RESEARCH ARTICLE

Convergent genetic aberrations in murine and human T lineage acute lymphoblastic leukemias

Benjamin J. Huang1, Anica M. Wandler1, Lauren K. Meyer1, Monique Dail2, Anneleen Daemen3, Deepak Sampath4, Qing Li5, Xinyue Wang1, Jasmine C. Wong1, Joy Nakitandwe6, James R. Downing6, Jinghui Zhang7, Barry S. Taylor8,9,10, Kevin Shannon1,11*

1 Department of Pediatrics, University of California San Francisco, San Francisco, CA, United States of America, 2 Department of Oncology Biomarker Development, Genentech, South San Francisco, CA, United States of America, 3 Department of Bioinformatics & Computational Biology, Genentech, South San Francisco, CA, United States of America, 4 Department of Translational Oncology, Genentech, South San Francisco, CA, United States of America, 5 Division of Hematology/Oncology, Department of Medicine, University of Michigan, Ann Arbor, MI, United States of America, 6 Department of Pathology, St Jude Children’s Research Hospital, Memphis, Tennessee, United States of America, 7 Department of Computational Biology, St Jude Children’s Research Hospital, Memphis, Tennessee, United States of America, 8 Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, United States of America, 9 Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, United States of America, 10 Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, United States of America, 11 Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, United States of America

* kevin.shannon@ucsf.edu

Abstract

The lack of predictive preclinical models is a fundamental barrier to translating knowledge about the molecular pathogenesis of cancer into improved therapies. Insertional mutagenesis (IM) in mice is a robust strategy for generating malignancies that recapitulate the extensive inter- and intra-tumoral genetic heterogeneity found in advanced human cancers. While the central role of “driver” viral insertions in IM models that aberrantly increase the expression of proto-oncogenes or disrupt tumor suppressors has been appreciated for many years, the contributions of cooperating somatic mutations and large chromosomal alterations to tumorigenesis are largely unknown. Integrated genomic studies of T lineage acute lymphoblastic leukemias (T-ALLs) generated by IM in wild-type (WT) and Kras mutant mice reveal frequent point mutations and other recurrent non-insertional genetic alterations that also occur in human T-ALL. These somatic mutations are sensitive and specific markers for defining clonal dynamics and identifying candidate resistance mechanisms in leukemias that relapse after an initial therapeutic response. Primary cancers initiated by IM and resistant clones that emerge during in vivo treatment close key gaps in existing preclinical models, and are robust platforms for investigating the efficacy of new therapies and for elucidating how drug exposure shapes tumor evolution and patterns of resistance.
A lack of predictive cancer models is a major bottleneck for prioritizing new anti-cancer drugs for clinical trials. We comprehensively profiled a panel of primary mouse T lineage leukemias initiated by insertional mutagenesis and found remarkable similarities with human T-ALL in regard to overall mutational burden, the occurrence of specific somatic mutations and large chromosomal alterations, and concordant gene expression signatures. We observed frequent duplication of the \textit{Kras} oncogene with loss of the normal allele, which has potential therapeutic implications that merit further investigation in human leukemia and in other preclinical models. Mutations identified in mouse leukemias that relapsed after \textit{in vivo} treatment with signal transduction inhibitors were also observed in relapsed human T-ALL, indicating that this model system can be utilized to investigate strategies for overcoming intrinsic and acquired drug resistance. Finally, preclinical models similar to the one described here that are characterized by a normal endogenous tumor microenvironment and intact immune system will become increasingly important for testing immunotherapy approaches for human cancer.

**Introduction**

Most new anti-cancer agents fail in the clinic [1]. Whereas cancer cell lines have been integral to the development of most anti-cancer drugs and exhibit genotype-specific responses to some targeted inhibitors, they fail to model many fundamental properties of primary tumors. Patient derived xenograft (PDX) models are promising platforms for testing anti-cancer drugs, but also have inherent limitations, including lack of an intact immune system or normal tumor microenvironment and failure to fully recapitulate the clonal heterogeneity of advanced human cancers [1–3]. Mouse models in which cancers arise due to mutations introduced into the germline or following exposure to chemical mutagens also have distinct advantages and liabilities. Genetically engineered mouse (GEM) models reproduce many morphologic features of advanced human cancers, but diverge substantially with respect to the overall mutational burden and pattern of copy number alterations [4–8]. Whereas models initiated by chemical carcinogens may more accurately reflect the mutational profile of some human cancers [4, 6], extensive genetic variability and the inability to perform technical replicates are substantial barriers to widespread preclinical use.

IM was developed as an unbiased strategy for cancer gene discovery [9]. This approach initially involved injecting mice with retroviruses, which integrated into the genome and promoted tumorigenesis by activating proto-oncogenes or disrupting tumor suppressors. The locations of exogenous viral DNA insertions also served as molecular sequence tags that facilitated identifying candidate cancer genes. An appealing aspect of IM models is that they recapitulate the inter- and intra-tumoral genetic heterogeneity that is a hallmark of advanced human cancers [2]. Furthermore, technical advances such as the PiggyBAC and \textit{Sleeping Beauty} transposon systems both expanded the spectrum of cancers that can be modeled and yielded tumors with insertions within many genes that are altered in corresponding human cancers [10–17].

We previously generated primary T-ALLs by injecting C57Bl/6 x 129Sv/Jae F1 mouse pups harboring a conditional mutant \textit{Kras} oncogene (\textit{Lox-STOP-Lox Kras}^{G12D}) and congenic WT animals with the MOL4070LTR retrovirus [18]. Subsequent activation of endogenous \textit{Kras}^{G12D} expression both accelerated disease onset and increased penetrance. Importantly, this system differs from other GEM models of RAS-driven cancers because the \textit{Kras}^{G12D}
mutation is a secondary event that cooperates with antecedent retroviral integrations [19, 20], and thus reflects the most common pathogenic sequence in patients [21]. T-ALLs initiated by IM typically exhibited 3–6 dominant clonal retroviral integrations, many of which occurred within or adjacent to known cancer genes. Interestingly, most of these leukemias also acquired somatic Notch1 mutations, which are highly prevalent in human T-ALLs [19] and have also been observed in other murine models of T-ALL [22–24]. This unexpected observation raised the possibility that non-insertional genetic mechanisms broadly contribute to tumorigenesis in cancer models driven by IM. In this study, we set out to more thoroughly define the genomic and transcriptional landscape of a cancer model initiated by IM and to assess the contribution that non-insertional genetic mechanisms have on tumorigenesis and clonal evolution upon treatment with signal transduction inhibitors.

Results

The mutational landscape of IM-induced T-ALL is similar to that of human T-ALL

We performed whole exome sequencing (WES) on 21 independent IM-induced T-ALLs that were tested in preclinical efficacy studies of MEK and PI3 kinase (PI3K) inhibitors [20]. These leukemias included 15 primary T-ALLs isolated from KrasG12D mice and 6 from WT mice (hereafter referred to as KrasWT). The mutational burden in T-ALLs induced by IM ranged from 0.2 to 0.8 somatic alterations per megabase of exome sequence (Fig 1A), which is similar to the mutation burden reported in human ALL [25].

WES identified all 14 known Notch1 mutations [19], and also uncovered somatic mutations in the Notch pathway genes Notch4 and Cntn1 in two additional leukemias (Fig 1B). Thus, the frequency of somatic mutations in genes linked to Notch signaling (16 of 21; 75%) was similar to what is observed in human T-ALL (Fig 1B), with NOTCH1/Notch1 targeted most frequently in both species.

IKZF1 encodes the Ikaros transcription factor that functions as a regulator of lymphocyte differentiation. Ten independent T-ALLs harbored a variety of Ikkζ1 alterations including retroviral integrations predicted to disrupt the coding region (n = 7), missense or truncating mutations (n = 3), and copy number losses (n = 2) (Fig 1B). Five additional genes that are mutated in human T-ALL and broadly regulate gene transcription or epigenetic programming were altered by retroviral integration (Myb, Lmo2, Runx1) or mutation (Kdm6a, Smarca4). Myb integrations and Kdm6a mutations were observed in three and two independent leukemias, respectively. Overall, 14 of 21 T-ALLs had one or more alterations of genes classified as transcriptional/epigenetic regulators, which is similar to the frequency observed in human T-ALL (Fig 1B). However, the involvement of specific genes varied between species. For example, while Ikkζ1 was affected far more often in murine T-ALL, IKZF1 is recurrently mutated or deleted in relapsed human ALL and in certain high-risk leukemia subtypes at diagnosis [26–28].

Somatic Ras/PI3K pathway alterations were detected in 6 of 6 KrasWT leukemias (100%), but in only 4 of 15 KrasG12D T-ALLs (27%). Rasgrp1, which encodes a guanine nucleotide exchange factor that promotes Ras activation, is a recurrent site of retroviral integrations in KrasWT leukemias [19, 29]. WES also identified mutations in five genes encoding protein components of the Ras/PI3K pathway in six additional T-ALLs. The low frequency of somatic Ras/PI3K pathway abnormalities in KrasG12D T-ALLs suggests that expression of endogenous oncogenic KrasG12D in leukemia initiating cells is sufficient to confer a clonal growth advantage in the absence of additional signaling mutations. Interestingly, however, mutant Kras allele frequencies were above 80% in 12 of 15 KrasG12D leukemias (Fig 1C). We considered two

Convergent genetics of mouse and human leukemias

PLOS Genetics | https://doi.org/10.1371/journal.pgen.1008168 June 14, 2019 3 / 17
potential explanations for this observation: a structural deletion of the WT \textit{Kras} allele or somatic uniparental disomy (UPD) resulting in loss of WT \textit{Kras} and duplication of \textit{Kras}^{G12D}. The presence of neutral \textit{Kras} copy number inferred from WES data combined with frequent loss of C57Bl/6-specific single nucleotide polymorphisms (SNPs) around the \textit{Kras} locus strongly implicated copy neutral loss of heterozygosity with increased \textit{Kras}^{G12D} dosage as the underlying genetic mechanism (S1 Fig). These data are indicative of substantial selective pressure favoring the outgrowth of leukemic clones that have increased Ras signaling through a genetic mechanism that precisely duplicates \textit{Kras}^{G12D} copy number, and are consistent with recent studies of mouse and human cancers \cite{30,31}. The overall proportion of mouse leukemias with Ras/PI3K mutations is substantially higher in this model system than in human T-ALL due to the presence of a \textit{Kras}^{G12D} knock in mutation in 15 of the 21 of them (Fig 1B).

Many somatic alterations occur in hotspots and are predicated to be pathogenic

To assess the functional consequence of single nucleotide variants (SNVs) and indels in murine T-ALLs, we mapped mutations to the corresponding human protein homologs and compared these data to somatic alterations reported previously. This analysis revealed
alterations within conserved functional motifs, including \( Pten^{R130W} \) and \( Raf1^{S259P} \) (Fig 2A) [32–34]. \( In \text{ silico} \) protein function prediction analysis revealed that mutations highlighted in Fig 1 are frequently deleterious and occur at residues that are conserved across multiple species.

**Fig 2.** Somatic alterations are frequently predicted to be pathogenic and correspond to mutations in human T-ALL. (A) \( Notch1 \), \( Pten \), \( Raf1 \), and \( Trp53 \) mutations in IM-induced T-ALLs (top panels) occur within hotspot locations that correspond to recurrent mutations in human cancer (bottom panels). (B) \( In \text{ silico} \) protein function prediction indicates that mutations highlighted in Fig 1 are frequently deleterious and occur at residues that are conserved across multiple species.

The transcriptional signatures of IM-induced T-ALL closely resemble human T-ALL

Based on the genetic similarities between human T-ALL and IM-induced T-ALL, we performed RNA-seq to assess the transcriptional landscape of 5 \( Kras^{G12D} \) and 4 \( Kras^{WT} \) T-ALLs. Gene expression signatures that reflect a T lymphoblastic immunophenotype were enriched across these samples, and we observed low-level expression of genes associated with B lymphoblastic and myeloid leukemia immunophenotypes (S2 Fig). Consistent with human T-ALL,
we also detected distinct and clonal T-cell receptor gene VDJ rearrangements in individual leukemias (S2 Fig) [35]. We then compared the gene expression of IM-induced leukemias with previously reported microarray data used to discriminate human T-ALL from various subtypes of B lineage acute lymphoblastic leukemia (B-ALL) [36, 37]. Expression profile signatures that most accurately distinguish T-ALL from B-ALL were used to compare the relative expression in IM-induced T-ALLs, which clustered with human T-ALL (Fig 3A). We also performed unbiased gene set enrichment analysis (GSEA) using the Molecular Signature Database (MSigDB) [38].

https://doi.org/10.1371/journal.pgen.1008168.g003

Fig 3. The transcriptional signature and pathway analysis of IM-induced T-ALL cluster with human T-ALL. (A) Expression profiling that discriminate human T-ALL from B-ALL clusters IM T-ALL closely with human T-ALL. (B) MYC, PI3K, and NOTCH1 pathway gene sets are significantly enriched in murine IM-induced and human T-ALLs compared to B-ALL.
Database Hallmark Gene Sets and C2 Curated Gene Sets to compare: (1) T-ALL versus B-ALL; and, (2) IM-induced T-ALL versus B-ALL. Genes that discriminate various subtypes of B-ALL from each other were included in the analysis [36, 37]. MYC, PI3K, NOTCH1, and TCR pathway gene sets were enriched across human and mouse T-ALLs (Fig 3B and S3A Fig), and we also observed negative enrichment scores for gene sets that are upregulated in subtypes of B-ALL (S3B Fig).

Clonal evolution occurs at the level of somatic alterations in response to targeted therapies

Treatment with the pan-PI3K inhibitor GDC-0941 alone and in combination with the MEK inhibitor PD0325901 (PD901) markedly extended the survival of recipient mice transplanted with Kras\(^ {WT} \) and Kras\(^ {G12D} \) T-ALLs [20], but these leukemias invariably relapsed despite continuous drug administration. As in relapsed human leukemias [38], these resistant murine T-ALLs were derived from pre-existing ancestral clones and they frequently showed loss of activated Notch1 expression and PI3K pathway activation [20]. To identify additional genetic alterations that might contribute to treatment resistance, we performed WES on parental/relapsed T-ALL pairs. This analysis revealed SNVs and indels that were enriched or depleted in individual relapsed leukemias, including two in which inactivating Pten mutations emerged at the cost of other drivers (Fig 4). In Kras\(^ {WT} \) T-ALL JW81, treatment with GDC-0941 and with the combination of GDC-0941/PD901 significantly prolonged survival. At relapse, this leukemia acquired a de novo Pten\(^ {H123R} \) mutation and also showed markedly decreased Notch1 mutant allele frequency (Fig 4). Conversely, Kras\(^ {G12D} \) T-ALL 73M, which harbored a loss of the WT Kras allele and a duplication of Kras\(^ {G12D} \), had a minimal response to treatment. Remarkably, however, this leukemia exhibited rapid clonal evolution upon in vivo drug exposure with outgrowth of cells showing decreased mutant Kras\(^ {G12D} \) allele frequency and substantial enrichment of a pre-existing, low frequency Pten\(^ {R130W} \) mutation (Fig 4) [32]. Consistent with these results, analysis of heterozygous SNPs revealed enrichment in the 129Sv/Jae SNP allele frequency across the chromosome containing the Pten mutant allele and the presence of both WT and mutant Kras associated SNPs in the resistant clone (S4 Fig).

Discussion

We describe the first comprehensive molecular characterization of a mouse cancer model initiated by insertional mutagenesis. Through integrated analyses, we demonstrate that T-ALLs initiated by insertional mutagenesis harbor extensive genomic diversity and also acquire a similar spectrum and overall burden of somatic mutations as the corresponding human leukemias. Specifically, many driver genes that are mutated in human T-ALL [39] are also recurrently altered in IM-induced T-ALL, including Notch1, Pten, Myb, and Ikzf1. Data from both species implicate deregulated transcriptional programs in disease initiation with subsequent Ras/PI3K and Notch pathway mutations cooperating in leukemic transformation [19, 26, 39].

Genome-wide analysis of solid cancers arising in GEM models expressing one or more oncogenic driver mutations has consistently revealed markedly fewer pathogenic point mutations than in the corresponding human tumors and a higher frequency of large chromosomal alterations [4–8]. This likely reflects the ability of alterations in potent cancer genes such as Kras, Trp53, Rb, and Pten to bypass the need for environmentally-induced and age-related mutations that contribute to tumorigenesis. For example, genetic models of Kras-driven lung cancer and skin squamous cell carcinoma acquired an average of six and four additional somatic alterations, respectively [4, 6]. By contrast, in parallel analyses, carcinogen-induced
models of the same two cancer types acquired an average of 192 and 846 somatic alterations, respectively, with mutational profiles that more accurately reflect human carcinomas [4, 6].

Acute leukemias are distinct from non-hematopoietic cancers with respect to a lower overall somatic mutational burden [25] and also exhibit characteristic patterns of initiating and cooperating events. Deep sequencing studies have identified epigenetic and transcription factor alterations as early events in both lymphoblastic and myeloid acute leukemias, with subsequent signal transduction pathway mutations playing a central role in driving leukemic outgrowth [26, 38, 40, 41]. Molecular analysis of IM-induced T-ALLs suggest that they recapitulate this pathogenic sequence with disease-initiating retroviral integrations that deregulate transcriptional programs generating preleukemic clones [19], that subsequently acquired an average of 13 additional somatic alterations, including recurring mutations in Ras/PI3K pathway genes in Kras\(^{WT}\) leukemias. This latter observation is consistent with the fact that activating Kras\(^{G12D}\) expression as a secondary event increased leukemia penetrance and also reduced latency [19]. In addition, resistant leukemias that emerged during in vivo treatment with MEK and PI3K inhibitors invariably retained multiple retroviral integrations found in the corresponding drug-sensitive parental T-ALL, but frequently showed loss of Notch1 mutation [20].

Frequent uniparental disomy resulting in loss of WT Kras and duplication of Kras\(^{G12D}\) in 14 of 17 IM-induced T-ALLs is consistent with recent work demonstrating that KRAS

---

**Fig 4. Clonal evolution following combination treatment with PI3K and MEK inhibitors.** Leukemias were transplanted and treated with GDC-0941 and PD901. (A) Variant allele frequencies of primary versus relapsed leukemias were determined based on whole exome sequencing. Treatment resulted in the emergence of a resistant clone that acquired a de novo Pten mutation in T-ALL JW81 and the outgrowth of a subclone of T-ALL 73M with a pre-existing Pten mutation at the cost of other drivers. (B) Graphical representation of clonal evolution of each primary/relapsed leukemia pair.

https://doi.org/10.1371/journal.pgen.1008168.g004
mutations in acute myeloid leukemia and colorectal cancer frequently exhibit allelic imbalance that modulate competitive fitness and MEK dependency [30]. Similarly, a comprehensive genomic characterization of 264 primary T-ALLs showed that KRAS mutations, but not NRAS mutations, were specifically associated with mutant allele fractions of greater than 0.7 in a subset of patients at diagnosis [39]. Another study that investigated large numbers of murine KrasG12D pancreatic and lung cancers also identified frequent Kras allelic imbalance that was modulated by treatment with a MEK inhibitor [8]. The panel of primary mouse leukemias described here are novel reagents for investigating the effects of restoring WT Kras expression on fitness and for interrogating K-Ras protein dimerization and other candidate molecular mechanisms of growth inhibition [42, 43].

MYC is a downstream target of activated Notch1 signaling that contributes to T-ALL pathogenesis [44–47]. As the mouse Myc gene is located on chromosome 15, it is intriguing that 13 of 21 T-ALLs analyzed in this study exhibited trisomy 15. However, Myc expression was similar in leukemias with and without trisomy 15 and we did not observe any with focal Myc amplification. Similarly, Mullighan, et al. [39] identified 12 human T-ALLs with trisomy 8 or large gains of 8q, which is the chromosome arm that contains MYC, that were not characterized by elevated MYC expression and did not harbor focal MYC amplification. Thus, the role of MYC/Myc in promoting the outgrowth of T-ALLs with trisomy 8 (human) or 15 (mouse) is uncertain.

NRAS, KRAS, PTEN, and AKT mutations are associated with worse outcomes in patients newly diagnosed with T-ALL [39, 48, 49], are enriched in relapse specimens [38, 50], and have previously been shown to provide a proliferative advantage based on xenograft models [51]. While cure rates have improved substantially in T-ALL, relapsed disease is still associated with dismal outcomes [52]. Furthermore, whereas T-ALLs characterized by activated Notch signaling are vulnerable to inhibitors of the Notch processing enzyme gamma secretase, Pten inactivation overcomes this dependency [53]. This might, in part, underlie the disappointing response rates seen in early phase clinical trials of agents targeting gamma secretase in T-ALL [54, 55]. Together, these data indicate a need for both therapeutic approaches directed against hyperactive Ras/PI3K signaling in T-ALL and for robust preclinical models to test them. Along these lines, it is notable that the 21 primary leukemias characterized here exhibit frequent alterations in Ras and PI3K pathway genes, some of which were enriched at relapse. This panel of transplantable primary and relapsed/refractory leukemias with known genetic aberrations thus comprises a unique resource for testing novel therapeutic strategies for high-risk subsets of T-ALL and for delineating additional resistance mechanisms.

Materials and methods

Ethics statement

All animal experiments conformed to national regulatory standards and were approved by the University of California, San Francisco Committee on Animal Research (IACUC Protocol Approval Number AN136527).

Murine leukemias

All animal experiments conformed to national regulatory standards and were approved by the University of California, San Francisco Committee on Animal Research. C57Bl/6 x 129Sv/Jae Mx1-Cre x KrasLSL-G12D mice (KrasG12D) and wild-type littermates (KrasWT) were injected with MOL4070LTR at 3–5 days of age and KrasG12D expression was activated at 21 days of age by administering a single dose of polyinosinic-polycytidylic acid (pIpC) as previously described [19, 20]. For adoptive transfer of acute leukemias, 2 × 10⁶ cells were injected retro-orbitally
into 8–12 week old wild-type C57Bl/6 × 129Sv/Jae F1 recipient mice that had been irradiated with 450 cGy. Mice that appeared ill were euthanized and underwent full pathological examination. Flushed bone marrow cells were isolated to confirm relapse and to assess for tumor purity. Relapsed leukemia cells were cryopreserved for molecular and genetic analysis.

**Whole exome sequencing**

Genomic DNA was extracted from the bone marrows of mice transplanted with IM-induced leukemias and then sheared to generate 150 to 200 base pair fragments using a Covaris S2 focused-ultrasonicator. Indexed libraries were prepared using the Agilent SureSelect XT2 Reagent Kit for the HiSeq platform. Exomes were captured using the Agilent SureSelect XT2 Mouse All Exon bait library. Sample quality and quantity were assessed using the Agilent 2100 Bioanalyzer instrument. Paired-end 100 base pair reads were generated on an Illumina HiSeq 2000 platform. All sequence data including read alignment; quality and performance metrics; post-processing, somatic mutation and DNA copy number alteration detection; and variant annotation were performed as previously described [56, 57] using the mm10 build of the mouse genome. Briefly, reads were aligned with Burrows-Wheeler Aligner [58], and processed using Picard (http://broadinstitute.github.io/picard) tools and the Genome Analysis Toolkit (GATK) [59] to perform base quality recalibration and multiple sequence realignment. Single nucleotide variants and indels were detected with the MuTect [60] and Pindel [61] algorithms, respectively. Candidate somatic mutations were manually reviewed using Integrative Genomics Viewer [62] and were confirmed by performing orthogonal Sanger sequencing validation. Tumor-specific copy number alterations were inferred from depth of coverage and B allele frequencies using control normal bone marrow samples. Loss of heterozygosity analysis was performed using GATK UnifiedGenotyper to compare germline heterozygous SNPs that differ between C57Bl/6 and 129Sv/Jae mouse strains called at a depth of 20x or greater between leukemia and control bone marrow samples.

**MOL4070 integration cloning**

Restriction enzyme digestion of genomic DNA from mouse T-ALLs, gel electrophoresis, Southern blot analysis, and hybridization with a MOL4070LTR-specific probe was performed as described previously [20, 30]. Junctional fragments at sites of retroviral integration were identified as previously described using linker-based PCR amplification and sequencing [63, 64].

**Mutation mapping and in silico protein function prediction**

Mutations were mapped onto corresponding protein sequence using MutationMapper [65]. Associated mutational frequencies in human cancer were extracted from Catalogue of Somatic Mutations in Cancer (COSMIC, version 82) [66]. Pairwise alignments between homologous proteins were generated using NCBI HomoloGene and BLASTP. *In silico* protein function prediction was performed using Mutation Assessor [67], PolyPhen-2 [68], and Sorted Intolerant From Tolerant (SIFT) [69].

**Transcriptome sequencing**

Total RNA was extracted from the bone marrows of mice transplanted with IM-induced leukemias using TRIzol. Phenol-chloroform-isooamyl extraction and ethanol precipitation were performed. Total RNA (1 μg) was treated with DNase I at room temperature for 15 min. The integrity of the DNase I-treated RNA was analyzed on an Agilent 2100 Bioanalyzer before selection of poly(A) mRNA. Selection for poly(A) mRNA and subsequent cDNA synthesis and
library preparation were carried out using the Illumina TruSeq RNA sample preparation kit according to the manufacturer’s protocol. cDNA was fragmented (200-bp peak) with a Covaris E210 ultrasonicator prior to library preparation. The quality and size of the final library preparation were analyzed on an Agilent 2100 Bioanalyzer. Paired-end 100 base pair reads were generated on an Illumina HiSeq 2500 platform. Reads were aligned with Bowtie 2 [70] using the mm10 database, and transcript quantification was performed with RNA-seq by Expectation Maximization (RSEM) [71].

Gene expression analysis

Raw microarray gene expression profiling data on patients with T-ALL and B-ALL was downloaded from publicly accessible repositories as previously described [36, 37]. Microarray data was normalized using the robust multi-array average (RMA) method [72]. Probe identifiers were mapped to associated gene identifiers using the Affymetrix HG U95A Annotations Reference (Release 36). Gene identifiers generated from RNA-seq data on IM-induced T-ALLs were mapped to human orthologs using NCBI HomoloGene. FPKM values were normalized to the microarray data using the trimmed mean of M-values (TMM) method [73]. Normalization was assessed by plotting average versus fold change normalized expression (MA plot) between the two datasets and average normalized expression comparing the two datasets. Clustering and heatmap dendrogram analysis was performed using a correlation similarity metric and average linkage clustering as previously described [74]. Gene set enrichment analysis was performed as previously described [75]. Gene set permutation was performed to calculate p-values and FDR due to small sample size.

Supporting information

S1 Fig. Frequent copy neutral loss of heterozygosity at the *Kras* locus with duplication of the mutant *Kras* \(^{G12D}\) allele. Leukemias were generated on a C57Bl/6 x 129Sv/Jae F1 strain background allowing us to analyze single nucleotide polymorphisms (SNPs) that differ between the two strains. Raw SNP allele frequencies are plotted against relative position on chromosome 6. The location of *Kras* is indicated with a diamond. *Kras* mutant IM T-ALLs frequently have loss of heterozygosity (AF = allele frequency), which is copy neutral (LR = log ratio copy number) and does not occur in wild-type *Kras* counterparts. (TIF)

S2 Fig. IM-induced T-ALL antigen expression is highly similar to human T-ALL patterns of expression. RNA-seq was performed on 9 primary IM-induced T-ALLs. The gene expression of IM T-ALLs is lineage specific with high gene expression of T-ALL specific antigens and low expression of B-ALL and/or myeloid specific antigens. TCR gene rearrangements are variable based on the primary leukemia (gray). (EPS)

S3 Fig. IM-induced T-ALL and human T-ALL have similar pathway enrichment compared to B-ALL. (A) A TCR pathway gene set are significantly enriched in murine IM-induced and human T-ALLs compared to B-ALL. (B) Gene sets associated with various subtypes of B-ALL are enriched in the respective subtype of B-ALL compared to either T-ALL or IM-induced T-ALL. (EPS)

S4 Fig. Clonal evolution following treatment with combination MEK and PI3K inhibitors. (A) Sequence reads aligned to *Pten* locus generated from WES of JW81 primary (top panel) versus relapsed leukemia (bottom panel). The *Pten* mutation is not detectable within the
primary tumor, whereas it is present at a 0.547 (read depth 137) allele frequency after treat-
ment with combination MEK and PI3K inhibitors. (B) $Kras^{G12D}$ allelic frequencies were deter-
mined based on Sanger sequencing and relative peak intensities of mutant versus wild-type
alleles. Loss of heterozygosity during leukemogenesis is a common event resulting in near loss
of wild-type $Kras$ in multiple independent leukemias. Outlier leukemia T-ALL 73M initially
lost WT $Kras$, but re-acquired $Kras^{G12D}$ heterozygosity following treatment with combination
MEK and PI3K inhibitors (red line). (C) Sequence reads aligned to $Kras$ and $Pten$ locus gener-
ated from WES of 73M primary versus relapsed leukemias. The $Kras$ allele frequency decreases
from 0.920 (read depth 138) to 0.547 (read depth 150), whereas the $Pten$ allele frequency
increases from 0.108 (read depth 102) to 0.618 (read depth 137) after treatment with combina-
tion MEK and PI3K inhibitors. (D) SNP allele frequencies plotted against relative position on
chromosome 6 ($Kras$) and 19 ($Pten$) for parental and relapsed 73M and (E) copy neutral num-
ber for both parental and relapsed 73M support uniparental disomy as the underlying mecha-
nism for increased mutant $Kras$ and $Pten$ allele frequencies.

Acknowledgments

We gratefully acknowledge the members of the Marie-Josée and Henry R. Kravis Center for
Molecular Oncology at Memorial Sloan Kettering Cancer Center, the DNA Sequencing Lab
and Departments of Bioinformatics & Computational Biology and Translational Oncology at
Genentech, and the Department of Computational Biology at St. Jude Children’s Research
Hospital. We obtained $Kras^{G12D}$ mice from Drs. David Tuveson and Tyler Jacks, and Dr.
Linda Wolff provided the MOL4070 retrovirus.

Author Contributions

Conceptualization: Benjamin J. Huang, Kevin Shannon.
Data curation: Benjamin J. Huang.
Formal analysis: Benjamin J. Huang.
Funding acquisition: James R. Downing, Jinghui Zhang, Barry S. Taylor, Kevin Shannon.
Investigation: Benjamin J. Huang, Anica M. Wandler, Lauren K. Meyer, Anneleen Daemen,
Xinyue Wang, Joy Nakitandwe.
Methodology: Benjamin J. Huang, Anica M. Wandler, Monique Dail, Qing Li, Jasmine C.
Wong.
Project administration: Benjamin J. Huang.
Resources: Deepak Sampath, Jinghui Zhang, Barry S. Taylor, Kevin Shannon.
Software: Benjamin J. Huang, Barry S. Taylor.
Supervision: Deepak Sampath, Jinghui Zhang, Barry S. Taylor, Kevin Shannon.
Validation: Benjamin J. Huang.
Visualization: Benjamin J. Huang.
Writing – original draft: Benjamin J. Huang, Kevin Shannon.
Writing – review & editing: Benjamin J. Huang, Anica M. Wandler, Lauren K. Meyer, Anne-
leen Daemen, Deepak Sampath, Barry S. Taylor, Kevin Shannon.
References

1. Day CP, Merlino G, Van Dyke T. Preclinical mouse cancer models: a maze of opportunities and challenges. Cell. 2015; 163(1):39–53. https://doi.org/10.1016/j.cell.2015.08.068 PMID: 26406370; PubMed Central PMCID: PMC4583714.

2. McGarahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. 2017; 168(4):613–28. https://doi.org/10.1016/j.cell.2017.01.018 PMID: 28187284.

3. Klco JM, Spencer DH, Miller CA, Griffith M, Lamprecht TL, O’Laughlin M, et al. Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. Cancer Cell. 2014; 25(3):379–92. https://doi.org/10.1016/j.ccr.2014.01.031 PMID: 24613412; PubMed Central PMCID: PMC3983786.

4. McFadden DG, Papagiannakopoulos T, Taylor-Weiner A, Stewart C, Carter SL, Cibulskis K, et al. Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell. 2014; 156(6):1298–311. https://doi.org/10.1016/j.cell.2014.02.031 PMID: 24630729; PubMed Central PMCID: PMC4040459.

5. Chung WJ, Daemen A, Cheng JH, Long JE, Cooper JE, Wang BE, et al. Kras mutant genetically engineered mouse models of human cancers are genomically heterogeneous. Proc Natl Acad Sci U S A. 2017; 114(51):E10947–E55. Epub 2017/12/06. https://doi.org/10.1073/pnas.1708391114 PMID: 29203670; PubMed Central PMCID: PMC5754767.

6. Uren AG, Kool J, Berns A, van Lohuizen M. Retroviral insertional mutagenesis: past, present and future. Oncogene. 2005; 24(52):7656–72. Epub 2005/11/22. https://doi.org/10.1038/sj.onc.1209043 PMID: 16299527.

7. Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature. 2005; 436(7048):221–6. https://doi.org/10.1038/nature03691 PMID: 16015321.

8. Mann KM, Ward JM, Yew CC, Kovochich A, Dawson DW, Black MA, et al. Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. Proc Natl Acad Sci U S A. 2012; 109(16):5934–41. https://doi.org/10.1073/pnas.1202490109 PMID: 22421440; PubMed Central PMCID: PMC3341075.

9. Friedrich MJ, Rad L, Bronner IF, Strong A, Wang W, Weber J, et al. Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. Nat Protoc. 2017; 12(2):289–309. https://doi.org/10.1038/nprot.2016.164 PMID: 28079877.

10. March HN, Rust AG, Wright NA, ten Hoeve J, de Ridder J, Eldridge M, et al. Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. Nat Genet. 2011; 43(12):1202–9. Epub 2011/11/08. https://doi.org/10.1038/ng.990 PMID: 22057237; PubMed Central PMCID: PMC3233530.

11. de Jong J, Akhtar W, Badhai J, Rust AG, Rad R, Hilkens J, et al. Chromatin landscapes of retroviral and transposon integration profiles. PLoS Genet. 2014; 10(4):e1004250. Epub 2014/04/12. https://doi.org/10.1371/journal.pgen.1004250 PMID: 24721906; PubMed Central PMCID: PMC3983033.

12. Kool J, Berns A. High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. Nat Rev Cancer. 2009; 9(6):389–99. Epub 2009/05/23. https://doi.org/10.1038/nrc2647 PMID: 19461666.

13. Wolff L, Koller R, Hu X, Arner MR. A Moloney murine leukemia virus-based retrovirus with 4070A long terminal repeat sequences induces a high incidence of myeloid as well as lymphoid neoplasms. J Virol. 2003; 77(8):4965–71. https://doi.org/10.1128/JVI.77.8.4965-4971.2003 PMID: 12663802.
19. Dail M, Li Q, McDaniel A, Wong J, Akagi K, Huang B, et al. Mutant Ikaros deficiency in T-cell leukemia. Nature. 2010; 467(7313):1055–60. PMID: 20419399; PubMed Central PMCID: PMC2884187.

20. Dail M, Wong J, Lawrence J, O’Connor D, Nakitandwe J, Chen SC, et al. Loss of oncogenic Notch1 with resistance to a PI3K inhibitor in T-cell leukemia. Nature. 2014; 513(7519):512–6. https://doi.org/10.1038/nature13495 PMID: 25043004; PubMed Central PMCID: PMC4213126.

21. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012; 150(2):264–78. https://doi.org/10.1016/j.cell.2012.06.023 PMID: 22817890; PubMed Central PMCID: PMC3407563.

22. Jeannet R, Mastio J, Macias-Garcia A, Oravecza A, Ashworth T, Geimer Le Lay AS, et al. Oncogenic activation of the Notch1 gene by deletion of its promoter in Ikaros-deficient T-ALL. Blood. 2010; 116 (25):5443–54. Epub 2010/09/11. https://doi.org/10.1182/blood-2010-05-286658 PMID: 20829372; PubMed Central PMCID: PMC3100247.

23. Taterek J, Cuillon K, Ashworth T, Gerstein R, Aster JC, Kelllerher MA. Notch1 inhibition targets the leukemia-initiating cells in a Tfl1/Lmo2 mouse model of T-ALL. Blood. 2011; 118(6):1579–90. Epub 2011/06/15. https://doi.org/10.1182/blood-2010-08-300343 PMID: 21670468; PubMed Central PMCID: PMC3156046.

24. Mansour MR. Oncogenic Kras and Notch-1 cooperate in T-cell acute lymphoblastic leukemia/lymphoma. Expert Rev Hematol. 2009; 2(2):133–6. Epub 2009/04/01. https://doi.org/10.1588/ehm.09.3 PMID: 20183447.

25. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500(7463):415–21. https://doi.org/10.1038/nature12477 PMID: 23945592; PubMed Central PMCID: PMC3776390.

26. Zhang J, Dier L, Holmfield L, Wu G, Heateley SL, Payne-Turner D, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukemia. Nature. 2012; 481(7380):157–63. https://doi.org/10.1038/nature10725 PMID: 22327106; PubMed Central PMCID: PMC3287575.

27. Mullighan CG, Phillips LA, Su X, Ma J, Miller CB, Shurtleff SA, et al. Genomic analysis of the clonal origin of relapsed acute lymphoblastic leukemia. Science. 2008; 322(5906):1377–80. https://doi.org/10.1126/science.1164266 PMID: 19039135; PubMed Central PMCID: PMC2746051.

28. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LA, Miller CB, et al. Deletion of Ikaros and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009; 360(5):470–80. https://doi.org/10.1056/NEJMoa089253 PMID: 19138520; PubMed Central PMCID: PMC2674612.

29. Ksionda O, Melton AA, Bache J, Tenhagen M, Baikker J, Harvey R, et al. RasGRP1 overexpression in T-ALL increases basal nucleotide exchange on Ras rendering the Ras/PI3K/Akt pathway responsive to protumorigenic cytokines. Oncogene. 2016; 35(28):3658–68. https://doi.org/10.1038/onc.2015.431 PMID: 26549032; PubMed Central PMCID: PMC4868787.

30. Burgess MR, Hwang E, Mroue R, Bielski CM, Wangler AM, Huang BJ, et al. KRAS Allelic Imbalance Enhances Fitness and Modulates MAP Kinase Dependence in Cancer. Cell. 2017; 168(5):817–29 e15. https://doi.org/10.1016/j.cell.2014.03.027 PMID: 24766807; PubMed Central PMCID: PMC4098792.

31. Kerr EM, Guade E, Turrell FK, Frezza C, Martins CP. Mutant Kras copy number defines metabolic reprogramming and therapeutic susceptibilities. Nature. 2016; 531(7592):110–3. https://doi.org/10.1038/nature16967 PMID: 26909577; PubMed Central PMCID: PMC4780242.

32. Papa A, Wan L, Bonora M, Salmena L, Song MS, Hobbs RM, et al. Cancer-associated PTEN mutants act in a dominant-negative manner to suppress PTEN protein function. Cell. 2014; 157(3):595–610. https://doi.org/10.1016/j.cell.2014.03.027 PMID: 24766807; PubMed Central PMCID: PMC4098792.

33. Cauwelber B, Dastugue N, Cools J, Poppe B, Herens C, De Paepe A, et al. Molecular cytogenetic study of 126 unselected T-ALL cases reveals high incidence of TCRbeta locus rearrangements and putative new T-cell oncogenes. Leukemia. 2006; 20(7):1238–44. https://doi.org/10.1038/sj.leu.2404243 PMID: 16673021.

34. Ross ME, Zhou X, Song G, Shurtleff SA, Girtman K,Williams WK, et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. Blood. 2003; 102(8):2951–9. https://doi.org/10.1182/blood-2003-01-0338 PMID: 12730115.

Convergent genetics of mouse and human leukemias
48. Petit A, Trinquand A, Chevret S, Ballerini P, Cayuela JM, Grardel N, et al. Oncogenetic mutations combined with MRD improve outcome prediction in pediatric T-cell acute lymphoblastic leukemia. Blood. 2014; 131(3):289–300. Epub 2017/10/21. https://doi.org/10.1182/blood-2017-04-778829 PMID: 29051182.

49. Trinquand A, Tanguy-Schmidt A, Ben Abdelali R, Lambert J, Beldjord K, Lengline E, et al. Toward a NOTCH1-FBXW7/RAS/PTEN-based oncogenetic risk classification of adult T-cell acute lymphoblastic leukemia: a Group for Research in Adult Acute Lymphoblastic Leukemia (GALLO) study. J Clin Oncol. 2013; 31 (34):4333–42. Epub 2013/10/30. https://doi.org/10.1200/JCO.2012.48.5292 PMID: 24166518.

50. Lu SX, Abdel-Wahab O. Genetic drivers of vulnerability and resistance in relapsed acute lymphoblastic leukemia. Proc Natl Acad Sci U S A. 2016; 113(40):11071–3. Epub 2016/09/25. https://doi.org/10.1073/pnas.1613836113 PMID: 27663730; PubMed Central PMCID: PMC5056046.

51. Clappier E, Gerby B, Sigaux F, Delord M, Touzri F, Hernandez L, et al. Clonal selection in xenografted human T cell acute lymphoblastic leukemia recapitulates gain of malignancy at relapse. J Exp Med. 2011; 208(4):653–61. Epub 2011/04/06. https://doi.org/10.1084/jem.20110105 PMID: 21464223; PubMed Central PMCID: PMC3135355.

52. Marks DJ, Rowntree C. Management of adults with T-cell lymphoblastic leukemia. Blood. 2017; 129 (9):1134–42. Epub 2017/01/25. https://doi.org/10.1182/blood-2016-07-692608 PMID: 28115371.

53. Herranz D, Ambesi-Impimbiotta A, Suderitch J, Sanchez-Martín M, Belver L, Tosello V, et al. Metabolic reprogramming induces resistance to anti-NOTCH1 therapies in T cell acute lymphoblastic leukemia. Nat Med. 2015; 21(10):1182–9. Epub 2015/09/22. https://doi.org/10.1038/nm.3955 PMID: 26390244; PubMed Central PMCID: PMC4598309.

54. Papayanidis C, DeAngelio DJ, Stock W, Huang B, Shaik MN, Cesari R, et al. A Phase 1 study of the novel gamma-secretase inhibitor PF-0304014 in patients with T-cell acute lymphoblastic leukemia and T-cell lymphoblastic lymphoma. Blood Cancer J. 2015; 5:e350. Epub 2015/09/26. https://doi.org/10.1038/bcj.2015.80 PMID: 26407235; PubMed Central PMCID: PMC4648526.
55. Knoechel B, Bhatt A, Pan L, Pedamallu CS, Severson E, Gutierrez A, et al. Complete hematologic response of early T-cell progenitor acute lymphoblastic leukemia to the gamma-secretase inhibitor BMS-906024: genetic and epigenetic findings in an outlier case. Cold Spring Harb Mol Case Stud. 2015; 1(1):a000539. Epub 2016/05/06. https://doi.org/10.1101/mcs.a000539 PMID: 27148573; PubMed Central PMCID: PMC4850884.

56. Al-Ahmadie H, Iyer G, Hohi M, Asthana S, Inagaki A, Schultz N, et al. Synthetic lethality in ATM-deficient RAD50-mutant tumors underlies outlier response to cancer therapy. Cancer Discov. 2014; 4(9):1014–21. https://doi.org/10.1158/2159-8290.CD-14-0380 PMID: 24934408; PubMed Central PMCID: PMC4155059.

57. Iyer G, Hanrahan AJ, Milowsky MI, Al-Ahmadi H, Scott SN, Janakiraman M, et al. Genome sequencing identifies a basis for everolimus sensitivity. Science. 2012; 338(6104):228–31. https://doi.org/10.1126/science.1226344 PMID: 22923433; PubMed Central PMCID: PMC3633467.

58. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324 PMID: 19451168; PubMed Central PMCID: PMC2705234.

59. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43(5):491–8. https://doi.org/10.1038/ng.806 PMID: 21478889; PubMed Central PMCID: PMC3083463.

60. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013; 31(3):213–8. https://doi.org/10.1038/nbt.2514 PMID: 23360613; PubMed Central PMCID: PMC3833702.

61. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009; 25(21):2865–71. https://doi.org/10.1093/bioinformatics/btp99 PMID: 19561018; PubMed Central PMCID: PMC2781750.

62. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29(1):24–6. https://doi.org/10.1038/nbt.1754 PMID: 21221095; PubMed Central PMCID: PMC3346182.

63. Akagi K, Suzuki T, Stephens RM, Jenkins NA, Copeland NG. RTCGD: retroviral tagged cancer gene database. Nucleic Acids Res. 2004;32(Database issue):D523-7. https://doi.org/10.1093/nar/gkh013 PMID: 14681473; PubMed Central PMCID: PMC308748.

64. Curtins NP, Bonifas JM, Lauchle JO, Balkman JD, Kratz CP, Emerling BM, et al. Isolation and analysis of candidate myeloid tumor suppressor genes from a commonly deleted segment of 7q22. Genomics. 2005; 85(5):600–7. https://doi.org/10.1016/j.ygeno.2005.01.013 PMID: 15820312.

65. Vohra S, Biggin PC. Mutationmapper: a tool to aid the mapping of protein mutation data. PLoS One. 2013; 8(8):e71711. https://doi.org/10.1371/journal.pone.0071711 PMID: 23951226; PubMed Central PMCID: PMC3739722.

66. Forbes SA, Beare D, Boutsalakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2011; 39(Database issue):D777–D83. https://doi.org/10.1093/nar/gkw1121 PMID: 27899578; PubMed Central PMCID: PMC3177186.

67. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011; 39(17):e118. https://doi.org/10.1093/nar/gkr1074 PMID: 21221095; PubMed Central PMCID: PMC3177186.

68. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7(4):248–9. https://doi.org/10.1038/nmeth0410-248 PMID: 20354512; PubMed Central PMCID: PMC2855889.

69. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012; 40(Web Server issue):W452–7. https://doi.org/10.1093/nar/gks539 PMID: 22689647; PubMed Central PMCID: PMC3394338.

70. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9(4):357–9. https://doi.org/10.1038/nmeth.1923 PMID: 22388286; PubMed Central PMCID: PMC322381.

71. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12:323. https://doi.org/10.1186/1471-2105-12-323 PMID: 21816040; PubMed Central PMCID: PMC3163565.

72. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4(2):249–64. https://doi.org/10.1093/biostatistics/4.2.249 PMID: 12925520.

73. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010; 11(3):R25. https://doi.org/10.1186/gb-2010-11-3-r25 PMID: 20196867; PubMed Central PMCID: PMC2864565.
74. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998; 95(25):14863–8. https://doi.org/10.1073/pnas.95.25.14863 PMID: 9843981; PubMed Central PMCID: PMC24541.

75. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102(43):15545–50. https://doi.org/10.1073/pnas.0506580102 PMID: 16199517; PubMed Central PMCID: PMC1239896.