


SOFTWARE

Open Access



SEESAW: detecting isoform-level allelic imbalance accounting for inferential uncertainty

Euphy Y. Wu¹, Noor P. Singh², Kwangbom Choi³, Mohsen Zakeri², Matthew Vincent³, Gary A. Churchill³, Cheryl L. Ackert-Bicknell⁴, Rob Patro² and Michael I. Love^{1,5*} 

*Correspondence:
michaelsaiahlove@gmail.com

¹ Department of Biostatistics,
University of North Carolina-
Chapel Hill, Chapel Hill, NC, USA

² Department of Computer
Science, University of Maryland,
College Park, MD, USA

³ The Jackson Laboratory, Bar
Harbor, ME, USA

⁴ Department of Orthopedics,
School of Medicine, University
of Colorado, Anschutz Campus,
Aurora, CO, USA

⁵ Department of Genetics,
University of North Carolina-
Chapel Hill, Chapel Hill, NC, USA

Abstract

Detecting allelic imbalance at the isoform level requires accounting for inferential uncertainty, caused by multi-mapping of RNA-seq reads. Our proposed method, SEESAW, uses Salmon and Swish to offer analysis at various levels of resolution, including gene, isoform, and aggregating isoforms to groups by transcription start site. The aggregation strategies strengthen the signal for transcripts with high uncertainty. The SEESAW suite of methods is shown to have higher power than other allelic imbalance methods when there is isoform-level allelic imbalance. We also introduce a new test for detecting imbalance that varies across a covariate, such as time.

Background

Genome-wide association studies (GWAS) have identified tens of thousands of genomic loci that are associated with complex traits or diseases, many of which are located in non-coding regulatory regions [1]. One potential mechanism by which allelic variation in these non-coding regions may affect phenotype is that the variants reside in transcription factor (TF) binding sites and influence the activities of TFs and transcription. Such a non-coding region may be referred to as a *cis*-regulatory element (CRE). Individuals that are heterozygous at such a variant may exhibit imbalanced allelic expression at any genes regulated by the CRE. With RNA-sequencing (RNA-seq) experiments, it is possible to observe such imbalance in allelic expression in the sequenced reads for those individuals also heterozygous for a variant in the exons of a regulated gene; other mechanisms of allelic imbalance, e.g., sequences affecting splicing or post-transcriptional regulation, are also possible to be detected. Recent advances in long-read technologies enable reconstruction of individual diploid genomes/transcriptomes, which leads to more accurate analysis at allele and isoform resolution. Analysis of allelic imbalance (AI) has the potential for higher power to detect *cis*-genetic regulation than analysis of total expression, as *trans*-regulatory and non-genetic effects on expression level are controlled for when comparing the



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

two alleles within samples [2–8]. Such effects controlled for in AI analysis include both biological variability and technical artifacts that may distort total expression levels across genes. AI is also a powerful analysis to reveal *cis*-genetic regulation in heterozygous individuals that varies across samples representing different conditions, tissues, spatial contexts, or time periods [9–13].

AI can be isoform-specific, but due to low statistical power and challenges to statistical inference caused by multi-mapping, AI is often measured at the exon-level or gene-level. If different isoforms are subject to regulation from different sets of CRE, and these harbor genetic variants for which the individual is heterozygous, then such isoforms may exhibit different strength or direction of AI, as has been observed recently in an analysis of genomic imprinting at isoform level [14] and in a survey of expression in GTEx using long-read technology [15]. Being able to detect AI at the isoform level could provide insight into mechanisms of complex traits and diseases. One challenge is that AI can only be observed when the individuals under study are heterozygous at an exonic variant. Furthermore, only a subset of reads that can be aligned or probabilistically assigned to a transcript or gene will provide allelic information. As described by Raghupathy et al. [16], an RNA-seq read can fall into various categories of multi-mapping with respect to gene, isoform, and allele, providing different information for expression estimation or “quantification” at different levels. The uncertainty in measuring the expression level from multi-mapping is referred to here as inferential uncertainty. When variants lie in exons that are not constitutive, reads overlapping these exonic variants provide information for isoform-level AI. In the other case, if exonic variants lie in constitutive exons, or in exons of dominant isoforms, AI can be effectively detected by existing methods such as *phASER* [8] or *WASP* [17], which count reads aligning to gene haplotypes. *WASP* for example examines imbalance in the pileup of reads mapped to the genome, after correction for technical bias due to differential mapping rates of the two alleles [18]. *WASP* can be followed up by methods such as *ASEP* for statistical inference of gene-level AI across a population of individuals that may be homozygous or heterozygous for regulatory variants [19].

A subset of existing methods are able to detect *cis*-genetic regulation at sub-gene resolution from short-read RNA-seq. Paired Replicate Analysis of Allelic Differential Splicing Events (*PAIRADISE*) can extract more information about allelic exon inclusion events, by counting reads that overlap both an informative splice junction and an exonic variant for which a subject is heterozygous [20]. Within the *PAIRADISE* framework, reads are mapped to personalized genomes based on phased genotypes. *PAIRADISE* provides a statistical model for detecting allele-specific splicing events, by aggregating allelic exon inclusion within individuals, and builds upon their previous method *GLiMMPS* to detect splicing quantitative trait loci (QTL) across donors of all genotypes [21]. *IDP-ASE* combines counts of reads from short-read RNA-seq falling along regions of exons with better resolved isoforms and alleles using long reads [22]. A potential limitation for approaches that count reads overlapping specific regions of genes is that these may not be able to fully aggregate allelic information from paired-end reads overlapping multiple informative features along the length of a transcript. For *PAIRADISE*, some cases of isoform-level AI may be missed when focusing on reads overlapping splice junctions, such as allele-specific differences in length of 5′ or 3′ untranslated regions (UTR).

Other method publications that have demonstrated quantification of expression at allelic- and isoform-level include *EMASE* [16], *kallisto* [23], *mmseq* [24], and *RPVG* [14]. *EMASE* proceeds in a similar manner to *PAIRADISE*, by first constructing a diploid reference; however in this case, *EMASE* aligns reads to a diploid transcriptome, constructed via the *g2gtools* software. The *EMASE* authors found that hierarchical assignment of reads based on their information content in some cases outperformed equal apportionment as would occur using EM-based algorithms such as *RSEM* [25], *kallisto* [23], and *Salmon* [26] with a diploid reference transcriptome. *mmseq* allows for alignment of reads to a diploid reference transcriptome using *Bowtie* [27] and additionally can take into account gene-, isoform-, and allelic-multi-mapping when performing inference across alleles in its *mmdiff* step [28]. *mmdiff* computes posterior distributions of expression of each feature via Gibbs sampling. Features can be aggregated at various levels of resolution by summing the posterior expression estimates within each sample. Aggregation also has proved an effective strategy in non-allelic contexts, as demonstrated in *tximport* [29], *SUPPA* [30], and *txrevise* [31]. *mmseq* also provides a method *mmcollapse* [28] to perform *data-driven* aggregation of features to reduce marginal posterior variance, although this procedure cannot currently be combined with differential analysis across alleles (i.e., AI analysis).

We introduce a suite of methods, Statistical Estimation of Allelic Expression using *Salmon* and *Swish* (*SEESAW*), for allelic quantification and inference of AI patterns across samples. With the objective of detecting isoform-level AI, we introduce a strategy to group isoforms based on their transcription start sites (TSS). Through simulation, we show that aggregating isoform-level expression estimates to the TSS level can have higher sensitivity than either gene- or isoform-level analysis. *SEESAW* utilizes *Salmon* [26] to estimate expression with respect to an allele-specific reference transcriptome, and a non-parametric test *Swish* [32] to test for AI. *Swish* incorporates inferential uncertainty into differential testing and makes no assumption of the distributional model of the data. *SEESAW* follows the general framework of *mmseq* and *mmdiff* for haplotype- and isoform-specific quantification and uncertainty-aware inference. Here, the *SEESAW* methods were applied to simulated data to benchmark against previously developed methods for detection of AI within heterozygous individuals, making use of multiple individuals as biological replicates. We applied *SEESAW* to an F1 mouse time course dataset, where it detected genes containing both gene-level AI and isoform-level AI. *SEESAW* can detect cases of AI that are consistent across all samples, differential AI across two groups of samples, or dynamic AI over a covariate, with a new correlation-based test. The statistical testing in *SEESAW* is available via the *Swish* function in the *fishpond* package [33] including a software vignette for allelic analysis.

Results

SEESAW

We first briefly describe the estimation and statistical testing steps in *SEESAW* (Fig. 1), which combines both existing and new functionality across a number of software packages, with further details provided in the “[Methods](#)” section. *SEESAW* assumes that phased genotypes are available, and is primarily designed for multiple

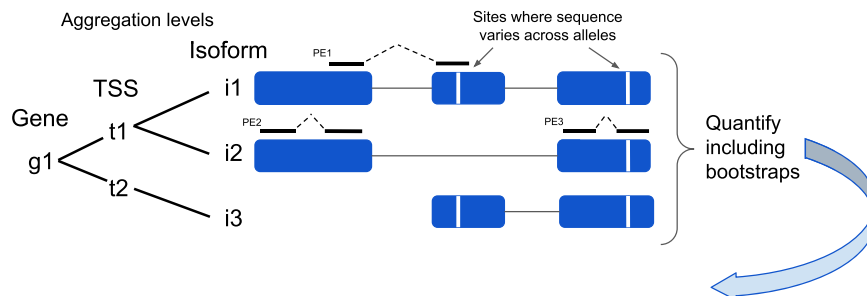
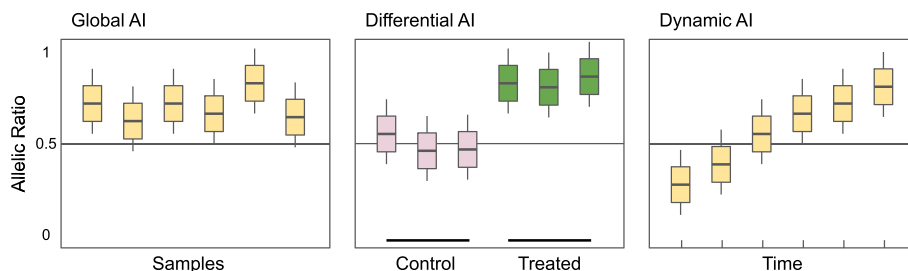
A Quantification over diploid transcriptome with *Salmon***B** Inference of AI: nonparametric testing with *Swish*

Fig. 1 *SEESAW* is a suite of tools for analysis of allelic imbalance across samples, first performing quantification and then statistical inference. **A** *Salmon* is used to quantify single-end (SE) or paired-end (PE) reads over a diploid transcriptome, and then estimates may be aggregated to various levels of resolution: isoform, TSS, or gene level. Different types of reads provide different types of information: PE1 contains both allelic and isoform-level information, PE2 contains only isoform-level information, and PE3 contains only allelic information. Information from all of these types of read data is included in quantification with *Salmon*. **B** *Swish* is then used to perform statistical testing of allelic imbalance across samples, taking into account multiple inferential replicates per sample (shown as boxes). *Swish* can test for global allelic imbalance, or differential or dynamic imbalance with respect to categorical or continuous covariates, respectively

replicates or multiple conditions of organisms with the same genotype. This can occur with multiple replicates of an F1 cross, or cell lines from individual human donors across developmental time points [34–36], or across conditions [37–40]. First, *g2gtools* is used to create a diploid transcriptome, which is indexed by *Salmon* [26] and used for estimating allele-specific expression with bootstrap replicate datasets to assess inferential uncertainty across genes, isoforms, and alleles (detailed *SEESAW* pipeline shown in Additional file 1: Fig. S1). This approach for allelic quantification, mapping reads to a custom diploid transcriptome, has been demonstrated as a successful strategy in previous work [16, 23, 24, 41], similarly for mapping reads to a spliced pangenome graph [14] or to a custom diploid genome for allelic read counting [7, 42–45]. Next, *SEESAW* facilitates importing the estimated allelic counts at various levels of aggregation: no aggregation (labelled hereafter “isoform,” or equivalently “transcript”/“txp”), transcription start site aggregation (“TSS”), or gene-level aggregation (“gene”). Finally, we leverage the *Swish* [32] tool for differential expression analysis to test across the two alleles within samples via the Wilcoxon signed rank test statistic, averaging over inferential replicates, and using the *qvalue* package [46] applied to permuted datasets for defining false discovery rate (FDR) bounded sets of

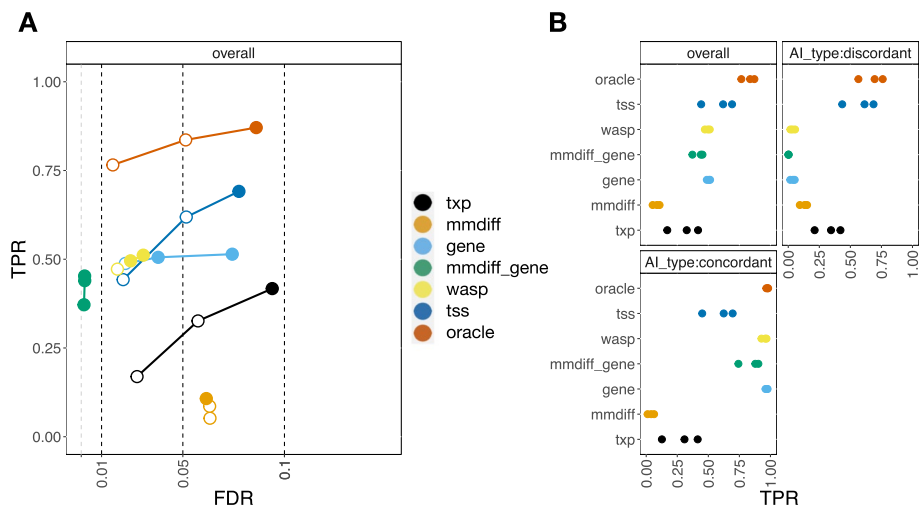


Fig. 2 Comparing results of *SEESAW* on *polyester* simulation with different levels of aggregation to *mmdiff* and *WASP*. *SEESAW* was applied at different levels of resolution including transcript (txp), aggregated-to-TSS (TSS), aggregated-to-gene (gene), and “oracle”, where oracle involved aggregating transcripts by the true AI signal direction, known only in simulation. *mmdiff* was applied at transcript (mmdiff) and gene level (mmdiff_gene), while *WASP* provided gene level analysis. **A** iCOBRA plot of sensitivity (true positive rate, or TPR) over achieved false discovery rate (FDR) with three circles indicating 1%, 5%, and 10% nominal FDR cutoffs, respectively. Filled circles indicate observed FDR less than nominal FDR. **B** Overall sensitivity for all cases of AI and sensitivity stratified by type of AI: “discordant” AI across isoforms within a gene (AI in different directions) or “concordant” AI within gene (AI in the same direction)

features. The steps for testing dynamic allelic imbalance are provided in the “[Methods](#)” section.

Simulation

Simulation of an F1 cross based on the *Drosophila melanogaster* transcriptome was used to assess method performance when the true AI status of each transcript was known. iCOBRA diagrams [47] were used to assess the sensitivity, or true-positive rate (TPR), and the FDR at nominal FDR thresholds of 1%, 5%, and 10%. Sensitivity was assessed per transcript, where detection of AI for a gene-level method was propagated to each of the gene’s expressed isoforms. We used Integrative Genomics Viewer (IGV) [48] to visualize the distribution of HISAT2 [49] aligned reads along the reference genome, after removing allelic-biased multi-mapping reads with *WASP* [17]. While *SEESAW* uses reads mapped to the diploid transcriptome with *Salmon*, examining genome-aligned reads with IGV allowed us to identify examples of reads that contained both allelic- and isoform-level information (Additional file 1: Fig. S2).

Notably, *SEESAW* with TSS aggregation had the highest overall sensitivity at 5% and 10% nominal FDR, above any of the gene- or isoform-level methods (Fig. 2). The reason behind the higher overall sensitivity can be seen when stratifying by types of AI, as in Fig. 2B; *SEESAW* with TSS aggregation was able to detect discordant AI on isoforms within a gene that could be masked after aggregation to the gene level. Discordant AI refers to the case where isoforms within a gene have opposite directions of AI, while concordant AI refers to the case where isoforms within a gene have the same direction of AI. Gene-level *SEESAW*, gene-level *mmdiff*, and *WASP* had loss

of sensitivity to detect these discordant cases of AI. In addition, *SEESAW* with TSS aggregation or gene-level aggregation, gene-level *mmdiff*, and *WASP* had similar sensitivity at 1% nominal FDR considering both discordant and concordant AI, and these methods had observed FDR for this nominal cutoff in the range of 0–2%.

Gene-level *SEESAW*, gene-level *mmdiff*, and *WASP* had higher sensitivity than *SEESAW* using TSS aggregation when AI was concordant across all isoforms of a gene. This is expected as aggregation at the appropriate level strengthens the AI signal while reducing inferential uncertainty, so increasing power. For example, *SEESAW* had the strongest power to detect AI when information about the grouping of transcripts by true AI signal was used to aggregate allelic counts (“oracle” in Fig. 2A). Both *SEESAW* and *mmdiff* at the isoform level did not have as high sensitivity as methods that aggregated signal. UpSet diagrams [50] of the sets of transcripts called by each method compared to the true AI transcripts indicated the highest overlap among the gene-level methods and TSS or oracle aggregation (Additional file 1: Fig. S3).

We demonstrated that error control could be lost if we used allelic log fold change (LFC) in the aggregation step as described in the “Methods” section (Additional file 1: Fig. S4). Aggregating transcripts by the LFC is a form of “double dipping” as it makes use of the counts across samples twice: once to determine allelic LFC for aggregation then again to test for allelic imbalance. Such a procedure can lead to loss of error control as described elsewhere [51]. LFC-based *p*-values did not follow a uniform distribution under the null hypothesis and would lead to increased FDR with increased sample size.

As *SEESAW* makes use of *Salmon* for quantification, with inferential uncertainty measured via bootstraps, we evaluated the accuracy of uncertainty estimation using $n = 10, 20, 30, 50$ and 100 bootstraps, respectively to define 95% bootstrap intervals for the estimated counts (see the “Methods” section). Coverage was evaluated by comparing intervals to the true, simulated counts. Overall these results indicated that the default of $n = 30$ inferential replicates (bootstrap samples) was sufficient for estimation of inferential uncertainty, with coverage rates close to the target 95%. The coverage rate slightly increased with number of inferential replicates, with a range of $\sim 2\%$. The coverage rate increased more across aggregation level, with a range of up to $\sim 10\%$ (Additional file 1: Table S1). The increments in coverage rate were less than 0.5% comparing 30 to 50 inferential replicates within each aggregation level and expression bin. The difference in coverage rate was still small when comparing between 30 and 100 inferential replicates, across all three expression bins (Additional file 1: Fig. S5). We find that 30 inferential replicates established a good compromise between precision of uncertainty information and computation time and storage.

We assessed if the use of inferential uncertainty by *Swish* resulted in better error control compared to AI analysis with a beta-binomial generalized linear model applied to *Salmon* estimated counts, without taking into account uncertainty. *Swish* had better control of the FDR at all levels of aggregation compared to a beta-binomial generalized linear model applied to estimated counts with Benjamini-Hochberg correction of *p*-values to control the FDR (Additional file 1: Fig. S6). *Salmon* estimated counts can have high uncertainty particularly across alleles and isoforms, which often have high similarity in terms of sequence. Collapsing isoforms to TSS or gene

level reduces uncertainty, but allelic multi-mapping remains for AI analysis [16]. This motivated our use of allelic testing methods such as *mmdiff* and *Swish* that take into account inferential uncertainty for estimated allelic counts.

We also assessed the performance of *SEESAW* compared to a new inference pipeline from the *WASP* developers, called *WASP2* (Additional file 1: Fig. S7). *WASP2* was equally sensitive as *WASP* in detecting gene-level AI, while it had less sensitivity than *SEESAW* to detect discordant AI signal as expected since it follows a similar approach to *WASP*. While we used the *locfdr* package [52] for multiple test correction for *WASP*, we found Benjamini-Hochberg [53] correction performed well for computing FDR-bounded sets for *WASP2*.

Osteoblast differentiation time course

We used *Swish* to test for AI at various levels (gene, isoform, TSS) in a time course experiment of differentiating osteoblasts from an F1 mouse, with C57BL/6J dams crossed with CAST/EiJ sires (see the “Methods” section). Following creation of the diploid reference transcripts and quantification steps of the *SEESAW* pipeline, we first tested for consistent AI across all nine time points (the “global AI test”). While exploring the osteoblast differentiation data, we observed that for isoforms of a gene with TSS that were near each other (within 50 bp), these isoforms often shared similar estimated allelic fold change as calculated with *SEESAW*. To facilitate data visualization, strengthen biological signal, and reduce inferential uncertainty, we grouped any transcripts with TSS within 50 bp of each other (referred to here as “fuzzy TSS groups”, to contrast with strict basepair-resolution TSS grouping). We tested at different levels of resolution: gene level, isoform level, and TSS level. To compare across these levels, we looked at genes in common: a gene was considered significant for global AI at isoform level or TSS group level if at least one isoform or TSS group within the gene was significant (nominal FDR < 5%). Isoform-level testing for global AI returned the most genes, with 6116 significant genes, followed by gene-level with 5701 genes, and TSS-level grouping with 5573 genes. The majority of genes (4625) were in common across all three levels of resolution (UpSet plot [50] provided in Additional file 1: Fig. S8).

Gene-level aggregation had high overlap with TSS-level indicating that, at least for global AI testing, most of the AI signal was not masked by discordant direction of AI among isoforms within a gene. Among genes displaying global AI under aggregation to the gene level, the TSS groups within those genes often had estimated imbalance in the same direction as the gene imbalance – 97.3% of significant genes had all of their TSS groups with significant AI having the same direction as the gene-level estimate. However, *SEESAW* was able to detect—among the 2.7% remaining genes—interesting examples of genes that had different direction of AI among its isoforms. A complete list of the 134 genes showing these significant and discordant patterns within gene is provided in Additional file 1: Table S2 and in the Zenodo deposition. For example, *Fuca2* exhibited discordant AI with the CAST/EiJ (CAST) allele more highly expressed than C57BL/6J (B6) for one of the two leftmost (more 5′) TSS but less expressed than B6 for the rightmost (more 3′) TSS, with both TSS groups significant at < 5% FDR (Fig. 3A).

Another gene of the 134 genes with discordant pattern was *Sparc*, the most highly expressed gene at the last time point in the osteoblast differentiation time course.

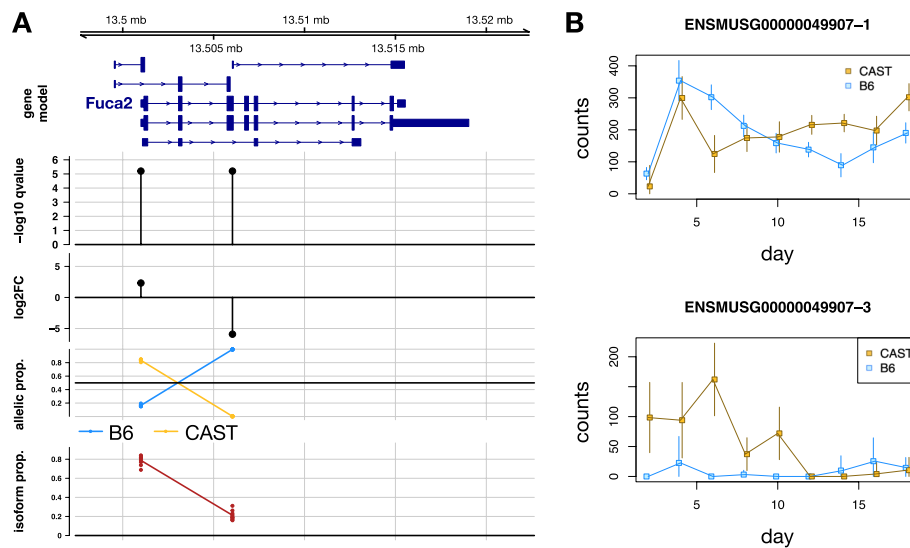


Fig. 3 SEESAW results for the mouse osteoblast differentiation dataset (TSS-level analysis). **A** Global AI results for the gene *Fuca2* where TSS groups showed discordant direction of imbalance. The computed statistics are plotted directly below the TSS group. B6 refers to the strain of C57BL/6J and CAST refers to the strain of CAST/EiJ, each parents in the F1 cross. Isoform proportion per TSS group was calculated by summing the estimated TPM (transcript per million) of the isoforms in the group and dividing by the gene-level TPM. Allelic proportion was calculated by dividing estimated allelic counts for each strain by the total counts from both alleles. **B** Dynamic AI revealed for two TSS groups of *Rasl1b*. Estimation uncertainty shown with error bars (95% intervals based on bootstrap variance)

Sparc is known to be critical for bone development. With fuzzy TSS aggregation and count filtering, *Sparc* displayed four transcripts groups, where group 5 (ENSMUSG00000018593-5) had positive allelic LFC (CAST > B6) and the other groups had negative allelic LFC (Fig. 4).

We additionally tested for “dynamic AI” using the correlation test implemented in *Swish* for testing changes in allelic log fold change over a continuous covariate. We again tested at gene level, isoform level, and TSS level. Gene-level dynamic AI testing returned the largest number of significant genes (nominal FDR < 5%): 57 genes displayed dynamic AI at gene level, 49 genes at TSS level, and 23 at isoform level (Additional file 1: Fig. S9). Those significant genes shared across all levels only represented a third of those detected at gene level, where another third were shared only between gene level and TSS level. Thus TSS-level aggregation appeared to help recover signal that would be lost if only testing at the isoform level.

Interestingly, we detected genes such as *Rasl1b* that had isoform-level AI trending in different directions over time (Fig. 3B, Additional file 1: Figs. S10 and S11). *Rasl1b* exhibited dynamic AI for two TSS groups, with the CAST allele more lowly expressed than the B6 allele for TSS group “1” from day 2 to day 6, roughly balanced from day 8 to day 10, and finally with CAST/EiJ more highly expressed from day 12 to day 18. The other TSS group “3” had almost the opposite allelic ratio behavior: CAST more highly expressed earlier in time but both alleles tending toward balanced, low expression at the end of the time course. While for *Rasl1b* this pattern was also significant when testing at the isoform level, other genes such as *Calcoco1* demonstrated the advantage of grouping features: *Calcoco1* exhibited dynamic AI for two TSS groups, “5” and “6”, which

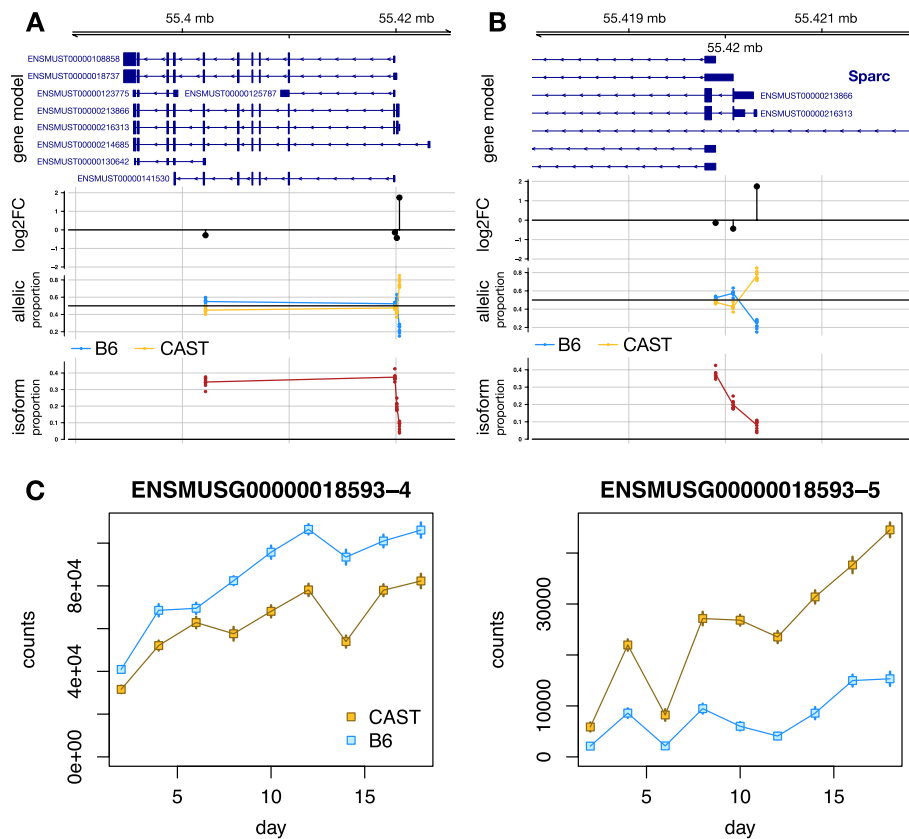


Fig. 4 *Sparc* gene results for osteoblast differentiation dataset (TSS-level global AI analysis). **A** Four transcript groups remained after TSS aggregation and count filtering. One group had positive allelic log fold change (LFC), with CAST expression higher than B6, and the other three groups had negative allelic LFC. **B** The 5' end of the *Sparc* transcripts in group 5 with positive allelic LFC, *ENSMUST00000213866* and *ENSMUST00000216313*. **C** Allelic counts for two discordant transcript groups of *Sparc*. Estimation uncertainty shown with error bars (95% intervals based on bootstrap variance)

differed in the direction of change in the imbalance (Additional file 1: Figs. S12 to S14). Here, the *p*-value and *q*-value for TSS group “6” was reduced when aggregating counts from the isoform to TSS-group level.

Discussion

Our new suite of methods, *SEESAW*, can be used to obtain allele-specific abundance with bootstrap replicates used to capture inferential uncertainty across genes, isoforms, and alleles, and to perform statistical testing of global or dynamic AI. We propose to aggregate estimates of allelic expression of isoforms by their TSS to increase statistical power in testing for AI that is a result of heterozygous variants in the promoter or in CRE that affect a particular promoter. We introduced two different AI testing procedures: global AI to test for the existence of consistent allelic fold changes across samples, and dynamic AI to test for non-zero correlation between the log allelic fold change and a continuous covariate. *SEESAW* can also be used to test differential AI between two groups, as introduced in Zhu et al. [32], or more complex designs using a general regression framework. Differential AI testing and differential correlation AI testing are

shown as examples in the allelic analysis software vignette. The above tests utilize non-parametric testing, thus making no assumption on the distribution of the data itself. Nonparametric testing had better performance on the simulated data than a standard beta-binomial generalized linear model. In simulation, we demonstrated that *SEESAW* on TSS level had the highest sensitivity in the case that AI was discordant within gene, and achieved an FDR that was close to the nominal value at all levels of resolution (gene-, TSS-, or isoform-level testing), implying *SEESAW* can maintain error control despite high and heterogeneous levels of uncertainty. *SEESAW* at gene level performed comparably to existing methods such as *WASP* and gene-level *mmdiff*. For the osteoblast differentiation experiment, *SEESAW* was able to recover some genes with discordant isoform-level AI across all time points and was able to detect genes with isoform-level AI that changed over time in different directions.

Currently, *SEESAW* does not support alignment of haplotypes across individuals of different genotype. SNP-based analysis simplifies this problem, but at a loss of information, as evidence of AI may be distributed across multiple exonic variants within a transcript. A newly developed approach *RPVG* [14] maps RNA-seq reads to a spliced pangenome, and then provides haplotype-specific transcript abundance estimates for each individual. It would require further work for the methods presented here to group individuals by their haplotype combinations per gene and perform across-sample inference while accounting for estimation uncertainty using *Swish*. Another limitation of our current approach is that grouping isoforms together based on their TSS reveals shared promoter-based regulation, but may miss isoform-specific AI caused by intronic variation or variation that affects nonsense-mediated-decay. *IDP-ASE* and *PAIRADISE* provide inference on AI of splicing events, and these methods could be considered for detection of these cases. Alternatively, the framework of *SEESAW* can be adapted and used with other aggregation rules for different biological purposes, e.g., aggregating isoforms by various splicing events in a manner similar to *SUPPA* or *txrevise*. While *SEESAW* can be used at various levels of resolution, from transcript or TSS group up to gene level, if the focus of interest is gene-level AI, we found that *WASP* and *WASP2* were equally sensitive and had good control of false discoveries, using *locfdr* and Benjamini-Hochberg correction, respectively. Additionally, the *ASEP* [19] method can be applied to allelic counts from *WASP* and allows for detection of gene-level AI across a population using a mixture model to account for the unobserved regulatory variants—individuals that are heterozygous for exonic variants may be either homozygous or heterozygous for regulatory variants. The analytical consequences of multiple regulatory SNPs and varying degree of linkage disequilibrium (LD) of these to the exonic SNP, with respect to detection of AI, have been described previously by Xiao and Scott [4].

While here we relied on gene annotation to group together transcripts and reduce inferential uncertainty of allelic expression estimates, another approach would be to use data-driven aggregation methods such as *mmcollapse* and *Terminus* [54]. We were not able to perform differential testing across alleles with *mmdiff* after aggregation with *mmcollapse*. A future direction that may improve performance with the inclusion of *Terminus* in the *SEESAW* pipeline would be stratification of different null distributions for test statistics in *Swish* based on aggregation level (transcript, transcript-group, or gene level).

After having detected AI, a natural next step is to try to understand the mechanism of *cis*-genetic regulation. It is possible to associate the AI seen on transcripts or genes with one or more regulatory variants, either through phasing or usage of population-level LD to establish the search space. The list of candidate regulatory SNPs can be further refined by integrating allelic signal within CRE at the epigenomic level, including allelic binding of proteins [42, 55–57], allelic accessibility [58, 59], or allelic methylation [60]. Alternatively, search for altered transcription factor binding motifs can be combined with RNA or protein abundance of potential regulators to winnow down the list of candidate causal regulatory variants [11, 61, 62]. It may also be of interest to detect in which cell types the allelic signal may be strongest or exclusively present, as has been investigated in recent methods for single cell allelic expression or accessibility datasets [63–65]. Finally, we note that a number of methods have shown that AI can be effectively integrated with total expression across individuals of all three genotypes [5, 13, 66–69]. This approach uses more information and so should produce a gain in sensitivity, as well as extending beyond genes harboring exonic variants, which is a limitation for AI-based methods.

Conclusions

Here we present a new suite of methods, called *SEESAW*, for quantifying and testing AI. *SEESAW* offers analysis at various levels of resolution (isoform, TSS, gene-level) and has significantly improved performance compared to existing methods for detecting when there is isoform-level imbalance. *SEESAW* provides statistical testing for global AI (across all samples), dynamic AI (differences along a continuous axis such as time), or differential AI (across groups of samples). The statistical testing in *SEESAW* is available in an R/Bioconductor package, *fishpond* [33], with an associated software vignette and visualizations designed specifically for allelic analysis.

Methods

SEESAW

The following steps were applied to analyze isoform-, TSS-, or gene-level AI. A sample-specific diploid transcriptome was constructed using *g2gtools* with the following input data: a reference genome (FASTA file); haplotype-specific variants (SNP and indel VCF files); and a catalog of all possible transcripts in the reference genome (GFF or GTF file). *g2gtools* was used to patch and transform the reference genome using the SNPs and indels from the the VCF file, and to extract transcripts from each haplotype of the custom diploid genome. Combined, these transcripts form the custom diploid transcriptome used to quantify RNA-seq reads. *Salmon* [26] was used to quantify expression at the level of allelic transcripts, where both alleles are kept in the reference during indexing (`--keepDuplicates`). During the quantification step, 30 bootstrap replicates were generated to capture inferential uncertainty across genes, isoforms, and alleles (`--numBootstraps 30`).

To increase statistical power in testing for AI at sub-gene resolution, we recommend to group isoforms at a resolution that prioritizes discovery of *cis*-genetic regulation effects from non-coding variation in the promoter or in CRE that affect a particular promoter. Aggregation of isoform expression to higher levels has been shown to reduce inferential uncertainty and may improve detection power as long as the signal of interest

is not at the same time lost or diminished through aggregation [28, 29, 54]. Other related approaches include performing inference at the level of equivalence classes (partitions of reads) [70, 71], although here we focus on analysis performed over sets of one or more transcripts for their biological interpretability. During aggregation, both estimated counts (point estimates) and bootstrap replicate counts were summed across isoforms within a TSS-based group for each allele. TSS groups can be defined strictly (identical start position) or with some basepair (bp) tolerance (“fuzzy TSS groups”). After aggregation, every aggregate feature will have a point estimate of abundance as well as a vector representing the bootstrap distribution for each of the two alleles. Likewise, we explored gene-level aggregation, summing across all isoforms for a gene.

To evaluate the extent of loss of error control, we also assessed another aggregation approach: grouping transcripts after having “double dipped” on the expression data for AI analysis [51]. For this test, we first performed AI analysis (described below) on transcript-level estimated allelic counts and then aggregated the transcript-level data within gene based on the sign of the transcript-level allelic log fold change (LFC), referring to this as “LFC-based” aggregation. This is not a recommended aggregation strategy within *SEESAW*, but only used here for evaluation of error control.

In the *fishpond* Bioconductor package, we used a convenience function `makeTx2Tss` for generating TSS groups (with an optional parameter to group nearby TSS) and then used `importAllelicCounts` to import the estimated counts, abundances, and bootstrap replicates, producing a *SummarizedExperiment* object and leveraging the *tximeta* package [72, 73]. In the case that there was no read information to distinguish the two alleles, e.g., identical sequence, or no reads covering any sequence differences, *Salmon* splits the total counts equally among the two alleles, so the estimated allelic fold change was equal to 1. Such features were filtered out of the dataset before differential testing, as demonstrated in the software vignette. Prior to differential testing, features that did not have a minimum count of 10 for three or more samples were filtered out.

Swish [32] was used here to detect AI across biological replicates or conditions while taking into account inferential uncertainty. *Swish* is a nonparametric method originally designed for isoform-level differential expression that extends the gene-level method *SAMseq* [74]. We tested for the existence of AI (allelic fold change not equal to 1) for a given feature across all samples by specifying a paired analysis with `x="allele"`, `pair="sample"`, which was referred to as “Global AI testing.” Reviewing the paired-sample method developed in Zhu et al. [32], *Swish* for global AI used the Wilcoxon signed-rank test [75], where the paired data in this context are the allelic counts for each sample and each bootstrap. The signed-rank test statistic was averaged over bootstraps (in this case, 30 bootstraps), and the allelic labels were swapped to generate a permutation null distribution. The framework of averaging nonparametric test statistics over replicate datasets, followed by permutation-based *q*-value computation, was derived from *SAMseq* [74].

We further extended *Swish* to test for changes in AI along a continuous covariate. We tested for non-zero correlation between the log allelic fold change within paired samples and a continuous covariate by specifying, e.g., `x="allele"`, `pair="sample"`, `cov="day"`, which was referred to as “Dynamic AI testing.” Either Pearson or Spearman correlations can be computed between the pairwise log fold changes and the

continuous covariate, and this statistic was then averaged over bootstrap samples. To stabilize the log fold change, a pseudocount was added to the numerator and denominator of the allelic fold change, with a default value of 5, as has been used previously for bulk RNA-seq [32]. The continuous covariate was then permuted and correlations recalculated to generate a null distribution over all permutations and all features, followed by *q*-value computation with the *qvalue* package [46] to obtain the significance of the relationship between the allelic fold change with the continuous covariate. Testing for changes in AI across a categorical covariate was already available in *Swish* as an “interaction test” and is demonstrated in the software vignette on allelic analysis. Additional tests can also be performed, following the *Swish* framework of averaging test statistics over bootstrap replicates and performing permutations to compute the *q*-values. An example of testing for differences in correlation of AI with a continuous covariate across two groups using a general regression framework is demonstrated in the software vignette for allelic analysis, under the heading, “More complex designs”.

For statistical methods designed for comparing gene expression across samples, it is common practice that measures of gene expression are scaled or an offset is included in the model to account for a well-known technical bias: differences in sequencing depth across samples affect the observed or estimated counts and if left unadjusted the across-condition estimates would be biased. Since *SEESAW* focuses on testing differences in expression between the two alleles within the same sample, the sequencing depth bias affected both allelic counts equally, and scaling/offsets were not needed. Thus, this scaling step should not be performed.

A number of plotting functions in *fishpond* were used to facilitate visualization of allelic expression changes across samples, isoforms, and covariates. `plotInfReps` was used to visualize allelic expression estimates and inferential uncertainty across samples and conditions/time points [32]. `plotAllelicGene` was used to visualize isoform- or TSS-group-level allelic expression data along with a diagram of a given gene model, using the *Gviz* package [76]. `plotAllelicHeatmap` was used to visualize allelic expression across isoforms or TSS groups and samples, leveraging the *pheatmap* package [77]. In *SEESAW*, transcript ranges were represented using `GRanges` [78] objects generated from `TxDb` or `EnsDb` databases [79] and attached as `rowRanges` to the main dataset object with estimated counts, abundance and bootstrap counts, facilitating downstream plotting and data exploration.

Simulation

To assess the performance of different methods in recovering gene-level and isoform-level AI, we simulated RNA-seq reads from a diploid transcriptome derived from the *Drosophila melanogaster* reference transcriptome, restricted to chromosomes 2, 3, 4, and X, simulating RNA transcripts from female flies. The simulation contained a total of 10 samples, with an average sequencing depth of 50 million paired-end reads per sample. The maternal reference transcripts included the protein-coding and non-coding RNA from Ensembl [80] (release 100), and the paternal reference transcripts were created from maternal transcripts by adding single nucleotide variants. To create paternal alleles, we randomly selected 5 exons from each gene, and the mid-position nucleotides of the selected exons were substituted with their complement nucleotides. Genes

overlapping simple tandem repeats of size 50 bp or larger were excluded from the simulation, leaving 14,821 genes.

While the majority of genes (93.3%) were simulated to not have AI, two types of AI were simulated: “concordant AI,” where all isoforms had concordant allelic fold change, and “discordant AI,” where there were discordant allelic fold changes among the isoforms of the gene. In the discordant AI case, the RNA abundance was balanced across the two alleles when summed across all the isoforms of the gene. We randomly selected 1000 genes using the following criterion: (1) the number of annotated isoforms in the selected gene was between three and six, (2) the gene had at least two isoforms sharing the same transcription start site (TSS), and (3) the gene had at least two distinctive TSS. Half of the selected genes (500) were simulated to have concordant AI and the other half (500) were simulated to have discordant AI, and the remaining 13,821 genes were simulated to have allelic balance.

The abundance of the maternal allele was set to a constant value, and the paternal allelic abundance was altered to generate AI. For genes with concordant AI, the paternal allele was either 25% upregulated or downregulated, chosen at random per gene. Within each gene with discordant isoform-level AI, one TSS was randomly chosen and isoforms sharing the selected TSS had abundance increased on the paternal allele. Abundance was increased such that the upregulation fold change was equal to $1 + \frac{1}{2n}$, where n is the number of isoforms with the selected TSS. The other isoforms of the gene had paternal abundance decreased at an equal rate such that the gene-level abundance was kept constant. Expected count values were then generated from the alleles of all transcripts by multiplying abundance by the transcript length and scaling up to the desired library size. Reads were generated using *polyester* [81], with the following settings: fragments with a mean size of 400 bp, paired-end reads of 150 bp, and Negative Binomial dispersion parameter `size=100`, such that the simulation contained across-sample biological variation on the allelic counts and the allelic fold change. Paired-end reads were shuffled so that the reads were listed in a random order. Read generation and all subsequent analysis steps for *SEESAW* and other methods were automated using a *Snakemake* workflow [82] available at the `ase-sim` repository [83].

Additional calculations were performed to evaluate the number of inferential replicates needed to capture inferential uncertainty. The calculation was performed on three levels of aggregation: isoform level, TSS level, and gene level, and across $n = 10, 20, 30, 50$ and 100 , the number of inferential replicates. Within each aggregation level, features were divided into three expression bins (low/medium/high tertiles) based on true, simulated counts. 95% bootstrap intervals were computed using the bootstrap mean and plus or minus the bootstrap standard deviation multiplied by the 97.5% quantile of the normal distribution. Mean and standard deviation of the rate of interval coverage were calculated across samples, within each aggregation level and expression bin.

A beta-binomial generalized linear model was fit to the same *Salmon* estimated allelic counts as provided to *Swish*, in order to compare the performance of *Swish* to a method that does not make use of information about the uncertainty in quantification. Maximum likelihood estimates of over-dispersion and generalized linear model coefficients were computed using *apeglm* [84] without coefficient shrinkage, and p -values computed

from the Wald statistics for the allelic log odds ratio. p -values were corrected for multiple testing using the Benjamini-Hochberg method [53].

We applied the *SEESAW* pipeline as well as existing methods *mmseq* and *WASP* to the simulated data. As our interest in this work was in identifying isoforms affected by *cis*-genetic regulation, we focused our comparisons on methods that aim to aggregate allelic information along the entire extent of the transcribed region, whereas other existing methods have focused on allelic differences in internal splice events (splicing quantitative trait loci, or sQTL); we chose a method that attempts to resolve AI at the isoform level (*mmseq*), as well as a method that resolves bias from genomic multi-mapping reads (*WASP*) and is primarily focused on gene-level AI. *SEESAW* and other methods were provided with the complete exonic sequence of the two alleles, and the gene annotation, either as FASTA (for *Salmon* and *mmseq*) or VCF files with known phasing (for *WASP*). We also utilized the ground truth of simulation to obtain an optimal isoform-grouping strategy, called “oracle.” For “oracle” grouping, upregulated or downregulated isoforms within a gene were grouped in cases of discordant AI. Otherwise, all isoforms within a gene were grouped together.

The following steps were used to apply *mmseq* (version 1.0.10a) and its differential testing step *mmdiff* to the simulated data. *Bowtie* (version 1.3.1) [27] was used to index the diploid reference transcripts and to align the reads. During the alignment, only the alignments that fell into the best stratum were reported if the alignments fell into multiple stratum using `--best --strata`. If more than 100 reportable alignments existed for a particular read, then all alignments were suppressed using the `-m` option. After obtaining *mmseq* expression estimates at gene and isoform level, we manually separated the estimates for the maternal and paternal alleles and subsequently used *mmdiff* to test for differential expression between the two alleles, using the flags `-de 10 10 <maternal files> <paternal files>`. Posterior probability of equal expression was used to threshold and define significant sets of transcripts or genes.

We ran *WASP* on the simulated data according to its recommended usage. First, HISAT2 (version 2.2.1) [49] was used to align reads to the bdgp6 reference genome, downloaded from the HISAT2 website. An h5 database was created from the simulation VCF file containing the location of the exonic SNPs and the known phasing information. HISAT2 was used to re-align the reads with flipped nucleotides, and genomic multi-mapping reads that would otherwise bias allelic ratios were filtered. Read counts and heterozygous probabilities were adjusted using *WASP* scripts. Finally the Combined Haplotype Test (CHT) [17] was applied with recommended defaults `--min_counts 50` and `--min_as_counts 10` to generate a p -value per gene for AI across samples. Multiple testing was controlled via the *locfdr* package [52], applied to z -scores derived from *WASP* p -value output.

To visually compare the AI simulation results across methods, we used the iCOBRA [85] Bioconductor package. We assessed the performance according to the true and reported allelic status of the transcripts (balanced or imbalanced), where reported significance of AI of a gene or TSS group was propagated to its isoforms. As the simulation consisted only of genes in which all isoforms or no isoforms exhibited true AI, this approach to compare methods at the transcript level should not unfairly impact the performance of the aggregated (gene- or TSS-level) AI tests.

Osteoblast differentiation time course

We applied *SEESAW* to an RNA-seq dataset of primary mouse osteoblasts undergoing differentiation, from F1 C57BL/6J x CAST/EiJ (B6xCAST) mice, to assess both global and dynamic AI. To obtain the B6xCAST F1 animals, pOBCol3.6GFPtpz transgenic dams [86] were bred with CAST/EiJ (JAX Stock number 000928) sires. Pre-osteoblast-like cells were collected from the parietal bones of neonatal mice as previously described [86]. The B6xCAST cultures were conducted at the same time, using the same lots of Fetal Bovine Serum to avoid batch effects due to culture conditions. The pOBCol3.6GFPtpz transgenic dams possess the pOBCol3.6GFPtpz allele on a C57BL/6J background (> 10 generations of backcrossing). For the culture conditions, collection of RNA and transcriptomic sequencing was conducted as previously described [87].

In the differentiation experiment, which has been described previously [88, 89], pre-osteoblast-like cells were extracted from neonatal calvaria, and cells were FACS sorted based on expression of CFP, as driven by the osteoblast Col3.6 promoter. Differentiation was induced with an osteoblast differentiation cocktail in sorted cells and RNA was collected every 2 days from day 2 to day 18 post differentiation (nine time points). Three technical replicates per time point were combined and quantified together as one biological replicate, after quality checking with *FASTQC* and *MultiQC* [90]. Expression data for osteoblasts from C57BL/6J mice of the same experiment are publicly available on the Gene Expression Omnibus at accession GSE54461 [91].

Reference transcripts were generated via *g2gtools* using a reference genome, strain-specific VCF files, and the reference gene annotation. The GRCh38 primary assembly for *Mus musculus* was downloaded from Ensembl (release 102) [80], and strain-specific VCF files CAST_EiJ.mgp.v5.snps.dbSNP142.vcf (SNP) and CAST_EiJ.mgp.v5.indels.dbSNP142.normed.vcf (indel) for mm10 were downloaded from the Mouse Genomes Project [92]. Reference transcripts (Mus_musculus.GRCh38.102.gtf) were downloaded from Ensembl (release 102) [80] and subsequently were patched and transformed for the CAST/EiJ strain using *g2gtools* (version 0.2.7). All code including a *Snakemake* workflow [82] for generation of the diploid transcriptome is provided in the `diploid_txomes` directory of the `osteoblast-quant` repository [93].

Salmon was used to quantify the RNA-seq reads against the custom diploid transcriptome with 30 bootstrap replicates, and these data were imported into Bioconductor and analyzed with *Swish* as described in the “*SEESAW*” section above. *Swish* with global AI test was performed on isoform level, TSS-group level, and gene level, and results were compared at various levels of resolution. In addition, we used the newly developed feature in *Swish* to test for dynamic AI: we tested the correlation between the log fold change comparing across alleles within a sample and the day of differentiation, using the Pearson correlation.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-023-03003-x>.

Additional file 1: Supplementary figures and tables. Additional file 1 includes 14 figures and 2 tables. The supplementary figures describe the pipeline of *SEESAW*, details on both simulated and osteoblast datasets, as well as

benchmark results with existing methods. The tables display bootstrap coverage and the list of genes with discordant AI in the mouse F1 time course.

Additional file 2. Review history.

Acknowledgements

We thank Dr. Matthew Hibbs for his contributions to the generation of the original dataset and Dr. Anuj Srivastava and Sai Lek for their assistance with data deposition. We thank the following individuals for feedback and suggestions during the development of the method and manuscript: Dr. HIRAK SARKAR, Dr. Andrew Parker Morgan, Dr. Fernando Pardo-Manuel de Villena, Dr. Terry Furey, Dr. Nil Aygun, Dr. Jason Stein, Dr. Kevin Currin, Dr. Karen Mohlke, Dr. Ina Hoeschele, Dr. Ernest Turro, Mr. Aaron Ho, Dr. Graham McVicker, and Dr. Bryce van de Geijn.

Peer review information

Andrew Cosgrove was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 2.

Authors' contributions

EYW, NPS, KC, RP, and MIL conceived of the quantification and statistical analysis methods. EYW and MIL wrote the new code and documentation for *fishpond*. EYW, NPS, RP, and MIL developed the simulation framework. EYW and MIL analyzed the performance of methods in simulation. KC, MV, GAC, and CLA-B provided data and expertise for analysis of the mouse F1 cross dataset. EYW analyzed the mouse F1 dataset and generated results tables and diagrams. EYW, NPS, KC, MZ, and MIL made major contributions to the writing of the manuscript. All authors read and approved the final manuscript.

Funding

- NSF CCF-1750472 and CNS-1763680: Rob Patro, Noor P Singh, Mohsen Zakeri.
- NIH R01 HG009937: Rob Patro, Noor P Singh, Mohsen Zakeri, Michael I Love.
- NIH T32 CA106209: Euphy Wu.
- NIH R01 GM070683: Gary A Churchill, Matthew Vincent.

Availability of data and materials

SEESAW is available in the *fishpond* [33] Bioconductor package, distributed with a GPL-2 open source license. Version 2.4.1 of *fishpond* was used in this manuscript, available from Bioconductor (release 3.16) Bioconductor [94]. The allelic analysis software vignette can be viewed at the package website. Code for quantifying the osteoblast data (using *Salmon* version 1.5.2) is provided in the `osteoblast-quant` repository [93] and code for AI testing and compiling results across different levels of resolution is provided at the `osteoblast-test` repository [95]. Code for generating the simulated data is provided at the `ase-sim` repository [83] and code for running *Swish* and compiling the results from other methods is provided at the `swish-ase-assessment` repository [96].

The B6xC57BL mouse osteoblast RNA-seq experiments are available at the SRA under project accession SRP036025 [97]. For the mouse osteoblast dataset, R data objects containing the total and allelic counts at gene and isoform level, diploid transcriptome sequences, *Salmon* quantification directories for all samples, global and dynamic AI test results at all three levels of resolution, and the discordant global AI gene list are provided on Zenodo [98]. For the *Drosophila melanogaster* simulated samples, R data objects containing the simulated abundances and status of each transcript, simulated transcriptome sequences, *Salmon* quantification directories for all samples, SummarizedExperiment objects at various levels of resolution, and results tables for each method are provided on Zenodo [99].

Declarations

Ethics approval and consent to participate

All animals were raised at The Jackson Laboratory. Mouse husbandry and all experiments were performed in accordance with The Jackson Laboratory Institutional Animal Care and Use Committee-approved protocols (Animal Care and Use Committee, ACUC).

Consent for publication

Not applicable.

Competing interests

R.P. is a co-founder of Ocean Genomics Inc.

Received: 1 September 2022 Accepted: 29 June 2023

Published online: 12 July 2023

References

1. Loos RJF. 15 years of genome-wide association studies and no signs of slowing down. *Nat Commun.* 2020;11(1):5900. <https://doi.org/10.1038/s41467-020-19653-5>.

2. Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in cis and trans gene regulation. *Nature*. 2004;430(6995):85–8. <https://doi.org/10.1038/nature02698>.
3. Fogarty MP, Xiao R, Prokunina-Olsson L, Scott LJ, Mohlke KL. Allelic expression imbalance at high-density lipoprotein cholesterol locus MMAB-MVK. *Hum Mol Genet*. 2010;19(10):1921–9. <https://doi.org/10.1093/hmg/ddq067>.
4. Xiao R, Scott LJ. Detection of cis-acting regulatory SNPs using allelic expression data. *Genet Epidemiol*. 2011;35(6):515–25. <https://doi.org/10.1002/gepi.20601>.
5. Sun W. A statistical framework for eQTL mapping using RNA-seq data. *Biometrics*. 2012;68(1):1–11. <https://doi.org/10.1111/j.1541-0420.2011.01654.x>.
6. Hasin-Brumshtein Y, Hormozdiari F, Martin L, van Nas A, Eskin E, Lusis AJ, et al. Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics*. 2014;15(1):471. <https://doi.org/10.1186/1471-2164-15-471>.
7. Crowley JJ, Zhabotynsky V, Sun W, Huang S, Pakatci IK, Kim Y, et al. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat Genet*. 2015;47(4):353–60. <https://doi.org/10.1038/ng.3222>.
8. Castel SE, Mohammadi P, Chung WK, Shen Y, Lappalainen T. Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nat Commun*. 2016;7:12817. <https://doi.org/10.1038/ncomms12817>.
9. Moyerbrailean GA, Richards AL, Kurtz D, Kalita CA, Davis GO, Harvey CT, et al. High-throughput allele-specific expression across 250 environmental conditions. *Genome Res*. 2016;26(12):1627–38. <https://doi.org/10.1101/gr.209759.116>.
10. Knowles DA, Davis JR, Edgington H, Raj A, Favé MJ, Zhu X, et al. Allele-specific expression reveals interactions between genetic variation and environment. *Nat Methods*. 2017;14(7):699–702. <https://doi.org/10.1038/nmeth.4298>.
11. Combs PA, Fraser HB. Spatially varying cis-regulatory divergence in *Drosophila* embryos elucidates cis-regulatory logic. *PLoS Genet*. 2018;14(11):e1007631. <https://doi.org/10.1371/journal.pgen.1007631>.
12. Castel SE, Aguet F, Mohammadi P, Consortium G, Ardlie KG, Lappalainen T. A vast resource of allelic expression data spanning human tissues. *Genome Biol*. 2020;21(1):234. <https://doi.org/10.1186/s13059-020-02122-z>.
13. Zhabotynsky V, Huang L, Little P, Hu YJ, Pardo-Manuel de Villena F, Zou F, et al. eQTL mapping using allele-specific count data is computationally feasible, powerful, and provides individual-specific estimates of genetic effects. *PLoS Genet*. 2022;18(3):e1010076. <https://doi.org/10.1371/journal.pgen.1010076>.
14. Sibbesen JA, Eizenga JM, Novak AM, Siren J, Chang X, Garrison E, et al. Haplotype-aware pantranscriptome analyses using spliced pangenome graphs. *BioRxiv*. 2021. <https://doi.org/10.1101/2021.03.26.437240>.
15. Glinos DA, Garborcauskas G, Hoffman P, Ehsan N, Jiang L, Gokden A, et al. Transcriptome variation in human tissues revealed by long-read sequencing. *Nature*. 2022;608(7922):353–9. <https://doi.org/10.1038/s41586-022-05035-y>.
16. Raghupathy N, Choi K, Vincent MJ, Beane GL, Sheppard KS, Munger SC, et al. Hierarchical analysis of RNA-seq reads improves the accuracy of allele-specific expression. *Bioinformatics*. 2018;34(13):2177–84. <https://doi.org/10.1093/bioinformatics/bty078>.
17. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods*. 2015;12(11):1061–3. <https://doi.org/10.1038/nmeth.3582>.
18. Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. *Genome Biol*. 2015;16(1):195. <https://doi.org/10.1186/s13059-015-0762-6>.
19. Fan J, Hu J, Xue C, Zhang H, Susztak K, Reilly MP, et al. ASEP: Gene-based detection of allele-specific expression across individuals in a population by RNA sequencing. *PLoS Genet*. 2020;16(5):e1008786. <https://doi.org/10.1371/journal.pgen.1008786>.
20. Demirdjian L, Xu Y, Bahrami-Samani E, Pan Y, Stein S, Xie Z, et al. Detecting allele-specific alternative splicing from population-scale RNA-Seq data. *Am J Hum Genet*. 2020;107(3):461–72. <https://doi.org/10.1016/j.ajhg.2020.07.005>.
21. Zhao K, Lu ZX, Park JW, Zhou Q, Xing Y. GLiMMPs: robust statistical model for regulatory variation of alternative splicing using RNA-seq data. *Genome Biol*. 2013;14(7):R74. <https://doi.org/10.1186/gb-2013-14-7-r74>.
22. Deonovic B, Wang Y, Weirather J, Wang XJ, Au KF. IDP-ASE: haplotyping and quantifying allele-specific expression at the gene and gene isoform level by hybrid sequencing. *Nucleic Acids Res*. 2016;45(5):e32–e32. <https://doi.org/10.1093/nar/gkw1076>.
23. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34(5):525–7. <https://doi.org/10.1038/nbt.3519>.
24. Turro E, Su SY, Gonçalves Â, Coin LJM, Richardson S, Lewin A. Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biol*. 2011;12(2):R13. <https://doi.org/10.1186/gb-2011-12-2-r13>.
25. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323. <https://doi.org/10.1186/1471-2105-12-323>.
26. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417–9. <https://doi.org/10.1038/nmeth.4197>.
27. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
28. Turro E, Astle WJ, Tavaré S. Flexible analysis of RNA-seq data using mixed effects models. *Bioinformatics*. 2014;30(2):180–8. <https://doi.org/10.1093/bioinformatics/btt624>.
29. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*. 2015;4:1521. <https://doi.org/10.12688/f1000research.7563.2>.
30. Trincado JL, Entizne JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, et al. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol*. 2018;19(1):40. <https://doi.org/10.1186/s13059-018-1417-1>.
31. Alasoo K, Rodrigues J, Danesh J, Freitag DF, Paul DS, Gaffney DJ. Genetic effects on promoter usage are highly context-specific and contribute to complex traits. *eLife*. 2019;8. <https://doi.org/10.7554/eLife.41673>.
32. Zhu A, Srivastava A, Ibrahim JG, Patro R, Love MI. Nonparametric expression analysis using inferential replicate counts. *Nucleic Acids Res*. 2019;47(18):e105. <https://doi.org/10.1093/nar/gkz622>.
33. Zhu A. The 'fishpond' Bioconductor package. *Bioconductor*. 2019. <https://doi.org/10.18129/B9.bioc.fishpond>.

34. Perrin HJ, Currin KW, Vadlamudi S, Pandey GK, Ng KK, Wabitsch M, et al. Chromatin accessibility and gene expression during adipocyte differentiation identify context-dependent effects at cardiometabolic GWAS loci. *PLoS Genet.* 2021;17(10):e1009865. <https://doi.org/10.1371/journal.pgen.1009865>.
35. Strober BJ, Elorbany R, Rhodes K, Krishnan N, Tayeb K, Battle A, et al. Dynamic genetic regulation of gene expression during cellular differentiation. *Science.* 2019;364(6447):1287–90. <https://doi.org/10.1126/science.aaw0040>.
36. Elorbany R, Popp JM, Rhodes K, Strober BJ, Barr K, Qi G, et al. Single-cell sequencing reveals lineage-specific dynamic genetic regulation of gene expression during human cardiomyocyte differentiation. *PLoS Genet.* 2022;18(1):e1009666. <https://doi.org/10.1371/journal.pgen.1009666>.
37. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nature Genetics.* 2018;50(3):424–31. <https://doi.org/10.1038/s41588-018-0046-7>.
38. Knowles DA, Burrows CK, Blischak JD, Patterson KM, Serie DJ, Norton N, et al. Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. *eLife.* 2018;7. <https://doi.org/10.7554/eLife.33480>.
39. Ward MC, Banovich NE, Sarkar A, Stephens M, Gilad Y. Dynamic effects of genetic variation on gene expression revealed following hypoxic stress in cardiomyocytes. *eLife.* 2021;10. <https://doi.org/10.7554/eLife.57345>.
40. Findley AS, Monziani A, Richards AL, Rhodes K, Ward MC, Kalita CA, et al. Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. *eLife.* 2021;10. <https://doi.org/10.7554/eLife.67077>.
41. Munger SC, Raghupathy N, Choi K, Simons AK, Gatti DM, Hinerfeld DA, et al. RNA-Seq alignment to individualized genomes improves transcript abundance estimates in multiparent populations. *Genetics.* 2014;198(1):59–73. <https://doi.org/10.1534/genetics.114.165886>.
42. Rozowsky J, Abyzov A, Wang J, Alves P, Raha D, Harmanci A, et al. AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol Syst Biol.* 2011;7:522. <https://doi.org/10.1038/msb.2011.54>.
43. Nariai N, Kojima K, Mimori T, Kawai Y, Nagasaki M. A Bayesian approach for estimating allele-specific expression from RNA-Seq data with diploid genomes. *BMC Genomics.* 2016;17(Suppl 1):2. <https://doi.org/10.1186/s12864-015-2295-5>.
44. Miao Z, Alvarez M, Pajukanta P, Ko A. ASElux: an ultra-fast and accurate allelic reads counter. *Bioinformatics.* 2018;34(8):1313–20. <https://doi.org/10.1093/bioinformatics/btx762>.
45. Saukkonen A, Kilpinen H, Hodgkinson A. PAC: Highly accurate quantification of allelic gene expression for population and disease genetics. *BioRxiv.* 2021. <https://doi.org/10.1101/2021.07.13.452202>.
46. Storey JD. A direct approach to false discovery rates. *J R Stat Soc Ser B (Stat Methodol).* 2002;64(3):479–98. <https://doi.org/10.1111/1467-9868.00346>.
47. Soneson C, Robinson MD. iCOBRA: open, reproducible, standardized and live method benchmarking. *Nat Methods.* 2016;13(4):283. <https://doi.org/10.1038/nmeth.3805>.
48. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29(1):24–6. <https://doi.org/10.1038/nbt.1754>.
49. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
50. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. Upset: visualization of intersecting sets. *IEEE Trans Vis Comput Graph.* 2014;20(12):1983–92. <https://doi.org/10.1109/TVCG.2014.2346248>.
51. Gao LL, Bien J, Witten D. Selective inference for hierarchical clustering. *J Am Stat Assoc.* 2022;1–11. <https://www.tandfonline.com/doi/full/10.1080/01621459.2022.2116331?scroll=top&needAccess=true&role=tab>.
52. Efron B. Large-Scale Simultaneous Hypothesis Testing. *J Am Stat Assoc.* 2004;99(465):96–104. <https://doi.org/10.1198/016214504000000089>.
53. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol).* 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
54. Sarkar H, Srivastava A, Bravo HC, Love MI, Patro R. Terminus enables the discovery of data-driven, robust transcript groups from RNA-seq data. *Bioinformatics.* 2020;36(Suppl-1):i1102–10. <https://doi.org/10.1093/bioinformatics/btaa448>.
55. Reddy TE, Gertz J, Pauli F, Kucera KS, Varley KE, Newberry KM, et al. Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Res.* 2012;22(5):860–9. <https://doi.org/10.1101/gr.131201.111>.
56. Chen J, Rozowsky J, Galeev TR, Harmanci A, Kitchen R, Bedford J, et al. A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nat Commun.* 2016;7:11101. <https://doi.org/10.1038/ncomms11101>.
57. Zhang Q, Keles S. An empirical Bayes test for allelic-imbalance detection in ChIP-seq. *Biostatistics.* 2018;19(4):546–61. <https://doi.org/10.1093/biostatistics/kxx060>.
58. Zhang S, Zhang H, Zhou Y, Qiao M, Zhao S, Kozlova A, et al. Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science.* 2020;369(6503):561–5. <https://doi.org/10.1126/science.aay3983>.
59. Liang D, Elwell AL, Aygün N, Krupa O, Wolter JM, Kyere FA, et al. Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. *Nat Neurosci.* 2021;24(7):941–53. <https://doi.org/10.1038/s41593-021-00858-w>.
60. Orjuela S, Machlab D, Menigatti M, Marra G, Robinson MD. DAMEfinder: a method to detect differential allele-specific methylation. *Epigenetics Chromatin.* 2020;13(1):25. <https://doi.org/10.1186/s13072-020-00346-8>.
61. Tognon M, Bonnici V, Garrison E, Giugno R, Pinello L. GRAFIMO: Variant and haplotype aware motif scanning on pangenome graphs. *PLoS Comput Biol.* 2021;17(9):e1009444. <https://doi.org/10.1371/journal.pcbi.1009444>.
62. Flynn ED, Tsu AL, Kasela S, Kim-Hellmuth S, Aguet F, Ardlie KG, et al. Transcription factor regulation of eQTL activity across individuals and tissues. *PLoS Genet.* 2022;18(1):e1009719. <https://doi.org/10.1371/journal.pgen.1009719>.

63. Fan J, Wang X, Xiao R, Li M. Detecting cell-type-specific allelic expression imbalance by integrative analysis of bulk and single-cell RNA sequencing data. *PLoS Genet.* 2021;17(3):e1009080. <https://doi.org/10.1371/journal.pgen.1009080>.
64. Heinen T, Secchia S, Reddington JP, Zhao B, Furlong EEM, Stegle O. scDALL: modeling allelic heterogeneity in single cells reveals context-specific genetic regulation. *Genome Biol.* 2022;23(1):8. <https://doi.org/10.1186/s13059-021-02593-8>.
65. Wu W, Lovett JL, Shedden K, Strassmann BI, Vincenz C. Targeted RNA-seq improves efficiency, resolution, and accuracy of allele specific expression for human term placentas. *BioRxiv.* 2021. <https://doi.org/10.1101/2021.01.25.428155>.
66. Kumasaka N, Knights AJ, Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat Genet.* 2016;48(2):206–13. <https://doi.org/10.1038/ng.3467>.
67. Zou J, Hormozdiari F, Jew B, Castel SE, Lappalainen T, Ernst J, et al. Leveraging allelic imbalance to refine fine-mapping for eQTL studies. *PLoS Genet.* 2019;15(12):e1008481. <https://doi.org/10.1371/journal.pgen.1008481>.
68. Vigorito E, Lin WY, Starr C, Kirk PD, White SR, Wallace C. BaseQTL: a Bayesian method to detect eQTLs from RNA-seq data with or without genotypes. *BioRxiv.* 2020. <https://doi.org/10.1101/2020.07.16.203851>.
69. Liang Y, Aguet F, Barbeira AN, Ardlie K, Im HK. A scalable unified framework of total and allele-specific counts for cis-QTL, fine-mapping, and prediction. *Nat Commun.* 2021;12(1):1424. <https://doi.org/10.1038/s41467-021-21592-8>.
70. Ntranos V, Kamath GM, Zhang JM, Pachter L, Tse DN. Fast and accurate single-cell RNA-seq analysis by clustering of transcript-compatibility counts. *Genome Biol.* 2016;17(1):112. <https://doi.org/10.1186/s13059-016-0970-8>.
71. Cmero M, Davidson NM, Oshlack A. Using equivalence class counts for fast and accurate testing of differential transcript usage. *F1000Research.* 2019;8:265. <https://doi.org/10.12688/f1000research.18276.2>.
72. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods.* 2015;12(2):115–21. <https://doi.org/10.1038/nmeth.3252>.
73. Love MI, Soneson C, Hickey PF, Johnson LK, Pierce NT, Shepherd L, et al. Tximeta: Reference sequence checksums for provenance identification in RNA-seq. *PLoS Comput Biol.* 2020;16(2):e1007664. <https://doi.org/10.1371/journal.pcbi.1007664>.
74. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Stat Methods Med Res.* 2013;22(5):519–36. <https://doi.org/10.1177/0962280211428386>.
75. Wilcoxon F. Individual comparisons by ranking methods. *Biom Bull.* 1945;1(6):80. <https://doi.org/10.2307/3001968>.
76. Hahne F, Ivanek R. Visualizing genomic data using gviz and bioconductor. *Methods Mol Biol.* 2016;1418:335–51. https://doi.org/10.1007/978-1-4939-3578-9_16.
77. Kolde R. Pheatmap: pretty heatmaps. *R Packag Version.* 2012;1(2):747.
78. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. *PLoS Comput Biol.* 2013;9(8):e1003118. <https://doi.org/10.1371/journal.pcbi.1003118>.
79. Rainer J, Gatto L, Weichenberger CX. ensemblDb: an R package to create and use Ensembl-based annotation resources. *Bioinformatics.* 2019;35(17):3151–3. <https://doi.org/10.1093/bioinformatics/btz031>.
80. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amodè MR, et al. Ensembl 2021. *Nucleic Acids Res.* 2021;49(D1):D884–91. <https://doi.org/10.1093/nar/gkaa942>.
81. Frazee AC, Jaffe AE, Langmead B, Leek JT. Polyester: simulating RNA-seq datasets with differential transcript expression. *Bioinformatics.* 2015;31(17):2778–84. <https://doi.org/10.1093/bioinformatics/btv272>.
82. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics.* 2018;34(20):3600. <https://doi.org/10.1093/bioinformatics/bty350>.
83. Wu EY. Love MI ase-sim repository GitHub. 2022. <https://doi.org/10.5281/zenodo.8046187>.
84. Zitovsky JP, Love MI. Fast effect size shrinkage software for beta-binomial models of allelic imbalance. *F1000Research.* 2019;8:2024. <https://doi.org/10.12688/f1000research.20916.2>.
85. Soneson C, Matthes KL, Nowicka M, Law CW, Robinson MD. Isoform prefiltering improves performance of count-based methods for analysis of differential transcript usage. *Genome Biol.* 2016;17:12. <https://doi.org/10.1186/s13059-015-0862-3>.
86. Kalajzic I, Kalajzic Z, Kaliterna M, Gronowicz G, Clark SH, Lichtler AC, et al. Use of type I collagen green fluorescent protein transgenes to identify subpopulations of cells at different stages of the osteoblast lineage. *J Bone Miner Res.* 2002;17(1):15–25.
87. Zheng HF, Forgetta V, Hsu YH, Estrada K, Rosello-Diez A, Leo PJ, et al. Whole-genome sequencing identifies EN1 as a determinant of bone density and fracture. *Nature.* 2015;526(7571):112–7.
88. Kemp JP, Medina-Gomez C, Estrada K, St Pourcain B, Heppè DHM, Warrington NM, et al. Phenotypic dissection of bone mineral density reveals skeletal site specificity and facilitates the identification of novel loci in the genetic regulation of bone mass attainment. *PLoS Genet.* 2014;10(6):e1004423. <https://doi.org/10.1371/journal.pgen.1004423>.
89. Kemp JP, Morris JA, Medina-Gomez C, Forgetta V, Warrington NM, Youtlen SE, et al. Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nat Genet.* 2017;49(10):1468–75. <https://doi.org/10.1038/ng.3949>.
90. Ewels P, Magnusson M, Lundin S, Källér M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016;32(19):3047–8. <https://doi.org/10.1093/bioinformatics/btw354>.
91. Ackert-Bicknell C. Temporal gene expression across osteoblastogenesis - GSE54461. GEO. 2014. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54461>.
92. Keane T, Goodstadt L, Danecek P, White M, Wong K, Yalcin B, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature.* 2011;477(7364):289–94. <https://doi.org/10.1038/nature10413>.
93. Wu EY. Love MI osteoblast-quant repository GitHub. 2022. <https://doi.org/10.5281/zenodo.8046205>.
94. Bioconductor. Release 3.16. Bioconductor website. https://bioconductor.org/packages/3.16/BiocViews.html#___Software. Accessed 07 July, 2023.
95. Wu EY. Love MI osteoblast-test repository GitHub. 2022. <https://doi.org/10.5281/zenodo.8057963>.
96. Wu EY. Love MI swish-ase-assessment repository GitHub. 2022. <https://doi.org/10.5281/zenodo.8046199>.
97. Ackert-Bicknell C. B6xCAST temporal gene expression across osteoblastogenesis - SRA accession SRP036025. Sequence Read Archive. 2022. <https://trace.ncbi.nlm.nih.gov/Traces/index.html?view=study&acc=SRP036025>.

98. Wu EY, Love MI. SEESAW quantification data for temporal gene expression across osteoblastogenesis (B6xCAST). Zenodo. 2022. <https://doi.org/10.5281/zenodo.6963809>.
99. Wu EY, Love MI. SEESAW quantification data for simulated *Drosophila melanogaster* F1 cross. Zenodo. 2022. <https://doi.org/10.5281/zenodo.6967130>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

