Using RNA-seq and targeted nucleases to identify mechanisms of drug resistance in acute myeloid leukemia

Susan K. Rathe1, Branden S. Moriarity1,2,3, Christopher B. Stoltenberg1, Morito Kurata1, Natalie K. Aumann1, Eric P. Rahrman1,2,3, Natasha J. Bailey1, Ellen G. Melrose1, Dominic A. Beckmann2,3, Chase R. Liska1 & David A. Largaespada1,2,3,4,5

1Masonic Cancer Center, University of Minnesota, Minneapolis, MN, USA, 2Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, USA, 3Center for Genome Engineering, University of Minnesota, Minneapolis, MN, USA, 4Brain Tumor Program, University of Minnesota, Minneapolis, MN, USA, 5Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA.

The evolution from microarrays to transcriptome deep-sequencing (RNA-seq) and from RNA interference to gene knockouts using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) and Transcription Activator-Like Effector Nucleases (TALENs) has provided a new experimental partnership for identifying and quantifying the effects of gene changes on drug resistance. Here we describe the results from deep-sequencing of RNA derived from two cytarabine (Ara-C) resistance acute myeloid leukemia (AML) cell lines, and present CRISPR and TALEN based methods for accomplishing complete gene knockout (KO) in AML cells. We found protein modifying loss-of-function mutations in Dck in both Ara-C resistant cell lines. CRISPR and TALEN-based KO of Dck dramatically increased the IC50 of Ara-C and introduction of a DCK overexpression vector into Dck KO clones resulted in a significant increase in Ara-C sensitivity. This effort demonstrates the power of using transcriptome analysis and CRISPR/TALEN-based KOs to identify and verify genes associated with drug resistance.

The 12,000+ patients diagnosed with acute myeloid leukemia (AML) in the United States each year face a dismal prognosis. The induction chemotherapy, which will most likely result in a remission, is typically not curative. However, induction chemotherapy can significantly reduce blast cells providing the clinician with additional time to try other therapies. Unfortunately, the additional therapies are generally not effective at achieving a long-term durable remission. At relapse, most patients will no longer respond to induction therapy, since the leukemic clones surviving the initial onslaught of induction chemotherapy have an innate resistance, and have therefore become the prevalent disease cells.

Arabinoside cytarabine (Ara-C) has been the primary component of induction chemotherapy for over 40 years. Ara-C, a cytidine analog, enters the cell via the dNTP salvage pathway, where it is metabolically activated by the addition of three phosphates in the same manner as cytidines. Each phosphate is added by a different kinase. The first kinase in the dNTP salvage pathway is deoxycytidine kinase (DCK), the rate limiting enzyme in the metabolic activation of Ara-C. Numerous studies have shown DCK expression is frequently downregulated in cells that are unresponsive to Ara-C2–7.

In a previous publication, we reported the results of a microarray gene expression analysis, which compared gene expression of two Ara-C resistant cell lines (B117H and B140H) with their respective Ara-C sensitive parental cells lines (B117P and B140P). The B117H and B140H cells tolerated concentrations of Ara-C 500–1000 times that of their parental lines. The most dramatic common change identified by the microarray study was the significant downregulation of Dck.

Here we report the results of a subsequent RNA sequencing of the transcriptome of the same four murine AML cell lines (B117P, B117H, B140P, and B140H). RNA-seq analysis uncovered evidence to the nature of the Dck functional impairment in both the B117H cells and the B140H cells: a large deletion of DNA spanning the splice acceptor of the last exon of Dck and a frameshift mutation in the fourth exon of Dck, respectively. Both mutations resulted in aberrant RNA transcripts for Dck. RNA-seq also identified gene expression changes not previously detected by gene expression microarray.
A CRISPR screen to knockout (KO) genes identified as down-regulated in the Ara-C resistant cell lines identified Dck as the primary contributor to Ara-C resistance. Total KO of Dck using Transcription Activator-Like Effector Nucleases (TALENs) in the B117P cells confirmed the loss of Dck expression was nearly sufficient for the high Ara-C IC50 levels found in the Ara-C resistant cell lines. Introduction of an inducible DCK overexpression vector in the B117P Dck KO clones restored most of the original Ara-C sensitivity.

This research demonstrates the value of using RNA-seq methods to identify changes in cells as they become resistant to drugs and provides two new methods for generating candidate drug resistant gene KOs in difficult-to-transfect AML cells using doxycycline inducible CRISPRs with puromycin selection and TALENs with single step drug selection.

Results
RNA-sequencing identifies more gene expression changes than microarray hybridization. Samples of RNA had previously been isolated from 2 murine BXH-2 AML cell lines and their Ara-C resistant derivatives, and then evaluated by microarray. Aliquots of RNA from the microarray experiment were submitted for RNA-sequencing (RNA-seq). TopHat was used to map the data to the mouse transcriptome (NCBI37/mm9), and the quality of the mapping was tested using Picard-tools. All samples had over 20 million paired reads with over 90% mapped and over 89% uniquely mapped (Supplementary Table S1). Cuffdiff was used to determine changes common to both Ara-C resistant cell lines (B117H and B140H) when compared to their parental lines (B117P and B140P). To avoid division by zero, a minimum FPKM was established at 0.001 based on FPKM distribution patterns (Supplementary Figure S1). These patterns also showed genes expressed in just one sample, a phenomenon not seen when studying microarray expression data due to the presence of background noise. Genes where both the parental and its Ara-C resistant derivative had FPKM levels less than 0.5 were excluded from the analysis, since even technical replicates display a high degree of variability at these lower expression levels. Integrated Genomic Viewer (IGV; http://www.broadinstitute.org/igv/) was then used to eliminate false positives, which included distortions due to reads mapping outside the normal transcription area, a high abundance of non-unique reads, and projected non-protein coding RNA sequences.

The previous microarray analysis identified 8 genes with expression levels with 2X or more fold changes. In comparison the RNA-seq method identified 60 genes. Seven genes appeared in both lists (Figure 1a). Genes identified by RNA-seq with a 3X or more fold change in both sets of cells (B117H vs. B117P and B140H vs. B140P) are listed in Table 1, while the greater than 2-fold and less than 3-fold change genes are included in Supplementary Table S2. Genes in bold were also identified by gene expression microarray with 2-fold + changes in expression. The only gene identified by microarray and not by RNA-seq was Psph, where the expression did not meet the 2-fold threshold in the RNA-seq analysis. The RNA-seq list includes Dck, the only gene appearing in the microarray data as being changed by more than 5-fold. The expression levels of Dck were verified by qPCR (Figure 1b). Of the 53 genes identified by RNA-seq but not by microarray, 3 genes did not have a probe designed on the microarray chip (2310007A19Rik, 2210417A02Rik, and AI427809), 1 gene had expression levels so high it probably saturated the microarray chips (Mpo), 10 genes had probes lacking specificity, 22 had expression levels too low for the microarray chips to distinguish significant differences (including Dup2lip), and 15 just missed the 2-fold cutoff in the microarray analysis, most likely due to the distortive effect of background noise. As for the last 2 genes, microarray was unable to distinguish between Ly6c2 and Ly6c1 expression levels due to sequence similarities between these two genes, or between Gng5 and its family members.

We next looked at changes unique to either the B117H or the B140H cells when compared to their individual parental cell lines, B117P and B140P, respectively. These lists were significantly longer than those generated by microarray, so we confined our analysis to

Figure 1 | Dck expression patterns and protein levels verified by IGV, qPCR and Western blot. (a) Venn diagram depicting the overlap in the genes identified as having a greater than 2-fold expression change in both sets of cell lines (B117H vs. B117P and B140H vs. B140P) when evaluated by gene expression microarray and RNA-seq. (b) Reduced expression of Dck in the BXH-2 cell lines was confirmed by qPCR using primers designed to span exons 5 and 6 of Dck. Error bars depict range. Two tailed T-test used to determined p-values (n = 3). (c) Sanger sequencing of the RNA in B140H cells verified an insertion mutation in exon 4. (d) Western blot of Dck protein levels in the BXH-2 cell lines. Cropped images presented in figure. Full-length blots can be found in Supplementary Figure S6a.
the gene expression changes of 100-fold or more (Supplementary Tables S3 and S4). Again due to the distortion caused by background noise only one gene had a 100X change within the microarray data (Ddx3y in B117 cells).

RNA-sequencing identifies mutations in Dck in the B117H cells and B140H cells. Missense Mutation and Frameshift Location Finder (MMuFLR)\(^1\), a Galaxy\(^1^4\)–16 based workflow developed to look for frameshift and missense mutations, was used to identify mutations in the Ara-C resistant cell lines that were not present in their respective parental cell lines. MMuFLR identified a single thymidine insertion in exon 4 of Dck following the 462\(^{nd}\) nt from the translational start site in the B140H cells, which would result in a severely truncated protein. Sanger sequencing showed the insertion was present in all expressed Dck transcripts (Figure 1c) and homozygous in the genomic DNA (Supplementary Figure S2a). The nearly complete elimination of Dck protein was verified by Western blot (Figure 1d). The genomic insertion would result in the translation of a protein with the 8 amino acids that would have been translated from exon 7 (Figure 1d), but the WB technique was not sensitive enough to detect a size change.

Mutation analyses tools identify other mutations acquired in Ara-C resistant cells. In addition to the frameshift mutation identified in Dck of B140H, another frameshift was identified by MMuFLR as being introduced in Ccdc88b of B140H (Supplementary Table S6). The Ccdc88b frameshift was shown to be heterozygous by Sanger sequencing. No frameshifs were detected in the B117H cells that were not also present in the B117P cells (Supplementary Table S7).

MMuFLR also identified 21 mutations introduced into either the B117H cells or the B140H cells (Table 2). The potential for functional changes to comparable proteins in human cells was examined using both PolyPhen-2\(^1^7\) and PROVEAN Protein\(^2^0\). The genes identified by PolyPhen-2 as “probably damaging” or by PROVEAN Protein as “deleterious” appear in bold in Table 2.

deFuse\(^2^1\) was used to look for the introduction of any protein modifying fusions in the Ara-C resistant cell lines, when compared to their respective parental lines. No protein modifying fusions were found as newly introduced into the Ara-C resistant cells (Supplementary Table S8).

CRISPR screen identifies loss of Dck as the primary contributor to Ara-C resistance. To determine which changes contribute to Ara-C resistance, a CRISPR screen was conducted on 7 genes identified as downregulated by more than 3-fold in both the B117H and B140H (Ly6c2, Dab2ip, Dck, Ksr1, Riad1, Cd14, and Mpo), as well a gene containing a frameshift mutation (Ccdc88b) and 8 genes containing potentially deleterious missense mutations (Kdm5c, Prkacb, Pus7l, Rasgrfp2, Vps33b, Atp6vd4, Pbrm1, and Smarca4). The target sequences for the gRNAs are listed in Supplementary Table S9.
The Ly6c2 gRNAs would also target Ly6c1. The CRISPR-Cas9 cloning vector is described in Figure 3a. Only Dck demonstrated a shift in response to Ara-C (Figure 3b). A CEL-I assay was performed on DNA from the Dck CKO-2 cells to confirm doxycycline inducible Cas9 activity (Supplementary Figure S5). Partial suppression of Dck using RNAi results in an increase of the IC₅₀ for Ara-C. To test whether the downregulation of Dck alone can change a cell’s response to Ara-C, knockdowns of Dck were performed in the parental cell lines, B117P and B140P, using OpenBiosystems shRNA constructs. Two TRC constructs for Dck were used, one targeted exon 6 (KD1) and the other targeted a sequence spanning exon 2 and 3 (KD2). The Dck knockdowns were verified by qPCR (Figure 3c). Drug assays were used to determine the Ara-C IC₅₀ in the knockdown cell lines. The Ara-C IC₅₀’s were higher in the cell lines with the greater downregulation of Dck (Figure 3c). As controls, knockdowns of Nfkb1 and p53, as well as introduction of GFP and empty vectors, were performed on the B117P cells. No significant change in IC₅₀ for Ara-C was observed (Supplementary Table S10).

Total KO of Dck using TALENs results in a significant increase of the IC₅₀ for Ara-C. TALENs targeting Dck were generated and used to knock-out (KO) Dck in B117P cells (Figure 3d). Single cell clones were grown out, selected for homozygous deletion mutations, and tested for Ara-C sensitivity. Ara-C IC₅₀s in the Dck knockouts were comparable to the Ara-C IC₅₀s in the Ara-C resistant cell lines (Figure 3c). The location of the deletion in the DNA of each of the B117P KO clones (T2A, T6B, and T11A) was determined by PCR amplification of the TALEN target site and Sanger sequencing.

Figure 2 | Sequence abnormalities in B117H Ara-C resistant cells verified by Sanger sequencing. (a) RNA-seq reads were mapped to the NCBI reference genome. Visualization of Dck expression using IGV. (b) Transcripts were assembled independently of a reference transcriptome using Cufflinks, then mapped to the mm9 mouse genome using Cuffcompare, and the resulting gtf file was visualized by IGV. (c) Sanger sequencing of DNA in B117H cells identified an 878 nt deletion spanning the splice acceptor of intron 6 and the translated portion of exon 7. Sanger sequencing of RNA verified a transcript matching the configuration identified by TopHat and IGV.
than the parental cell lines from which they were derived. We the-

Ara-C, tolerating concentrations of Ara-C 500–1000 times greater

The B117H and B140H cells used in this study are highly resistant to

Discussion

Rescue of Dck expression in Dck KO clones results in a decrease of

Western blot (Figure 3g).

and the second band was an off target amplification of another

alternatively spliced version of the T6B clone were sequenced. The top band revealed an

which resulted in multiple light bands (Figure 3f). The top two bands

clones to look for transcript changes within the first 3 exons of

(Figure 3e). RT-PCR was performed on RNA-derived cDNA of KO

doxycycline, the cells exhibited an Ara-C IC50 slightly lower than the

Dck KO cells (Figure 4b). Inducing DCK overexpression vector (Figure 4c) was stably integrated into to the

drug response would allow us to focus

on the most prominent change in the cells. RNA samples, previously

array probes were designed to detect. However, the IGV visualization

9

-UTR was missing, but only in the areas where the micro-

array probes were designed to detect. However, the IGV visualization

RNA-seq's unbiased approach to expression analysis has the poten-
tial to identify expression changes in genes not represented on micro-

array chips, and to detect expression levels at an mRNA isoform level.

RNA-seq has found at least one other critical advantage over microarray analysis. It has the potential to identify RNA variants, such as unusual transcripts, fusions, frameshifts and missense muta-
tions. The RNA-seq results were instrumental in identifying the

unusual changes to the Dck locus in both the B117H and B140H cell

lines. In the B117H cells, microarray data had previously shown part of the 3'-UTR was missing, but only in the areas where the micro-

array probes were designed to detect. However, the IGV visualization of the RNA-seq data specifically showed expression in the first part of

intron 6, and a small section of expression in the 3'-UTR, as well as a large transcribed section beyond the exon encoding the 3'-UTR. The

sequence of amino acids at the C-terminus of Dck forms an alpha-

helix, which is highly conserved across various species. The end of the C-terminus is in close proximity to a number of residues (Ile24, A119, and Pro122) important for Dck kinase activity. The

replacement of 8 amino acids by 20 amino acids in the case of the

transcript in B117H may result either in instability of

Dck transcript or interference with residues important to Dck's function. The frameshift mutation identified by MMuFLR in exon 4 of Dck in the B140H cells would result in a severely truncated

version of the Dck protein.

Although RNA-seq is clearly a technological advancement from microarrays, it is not without problems or limitations. For example, the sequencing process has difficulty determining the correct num-

ber of nucleotides when reading through a poly-A or poly-T

sequence, which can lead to the identification of small indels that

do not really exist, an artifact referred to as "stuttering". MMuFLR

(4:6048). The B117H and B140H cells used in this study are highly resistant to Ara-C, tolerating concentrations of Ara-C 500–1000 times greater than the parental cell lines from which they were derived. The top band revealed an

alternatively spliced version of Dck (Supplementary Figure S4), and the second absence of Dck protein in the KO clones was verified by Western blot (Figure 3g).

Rescue of Dck expression in Dck KO clones results in a decrease of the IC50 for Ara-C. A doxycycline inducible human DCK overexpression vector (Figure 4c) was stably integrated into the three B117P Dck KO cell lines (T2A, T6B, and T11A) using the piggyBac transposon system. Doxycycline induction of Dck expression was confirmed by qPCR (Figure 4a). In the absence of
doxycycline, the cells exhibited an Ara-C IC50 slightly lower than the

Dck KO cells (Figure 4b). Inducing DCK with doxycycline resulted in a significant reduction in the Ara-C IC50. Gene expression levels were measured by qPCR and the presence of DCK protein was confirmed by Western Blot (Figure 4d).

Table 2 | Protein modifying missense mutations found in the B117H and B140H cells, but not found in the respective parental cell lines. Predicted detrimental mutations in bold

| Cell Line | Gene | Variant | Position | Ref. | Var. | Prevalence | Seq. | Protein |
|-----------|------|---------|----------|------|------|------------|------|---------|
| B117H     | Commd5 | chr15:7630961 | T | G | 0.37 | 38 | L43V |
|          | Dram1 | chr10:8780336 | G | A | 0.53 | 17 | V91I |
|          | Kdm3b | chr18:34987086 | T | C | 0.40 | 30 | S1157P |
|          | Kdm5c | chrX:148705903 | G | T | 0.33 | 39 | V1222L |
|          | Pigt | chr2:164125988 | C | T | 0.47 | 108 | P385S |
|          | Prkacb | chr3:146409579 | G | T | 0.44 | 94 | E209D |
|          | Pus7l | chr15:94358455 | C | T | 0.37 | 19 | R499C |
|          | Pyrox1 | chr1:142303201 | A | G | 0.43 | 7 | N179S |
|          | Rasgrp2 | chr19:64607273 | G | T | 0.41 | 99 | S349I |
|          | Suda | chr4:59328600 | T | A | 0.35 | 72 | S748T |
|          | Vps33b | chr7:87432008 | G | A | 0.37 | 27 | D414N |
|          | Atpdb4 | chr2:114470474 | C | G | 0.45 | 33 | L99V |
|          | Chad7 | chr4:8678934 | A | G | 0.50 | 6 | N95D |
|          | Map3k11 | chr19:5700807 | G | A | 0.38 | 52 | M684I |
|          | Mefl16 | chr11:74609506 | G | A | 0.34 | 67 | R203Q |
|          | Pbrm1 | chr14:31863259 | A | C | 0.37 | 62 | S375R |
|          | Rnf149 | chr1:39609321 | G | A | 0.49 | 142 | C94Y |
|          | Smarca4 | chr9:21445509 | G | A | 0.49 | 164 | E525K |
|          | Spag5 | chr11:78118221 | C | T | 0.42 | 73 | S284L |
|          | Usp34 | chr11:23321104 | C | G | 0.50 | 38 | L1812V |
|          | Zfp668 | chr7:135009904 | G | A | 0.43 | 37 | R542Q |
has parameters that can be set to ignore this type of error. On the analysis side, software tools to interpret the RNA-seq data are still in their early developmental stages, as are the tables used to characterize the data, such as the tables identifying SNPs and isoforms. Normalizing RNA-seq data between samples is also providing a challenge, and many efforts are underway to improve normalization techniques.\textsuperscript{10,24} Using the FPKM normalization technique provided by Cuffdiff was adequate for comparing drug resistance derivatives to parental cell lines, as in this study, where the samples were all prepared at the same time using the same technique.\textsuperscript{20} Although there is no agreement on which approach is more accurate for measuring differential expression changes (microarray, RNA-seq, or qPCR), we
did find RNA-seq, based on the quality measurement of the reads, perfectly and uniquely mapped reads to genes that could not be verified by qPCR due to the inability to create primers specific to the gene in question. It is also interesting to note that despite the exposure to extreme levels of Ara-C and the mutagenic nature of the BXH-2 cells being used in the study, there were no protein modifying fusions introduced to the Ara-C resistant cell lines.

We elected to use CRISPRs to test the candidate genes for their involvement in Ara-C resistance. The generation of CRISPR gRNAs is an easier and cheaper technique than creating TALENs, but it is generally agreed TALENs are more specific than CRISPRs25,26, so we used TALENs to generate Dck KOs without the concern of off-target modification. It was not surprising to find mutations in Dck were the primary common changes to the Ara-C resistant cell lines and the level of Dck expression correlated to the Ara-C IC_{50} level, since Dck is the rate limiting enzyme in the dNTP salvage pathway, which is required for the metabolic activation of Ara-C. An alternatively spliced version of DCK was also found in a study of Ara-C resistant human acute lymphoblastic cell lines2, and downregulation of DCK was discovered in Ara-C resistant human acute myeloid cell lines5. In the clinical setting, the presence of DCK SNPs and alternatively spliced versions of DCK have been correlated to patient response to chemotherapy6,7. Since Dck was both downregulated and mutated in the Ara-C resistant B117H and B140H cell lines, we suspect during the process of creating the Ara-C resistant cell lines, by adding increasing concentrations of Ara-C, the cells initially responded by downregulating expression of Dck. Eventually the Ara-C concentrations became so high only the cells with defective Dck were able to survive.

Although changes in DCK expression have been associated with Ara-C resistance, the importance of DCK regulation in cells exposed to Ara-C has not been quantified. The significance of DCK regulation alone to Ara-C resistance was demonstrated by the knockdown and KO of Dck in this study. The use of new and powerful nuclease-based techniques to specifically and completely KO Dck in the B117P cells proved this single modification can account for over 85% of the Ara-C resistance present in the B117H cells. Rescue of DCK expression significantly reduced Ara-C resistance. The failure of the DCK overexpression to return the B117P Dck KO cells to its original Ara-C IC_{50} level may be due to the use of a human version of DCK, which may vary from mouse Dck in its kinetics or activation/repression methods. As evidence to this, the DCK protein levels in the dox-induced DCK overexpression cell lines were significantly less than the Dck levels found in the B117P cells. The ability to reintroduce DCK to the cells to restore most of the Ara-C sensitivity also indicated most of the changes, which took place when Dck was lost, were not permanent. However, it is possible other gene alterations or gene expression changes partially contributed to the residual Ara-C resistance in these cell lines.

The effect of loss of Dck function has yet to be fully quantified in human patient samples. Review of the DCK expression in a microarray study of 461 de novo AML patients under the age of 61 showed.

Figure 4 | Rescue of Dck expression in the B117P KO clones results in a decrease of the IC_{50} value for Ara-C. (a) qPCR of DCK expression in Dck KO clones plus DCK overexpression (OE) with and without DCK expression activation by doxycycline. Error bars depict range. Two tailed T-test used to determined p-values (n=3). (b) Ara-C IC_{50} levels in Dck KO clones plus DCK overexpression (OE) with and without DCK expression activation by doxycycline, as determined by MTS-Tetrazolium assay. Error bars depict range. Two tailed T-test used to determined p-values (n=3). (c) Structure of the doxycycline inducible DCK overexpression vector. (d) Western blot of Dck/DCK proteins levels in the TALEN-based Dck KO clones with doxycycline induced DCK overexpression. Protein levels shown with and without doxycycline activation of DCK. Cropped images presented in figure. Full-length blots can be found in Supplementary Figure S6c.
10% of the patients had DCK expression levels lower than 2-fold from the median expression level. Since AML samples consist of a heterogeneous population of cells, it is conceivable that in the samples with reduce DCK expression, there exist cells will little or no DCK expression. Unfortunately, there have been no large scale studies quantifying the expression levels of DCK or the presence of DCK mutations in refractory AML patients. Since DCK forms a homodimer, a mutation affecting protein function in just one allele could have the same effect as a 4-fold downregulation of the transcript.

This study illustrates normal Dck functionality is critical for Ara-C responsiveness in murine AML cell lines. These cell lines provide a model for understanding the clinical response to Ara-C and the development of Ara-C resistant AML. It demonstrates the many ways by which Dck function can be altered by mutation. Further analysis of transcriptome changes in the Dck knockout cell lines will provide a better understanding of the changes taking place in the cells to compensate for the loss of Dck. This will be crucial in identifying drug targets to prevent the expansion of AML cells with defective Dck function. The CRISPR/TALEN-based KO techniques described here are especially suited to test each of the gene targets identified in these subsequent research efforts.

Methods

Cell culture. The B117P, B117H, B140P, and B140H (murine BXH-2) cell lines were maintained at 37°C in 10% CO2 in AS4 in the absence of Ara-C. Cells were passaged three times each week and were replaced from frozen stocks every 3–4 months, to minimize genetic drift. Knockdown cells (KDs) were maintained in the same manner, but in the presence of selective doses of puromycin (1.5 µg/ml for the B117P KDs and 0.6 µg/ml for the B140P KDs). Ara-C was acquired from Bedford Laboratories (Bedford, OH).

Transcriptome deep sequencing and analysis. Aliquots of RNA derived from the cells designated “passage B” of B117P, B117H, B140P, and B140H (GSM457359, GSM457362, GSM457365, and GSM457368, respectively) for the previously published microarray experiment were submitted for transcriptome sequencing. The RNA was quality tested using a Bioanalyzer 2100 (Agilent Technologies, Santa Clarra, CA). cDNA was created by reverse transcription of oligo-dT purified polyadenylated RNA. The cDNA was fragmented, blunt-ended, and then ligated to barcoded adaptors. Lastly, the library was size selected, and the selection process was validated and quantified by capillary electrophoresis and qPCR, respectively. Sequencing was accomplished on the HiSeq 2000 (Illumina Inc., San Diego, CA), with the goal of generating a minimum of 20 million paired-end 76 bp reads. The resulting data was loaded into Galaxy. TopHat 2.0.5 was used to map the paired reads to the NCBI37/mm9 assembly of the mouse genome. The mean inner distance was determined using the insert size metrics feature of Picard-tools (http://picard.sourceforge.net). Other than stipulating the use of NCBI mouse genes for the gene annotation, the default parameters (as defined by the University of Minnesota’s Galaxy implementation) were used. The resulting TopHat data served as input to other analytical tools, which compared data from B117P to B117H, and B140H to B140P. Visualization of the mapped reads was accomplished using the Integrative Genomic Viewer (http://www.broadinstitute.org/igv). Gene expression was calculated using Cufflinks tools. Cuffdiff was executed from Galaxy using default parameters. Transcripts were also visualized and quantified by capillary electrophoresis and qPCR, respectively. Sequencing was performed using a DNeasy Blood & Tissue Kit (QIAGEN). The resulting samples were quantified using a NanoDrop™ 1000 Spectrometer (Thermo Fisher Scientific Inc.). DNA samples were stored at −20°C.

cDNA was prepared from RNA using the Invitrogen™ Superscript™ III First-Strand Synthesis System (Life Technologies Corporation, Carlsbad, CA). DNA (or cDNA) was amplified using Taq DNA Polymerase (QIAGEN), and separated by gel electrophoresis on a 1% agarose gel. The DNA was extracted from the resulting bands using the UltraClean™ GelSpin DNA Extraction Kit (MO BIO Laboratories, Inc., Carlsbad, CA). Classic Sanger sequencing was done using the ABI PRISM® 3730xl DNA Analyzer (Life Technologies Corporation).

The DNA samples from the B117P and B117H cell lines were amplified using the Expand Long Template PCR System (Roche Applied Systems, Indianapolis, IN). Primers are described in Supplementary Table S11.

Quantitative PCR (qPCR) was performed using SYBR® Green PCR Master Mix (Life Technologies Corporation) on a Mastercycler® ep realpex device (Eppendorf, Hamburg, Germany). Primers are described in Supplementary Table S11.

RNA experiments. The Open Biosystems’ shRNA TRC constructs for Dck, 23582 (KD1) and 25383 (KD2), Nfrk (9514 and 9511), Tp53 (12335 and 12356), GFP and empty vector (Thermo Fisher Scientific Inc.), were provided in E. coli and plated on carboniccellin media to isolate single clones. Next, the plasmids containing the shRNA constructs were isolated from the E. coli using the Invitrogen™ PureLink® Quick Plasmid Miniprep Kit (Life Technologies Corporation). The plasmids were then transfected into Open Biosystems’ packaging cells, TLA-HEK293T, using the Open Biosystems’ Trans-Lentiviral™ Packaging System (Thermo Fisher Scientific Inc.). The TLA-HEK293T cell lines were in the recommended growth media. Viral particles were then collected and concentrated using PEG–DIH Virus Precipitation Solution (System Biosciences, Mountain View, CA). The viral particles were transfected into the B117P and B140P cell lines by adding virus (MOI of 100) and 8 µg/ml of polybrene to the cells and incubating for 2 hours at 37°C followed by spinoculation (30 min, 300 g).

CRISPR knockouts. Candidate target sequences for CRISPR were designed using ZIFIT Targeter Version 4.2 (http://zifit.partners.org/ZIFIT/). The sequences of guide RNA were placed in pENTR221-U6-gRNA by inverse PCR, as previously described, hCas9 was purchased from addgene (Plasmid #41815). hCas9 was PCR amplified and transferred to PENT221 by standard BP Clonase reaction (Invitrogen), following manufacturers protocol. hCas9 was then transferred to PB-TRE-DEST-EF1a-rTA-RES-Puro by standard LR Clonase reaction (Invitrogen), following manufacturers protocol, to generate PB-TRE-Cas9-EF1a-rTA-RES-Puro. The Gateway DEST cassette was then PCR amplified with Nhcl site engineered into the primers and cloned into a unique Nhcl site upstream of the TRE promoter to generate PB-DEST-TRE-Cas9-EF1a-rTA-RES-Puro. The guide RNAs were then transferred to PB-DENT-Cas9-EF1a-rTA-RES-Puro via standard LR Clonase reaction (Invitrogen), following manufacturers protocol. Two micrograms of each PB-U6-gRNA-TRE-Cas9-EF1a-rTA-RES-Puro and Super piggyBac transposase (System Biosciences) were transfected into B117P by NEON® Transfection System (Life Technologies Corporation, Carlsbad, CA) using 1,400 volts and 20 milliseconds for 2 pulses. Cells were selected with 1.5 µg/ml of doxycycline and 1.0 µg/ml of doxycycline for more than 3 weeks to generate stable knockout cell lines. DNA was collected using standard Phenol:Chlorormethane extraction. A CEL-1 assay was performed (Supplementary Figure S5) and the gene modification ratio calculated, as previously described.

TALEN assembly and generation of KO cells. Candidate Dck TALENs were designed using TALE-NT (http://boglab.plp.iastate.edu/node/add/talen). From the list of candidate TALENs generated using TALE-NT, three were chosen based on methods previously described. TALENs were assembled using Golden Gate cloning. Clones were analyzed for transfection cells were selected for Dck KO using 50 µg/ml of Ara-C for 5 days. From this pool of TALEN modified cells single cells clones were isolated by limiting dilution cloning. Clones were analyzed for Dck KO by direct PCR and sequencing of the TALEN targeted region of Dck, using primers described in Supplementary Table S11.

Inducible DCK overexpression vector. Full-length human DCK cDNA was purchased from GeneCopoeia (Rockville, MD) in a ready entry ORFEXPRESS™ Gateway® Plus Shuttle (Cat#GC-C0081). The DCK cDNA was then transferred to PB-TRE-DEST-EF1a-rTA-RES-Puro via the Invitorgen™ Gateway® LR Clonase reaction (Life Technologies Corporation), following manufacturers instructions. Two micrograms of PB-TRE-Dck-EF1a-rTA-RES-Puro was electroporated with 2 µg of Super piggyBac transposase (System Biosciences) into B117P using the NEON® Transfection System (Life Technologies Corporation), following manufacturer’s instructions. Two days after transfection, cells were selected with 1.5 µg/ml puromycin for 5 days to generate stable cell lines. Overexpression was activated by adding 1.0 µg/ml of doxycycline.

Western blot analysis. Protein lysate was isolated from cells using RIPA lite buffer (50 mM Tris-HCL pH7.6, 150 mM NaCl, 1% NP40, 5 mM NaF, 1 mM EDTA)
supplemented with protease inhibitors (Roche Applied Systems) and phosphatase inhibitors (Sigma-Aldrich, St. Louis, MO). 200 μg of protein lysate was separated on a NuPAGE® Novex® 10% Bis-Tris Gel (Life Technologies Corporation) and transferred to a PVDF membrane. The membrane was blocked with 5% milk in 1XTBST (4hrs RT) and incubated overnight at 4°C with the primary antibodies anti-DCX (1: 500, Proteintech, Chicago, IL) and anti-Gapdh (1: 10,000, Cell Signaling Technology, Danvers, MA). Goat anti-Rabbit HRP conjugated secondary antibodies were utilized at 1: 5,000 dilution (Santa Cruz Bioteclondigies, Dallas, TX).

Membranes were developed with WesternBright Ecl kit (BioExpress, Kaysville, UT).

**Drug assays.** Drug assays were performed using the CellTiter 96® Aqueous Non-Radioactive Cell Proliferation Assay (Promega, Madison, WI), as described previously. For each cell line the drug was tested using 10 different concentrations, and each drug concentration was tested in quadruplicate. The drug concentrations were selected to maximize the number of data points between IC₅ and IC₉⁵. For the results to be acceptable there needed to be data points on both sides of the IC₅ and the r-value needed to be greater than 0.85. Inhibitory concentrations (IC) values and r-values were calculated using CalcuSyn 2.0 (Biosoft, Cambridge, UK).

1. Estey, E. H. Acute myeloid leukemia 2012 update on diagnosis, risk stratification, and management. *Am J Hematol* 87, 89–99 (2012).

2. Cai, J. et al. Two distinct molecular mechanisms underlying cytarabine resistance in human leukemic cells. *Cancer Res* 68, 2349–57 (2008).

3. Shi, J. Y. et al. Association between single nucleotide polymorphisms in deoxycytidine kinase and treatment response among acute cytothroid leukemia patients. *Pharmacogenet Genomics* 14, 759–68 (2004).

4. Veuger, M. J., Honders, M. W., Landegent, J. E., Willemez, R. & Rarge, R. M. High incidence of alternatively spliced forms of deoxycytidine kinase in patients with resistant acute myeloid leukemia. *Blood* 96, 1517–24 (2000).

5. Song, J. H. et al. Defective expression of deoxycytidine kinase in cytarabine-resistant acute myeloid leukemia cells. *J Oncol* 34, 1165–71 (2009).

6. Rathe, S. K. & Largaespada, D. A. Deoxycytidine kinase is downregulated in Ara-C-resistant acute myeloid leukemia murine cell lines. *Leukemia* 24, 1513–5 (2010).

7. Veuger, M. J., Honders, M. W., Willemez, R. & Rarge, R. M. Deoxycytidine kinase expression and activity in patients with resistant versus sensitive acute myeloid leukemia. *Eur J Haematol* 69, 171–8 (2002).

8. Yin, B., Kogan, S. C., Dickins, R. A., Lowe, S. W. & Largaespada, D. A. Trp5 loss during in vitro selection contributes to acquired Ara-C resistance in acute myeloid leukemia. *Exp Hematol* 34, 631–41 (2006).

9. Roberts, A., Pimentel, H., Trapnell, C. & Pachtet, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* 27, 2325–9 (2011).

10. Roberts, A., Trapnell, C., Donaghey, J., Rinn, J. L. & Pachtet, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol* 12, R22 (2011).

11. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28, 511–5 (2010).

12. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5, 621–8 (2008).

13. Rathe, S. K. et al. MMuFLR: Missense Mutation and Frameshift Location Reporter. *Bioinformatics* 29, 2353–4 (2013).

14. Blankenberg, D. et al. Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol Chapter* 19, Unit 19 10 1–21 (2010).

15. Giardine, B. et al. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* 15, 1451–5 (2005).

16. Goecks, J., Nekrutenko, A. & Taylor, J. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11, R86 (2010).

17. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).

18. Trapnell, C., Pachtet, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25, 1105–11 (2009).

19. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat Methods* 7, 248–9 (2010).

20. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7, e46688 (2012).

21. McPherson, A. et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *PLoS Comput Biol* 7, e1001138 (2011).

22. Kocabas, N. A. et al. Gemcitabine pharmacogenomics: deoxycytidine kinase and cytidylate kinase gene resequencing and functional genomics. *Drug Metab Dispos* 36, 1951–9 (2008).

23. Fazekas, A., Steeves, R. & Newmaster, S. Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques* 48, 277–85 (2010).

24. Bullard, J. H., Purdom, E., Hansen, K. D. & Dudoit, S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* 11, 94 (2010).

25. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nuclease in human cells. *Nat Biotechnol* 31, 822–6 (2013).

26. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 31, 833–8 (2013).

27. Verhaak, R. G. et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica* 94, 131–4 (2009).

28. Largaespada, D. A., Shaughnessy, J. D., Jr., Jenkins, N. A. & Copeland, N. G. Retroviral integration at the Evi-2 locus in BXH-2 myeloid leukemia cell lines disrupts NF1 expression without changes in steady-state Ras-GTP levels. *J Virol* 69, 5095–102 (1995).

29. Moriatry, B. S. et al. Simple and efficient methods for enrichment and isolation of endonuclease modified cell lines. *PLoS One* 9, e96114 (2014).

30. Guschin, D. Y. et al. A rapid and general assay for monitoring endogenous gene modifications. *Methods Mol Biol* 649, 247–56 (2010).

31. Cermak, T. et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39, e82 (2011).

32. Carlson, D. F. et al. Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A* 109, 17382–7 (2012).

33. Moriatry, B. S. et al. Modular assembly of transposon integratable multigene vectors using RecW assembly. *Nucleic Acids Res* 41, e92 (2013).

**Acknowledgments**

Funding for this project was provided by The Leukemia & Lymphoma Society (grant 7019-04). We extend our thanks to the many University of Minnesota resources involved in our project. The BioMedical Genomics Center provided services for RNA sequencing, oligo preparation, and Sanger sequencing. The Minnesota Supercomputing Institute maintains the Galaxy Software, as well as provides data management services and training. The Masonic Cancer Center Bioinformatics Core provided guidance in the analysis of the RNA-seq data. We would also like to express our thanks to the assistance provided by Willemijn Veldhuizen.

**Author contributions**

D.A.L. and S.K.R. designed the study. S.K.R. analyzed the RNA-seq data, and wrote most of the manuscript. B.S.M. designed the CRISPR and TALEN KO, and the OE experiments, and wrote the sections related to the KO. S.K.R., B.S.M., C.B.S., M.K., N.K.A., E.P.R., N.J.B., E.G.M., D.A.B. and C.R.L. participated in experiments, analyzed results, and contributed to the Methods/ Figures. All authors reviewed the manuscript.

**Additional information**

Supplementary information accompanies this paper at http://www.nature.com/scientificreports.

**Competing financial interests:** Yes there is potential Competing Interest. Dr. Largaespada is co-owner and advisor to NeoClone Biotechnology, Inc. and Discovery Genomics, Inc. No resources or personnel from either company were involved in this research in any way. The remaining authors declare no competing financial interest.

**How to cite this article:** Rathe, S.K. et al. Using RNA-seq and targeted nucleases to identify mechanisms of drug resistance in acute myeloid leukemia. *Sci. Rep.* 4, 6048; DOI:10.1038/srep06048 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/