic leukemia (JMML). *Leukemia* **13**, 32–37 (1999).
9. Tartaglia, M. *et al.* Somatic mutations in *PTPN11* in juvenile myelomonocytic leukemia, myelodysplastic syndromes and acute myeloid leukemia. *Nat. Genet.* **34**, 148–150 (2003).
10. Loh, M.L. *et al.* Mutations in *PTPN11* implicate the SHP-2 phosphatase in leukemogenesis. *Blood* **103**, 2325–2331 (2004).
11. Loh, M.L. *et al.* Mutations in *CBL* occur frequently in juvenile myelomonocytic leukemia. *Blood* **114**, 1859–1863 (2009).
12. Sakaguchi, H. *et al.* Exome sequencing identifies secondary mutations of *SETBP1* and *JAK3* in juvenile myelomonocytic leukemia. *Nat. Genet.* **45**, 937–941 (2013).
13. Stieglitz, E. *et al.* Subclonal mutations in *SETBP1* confer a poor prognosis in juvenile myelomonocytic leukemia. *Blood* **125**, 516–524 (2015).
14. Niemeyer, C.M. *et al.* Germline *CBL* mutations cause developmental abnormalities and predispose to juvenile myelomonocytic leukemia. *Nat. Genet.* **42**, 794–800 (2010).
15. Stephens, K. *et al.* Interstitial uniparental isodisomy at clustered breakpoint intervals is a frequent mechanism of *NF1* inactivation in myeloid malignancies. *Blood* **108**, 1684–1689 (2006).
16. Flex, E. *et al.* Activating mutations in *RRAS* underlie a phenotype within the RASopathy spectrum and contribute to leukaemogenesis. *Hum. Mol. Genet.* **23**, 4315–4327 (2014).
17. Graham, S.M. *et al.* TC21 and Ras share indistinguishable transforming and differentiating activities. *Oncogene* **18**, 2107–2116 (1999).
18. Patmore, D.M. *et al. In vivo* regulation of TGF-β by R-Ras2 revealed through loss of the RasGAP protein NF1. *Cancer Res.* **72**, 5317–5327 (2012).

19. Chan, A.M., Miki, T., Meyers, K.A. & Aaronson, S.A. A human oncogene of the *RAS* superfamily unmasked by expression cDNA cloning. *Proc. Natl. Acad. Sci. USA* **91**, 7558–7562 (1994).
20. Locatelli, F. & Niemeyer, C.M. How I treat juvenile myelomonocytic leukemia (JMML). *Blood* **125**, 1083–1090 (2015).
21. Tinschert, S. *et al.* Segmental neurofibromatosis is caused by somatic mutation of the neurofibromatosis type 1 (*NF1*) gene. *Eur. J. Hum. Genet.* **8**, 455–459 (2000).
22. Side, L.E. *et al.* Mutations of the *NF1* gene in children with juvenile myelomonocytic leukemia without clinical evidence of neurofibromatosis, type 1. *Blood* **92**, 267–272 (1998).
23. Kotecha, N. *et al.* Single-cell profiling identifies aberrant STAT5 activation in myeloid malignancies with specific clinical and biologic correlates. *Cancer Cell* **14**, 335–343 (2008).
24. Rudd, C.E. Lnk adaptor: novel negative regulator of B cell lymphopoiesis. *Sci. STKE* **2001**, pe1 (2001).
25. Oh, S.T. *et al.* Novel mutations in the inhibitory adaptor protein LNK drive JAK-STAT signaling in patients with myeloproliferative neoplasms. *Blood* **116**, 988–992 (2010).
26. Lasho, T.L., Pardanani, A. & Tefferi, A. *LNK* mutations in *JAK2* mutation–negative erythrocytosis. *N. Engl. J. Med.* **363**, 1189–1190 (2010).
27. Hurtado, C. *et al. LNK* can also be mutated outside PH and SH2 domains in myeloproliferative neoplasms with and without V617F*JAK2* mutation. *Leuk. Res.* **35**, 1537–1539 (2011).
28. Perez-Garcia, A. *et al.* Genetic loss of *SH2B3* in acute lymphoblastic leukemia. *Blood* **122**, 2425–2432 (2013).
29. McMullin, M.F., Wu, C., Percy, M.J. & Tong, W. A nonsynonymous *LNK* polymorphism associated with idiopathic erythrocytosis. *Am. J. Hematol.* **86**, 962–964 (2011).
30. Gambacorti-Passerini, C.B. *et al.* Recurrent *ETNK1* mutations in atypical chronic myeloid leukemia. *Blood* **125**, 499–503 (2015).
31. Yang, Z. *et al.* Increased c-Jun expression and reduced GATA2 expression promote aberrant monocytic differentiation induced by activating *PTPN11* mutants. *Mol. Cell. Biol.* **29**, 4376–4393 (2009).
32. Pasquet, M. *et al.* High frequency of *GATA2* mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia. *Blood* **121**, 822–829 (2013).
33. Kazenwadel, J. *et al.* Loss-of-function germline *GATA2* mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature. *Blood* **119**, 1283–1291 (2012).
34. Hsu, A.P. *et al.* Mutations in *GATA2* are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* **118**, 2653–2655 (2011).
35. Hahn, C.N. *et al.* Heritable *GATA2* mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat. Genet.* **43**, 1012–1017 (2011).
36. Ostergaard, P. *et al.* Mutations in *GATA2* cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome). *Nat. Genet.* **43**, 929–931 (2011).
37. Stieglitz, E. *et al.* Mutations in *GATA2* are rare in juvenile myelomonocytic leukemia. *Blood* **123**, 1426–1427 (2014).
38. Fasan, A. *et al. GATA2* mutations are frequent in intermediate-risk karyotype AML with biallelic *CEBPA* mutations and are associated with favorable prognosis. *Leukemia* **27**, 482–485 (2013).
39. Papaemmanuil, E. *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* **122**, 3616–3627, quiz 3699 (2013).
40. Bauer, D.E., Loh, M.L., Bhagat, G., Cantor, A.B. & Kung, A.L. Potential role of RUNX1 in the pathogenesis of juvenile myelomonocytic leukemia (JMML). *Blood* **122**, 45 (2013).
41. Osato, M. Point mutations in the *RUNX1*/*AML1* gene: another actor in RUNX leukemia. *Oncogene* **23**, 4284–4296 (2004).
42. Sano, H. *et al.* Wiskott-Aldrich syndrome with unusual clinical features similar to juvenile myelomonocytic leukemia. *Int. J. Hematol.* **96**, 279–283 (2012).
43. Yoshimi, A. *et al.* Wiskott-Aldrich syndrome presenting with a clinical picture mimicking juvenile myelomonocytic leukaemia. *Pediatr. Blood Cancer* **60**, 836–841 (2013).
44. Albert, M.H. *et al.* X-linked thrombocytopenia (XLT) due to *WAS* mutations: clinical characteristics, long-term outcome, and treatment options. *Blood* **115**, 3231–3238 (2010).
45. Pérez, B. *et al.* Genetic typing of *CBL*, *ASXL1*, *RUNX1*, *TET2* and *JAK2* in juvenile myelomonocytic leukaemia reveals a genetic profile distinct from chronic myelomonocytic leukaemia. *Br. J. Haematol.* **151**, 460–468 (2010).
46. Sugimoto, Y. *et al.* Spectrum of molecular defects in juvenile myelomonocytic leukaemia includes *ASXL1* mutations. *Br. J. Haematol.* **150**, 83–87 (2010).
47. Kar, S.A. *et al.* Spliceosomal gene mutations are frequent events in the diverse mutational spectrum of chronic myelomonocytic leukemia but largely absent in juvenile myelomonocytic leukemia. *Haematologica* **98**, 107–113 (2013).
48. Ribeiro, A.F. *et al.* Mutant *DNMT3A*: a marker of poor prognosis in acute myeloid leukemia. *Blood* **119**, 5824–5831 (2012).
49. Ley, T.J. *et al. DNMT3A* mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
50. Madan, V. *et al.* Aberrant splicing of U12-type introns is the hallmark of ZRSR2 mutant myelodysplastic syndrome. *Nat. Commun.* **6**, 6042 (2015).
51. Hirabayashi, S. *et al.* Spliceosomal gene aberrations are rare, coexist with oncogenic mutations, and are unlikely to exert a driver effect in childhood MDS and JMML. *Blood* **119**, e96–e99 (2012).
52. Takita, J. *et al.* Novel splicing-factor mutations in juvenile myelomonocytic leukemia. *Leukemia* **26**, 1879–1881 (2012).
53. Jankowska, A.M. *et al.* Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: *UTX*, *EZH2*, and *DNMT3A*. *Blood* **118**, 3932–3941 (2011).
54. Furlan, I. *et al.* Intriguing response to azacitidine in a patient with juvenile myelomonocytic leukemia and monosomy 7. *Blood* **113**, 2867–2868 (2009).
55. Russler-Germain, D.A. *et al.* The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell* **25**, 442–454 (2014).
56. Archambeault, S. *et al.* Development of an allele-specific minimal residual disease assay for patients with juvenile myelomonocytic leukemia. *Blood* **111**, 1124–1127 (2008).
57. Kothari, A., Hulbert, M.L., Cottrell, C.E. & Nguyen, T.T. Aggressive congenital juvenile myelomonocytic leukemia associated with somatic *KRAS* p.G13D mutation and concurrent germline *IGF1R* duplication. *Leuk. Lymphoma* 4 **56**, 1175–1178 (2015).
58. Lee, M.L. *et al.* Juvenile myelomonocytic leukemia in a premature neonate mimicking neonatal sepsis. *Pediatr. Neonatol.* doi:10.1016/j.pedneo.2013.06.009 (25 October 2013).
59. Greaves, M. & Maley, C.C. Clonal evolution in cancer. *Nature* **481**, 306–313 (2012).
60. Chang, T. *et al.* Sustained MEK inhibition abrogates myeloproliferative disease in *Nf1* mutant mice. *J. Clin. Invest.* **123**, 335–339 (2013).
61. Lyubynska, N. *et al.* A MEK inhibitor abrogates myeloproliferative disease in *Kras* mutant mice. *Sci. Transl. Med.* **3**, 76ra27 (2011).

ONLINE METHODS

Patients. Twenty-nine patients diagnosed with JMML were included in the discovery cohort and had germline-tumor pairs evaluated by whole-exome sequencing. Seven of these patients also had relapse samples included in the exome analysis (**Supplementary Table 1**). A validation cohort comprising 71 distinct patients was tested for the 16 genes found to be mutated (including *JAK3*, for which we did not identify lesions by exome sequencing) during the discovery phase. Ninety-eight patients with JMML confirmed on the basis of internationally accepted criteria[62] were included in the survival analysis. In addition to date of diagnosis, follow-up and sequencing data, 12 potential clinical and laboratory prognostic factors were analyzed (**Supplementary Table 8**). Ten of these patients had relapse tissue included in targeted resequencing. Fifty of the validation cohort samples were from patients who were enrolled in clinical trial AAML0122, which evaluated the safety and efficacy of a farnesyltransferase inhibitor as a window treatment before HSCT[63]. Eleven of the 81 validation samples were previously included in the discovery cohort. Fourteen patients from the combined cohorts also had RNA-seq performed. Twelve patients had genome-wide DNA methylation analysis performed. Approvals for these studies were obtained from the University of California San Francisco (UCSF) Committee on Human Research. All participants or their guardians provided informed consent in accordance with the Declaration of Helsinki.

Tissue. Germline DNA was extracted using standard methods from buccal swabs, cord blood, skin fibroblasts, bone marrow fibroblasts or EBV-immortalized cell lines, as available. Tumor DNA was extracted using standard methods from bone marrow or peripheral blood mononuclear cells obtained at diagnosis and at relapse, when available. RNA was extracted using a TRIzol chloroform method.

Discovery. Whole-exome sequencing was performed using the Illumina platform as described previously. In brief, DNA was fragmented by sonication (Covaris) to 150 bp in size and further purified using Agencourt AMPure XP beads. Size-selected DNA (50 ng) was then ligated to specific adaptors during library preparation (Illumina TruSeq). Each library was constructed with sample-specific barcodes and quantified by quantitative PCR (Kapa Biosystems), and two libraries were pooled for a total of 500 ng for exome enrichment using the Agilent SureSelect hybrid capture kit (Whole Exome_v1.1, Agilent Technologies). Several captured libraries were pooled further and sequenced on one or more lanes to a final equivalent of two exomes per lane on a HiSeq 2500 instrument (Illumina).

Calling of somatic variants, germline variants, and small deletions and insertions was performed within the Firehose environment at the Broad Institute with the previously published MuTect[64], Haplotype[65,66] and Indelocator[67]. In addition, we manually reviewed all candidate driver events using the Integrative Genomics Viewer. Genome-wide copy-ratio profiles were inferred using CAPSEG[68]. Read depth at capture targets in tumor samples was calibrated to estimate copy ratio using the depths observed in a panel of normal genomes. We then segmented the copy-ratio profiles using the circular binary segmentation (CBS) algorithm[68]. Next, we performed allelic copy analysis using reference and alternate counts at germline heterozygous SNP sites. These data allow the identification of copy-neutral LOH events and inference of the contribution of each chromosome to observed copy profiles. We then combined the two data types to derive allelic copy-ratio profiles (**Supplementary Fig. 4**).

To determine which observed somatic variants were pathogenic, we first identified events that had a minor allele frequency less than 0.001 in the Exome Aggregation Cohort (ExAC) database and were marked as deleterious by Mutation Assessor. Next, we enriched for mutated genes that were previously implicated in myeloid malignancies. Finally, we compared each mutation's variant classification with expected gene function and selected inactivating mutations for tumor suppressors and activating mutations in regulatory domain hotspots for oncogenes as annotated in the Catalogue of Somatic Mutations in Cancer (COSMIC) database[69,70].

To assess tumor contamination of normal tissue (TiN), we applied deTiN (A.N.T.-W., C.S., D.A. Landau, K. Cibulskis and E. Rheinbay *et al.*, unpublished data) using both copy number events and mutations, when available. Briefly, deTiN estimates TiN using evidence for putative somatic events observed in the normal sample (**Supplementary Table 11**). The contamination estimates were then applied to our single-nucleotide variant (SNV) and indel calls to recover events previously rejected because of an impure germline source. We were then able to ensure higher sensitivity, even with non-optimal germline samples. We observed distinct differences in TiN content specific to each tissue type (**Supplementary Fig. 13**). Fibroblasts from either bone marrow or skin ($n = 4$) had the lowest TiN content with a mean of less than 1%. Sixty percent of the buccal samples ($n = 10$) collected at the time of diagnosis were infiltrated by leukemia cells and had a mean TiN content of 11%, ranging from 0 to 25%. A previous study cast doubt on the use of EBV-immortalized cell lines as a source of germline material in JMML[13]. Consistent with this report, EBV-transformed lymphoblasts from two of six patients with JMML had a TiN of <1%, whereas in four others there was clear involvement of mutations in a clonal fashion. Similarly, cord blood ($n = 5$) from patients contained >60% TiN in three samples, reflecting the presence of leukemia cells at birth[71].

To estimate TiN using mutations, deTiN fits the fraction of alternate reads in the normal sample at candidate somatic sites. When samples had fewer than three somatic SNVs identified with our standard mutation calling pipeline, we used alternate and reference allele counts at previously identified canonical events in those samples to build the contamination model. To construct this model, deTiN first calculates the probability that a given candidate mutation is a somatic or germline event over a range of potential TiN levels. Then, we iterate over TiN levels, each time constraining the fit using mutations. We repeat the same process using heterozygous SNPs in regions of allelic imbalance and constrain the estimate using the amount of allelic imbalance observed in the normal sample in relation to the tumor sample. We then combine each of these estimates by adding the likelihood curves.

Next, we use the combined TiN estimate to keep mutation calls from variants that were originally rejected as germline but were actually 1,000 times more likely to be explained by TiN.

Validation. A customized TruSeq amplicon kit was designed, using the online DesignStudio pipeline (Illumina), that targeted the entire coding regions of 16 genes (*ASXL1*, *CBL*, *DNMT3A*, *EZH2*, *GATA2*, *JAK3*, *KRAS*, *NF1*, *NRAS*, *PTPN11*, *RRAS*, *RRAS2*, *RUNX1*, *SETBP1*, *SH2B3* and *ZRSR2*) and covered a total of 49,710 bp with 184 amplicon regions. Libraries prepared from the DNA of each of 91 samples were indexed and subjected to 250-bp paired-end sequencing on the Illumina MiSeq platform, and data from three independent MiSeq runs were combined to generate an average of 761,678 reads/sample (median of 383,391 reads/sample; **Supplementary Fig. 14a**). While 95.6% of the total reads passed quality filtering, more than 97.4% of the mapped reads (99.8% mapping rate) were in the targeted regions, providing an average coverage depth of 1,380× across the targeted regions (**Supplementary Fig. 14b** and **Supplementary Table 12**). Hotspots not adequately covered on the MiSeq platform were analyzed by Sanger sequencing. Targeted sequencing data were analyzed as described previously[72]. In brief, after quality trimming, reads from each sample were aligned to the human genome reference sequence (GRCh37/hg19), and the CLC Genomic Workbench (v7.5, CLC Bio) was used for mapping and variant calling. An in-house tool was then used to filter through the variant list, and the consequences of the protein sequence changes resulting from the detected DNA sequence changes were predicted using the ANNOVAR functional annotation tool. Only variants with greater than 5% allelic fraction and at least 8× coverage of the mutated allele were included as a somatic alteration; additional detectable variants are reported for completeness (**Supplementary Table 3**).

DNA methylation analysis. Genomic DNA was bisulfite converted using the EZ DNA Methylation kit (Zymo Research) and processed on Infinium HumanMethylation450 BeadChip arrays (Illumina) according to the manufacturer's protocol. Probe-level signals for individual CpG sites were subjected to both background and global dye bias correction[73]. Probes that mapped to regions with known germline polymorphisms, to multiple genomic loci or to either sex chromosome were filtered out. Two-way unsupervised hierarchical clustering was performed using Euclidean distance and Ward linkage on the 1,486 most variable CpG sites across the cohort, with variability ranked by standard deviation. Four subjects with clonal *ASXL1* mutations at diagnosis were included for analysis. An additional seven patients were chosen as controls because they had similar mutational profiles in comparison to the subjects except they did not possess any known genetic mutations in genes for epigenetic regulators (**Supplementary Fig. 6**).

RNA sequencing. RNA was extracted, used to prepare mRNA libraries and sequenced on the Illumina HiSeq platform, resulting in paired 50-nt reads; resulting data were subjected to quality control as previously described[74,75]. RNA reads were aligned using TopHat. Gene expression was quantified for transcripts corresponding to GENCODE v12 genes using reads per kilobase of transcript per million mapped reads (RPKM). Detection of rearrangements was performed using dRanger as described previously[76]. Fusion detection was performed using TopHat-Fusion.

Analysis of RNA sequencing data. For RNA-seq analysis, paired-end sequencing was performed using the HiSeq platform with a 100-bp read length. Paired-end reads from RNA-seq were aligned to the following four database files using Burrows-Wheeler Aligner (BWA; 0.5.10): (i) the human GRCh37-lite reference sequence, (ii) RefSeq, (iii) a sequence file representing all possible combinations of non-sequential pairs in RefSeq exons and (iv) the AceView database flat file downloaded from the UCSC Genome Browser representing transcripts constructed from human ESTs. The mapping results from databases (ii) to (iv) were translated to human reference genome coordinates. In addition, they were aligned using STAR 2.3.0 to the human GRCh37-lite reference sequence without annotations. A BAM file was constructed by selecting the best alignment among the five mappings. Poor-quality mappings were improved using SIM4 when possible to generate the final BAM file. Coverage was calculated using an in-house pipeline. Structural variation detection was carried out using CICERO, a novel algorithm that uses *de novo* assembly to identify structural variation in RNA-seq data (Y.L., T. Bo, M. Rusch, J. Easton and K. Boggs *et al.*, unpublished data).

Digital gene expression profiling. Transcript expression levels were estimated as fragments per kilobase of transcript per million fragments (FPKM); gene FPKM values were computed by summing the transcript FPKM values for each gene using Cuffdiff2 (ref. 77). A gene was considered to be expressed if the FPKM value was ≥0.5 according to the distribution of FPKM gene expression levels. Genes that were not expressed in any sample were excluded from the final data matrix for downstream analysis.

Sanger sequencing. Sanger sequencing of the discovery cohort including of *ASXL1* (p.Tyr591* and p.His630fs), *DNMT3A* (p.Arg882Cys), *EZH2* (p.Val674Leu), *GATA2* (p.Asn317Ser), *RRAS* (p.Gln87Leu), *RRAS2* (p.Gln72Leu), *RUNX1* (p.Arg349fs), *SETBP1* (p.Asp868Asn, p.Gly870Asp and p.Gly870Ser), *SH2B3* (p.Gln258*, p.Trp262*, p.Phe390fs, p.Glu400Lys and p.His414_splice) and *ZRSR2* (p.Gly179Glu) was performed using the primers listed in **Supplementary Table 13**. All mutations were tested in both germline and tumor tissue to determine germline versus somatic acquisition (**Supplementary Fig. 15**). Sequencing was carried out using Hot-Start Polymerase (Promega) and the following PCR conditions: incubation at 95 °C for 2 min; 35 cycles of 94 °C for 45 s, 60 °C for 30 s and 72 °C for 40 s; and incubation at 75 °C for 5 min. Sequences were aligned using CLC software (CLC Bio).

Single-colony analysis. Mononuclear cells were isolated from fresh bone marrow or peripheral blood samples and resuspended in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 2% FBS. Cells were suspended at a concentration of 200,000 cells/ml. The cell suspension (154 µl) was added to a tube containing 1.2 ml of human methylcellulose (R&D Systems Human Methylcellulose Complete Media, HSC003), 15 µl of 100× penicillin-streptomycin, diluted in water, and IMDM to bring the final volume to 1.54 ml. The cytokines used included erythropoietin, granulocyte-macrophage colony-stimulating factor (GM-CSF), interleukin-3 (IL-3) and stem cell factor (SCF). The solution was vortexed for 15 s and rested for 15 min; 1.1 ml was plated into a 35 × 10 mm Petri dish (BD Falcon, 351008) placed in a 150 × 15 mm Petri dish (BD Falcon, 351058), with another dish containing sterile water, and plates were placed in an incubator at 37 °C with 5% CO_2. After 14 d, plates were removed from the incubator, and 2 µl of colonies (clusters of 50 cells or more) were plucked under a microscope at 40× magnification and resuspended in 20 µl of TE buffer. PCR amplification was then carried out for each of the mutations using 1 ml of the vortexed colony suspension and the primers listed in **Supplementary Table 13**.

Fibroblast cultures. Bone marrow or skin cells were cultured in Chang media (Irvine Scientific, T105), which was changed every 3 d. Cells were split when the fibroblasts became 70% confluent and were collected for DNA extraction using standard methods when the fibroblasts began to have a flattened morphological appearance.

Immunoblotting. Whole-cell extracts were prepared from previously cryopreserved samples in NP-40 lysis buffer supplemented with protease inhibitors. Blots were blocked with 5% BSA in TBS with 0.1% Tween-20 for 1 h at room temperature and probed with primary antibody to LNK (H-129, Santa Cruz Biotechnology) and secondary horseradish peroxidase (HRP)-conjugated antibody (GE Healthcare) for 1 h each. Protein quantification for immunoblots was performed using Image Lab software (v.5.2, Bio-Rad).

Statistical analyses. Data from patient AAML0122 were current as of 8 June 2010. Data from the remaining patients were current as of 1 October 2014. The number of somatic alterations at diagnosis, categorized as zero or one versus two or more, was the predictor variable of interest in this analysis. Somatic alterations were defined as pathogenic mutations in genes or monosomy 7. Clinical and biological features were compared between the groups. The significance of observed differences in proportions was tested using the χ^2 test (for categorical variables) or the *t* test (for continuous variables), and Fisher's exact test was used when data were sparse. The Mann-Whitney test was used to determine the difference in medians. The Kaplan-Meier method[78] was used to estimate the probabilities of event-free survival and overall survival. Event-free survival was defined as the time from diagnosis to first occurrence of relapse, treatment-related mortality or death; overall survival was defined as the time from diagnosis to death. Event-free and overall survival distributions were compared according to the number of somatic alterations at diagnosis using a two-sided log-rank test. Cox proportional hazards models[62] were used to estimate the hazard ratio for defined groups of patients in univariate and multivariate analyses of overall and event-free survival. Relapse incidence was defined as the probability of having a relapse before time *t*; death without experiencing a relapse was considered a competing event. Cumulative incidence was compared using Gray's test[63]. *P* values less than 0.05 were considered statistically significant. All statistical analyses were performed using STATA.

62. Chan, R.J., Cooper, T., Kratz, C.P., Weiss, B. & Loh, M.L. Juvenile myelomonocytic leukemia: a report from the 2nd International JMML Symposium. *Leuk. Res.* **33**, 355–362 (2009).
63. Stieglitz, E. *et al.* Phase II/III trial of a pre-transplant farnesyl transferase inhibitor in juvenile myelomonocytic leukemia: a report from the Children's Oncology Group. *Pediatr. Blood Cancer* **62**, 629–636 (2015).
64. Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
65. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
66. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
67. Chapman, M.A. *et al.* Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**, 467–472 (2011).
68. Venkatraman, E.S. & Olshen, A.B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657–663 (2007).
69. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* **39**, e118 (2011).
70. Forbes, S.A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–D811 (2015).
71. Triche, T.J. Jr., Weisenberger, D.J., Van Den Berg, D., Laird, P.W. & Siegmund, K.D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41**, e90 (2013).
72. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
73. Dulak, A.M. *et al.* Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat. Genet.* **45**, 478–486 (2013).
74. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
75. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
76. Kaplan, E.L. & Meier, P. Nonparametric-estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481 (1958).
77. Cox, D.R. Regression models and life-tables. *J. R. Stat. Soc., B* **34**, 187–220 (1972).
78. Gray, R.J. A class of *K*-sample tests for comparing the cumulative incidence of a competing risk. *Ann. Stat.* **16**, 1141–1154 (1988).

doi:10.1038/ng.3400