
Identification of human genetic variants controlling circular RNA expression

IKHLAK AHMED,^{1,2} THASNI KAREDATH,¹ FATIMA M. AL-DASIM,¹ and JOEL A. MALEK^{1,3}

¹Department of Genetic Medicine, Weill Cornell Medicine-Qatar, Doha, Qatar

²Biomedical Informatics, Sidra Medicine, Doha, Qatar

³Genomics Core, Weill Cornell Medicine-Qatar, Doha, Qatar

ABSTRACT

Circular RNAs (circRNAs) are abundant in eukaryotic transcriptomes and have been linked to various human disorders. However, understanding genetic control of circular RNA expression is in the early stages. Here we present the first integrated analysis of circRNAs and genome sequence variation from lymphoblastoid cell lines of the 1000 Genomes Project. We identified thousands of circRNAs in the RNA-seq data and show their association with local single-nucleotide polymorphic sites, referred to as circQTLs, which influence the circRNA transcript abundance. Strikingly, we found that circQTLs exist independently of eQTLs with most circQTLs having no effect on mRNA expression. Only a fraction of the polymorphic sites are shared and linked to both circRNA and mRNA expression with mostly similar effects on circular and linear RNA. A shared intronic QTL, rs55928920, of HMSD gene drives the circular and linear expression in opposite directions, potentially modulating circRNA levels at the expense of mRNA. Finally, circQTLs and eQTLs are largely independent and exist in separate linkage disequilibrium (LD) blocks with circQTLs highly enriched for functional genomic elements and regulatory regions. This study reveals a previously uncharacterized role of DNA sequence variation in human circular RNA regulation.

Keywords: circular RNA; eQTL; circQTL; 1000 Genomes; RNA-seq

INTRODUCTION

Circular RNAs (circRNAs) are an abundant class of regulatory transcripts primarily derived from protein-coding exons and widely expressed across eukaryotic organisms including *Homo sapiens* and *Mus musculus* (Danan et al. 2012; Salzman et al. 2012, 2013; Jeck et al. 2013; Memczak et al. 2013b; Wang et al. 2014). These transcripts are produced cotranscriptionally; the 5' and 3' are joined in a head-to-tail "backsplice" junction to form a circular molecule leaving no exposed ends. CircRNAs are known to be evolutionarily conserved across species, to be relatively stable in the cytoplasm, and to often show tissue- and developmental stage-specific expression patterns. CircRNAs have been shown to be involved in posttranscriptional regulation by functioning as decoys for the binding of miRNAs, reducing their ability to target mRNAs (Hansen et al. 2013; Memczak et al. 2013a). Specifically, the circRNA ciRS-7, also known as CDR1as, has been found to harbor more than 70 miR-7 binding sites

and functions as an miRNA sponge to repress miR-7 activity. CDR1as knockout mice suffer from defects in sensorimotor gating (Piwecka et al. 2017) and silencing of CDR1as in colorectal cancer, and hepatocellular carcinoma cell lines resulted in decreased tumor proliferation (Yu et al. 2016; Tang et al. 2017). Similarly, the circular RNA of the SRY gene contains 16 binding sites for miR-138, and miR-138-mediated knockdown of mRNA was shown to be attenuated with circular SRY overexpression (Hansen et al. 2013). Indeed, multiple studies have reported the sequestration of miRNAs by circRNAs, making them excellent candidates for competing endogenous RNA activity (Hansen et al. 2013; Memczak et al. 2013a; Guo et al. 2014; Jeck and Sharpless 2014; Li et al. 2015; Du et al. 2016; Han et al. 2017). Although multiple reports have emerged indicating certain circRNAs function as miRNA sponges, increasing evidence indicates that circular RNAs may have other potential functional roles like storing or sequestering transcription factors or RNA binding proteins (Hentze and Preiss 2013) or microRNA transport (Piwecka et al. 2017), and at least some are translatable

Corresponding author: jom2042@qatar-med.cornell.edu

Article is online at <http://www.rnajournal.org/cgi/doi/10.1261/rna.071654.119>. Freely available online through the RNA Open Access option.

© 2019 Ahmed et al. This article, published in RNA, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

into functional proteins (Wang and Wang 2015; Legnini et al. 2017; Pamudurti et al. 2017; Yang et al. 2017). Legnini et al. (2017) were able to show that circ-ZNF609 is translated into a protein in a splicing-dependent and cap-independent manner. The resultant protein has a functional role in muscle differentiation in Duchenne muscular dystrophy by regulating myoblast proliferation. circRNAs have been shown to be altered in a variety of pathological conditions, which has stimulated significant interest into their role in human disease and cancer (Li et al. 2015; Greene et al. 2017). Although their exact roles and mechanisms of gene regulation remain to be understood, circRNAs are being actively investigated for their association with diseases and their role as potential biomarkers and novel therapeutic targets.

Recent association studies on expression quantitative trait loci (eQTL) have provided information on genetic factors, especially single-nucleotide polymorphisms (SNPs), associated with variation in gene expression. These studies have demonstrated the regulatory role of SNPs on gene expression and the splicing patterns of mRNA (Lappalainen et al. 2013; GTEx Consortium 2015; Li et al. 2016; Takata et al. 2017). Expression QTL studies offer an excellent platform to link DNA sequence variation to changes in gene expression that may contribute to phenotypic diversity and disease susceptibility.

In this study, we present the first integrated analysis of genome sequence variation and circular RNA expression to identify a set of regulatory variants influencing circRNA expression. We identified thousands of circRNAs in RNA-seq data of lymphoblastoid cell lines (LCLs) from the 1000 Genomes Project (Lappalainen et al. 2013) and mapped their association with local SNP sites, referred to as circQTLs. These circQTLs associate with the circRNA abundance and exist independently of eQTLs, with most circQTLs having no effect on mRNA expression. We report that only a fraction of the polymorphic sites are shared and linked to both circRNA and mRNA expression with mostly similar effects on circular and linear RNA. We also show that circQTLs and eQTLs exist in separate linkage disequilibrium (LD) blocks, with circQTLs enriched for exonic and regulatory region variants. This study reveals a previously uncharacterized role of DNA sequence variation in human circular RNA regulation.

RESULTS

Assessing circRNA expression

We used RNA-sequencing data of 358 LCL samples of European ancestry sequenced in the framework of the Geuvadis RNA-sequencing project (Lappalainen et al. 2013) along with their genotype information from 1000 Genomes data (Auton et al. 2015). We analyzed the circRNA expression of these samples using our in-house

developed circular RNA detection pipeline (Ahmed et al. 2016). This pipeline identifies circular RNA structures by using Bowtie 2 (Langmead and Salzberg 2012) to align paired-end RNA-seq data to a custom reference exome containing all possible pairs of intragenic nonlinear combinations of exons, as well as single-exon “backsplices.” This process yielded a total of 95,675 unique circRNA candidates with a minimum of 50 independent junctional reads across all samples and expression in at least 60% of the samples (Supplemental_File_S1.xlsx). Of the 95,675 circRNAs as identified by the pipeline, 58,050 overlap with one or more circRNA loci defined in the circBase (Supplemental_File_S2.xlsx; Glažar et al. 2014).

We have validated several circRNA candidates identified in this study in human T-lymphocyte Jurkat cells and EBV-transformed lymphoblastoid B-cell lines (LCLs) using multiple approaches. Divergent primers were designed for a set of circRNAs to perform PCR amplifications of the backsplice exon junction from cDNA of the Human T lymphocyte Jurkat and LCL cells (Supplemental_File_S3.xlsx). An “outward-facing” strategy in the design of the divergent primers guarantees the amplifications are from a circular template (Shen et al. 2015). For all 17 circRNA candidates tested, a single distinct band of expected product size was obtained in a PCR assay, and Sanger sequencing of the PCR products confirmed the presence of the backsplice junction sequence (Fig. 1; Supplemental_Figure_1.pdf). As a control, we also designed convergent primers to amplify the linear RNA (mRNA) of the respective host genes of some of the circRNA candidates. Because the mRNA structure is dictated by the genomic order of exons, PCR primers designed to amplify the linear RNA can produce bands for both cDNA and genomic DNA (gDNA) fractions. On the other hand, the backsplice junction sequence of the correctly identified circRNAs should not exist in the genome; hence, circRNA-specific primers should be able to amplify only in the cDNA but not the gDNA fraction. Indeed, as expected, positive and negative results of amplifications were obtained for cDNA and gDNA fractions for circRNA-specific primers. Primers designed to amplify linear RNA produced bands of expected product sizes for both cDNA and gDNA (Fig. 1C). Further evidence of a circularized structure for these circRNA candidates came from the digestion of total RNA with an exoribonuclease enzyme RNase R. This exonuclease enzyme digests all linear RNA forms with a 3′ single-stranded region of >7 nt (Vincent and Deutscher 2006). The circRNAs have been shown to resist the RNase R-mediated digestion because of their lack of 3′ single-strand overhangs and hence show enrichment in the RNase R-treated samples. Indeed, there was ample enrichment of our tested circRNA candidates after the RNase R treatment compared to mRNA, confirming their resistance to the exoribonuclease digestion and strongly attesting to their nonlinear structure (Fig. 1D).

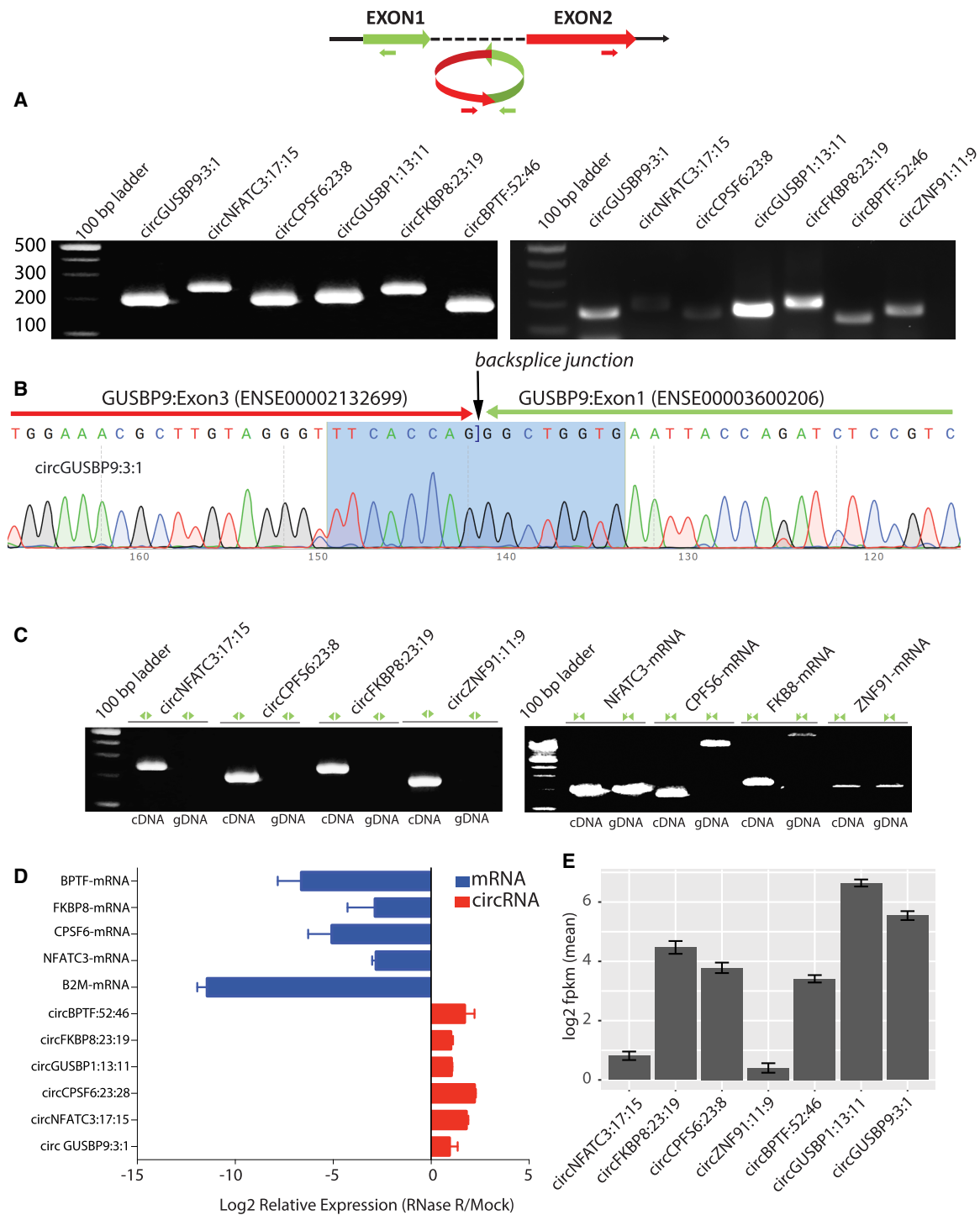


FIGURE 1. Validation of circRNA candidates through RT-PCR and Sanger sequencing. (Also see [Supplemental_Figure_1.pdf](#).) (A) Divergent primers with respect to genomic sequence were designed for 17 circRNA candidates. These become properly inward-facing and identify the backsplice junction sequence. The *left* and *right* panels of the gel show human Jurkat and LCL cells, respectively. (B) Sanger sequencing confirms the backsplice junction sequence of the PCR products. Arrows indicate the presence of backsplice junction. (C) Divergent primers with respect to the genomic sequence amplify the circRNA backsplice junction sequence in the cDNA but not in the gDNA fraction. Convergent primers with respect to the genomic sequence amplify mRNA in both cDNA and gDNA fractions. (D) Treatment with RNase R leads to enrichment of circRNAs and depletion of the host gene and Beta-2 microglobulin (B2M) mRNAs compared to mock. (E) Average expression levels of the tested circRNAs in 358 EUR LCL samples of the 1000 Genomes Project RNA-seq data. Error bars indicate 95% confidence intervals.

Assessing circRNA-associated genomic variants

Genotype calls corresponding to the 358 LCL samples were extracted from VCF files downloaded from Phase 3 release of the 1000 Genomes Project (Auton et al. 2015). Only common SNP variants with a minimum allele frequency of >5% and passing the Hardy–Weinberg equilibrium filter were taken for association testing. The significance of correlations between genotypes and circRNA expression levels were determined by linear regression on quantile normalized circRNA expression values, after correction for known and inferred technical covariates (see Materials and Methods), using Matrix eQTL (Shabalin 2012). The results of this additive linear genetic model association analysis are summarized in Figure 2. We tested approximately 6.2 million SNPs and 95,675 circRNAs for *cis* QTL associations (within 1 MB of gene boundaries) and plotted the genome-wide distribution of the test statistic against the expected null distribution to observe any inflation or resid-

ual deviations in the association statistic that may lead to an excessive false-positive rate. The observed distribution of *P*-values deviates from the expected null distribution in the extreme right tail only, indicating no evidence for systematic spurious associations (Fig. 2A). The nominal *P*-values obtained from Matrix-eQTL were adjusted for multiple testing using the eigenMT package (Davis et al. 2016). This multiple testing procedure estimates the number of independent tests for each gene using a genotype correlation matrix and then applies a Bonferroni procedure while accounting for LD structure among local variants. We defined “ecircRNAs” as circRNAs with at least one SNP in *cis* significantly associated (adjusted *P*-value of <0.05) with expression of that circRNA. The number of identified circQTL variants for each gene was plotted against the gene length to identify any potential biases of gene length contributing to the detection of ecircRNAs. Figure 2B shows a nearly uniform distribution of circQTL variants with respect to gene length, ruling out any potential biases

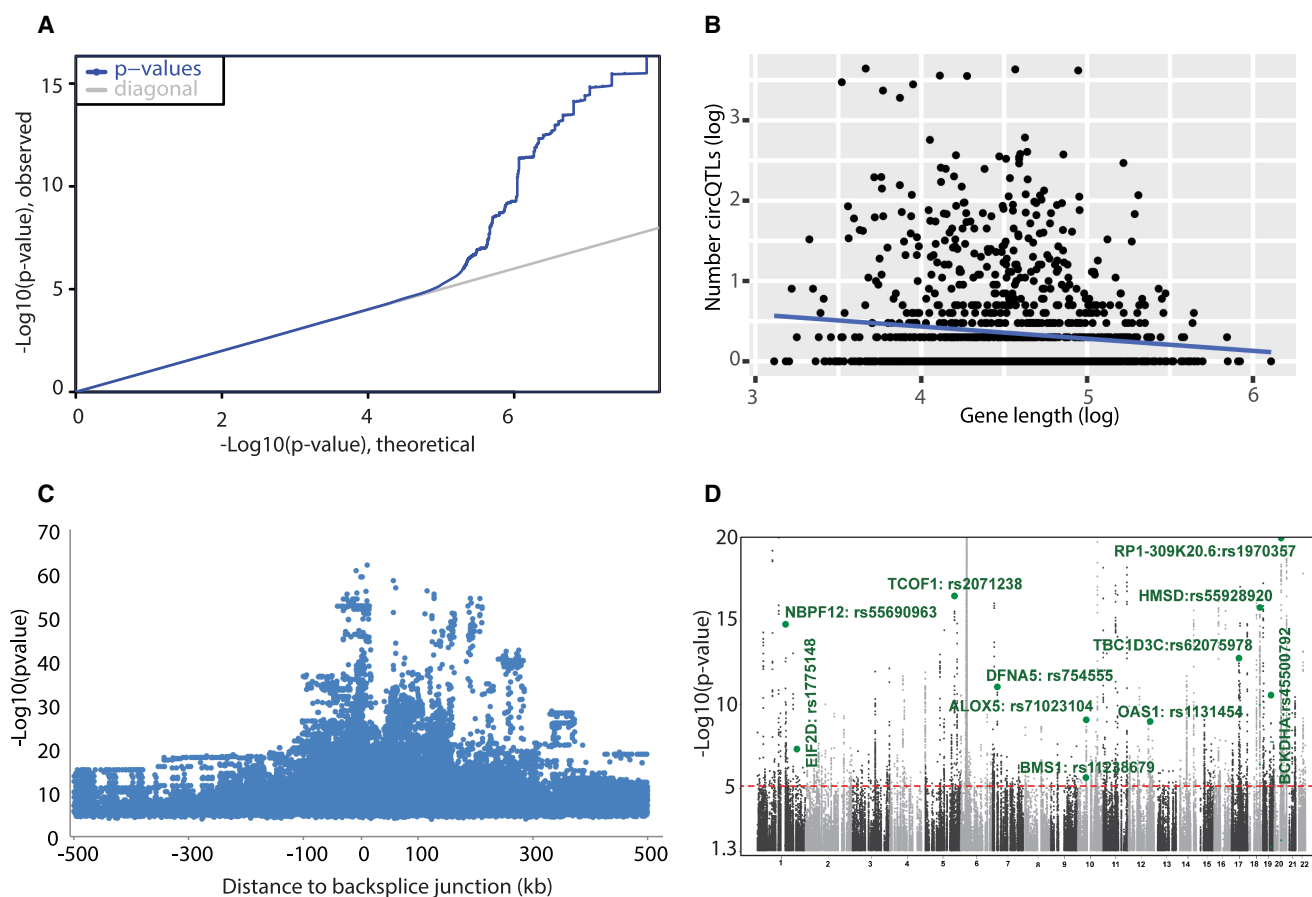


FIGURE 2. Summary plots of association statistics. (A) Q–Q plot distribution of all recorded *P*-values for chromosome 22. (B) Scatterplot of gene length versus number of identified circQTL SNPs for each gene. The number of identified circQTL SNPs shows a uniform distribution with respect to gene length as indicated by the line of best fit (blue line). (C) The distribution of statistical significance of association for circQTL SNPs against proximity to circular RNA. (D) Manhattan plot of association statistics. The $-\log_{10}(P\text{-value})$ is plotted against the physical positions of each SNP on each chromosome and have been capped to $P = 1 \times 10^{-20}$. The genome-wide significance threshold of 8.4×10^{-6} , obtained as the median of the unadjusted *P*-values for the best circQTL of each ecircRNA, is shown as a horizontal red line.

due to larger genes having a better chance to be detected than smaller genes. A genome-wide P -value limit of 8.4×10^{-06} , obtained as the median of the unadjusted P -values for the best circQTL in each gene, is used as a cutoff to identify all the circQTLs. This resulted in a total of 139,485 circQTLs for 2260 ecircRNAs in 1359 genes (Supplemental_File_S4.xlsx). The circRNA expression level of these genes does not seem to significantly influence the number of identified circQTLs (Supplemental_Figure_2.pdf). We also observed an enrichment of the circQTL variants at the proximity of the backsplice junction (Fig. 2C). Mapping of reads to measure the circRNA expression could show allelic biases due to sequencing polymorphisms within the backsplice forming exons and hence could act as a cofounder for the association signal (van de Geijn et al. 2015). Thus, if one allelic read maps better to a location compared to the other allelic read, a strong false-positive association would be induced, generally closer to the measured event. We found only 721 out of 139,485 circQTLs encompass the backsplicing exons, strongly indicating the absence of any allele-specific read mapping biases in our results.

Many of the genes containing ecircRNAs are associated with a disease phenotype (*ALOX5*, *TCOF1*, *DFNA5*, *OAS1*, *EIF2D*) or a physical trait like BMI and height (*BMS1*, *RP1-309K20.6*; Fig. 2D). For example, the gene Arachidonate 5-Lipoxygenase (*ALOX5*) has been associated with chronic bronchial asthma susceptibility, and we found several circQTL SNPs in the promotor of the *ALOX5* gene that are associated with the expression of circular RNA processed from exons 1 and 2 of this gene (Supplemental_Figure_3.pdf).

circRNA QTLs exist independently of eQTLs

We compared our results for ecircRNAs with those of eQTL associations obtained by Lappalainen et al. in the Geuvadis project (Lappalainen et al. 2013). Out of a total of 1359 genes containing ecircRNAs, 75% of these genes have no corresponding eQTLs in the Geuvadis data (representation factor = 0.3, $P < 2.2 \times 10^{-16}$; representation factor < 1 indicates less overlap than expected) and ~62% of the genes with only eQTLs have no evidence for circRNA expression (representation factor = 0.6, $P < 2.2 \times 10^{-16}$). Only one-quarter of the ecircRNA genes have one or more eQTLs associated with the linear expression of the gene (Fig. 3A). Of these genes, which contain both circQTL and eQTL SNP markers, only ~7% (representation factor = 0.3, $P < 2.2 \times 10^{-16}$) of the QTLs are shared, having significant associations with both circRNA and mRNA expression (Fig. 3B). Thus, in these genes, 74% of the circQTL markers and 90% of the eQTL markers are exclusively associated with the expression of their respective isoform. To determine whether circQTLs are associated with any measurable effect on mRNA expression of the host gene, we compared

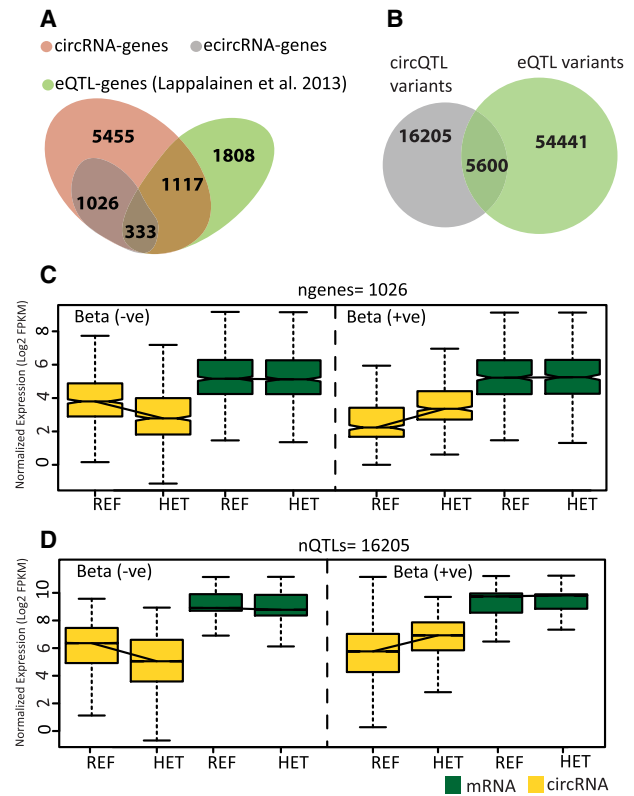


FIGURE 3. Comparison of circQTL and eQTL variants and their effects on gene expression. (A) Venn diagram showing an overlap of circRNA genes with eQTL genes from Lappalainen et al. (2013). (B) Venn diagram showing distribution of genomic variants associated with circRNA and mRNA expression in 333 genes that contain both circQTL and eQTL SNPs. Only ~7% (5600) of QTLs are shared and show association with both circRNA and mRNA expression in these genes. (C) Boxplots representing distribution of average expression from all individuals with a given genotype. The trends in circRNA and mRNA expression between reference homozygous and heterozygous genotypes for best circQTL SNP, in 1026 ecircRNA genes that do not contain any eQTL associations. (D) circRNA and mRNA expression trends for 16205 circQTL SNPs from genes having both circQTL and eQTL markers.

the circRNA and mRNA expression changes between homozygous (REF) and heterozygous genotypes of circQTL markers. Figure 3C shows circRNA and mRNA expression changes between homozygous and heterozygous genotypes for the best circQTL SNP of 1026 ecircRNA genes that have no corresponding eQTL. Although circRNA expression shows a clear trend for both positive ($\beta = 1.121$, $P < 2.2 \times 10^{-16}$) and negative ($\beta = -1.015$, $P < 2.2 \times 10^{-16}$) β values (effect size), there is no such trend in the mRNA expression ($\beta = 0.0294$, $P = 0.8654$; and $\beta = -0.0339$, $P = 0.8639$ for positive and negative β values, respectively). Similarly, in Figure 3D, there is only a marginal difference in mRNA expression ($\beta = 0.072$, $P < 2.2 \times 10^{-16}$; and $\beta = -0.13$, $P < 2.2 \times 10^{-16}$ for positive and negative β values, respectively) compared to circRNA expression ($\beta = -1.31$, $P < 2.2 \times 10^{-16}$; and $\beta = 1.16$, $P < 2.2 \times 10^{-16}$ for positive and

negative β values, respectively) between homozygous (REF) and heterozygous genotypes. These expression differences are for the 16,205 circRNA exclusively associated markers, in genes that contain both circQTL and eQTL SNPs. Thus, circQTL markers do not seem to have any notably detectable effect on the mRNA expression of their host genes, ruling out any artificial increase in spurious associations for ecircRNAs because of systematic or analytical biases from mRNA. These results also imply that QTL SNP markers exclusively associate with either circRNA or mRNA expression. Thus, a large fraction of ecircRNA genes exist without eQTLs, indicating the lack of total biological dependence of circRNA expression on eQTLs.

Shared QTLs have similar effects on circRNA and mRNA expression

To understand the effect of shared QTLs, we plotted circRNA and mRNA expression profiles for 5600 QTL markers that were significantly *cis*-associated with both circRNA and mRNA expression (Fig. 4A). The trends in expression from homozygous to heterozygous genotypes are analogous for circRNA and mRNA, as would be expected based on cotranscription. Figure 4B further evaluates the effect of the shared QTLs, by charting β values of circQTL and eQTL associations in a scatterplot. The effect size (β) defined as the slope of the linear regression indicates the direction of change in expression from reference to alternate allele with negative values implying higher expression for the reference allele. Although most shared QTLs have similar effects, a small subset has opposite effects on the circular and linear expression of their host gene. For example, the Histocompatibility Minor Serpin Domain Containing (*HMSD*) gene contains a cluster of 306 QTL markers that are associated with the expression of its linear transcript as well as a circRNA formed by the second and fourth exons of the *HMSD*-002 transcript (Fig. 4C). These shared QTL markers seem to drive the circular and linear expres-

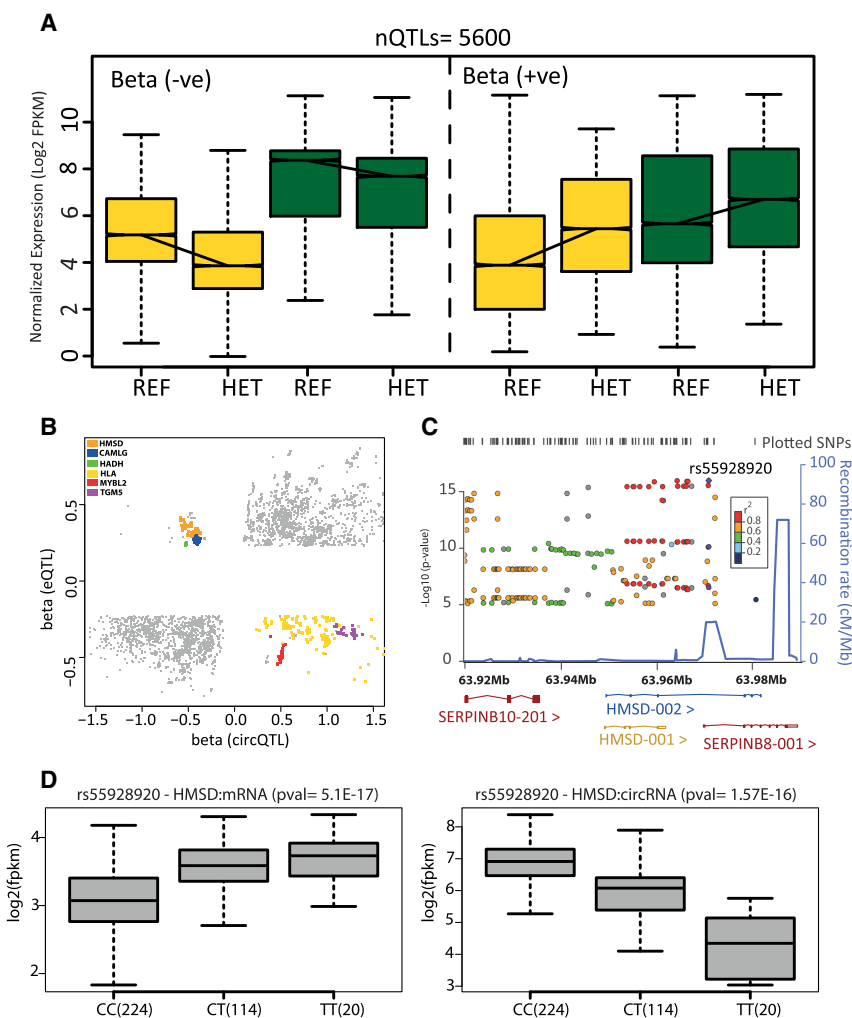


FIGURE 4. Effects of shared QTL variants on gene expression. (A) Boxplots representing distribution of average circRNA and mRNA expression from all individuals with a given genotype for shared QTLs. (B) Spearman rank correlation defines the effect size (β) and indicates the direction of change from reference to alternate allele. Each point in the plot is a shared QTL marker with the x-axis representing β for circRNA association and the y-axis representing β for mRNA association. The high density of points in the *upper right* and *bottom left* quadrants show that most shared QTLs have similar effects from their reference and alternate alleles on circular and linear expression. Colored points represent QTLs that have opposite effects on circular and linear expression of their host genes. (C) Region association plot for *HMSD* circRNA. The shared intronic QTL (rs55928920) is associated with *HMSD* circRNA ($P = 1.57 \times 10^{-16}$) as well as mRNA expression ($P = 5.1 \times 10^{-17}$). Values of r^2 are based on the CEU HapMap 2 samples. CEU HapMap 2 recombination rates are indicated in blue on the right y-axis as obtained using LocusZoom (<http://csg.sph.umich.edu/locuszoom>). (D) Reference allele of rs55928920 is associated with higher expression of circRNA from the second and fourth exons of the *HMSD* transcript (*HMSD*-002) and lower expression of the mRNA transcripts from *HMSD* and *SERPINB8* genes.

sion of the *HMSD* gene in opposite directions, suggesting potential modulation of the circRNA levels at the expense of mRNA. The reference allele of the most highly associated circQTL marker (rs55928920) alleviates the expression of the *HMSD* circRNA, whereas the alternate allele favors mRNA expression of *HMSD* and *SERPINB8* genes (Fig. 4D). We tested whether genes with shared QTLs having

opposite effects on circRNA and mRNA abundance levels are expressed at similar scales in comparison to other ecircRNA genes (Supplemental_Figure_4.pdf)—the results show no difference in mRNA expression but a statistically significant higher circRNA expression of these genes compared to other ecircRNA genes.

Sharp LD breakdown between circQTL and eQTL SNPs

Linkage disequilibrium (LD) is the nonrandom association between alleles at different loci and can be estimated using the correlation between SNP markers (Vos et al. 2017). To assess if circQTL and eQTL SNPs are largely independent, we asked how often a top eQTL SNP is also a circQTL SNP and vice versa. We found only 43 instances in which a top eQTL SNP was also a circQTL SNP for the gene and 72 instances in which a top circQTL SNP was also an eQTL SNP for the gene. We further explored the extent of LD structure between circQTL and eQTL markers by computing the average pairwise correlation for shared QTL, circQTL, and eQTL SNPs and compared these distributions with a cross-pair between circQTL and eQTL markers. Figure 5A indicates that eQTL SNPs tend to encompass weaker LD regions compared to circQTL and shared QTL markers. There is a significant decrease in the average pairwise correlation between a circQTL and eQTL cross-pair, when compared to intra-pairwise correlation distributions of shared QTL, circQTL and eQTL markers (median pairwise correlation drops from 0.8 to 0.3). This strongly suggests a sharp breakdown in the LD

structure between circQTL and eQTL SNP-containing regions. Sharp LD breakdown between circQTL and eQTL variants could indicate differential selection pressures shaping their molecular evolution.

To estimate the rate of LD decay for circQTL and eQTL markers and compare their LD patterns, we plotted the LOESS curves of average pairwise correlations against the physical distance between the SNP pair (Fig. 5B). LD for circQTL SNPs decayed faster than eQTL and shared QTL markers. The average pairwise correlation for circQTL variants drops sharply within a span of 50 kb, indicating that circQTLs are held in tightly linked and comparatively shorter LD blocks. For the cross-pair between circQTL and eQTL SNPs, LD shows very minimal change with distance, indicating the lack of genetic linkage between the two sets of these markers.

Functional characterization of circQTL variants

To assess the functional elements of the genome enriched for circQTL variants, we performed Monte Carlo simulations using 10,000 random sets of background eQTL SNPs with minor allele frequencies (MAFs) matched to the set of 23,769 circQTL variants (see Materials and Methods). By functionally classifying the circQTL and background eQTL SNPs as per the definitions in SnpEff (Cingolani et al. 2012), we found that circQTLs are significantly enriched among variants in 5'-untranslated (OR = 1.86; Monte Carlo *P*-value 2.3×10^{-02}), exonic (OR = 1.69; Monte Carlo *P*-value 8.61×10^{-03}), intergenic (OR = 1.73; Monte Carlo *P*-value $< 1 \times 10^{-04}$), and regulatory

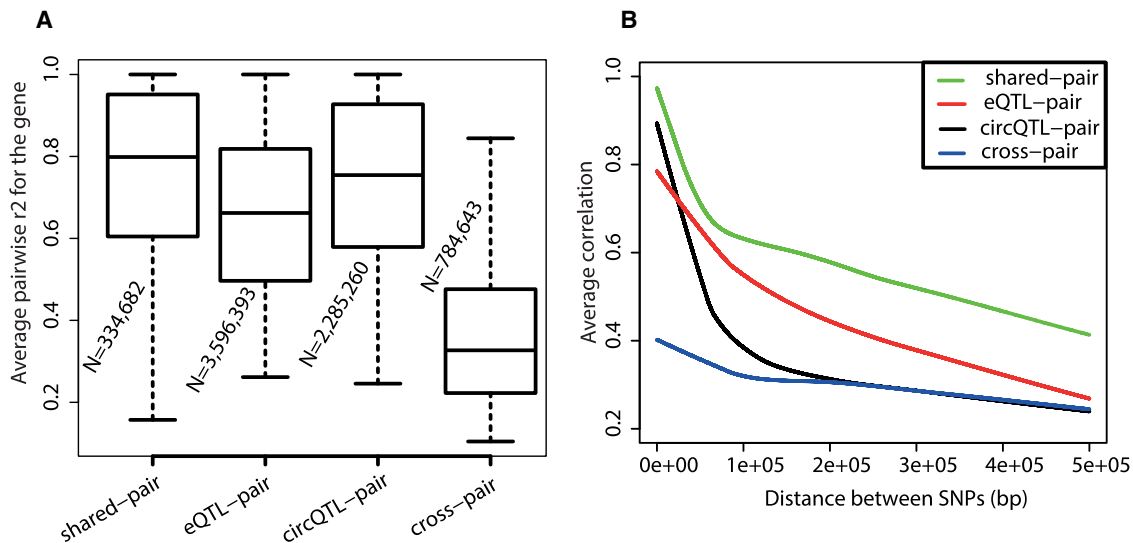


FIGURE 5. LD structure between circQTL and eQTL markers. (A) Boxplot distributions of the average pairwise correlation (r^2) between shared QTL, eQTL, circQTL, and cross-QTL SNPs for each gene. Average pairwise correlations were computed for each category of QTL markers for genes containing both circQTL and eQTL SNPs. Cross-pair is the pairwise correlation between a circQTL and an eQTL marker. For each boxplot distribution, *N* represents the total number of pairwise combinations from all genes; only combinations with $r^2 > 0.1$ were used for the analysis. (B) LOESS fit curves for average correlation (r^2) versus distance between the QTL markers.

(OR = 1.75; Monte Carlo P -value 9.94×10^{-04}) regions of the genome. In contrast, there is a significant drop (from 64% to 50%) of circQTL SNPs among intronic region variants (OR = 0.47, Monte Carlo P -value $< 1 \times 10^{-04}$) (Fig. 6A,B). Although circQTL variants were also found to be enriched for splice site regions (odds ratio OR = 1.55), the P -value for this comparison was insignificant ($P = 1.99 \times 10^{-01}$), possibly because of a low fraction of variants annotated as “splicing.”

To further explore if chromatin structure has a perceivable influence on circRNA expression, we analyzed the enrichment of circQTL variants among various genetic regulatory element sequences. These regulatory sequences correspond to cell-type-specific tracks of chromatin modifications, RNA polymerase (Pol II, Pol III)-bound regions, and transcription factor binding sites for the GM12878 (lymphoblastoid) cell line obtained from the ENCODE project (The ENCODE Project Consortium 2012) and are used to annotate circQTL and background eQTL variants. Using Monte Carlo simulations, 10^4 iterations were performed to annotate a randomly generated set of background eQTL SNPs with MAFs matched to circQTL SNPs, and the odds ratio and significance of the enrichment computed for each chromatin mark. A total of 42 regulatory marks were analyzed and significant enrichment of circQTL SNPs was found among variants contained in transcriptionally active chromatin regions. Histone modifications associated with active transcription that showed most significant enrichments were acetylated histone H3 lysine 9 (H3K9ac; OR = 1.79; $P = 1.1 \times 10^{-02}$), acetylated histone H3 lysine 27 (H3K27ac; OR = 1.63; $P = 4.3 \times 10^{-03}$), and trimethylated histone H3 lysine 4 (H3K4me3; OR = 1.49; $P = 7.1 \times 10^{-03}$). On the contrary, a significant depletion of circQTL SNPs was observed among trimethylated histone H3 lysine 36 (H3K36me3; OR = 0.44; $P = 3.2 \times 10^{-03}$) marked genomic regions. We also found

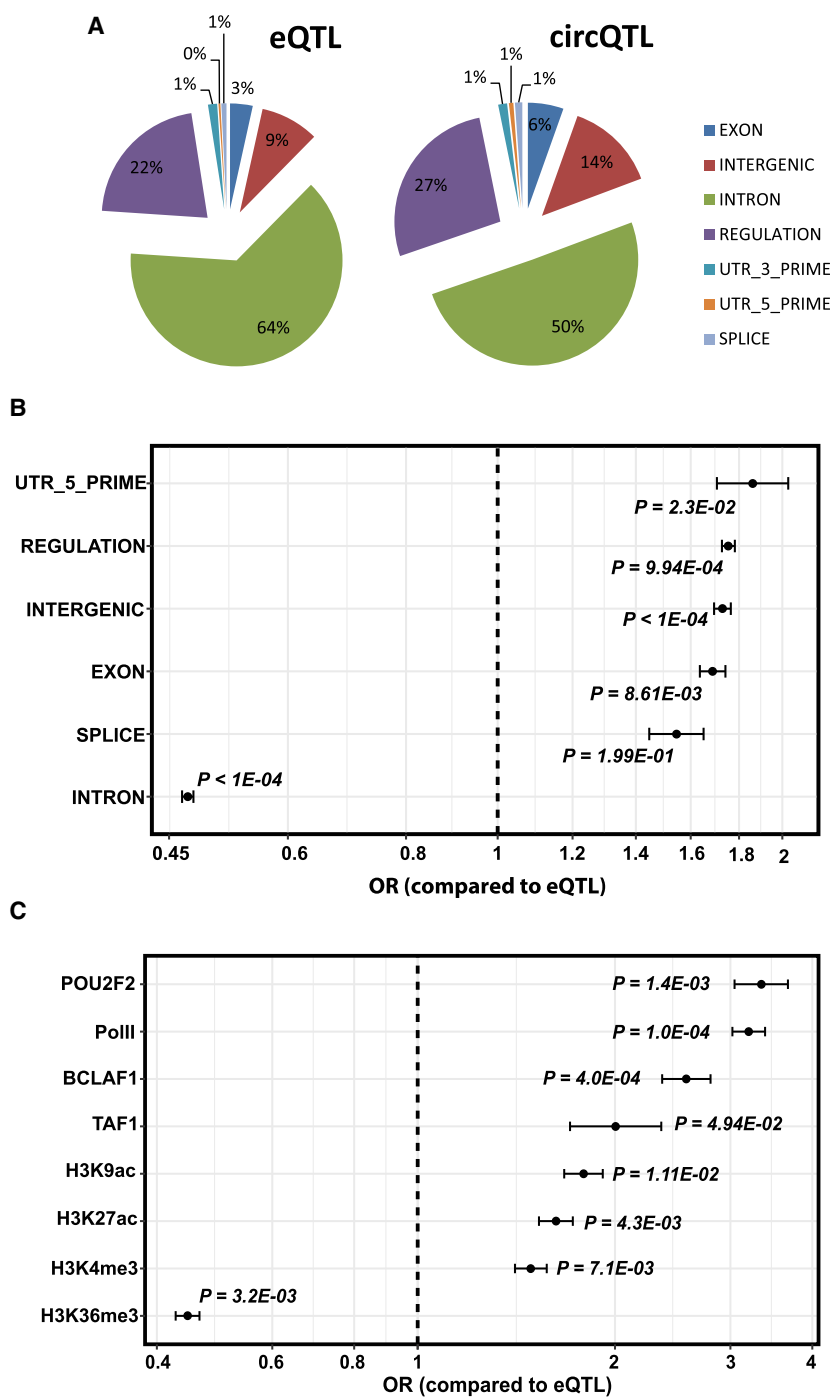


FIGURE 6. (A) Enrichment of circQTL variants in genomic sequences. Proportion of SNPs annotated for each functional category using SnpEff (Cingolani et al. 2012). (B) Odds ratio (OR) for each functional category compared to the eQTL background. (C) Enrichment of circQTL variants within genetic regulatory elements. OR > 1 indicates enrichment of circQTL variants. The significance of the OR was tested using 10,000 Monte Carlo permutations of the eQTL SNPs, and the significance level is reported as a Monte Carlo P -value. Bars indicate 95% confidence intervals.

significant enrichment of circQTL variants for Pol II RNA polymerase bound regions (Pol II; OR = 3.2; $P = 1.0 \times 10^{-04}$) and POU2F2 (OR = 3.34; $P = 1.4 \times 10^{-03}$), BCLAF1 (OR =

2.57; $P = 4.0 \times 10^{-04}$), and TAF1 (OR = 2; $P = 4.94 \times 10^{-02}$) transcription factor binding sites (Fig. 6C).

The potential of circQTL SNPs in the context of their contribution to disease risk and phenotype association was analyzed using the NHGRI-EBI GWAS Catalog (MacArthur et al. 2017), which is a curated collection of all the published genome-wide association studies (GWASs) for various human diseases and traits. We tested the enrichment of circQTL SNPs for various Experimental Factor Ontology (EFO) terms (Malone et al. 2010) defined in the catalog and representing a disease or a trait. The circQTL and a random set of background eQTL SNPs of matched MAFs were separately grouped into genomic loci based on their LD patterns and defined as the genomic region containing SNPs in LD with the index SNP at $r^2 > 0.8$. An associated genomic locus for a trait is identified if a GWAS risk SNP falls within the locus. Finally, the enrichment of circQTL SNPs among associated loci is evaluated by a one-tailed Fisher's exact test followed by Bonferroni correction for the total number of traits assessed. The top diseases associated with the genomic loci containing a significant enrichment for circQTL SNPs are schizophrenia (OR = 4.06, $P = 2.02 \times 10^{-23}$, one-tailed Fisher's exact test with Bonferroni correction), nephropathy (OR = 63.6, $P = 7.08 \times 10^{-08}$), plantar warts (OR = 50.7, $P = 8.27 \times 10^{-06}$), disc degeneration (OR = 14, $P = 3.4 \times 10^{-03}$), lung cancer (OR = 4.24, $P = 3.49 \times 10^{-03}$), and asthma (OR = 8.45, $P = 4.49 \times 10^{-03}$) (Supplemental_Figure_5.pdf).

DISCUSSION

In this study, we quantified the contribution of *cis*-acting genetic variants on circular RNA expression from RNA-seq data of the 1000 Genomes Project LCLs (Lappalainen et al. 2013). To our knowledge, this is the first study to comprehensively identify the set of regulatory variants influencing the circRNA expression in a single cell type.

Using our circRNA detection pipeline (Ahmed et al. 2016), we identified 95,675 circular RNAs that are likely to be adequately expressed in human peripheral blood transcriptomes and thus constitute an important resource. Our genome-wide analysis for association of circRNA expression with genotype identified thousands of previously uncharacterized *cis*-acting genetic variants influencing circRNA transcript abundance. We identified a total of 1359 genes with circQTL associations, referred to here as ecircRNA genes. These ecircRNA genes are enriched in canonical pathways controlling important cellular processes like cell cycle regulation and spliceosome formation. Many of the ecircRNA genes are linked to a disease phenotype (*ALOX5*, *TCOF1*, *DFNA5*, *OAS1*, *EIF2D*) or a physical trait like BMI and height (*BMS1*, *RP1-309K20.6*). The gene Arachidonate 5-Lipoxygenase (*ALOX5*), which contains several circQTL SNPs in its promoter region, has been linked to chronic bronchial asthma susceptibility and en-

codes a member of the lipoxygenase gene family that catalyzes the conversion of arachidonic acid to leukotriene A4. Leukotrienes are important mediators of a number of inflammatory and allergic conditions and mutations in the promoter region of this gene have been shown to lead to a diminished response to anti-leukotriene drugs used in the treatment of asthma (De Caterina and Zampolli 2004). The presence of several upstream circQTL SNPs exclusively associating with the circRNA transcript processed from exons 1 and 2 of this gene could indicate a causal connection of the circQTLs with drug response, acting thereby through modulation of circRNA levels.

By evaluating genes for the presence of circQTLs or eQTLs and the influence of these SNPs on host gene expression, we were able to show that most QTL SNP markers exclusively associate with either circRNA or mRNA expression. A large fraction of ecircRNA genes do not have eQTLs, indicating an arrangement at the level of DNA sequence for circRNA regulation that does not seem to influence the basal gene expression. Likewise, circQTL markers do not have any significantly measurable effect on the mRNA expression of their host genes, reinforcing the idea that circRNA and mRNA transcription could involve distinctive regulatory mechanisms.

In genes that contain both circQTLs and eQTLs, only a small fraction of the markers are shared between the two associations. These shared QTL markers frequently, but not always, have similar effects on the linear and circular RNA expressions of the gene. Some shared QTL markers, however, have opposite effects on gene expression potentially driving the circRNA levels at the expense of mRNA.

We compared the LD patterns for circQTL and eQTL markers and found that eQTL SNPs tend to encompass genomic regions of weaker LD structure compared to circQTL and shared QTL SNPs. The circQTL SNPs are held in tightly linked LD blocks with their LD structure decaying faster than eQTL and shared QTL markers. A sharp breakdown in the LD structure exists between circQTL and eQTL SNP-containing genomic regions. Thus differential selection pressures could be working on circQTL and eQTL SNPs to influence their distinct regulatory mechanisms on gene expression.

We further characterized the circQTL SNPs for a relationship to various annotated genomic features. We found an enrichment of circQTL SNPs among transcribed sequence variants including 5'-untranslated, exonic, intergenic, and regulatory regions of the genome. The enrichment of circQTL SNPs in coding sequences is in accordance with the notion that coding sequences in complex genomes can simultaneously encode amino acid and regulatory information including the potential to act as transcriptional enhancers or splicing signals (Stergachis et al. 2013). Coding exons have also been shown to function as tissue-specific enhancers of nearby genes operating as coding exons in one tissue and enhancers of nearby gene(s) in

another tissue (Birnbaum et al. 2012). In this context, circQTL SNPs in coding exons could have dual functional roles with the ability to either encode a transcript variant of its host gene or adjust a transcriptional regulatory signal for a nearby gene to alter its circRNA levels.

We also found that circQTL SNPs are enriched among genomic regions marked by chromatin modifications associated with active transcription. The enrichment of circQTL SNPs among histone modifications and transcription factors associated with an enhanced transcriptional activity could indicate a role for chromatin structure in regulating circRNA expression—possibly through aiding splicing machinery in coordinating splicing reactions. Several studies have suggested a cross talk between transcription and pre-mRNA splicing, emphasizing a functional connection between RNA Polymerase II (RNAP II) elongation and the splicing machinery (McCracken et al. 1998; Dower and Rosbash 2002; Rgio et al. 2012; Haque and Oberdoerffer 2014). There is accumulating evidence that both RNAP II and the landscape of the chromatin marks contribute to regulate the splicing patterns through a complex regulatory network that contains both feedforward and feedback loops. The kinetic model of transcription-coupled splicing regulation proposed by Kornblihtt (2007) states that chromatin structure influences the rate of RNAP II transcription elongation, and the rate of transcription elongation in turn influences alternative splicing decisions (Kornblihtt 2007; Schor et al. 2013). Weak exons lead to RNAP II pausing and increased inclusion of the exon in spliced mRNA, whereas increased elongation rate decreases exon inclusion in mRNA (Haque and Oberdoerffer 2014). The chromatin modifications H3K4me3, H3K9ac, and H3K27ac, which show an enrichment for circQTL SNPs and mark active gene promoters and enhancers, are known to play a role in alternative splicing through modulating the rate of RNAP II elongation (Luco et al. 2010; Rgio et al. 2012; Haque and Oberdoerffer 2014; Zhou et al. 2014). On the other hand, circQTL SNPs are underrepresented for H3K36me3, another histone mark characteristic of active transcription, particularly from lowly expressed genes, but known to cause hypoacetylation and exon inclusions in the mRNA, possibly through inducing RNAP II pausing (Wilhelm et al. 2011; Rgio et al. 2012). Therefore, chromatin modifications, probably through a combinatorial pattern, could provide RNAP II with signals that locally regulate elongation rate and alter the splicing decisions to favor either mRNA or circRNA expression. In this scenario certain combinatorial patterns of active chromatin modifications could induce a high RNAP II elongation rate that would translate into exon exclusions from mRNA and possibly their retention in the circRNA form.

As genetic variants can influence patterns of chromatin marks (Grubert et al. 2015), some of the circQTL SNPs could function by altering the epigenetic status of various histone modifications, thereby enhancing the

RNAP II elongation rate, which will promote a higher circRNA production. Alternatively, circQTL SNPs can have a more direct role through recruitment of transcription or splicing factors to influence the transcription rate, which in turn could alter the epigenetic state of the DNA template.

We observed significant enrichment of circQTL SNPs with strong to moderate associations with various human diseases indicating a role for circQTL variants in genetic disease etiologies. Thus, circQTL SNPs with mostly exclusive associations to circRNA expression could potentially act as causal variants or explain an additional part of the genetic architecture of the disease, linking GWAS phenotypes and diseases to putative target genes and regulatory elements. With the findings presented here, we provide a set of SNPs for future investigations of the role of circular RNAs in the genetic and molecular mechanisms underlying diseases and traits.

MATERIALS AND METHODS

Cell culture

Human T-lymphocyte Jurkat cells and LCL were cultured in RPMI (Sigma-Aldrich) supplemented with 10% fetal bovine serum (Life Technologies). The LCL was isolated and characterized from a healthy human subject (LCL-H2).

RNA isolation, RNase R treatment, and DNA isolation

Total RNA from Jurkat cells was isolated by using RNeasy Mini Kit (QIAGEN). For RNase R treatment, 20 units of RNase R (Epicentre), 1 μ L of RNase R buffer, and 1 unit/ μ L Murine Ribonuclease Inhibitor (New England Biolabs) were added to the 2 μ g of RNA samples and incubated for 30 min at 37°C. Total DNA was isolated from Jurkat cells using a DNeasy Kit (QIAGEN).

Sanger sequencing and real-time PCR

One microgram of total RNA was used for cDNA synthesis using the iScript Select cDNA Synthesis Kit (Bio-Rad). Outward-facing primers with respect to genomic sequence were designed to specifically amplify the backsplice junctions of representative circRNAs in a PCR assay. PCR reactions were performed for cDNA samples and genomic DNA using 28 cycles. PCR products were analyzed by 2.2% agarose gel electrophoresis and then purified using GenElute Gel Extraction Kit (Sigma-Aldrich). Gel purified PCR products were directly sequenced to identify the gel product.

For real-time PCR, Fast Start Universal SYBR Green Master Mix (Roche) was used to amplify the backsplice junctions using divergent primers; primers for mRNA were obtained from Primer3 software. Each real-time assay was done in triplicate using the StepOne-Plus Real-time PCR system (Life Technologies).

Processing of genotype data

The 1000 Genomes Project phase3 release of variant calls was downloaded from <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>. The variant set contained genotype data for 2504 individuals from 26 populations. The `vcf-subset` function from VCFtools (Danecek et al. 2011) was used to extract the genotype data for 358 European samples from the downloaded VCF files. These corresponded to 89 CEU, 92 FIN, 86 GBR, and 91 TSI samples. We recomputed the allele number and allele counts with `-recode -recode-INFO-all` parameters for the defined number of samples, and the variants with a resulting minimum allele frequency of >5% were taken for further analysis. Further variants that did not pass the HWE test were also filtered out. The original VCF files were GRCh37 reference assembly-mapped. We used the `All.vcf.gz` file downloaded from ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606_b142_GRCh38/VCF/ and containing GRCh38 mapped coordinates for 1000 Genome variants to remap the GRCh37 variant coordinates from the original VCF files to GRCh38 cords using custom Perl scripts.

Processing of RNA-seq data

RNA-seq data for 358 European individuals corresponding to our genotype samples were downloaded from <http://www.ebi.ac.uk/arrayexpress/experiments/E-GEUV-1/samples>, and its mapping with the genotype information was obtained from <http://www.ebi.ac.uk/arrayexpress/files/E-GEUV-1/E-GEUV-1.sdrf.txt>. CircRNA pipeline (Ahmed et al. 2016) was run on all 358 RNA-seq data sets on a Linux cluster managing batches of jobs as per the load on the cluster. Circular RNA structures are identified by using Bowtie 2 (Langmead and Salzberg 2012) to align paired-end RNA-seq data to a custom reference exome containing all possible pairs of intragenic nonlinear combinations of exons, as well as single-exon “backsplices.” A scrambled junction is inferred as a circRNA candidate when one mate of a paired-end read aligns at the backsplice junction with a minimum of 10 bp overlap on either side of the junction and the other mate aligns in the body of the backsplice forming exon(s). Supportive evidence is also considered when mates of a paired-end read align to the exons in divergent orientation with respect to the genomic sequence, suggesting a scrambled junction instead of a linear junction. These alignments are further filtered to remove any potential PCR duplicates, and only primary alignments are kept for final assessment of circRNA expression (Ahmed et al. 2016). The expression of the circRNA in each sample is then defined as the sum of junctional read-pairs and supportive read-pairs. As the supportive reads align in the body of the backsplice forming exon(s), the length of these exons could influence the levels of circRNA expression with longer exons likely to have a higher number of supportive read-pairs. The circRNA expression is thus corrected for the exon length using the FPKM method. Although FPKM seems the appropriate method to report circRNA expression, it may lead to underestimation of circRNA counts, as it is hard to distinguish whether reads falling within an exon belong to a circular or linear transcript.

With all 358 samples, aligned and circRNAs called, results were merged into one data table with rows as circRNA name and columns as counts from each sample. The pipeline initially reported

3,332,457 backsplice junctions with at least one read pair in 358 samples. We applied the first filtering cutoff for a minimum of 50 junctional reads across the samples, and this resulted in 323,900 circRNA candidates. As we also consider supportive evidence, the minimum of 50 junctional read-pairs could come from either one sample or multiple samples. Nonetheless, the requirement for a high number of minimally seen junctional reads mostly translates to have the junctional evidence coming from multiple samples rather than a single sample. A further filtering for expression in at least 60% of the samples and outliers removed (expression <500 or >50000 counts) resulted in 95,675 circRNAs that were taken for the final association analysis.

circQTL calling

We used Matrix eQTL (Shabalina 2012) to obtain the significance of the correlation between genotypes and circRNA expression after correcting for known and inferred technical covariates. The genotype VCF files were converted into Matrix_eQTL input genotype and snploc files using custom Perl scripts. The raw quantifications for circRNA expression were library size and length corrected (converted to FPKM), quantile normalized, and \log_2 transformed, and association with genotype was determined by linear regression on normalized circRNA expression values. For controlling potential confounding factors, we included the first three principal components from the genotype data—RNA library sizes, population ID, and RNA-sequencing lab ID—as covariates in the regression model. All SNPs that lie within a 1-Mb region of the gene boundaries were tested for cisQTL associations with the circRNA expression of the gene. The *P*-values obtained from Matrix-eQTL were adjusted for multiple testing using the eigenMT package, which uses a LD aware method for multiple testing correction in eQTL studies (Davis et al. 2016).

Matched control sets

The circQTL SNPs that are exclusively associated with the circRNA expression were used for all enrichment analysis (functional/regulatory/disease). The 139,485 circQTLs identified in this study correspond to 23,769 unique SNPs exclusively associated with circRNA expression. We used this list to generate BED files of random background sets of eQTL SNPs for comparing the enrichments. First, the set of eQTL SNPs from Lappalainen et al. (2013) was filtered using a custom script to include only SNPs that are not in LD with the circQTL SNPs ($r^2 < 0.2$). Finally, using circQTL and non-LD eQTL lists of SNPs as input in UES (MacArthur et al. 2017), we generated random background sets of eQTL SNPs for use in enrichment analysis.

Functional annotation enrichment analysis of circQTL SNPs

To conduct functional enrichment annotation analysis for circQTL SNPs, we generated 10,000 random background data sets of matching eQTL SNPs using the process described above. The random sets generated by UES (MacArthur et al. 2017) are in BED file format. We extracted the SNP IDs from BED files and used them in VCFtools (Danecek et al. 2011) to generate VCF files

for the background data. Similarly, using VCFtools, a VCF file was generated for the circQTL SNPs. We used SnpEff (Cingolani et al. 2012), with the options (-csvStats -spliceSiteSize 20 -ud 0 -no-intergenic -no-utr -v -reg) to annotate both circQTL and the eQTL background sets of VCF files for the following functional annotation terms defined for GRCh38.86: EXON, INTERGENIC, INTRON, REGULATION, SPLICE_SITE_ACCEPTOR, SPLICE_SITE_DONOR, SPLICE_SITE_REGION, UTR_3_PRIME, UTR_5_PRIME. The output csv files generated by SnpEff are processed using custom scripts to compute odds ratio (OR), 95% CI, and *P*-value of the enrichment. The results of the SPLICE_SITE_ACCEPTOR, SPLICE_SITE_DONOR, and SPLICE_SITE_REGION terms were combined and reported as "SPLICE." The Monte Carlo *P*-value for the enrichment or underrepresentation of each functional term was calculated as follows:

$$\text{For enrichment, } P = \frac{\# \% \text{background} \geq \# \% \text{circQTL}}{\# \text{ iterations}}$$

$$\text{For underrepresentation, } P = \frac{\# \% \text{background} \leq \# \% \text{circQTL}}{\# \text{ iterations}},$$

where

= number of times,

%background = fraction of SNPs for the annotation term in background eQTL sets,

%circQTL = fraction of SNPs for the annotation term in circQTL data set,

iterations = 10,000.

Regulatory elements enrichment analysis of circQTL SNPs

Encode annotations of the regulatory tracks corresponding to GRCh38.86 were downloaded through SnpEff using its download option. We generated 10,000 random background data sets of matching eQTL SNPs using the process described above and converted them into VCF files as described in the functional annotation enrichment analysis. We used SnpEff (Cingolani et al. 2012) to annotate both circQTL and the eQTL background sets of VCF files for these regulatory tracks of the lymphoblastoid (GM12878) cell line: ATF3-GM12878_enriched_sites BATF-GM12878_enriched_sites BCL11A-GM12878_enriched_sites BCL3-GM12878_enriched_sites BCLAF1-GM12878_enriched_sites CTCF-GM12878_enriched_sites DNase1-GM12878_enriched_sites EBF1-GM12878_enriched_sites Egr1-GM12878_enriched_sites ELF1-GM12878_enriched_sites ETS1-GM12878_enriched_sites Gabp-GM12878_enriched_sites H2AZ-GM12878_enriched_sites H3K27ac-GM12878_enriched_sites H3K27me3-GM12878_enriched_sites H3K36me3-GM12878_enriched_sites H3K4me1-GM12878_enriched_sites H3K4me2-GM12878_enriched_sites H3K4me3-GM12878_enriched_sites H3K79me2-GM12878_enriched_sites H3K9ac-GM12878_enriched_sites IRF4-GM12878_enriched_sites Jund-GM12878_enriched_sites MEF2A-GM12878_enriched_sites MEF2C-GM12878_enriched_sites Nrsf-GM12878_enriched_sites p300-GM12878_enriched_sites Pax5-GM12878_enriched_sites Pbx3-GM12878_enriched_sites PolII-GM12878_enriched_sites PolIII-GM12878_enriched_sites POU2F2-GM12878_enriched_sites PU1-GM12878_enriched_sites Rad21-GM12878_enri-

ched_sites RXRA-GM12878_enriched_sites SIX5-GM12878_enriched_sites SP1-GM12878_enriched_sites Srf-GM12878_enriched_sites TAF1-GM12878_enriched_sites Tcf12-GM12878_enriched_sites Tr4-GM12878_enriched_sites USF1-GM12878_enriched_sites Yy1-GM12878_enriched_sites ZBTB33-GM12878_enriched_sites ZEB1-GM12878_enriched_sites. The annotated VCF files generated by SnpEff were processed using custom scripts to compute the fraction of SNPs in circQTL and background eQTL data sets, odds ratio (OR), 95% CI, and *P*-value of the enrichment for each regulatory element. The Monte Carlo *P*-value for the enrichment or underrepresentation of each regulatory mark was computed as described in the functional enrichment annotation section.

Enrichment of circQTL SNPs among GWAS loci

The NHGRI-EBI catalog of published GWASs was downloaded from EBI (<http://www.ebi.ac.uk/gwas>; gwas_catalog_v1.0.1-associations_e91_r2018-01-01.tsv file). The genomic loci were separately created for circQTL and randomly generated background eQTL SNPs based on their LD patterns. To estimate LD patterns within a genomic region, PLINK (Purcell et al. 2007) version 1.9 was run with the options -r2 -ld-window-kb 1000 -ld-window-r2 0.8 and SNPs with $r^2 > 0.8$ with an index SNP were grouped together into a genomic locus. The genomic loci created for circQTL and background eQTL SNPs were sorted by chromosome and start position and converted into BED files. Similarly, columns DISEASE/TRAIT, REGION, CHR_ID, CHR_POS, MAPPED_TRAIT, and MAPPED_TRAIT_URI were extracted from the GWAS catalog file and processed into a BED file sorted by CHR_ID and CHR_POS. Finally, bedtools intersect (Quinlan and Hall 2010) was used to screen overlaps of GWAS SNPs with genomic loci BED files. An associated genomic locus for a trait as defined by the EFO (Malone et al. 2010) tag of MAPPED_TRAIT_URI field in the GWAS catalog is identified if a GWAS risk SNP falls within the locus. Finally, the enrichment of circQTL SNPs among associated loci is evaluated by a one-tailed Fisher's exact test with the following 2×2 table: columns, circQTL SNPs and control eQTL SNPs; rows, SNPs within and not within the disease-associated loci. A total of 306 traits were evaluated for the enrichment and the *P*-values obtained from Fisher's exact test were corrected for the multiple testing based on Bonferroni correction.

DATA DEPOSITION

All data generated or analyzed during this study are included in this published article (and its Supplemental Information Files).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work was funded by grants from the Basic Medical Research Program (BMRP) from the Qatar Foundation to WCM-Q. LCL-H2 and Jurkat cells were kind gifts from Dr. Said Dermime's laboratory at Hamad Medical Corporation in Qatar.

Received April 28, 2019; accepted September 11, 2019.

REFERENCES

- Ahmed I, Karedath T, Andrews SS, Al-Azwani IK, Ali Mohamoud Y, Querleu D, Rafii A, Malek JA. 2016. Altered expression pattern of circular RNAs in primary and metastatic sites of epithelial ovarian carcinoma. *Oncotarget* **7**: 36366–36381. doi:10.18632/oncotarget.8917
- Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, et al. 2015. A global reference for human genetic variation. *Nature* **526**: 68–74. doi:10.1038/nature15393
- Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, Pappalardo Z, Clarke SL, Wenger AM, Nguyen L, et al. 2012. Coding exons function as tissue-specific enhancers of nearby genes. *Genome Res* **22**: 1059–1068. doi:10.1101/gr.133546.111
- Cingolani P, Platts A, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w¹¹¹⁸*; *iso-2*; *iso-3*. *Fly (Austin)* **6**: 80–92. doi:10.4161/fly.19695
- Danan M, Schwartz S, Edelheit S, Sorek R. 2012. Transcriptome-wide discovery of circular RNAs in *Archaea*. *Nucleic Acids Res* **40**: 3131–3142. doi:10.1093/nar/gkr1009
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158. doi:10.1093/bioinformatics/btr330
- Davis JR, Fresard L, Knowles DA, Pala M, Bustamante CD, Battle A, Montgomery SB. 2016. An efficient multiple-testing adjustment for eQTL studies that accounts for linkage disequilibrium between variants. *Am J Hum Genet* **98**: 216–224. doi:10.1016/j.ajhg.2015.11.021
- De Caterina R, Zampolli A. 2004. From asthma to atherosclerosis—5-lipoxygenase, leukotrienes, and inflammation. *N Engl J Med* **350**: 4–7. doi:10.1056/NEJMp038190
- Dower K, Rosbash M. 2002. T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export. *RNA* **8**: 686–697. doi:10.1017/S1355838202024068
- Du WW, Yang W, Liu E, Yang Z, Dhaliwal P, Yang BB. 2016. Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res* **44**: 2846–2858. doi:10.1093/nar/gkw027
- The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74. doi:10.1038/nature11247
- Glažar P, Papavasileiou P, Rajewsky N. 2014. circBase: a database for circular RNAs. *RNA* **20**: 1666–1670. doi:10.1261/rna.043687.113
- Greene J, Baird A-M, Brady L, Lim M, Gray SG, McDermott R, Finn SP. 2017. Circular RNAs: biogenesis, function and role in human diseases. *Front Mol Biosci* **4**: 38. doi:10.3389/fmolb.2017.00038
- Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, Greenside P, Srivas R, Phanstiel DH, Pekowska A, et al. 2015. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* **162**: 1051–1065. doi:10.1016/j.cell.2015.07.048
- GTEx Consortium. 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**: 648–660. doi:10.1126/science.1262110
- Guo JU, Agarwal V, Guo H, Bartel DP. 2014. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol* **15**: 409. doi:10.1186/s13059-014-0409-z
- Han D, Li J, Wang H, Su X, Hou J, Gu Y, Qian C, Lin Y, Liu X, Huang M, et al. 2017. Circular RNA circMTO1 acts as the sponge of microRNA-9 to suppress hepatocellular carcinoma progression. *Hepatology* **66**: 1151–1164. doi:10.1002/hep.29270
- Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J. 2013. Natural RNA circles function as efficient microRNA sponges. *Nature* **495**: 384–388. doi:10.1038/nature11993
- Haque N, Oberdoerffer S. 2014. *Chromatin and splicing*, pp. 97–113. Humana Press, Totowa, NJ.
- Hentze MW, Preiss T. 2013. Circular RNAs: splicing's enigma variations. *EMBO J* **32**: 923–925. doi:10.1038/emboj.2013.53
- Jeck WR, Sharpless NE. 2014. Detecting and characterizing circular RNAs. *Nat Biotechnol* **32**: 453–461. doi:10.1038/nbt.2890
- Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. 2013. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**: 141–157. doi:10.1261/ma.035667.112
- Kornblihtt AR. 2007. Coupling transcription and alternative splicing. *Adv Exp Med Biol* **623**: 175–189. doi:10.1007/978-0-387-77374-2_11
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359. doi:10.1038/nmeth.1923
- Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, González-Porta M, Kurbatova N, Griebel T, Ferreira PG, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**: 506–511. doi:10.1038/nature12531
- Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, Fatima A, Santini T, Andronache A, Wade M, et al. 2017. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol Cell* **66**: 22–37.e9. doi:10.1016/j.molcel.2017.02.017
- Li J, Yang J, Zhou P, Le Y, Zhou C, Wang S, Xu D, Lin H-K, Gong Z. 2015. Circular RNAs in cancer: novel insights into origins, properties, functions and implications. *Am J Cancer Res* **5**: 472–480.
- Li Yi, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, Gilad Y, Pritchard JK. 2016. RNA splicing is a primary link between genetic variation and disease. *Science* **352**: 600–604. doi:10.1126/science.aad9417
- Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* **327**: 996–1000. doi:10.1126/science.1184208
- MacArthur J, Bowler E, Cerezo M, Gil L, Hall P, Hastings E, Junkins H, McMahon A, Milano A, Morales J, et al. 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res* **45**: D896–D901. doi:10.1093/nar/gkw1133
- Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, Zhukova A, Brazma A, Parkinson H. 2010. Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics* **26**: 1112–1118. doi:10.1093/bioinformatics/btq099
- McCracken S, Rosonina E, Fong N, Sikes M, Beyer A, O'Hare K, Shuman S, Bentley D. 1998. Role of RNA polymerase II carboxy-terminal domain in coordinating transcription with RNA processing. *Cold Spring Harb Symp Quant Biol* **63**: 301–309. doi:10.1101/sqb.1998.63.301
- Memczak S, Jens M, Elefsinioti A, Torti F. 2013a. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**: 333–338. doi:10.1038/nature11928
- Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, Maier L, Mackowiak SD, Gregersen LH, Munschauer M, et al. 2013b. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**: 333–338. doi:10.1038/nature11928
- Pamudurti NR, Bartok O, Jens M, Ashwal-Fluss R, Stottmeister C, Ruhe L, Hanan M, Wyler E, Perez-Hernandez D, Ramberger E,

- et al. 2017. Translation of CircRNAs. *Mol Cell* **66**: 9–21.e7. doi:10.1016/j.molcel.2017.02.021
- Piwecka M, Glažar P, Hernandez-Miranda LR, Memczak S, Wolf SA, Rybak-Wolf A, Filipchyk A, Klironomos F, Cerda Jara CA, Fenske P, et al. 2017. Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science* **357**: eaam8526. doi:10.1126/science.aam8526
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559–575. doi:10.1086/519795
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- Rgio S, De Almeida F, Carmo-Fonseca M. 2012. Design principles of interconnections between chromatin and pre-mRNA splicing. *Trends Biochem Sci* **37**: 248–253. doi:10.1016/j.tibs.2012.02.002
- Salzman J, Gawad C, Wang PL, Lacayo N, Brown PO. 2012. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS ONE* **7**: e30733. doi:10.1371/journal.pone.0030733
- Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. 2013. Cell-type specific features of circular RNA expression. *PLoS Genet* **9**: e1003777. doi:10.1371/journal.pgen.1003777
- Schor IE, Gómez Acuña LI, Kornblihtt AR. 2013. Coupling between transcription and alternative splicing. *Cancer Treat Res* **158**: 1–24. doi:10.1007/978-3-642-31659-3_1
- Shabalín AA. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**: 1353–1358. doi:10.1093/bioinformatics/bts163
- Shen T, Han M, Wei G, Ni T. 2015. An intriguing RNA species—perspectives of circularized RNA. *Protein Cell* **6**: 871–880. doi:10.1007/s13238-015-0202-0
- Stergachis AB, Haugen E, Shafer A, Fu W, Vernot B, Reynolds A, Raubitschek A, Ziegler S, LeProust EM, Akey JM, et al. 2013. Exonic transcription factor binding directs codon choice and affects protein evolution. *Science* **342**: 1367–1372. doi:10.1126/science.1243490
- Takata A, Matsumoto N, Kato T, Swertz MA, Kapushesky M. 2017. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat Commun* **8**: 14519. doi:10.1038/ncomms14519
- Tang W, Ji M, He G, Yang L, Niu Z, Jian M, Wei Y, Ren L, Xu J. 2017. Silencing CDR1as inhibits colorectal cancer progression through regulating microRNA-7. *Oncotargets Ther* **10**: 2045–2056. doi:10.2147/OTT.S131597
- van de Geijn B, McVicker G, Gilad Y, Pritchard JK. 2015. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods* **12**: 1061–1063. doi:10.1038/nmeth.3582
- Vincent HA, Deutscher MP. 2006. Substrate recognition and catalysis by the exoribonuclease RNase R. *J Biol Chem* **281**: 29769–29775. doi:10.1074/jbc.M606744200
- Vos PG, Paulo MJ, Voorrips RE, Visser RGF, van Eck HJ, van Eeuwijk FA. 2017. Evaluation of LD decay and various LD-decay estimators in simulated and SNP-array data of tetraploid potato. *Theor Appl Genet* **130**: 123–135. doi:10.1007/s00122-016-2798-8
- Wang Y, Wang Z. 2015. Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**: 172–179. doi:10.1261/rna.048272.114
- Wang PL, Bao Y, Yee M-C, Barrett SP, Hogan GJ, Olsen MN, Dinneny JR, Brown PO, Salzman J. 2014. Circular RNA is expressed across the eukaryotic tree of life. *PLoS One* **9**: e90859. doi:10.1371/journal.pone.0090859
- Wilhelm BT, Marguerat S, Aligianni S, Codlin S, Watt S, Bähler J. 2011. Differential patterns of intronic and exonic DNA regions with respect to RNA polymerase II occupancy, nucleosome density and H3K36me3 marking in fission yeast. *Genome Biol* **12**: R82. doi:10.1186/gb-2011-12-8-r82
- Yang Y, Fan X, Mao M, Song X, Wu P, Zhang Y, Jin Y, Yang Y, Chen L, Wang Y, et al. 2017. Extensive translation of circular RNAs driven by N⁶-methyladenosine. *Cell Res* **27**: 626–641. doi:10.1038/cr.2017.31
- Yu L, Gong X, Sun L, Zhou Q, Lu B, Zhu L. 2016. The circular RNA Cdr1as act as an oncogene in hepatocellular carcinoma through targeting miR-7 expression. *PLoS One* **11**: e0158347. doi:10.1371/journal.pone.0158347
- Zhou H-L, Luo G, Wise JA, Lou H. 2014. Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Res* **42**: 701–713. doi:10.1093/nar/gkt875