Computers in Biology and Medicine 196 (2025) 110890



Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed



Allele-specific expression analysis: pipelines, applications, challenges, and unmet needs

Stefania Mattevi [a], Francesco Mazzarotto [a,b,1], Paolo Martini [a,*,1]

[a] Department of Molecular and Translational Medicine, University of Brescia, Brescia, Italy
[b] National Heart and Lung Institute, Imperial College London, United Kingdom

A B S T R A C T

In diploid organisms, genes typically exhibit balanced expression of maternal and paternal alleles. However, exceptions exist, such as autosomal genes with allele-specific expression, where genetic and epigenetic variations can lead to the exclusive or preferential expression of a particular allele. In this context, allele-specific expression analysis serves as a powerful tool for understanding gene regulation, with significant functional and clinical implications.

Despite their increasing importance, current analysis pipelines face notable limitations including a lack of end-to-end solutions, restricted options for multi-omics integration, and insufficient support for single-cell sequencing technologies.

This review critically assesses 26 cutting-edge pipelines for allele-specific expression analysis, focusing on their input requirements, capabilities, and applications in the field. Pipelines are categorized based on their ability to handle various data types, support haplotype phasing, employ statistical approaches, and provide graphical outputs. Most pipelines fail to automate preprocessing, integrate multi-omic data, and support high-throughput single-cell sequencing. Future advancements should prioritize the development of automated multi-omic workflows, implementing visualization options, and enhancing compatibility with single-cell technologies. By addressing these gaps, next-generation allele-specific expression pipelines will offer insights into the mechanisms of allele-specific expression regulation, thereby advancing our understanding of its biological and clinical significance.

1. Introduction

Diploid organisms carry two highly similar copies of the same gene in almost all cells, with transcription of both alleles generally occurring. However, there are some exceptions. In addition to imprinted genes that are constitutively expressed from either the maternal or paternal allele, the entire X chromosome is mainly expressed as a single copy within each cell in females. This phenomenon is termed allele-specific expression (ASE) and encompasses a spectrum ranging from the imbalanced expression of the two alleles to the exclusive expression of one allele. The ability to detect such events generally relies on estimating the expression ratio between the two alleles.

The widespread adoption of increasingly efficient technologies for genomic analysis has facilitated the routine production of matched (epi) genetic and expression data, thus paving the way for new ASE analysis methods. These advances have contributed to the realization that biallelic expression is less pervasive than previously believed [1–3], highlighting the importance of ASE analysis in understanding how gene expression differs between alleles within individuals and across populations.

Allelic expression differences are often explained by genetic variants, such as SNV, copy numbers, and structural variants acting in cis or trans, as well as epigenetic factors. However, up to 88 % of autosomal genes undergo random monoallelic expression (RME, i.e., casual and selective expression of a single allele) in pre-gastrulation mouse embryos [4], highlighting the importance of ASE in early development. Furthermore, the advent of single-cell technologies such as single-cell RNA sequencing (scRNA-seq) has opened the possibility of investigating ASE at a more granular level, thereby providing new insights into cell-specific allelic expression patterns[5–8].

For these reasons, ASE analysis has far-reaching potential and can help shed light on largely underexplored phenomena, with important biological and clinical implications (Fig. 1). Cancer research is an example, as ASE analysis is used to identify driver mutations, oncogene activation, and tumor suppressor inactivation, which are paramount in determining tumor heterogeneity and clonal evolution[9–11]. In cancer and Mendelian diseases, ASE can also explain incomplete penetrance and variable expressivity, revealing situations in which loci harboring pathogenic variants are subjected to imbalanced expression[12–14]. An example is the application of genome-wide association study

* Corresponding author.
E-mail address: paolo.martini@unibs.it (P. Martini).
[1] These authors should be regarded as Joint Last Authors.

https://doi.org/10.1016/j.compbiomed.2025.110890
Received 5 February 2025; Received in revised form 29 July 2025; Accepted 31 July 2025
Available online 3 August 2025
0010-4825/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

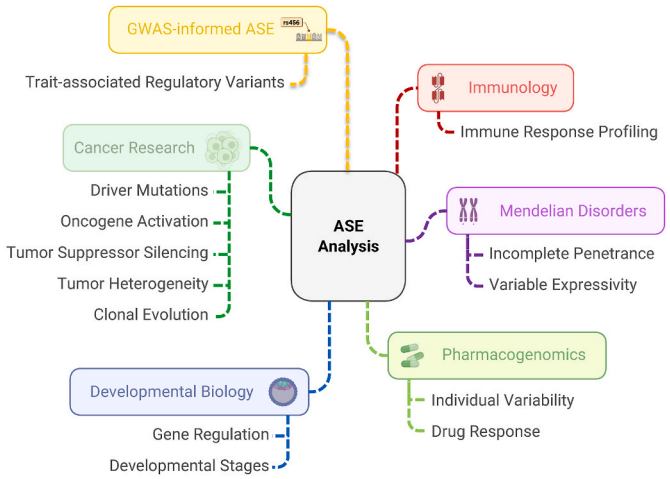


Fig. 1. ASE Analysis applications and domains. ASE analysis can help reveal new insights into underexplored phenomena with important biological and clinical implications. The image was partially created using biorender.com.

(GWAS)-informed ASE, in which ASE is used to investigate allelic imbalances potentially determined by numerous trait-associated variants located in regulatory regions[14–16].

In addition, ASE is relevant in developmental biology, immunology, and pharmacogenomics, and can be instrumental in understanding gene regulation during development, immune response, and individual responses to drugs[6,17–22].

In practice, ASE analysis of a given gene can be performed if it harbors at least one heterozygous SNP. According to a recent study, a heterozygous SNP is present in approximately half of the transcripts of an individual [23]. Such variants can be identified from RNA-sequencing (RNA-seq) data, but this approach suffers from the inherent limitation of being blind to cases of "pure" monoallelic expression (Fig. 2), because heterozygous and homozygous SNPs are indistinguishable from RNA-seq within a monoallelically expressed gene. Consequently, although RNA-seq is informative for skewed expression of the maternal or paternal allele, it cannot aid in the detection of pure monoallelic expression. Therefore, the identification of genes subject to monoallelic expression requires the availability of DNA-sequencing (DNA-seq) data alongside RNA-seq. All SNPs identified through DNA-seq must be phased (i.e., unambiguously mapped to either the maternal or paternal allele) so that RNA-seq reads can be mapped to a specific parental allele, enabling ASE estimation.

Notably, DNA-seq and RNA-seq data are not the only data types that allow for the possibility of performing ASE analysis. Other genomic readouts, such as chromatin immunoprecipitation sequencing (ChIP-seq) and DNA methylation sequencing (Methyl-seq), can be used to identify heterozygous SNPs and understand how multiple mechanisms contribute to ASE[24–26].

The variety of *omics* that can be exploited to perform ASE analysis has fostered the development of several tools for ASE identification that support different combinations of data types [27].

Currently, a comprehensive review of available state-of-the-art pipelines for ASE analysis is lacking. In this study, we address this issue by describing several ASE tools and pipelines that are currently available and discussing their requirements, applications, strengths, and weaknesses.

In doing so, we focused on end-to-end automatic pipelines that are easily accessible and well documented.

We categorized pipelines according to different features, including the presence of a haplotype phasing step, possibility of performing multi-sample analysis, capability to integrate data pertaining to different *omics,* and the ability to analyze scRNA-seq data.

In conclusion, our aim is to provide a useful compendium for those approaching ASE analysis and intending to develop new ASE pipelines in the future.

2. Pipeline inclusion, comparison criteria and implementation

We exclusively evaluated pipelines that were considered accessible (i.e., sufficiently well-maintained, well-documented, and easy to install) and readily useable (i.e., implemented using common bioinformatics programming languages). ASE analysis pipelines requiring custom scripts and cumbersome tailoring of the data were not included in this review, as we did not consider them sufficiently user friendly. To

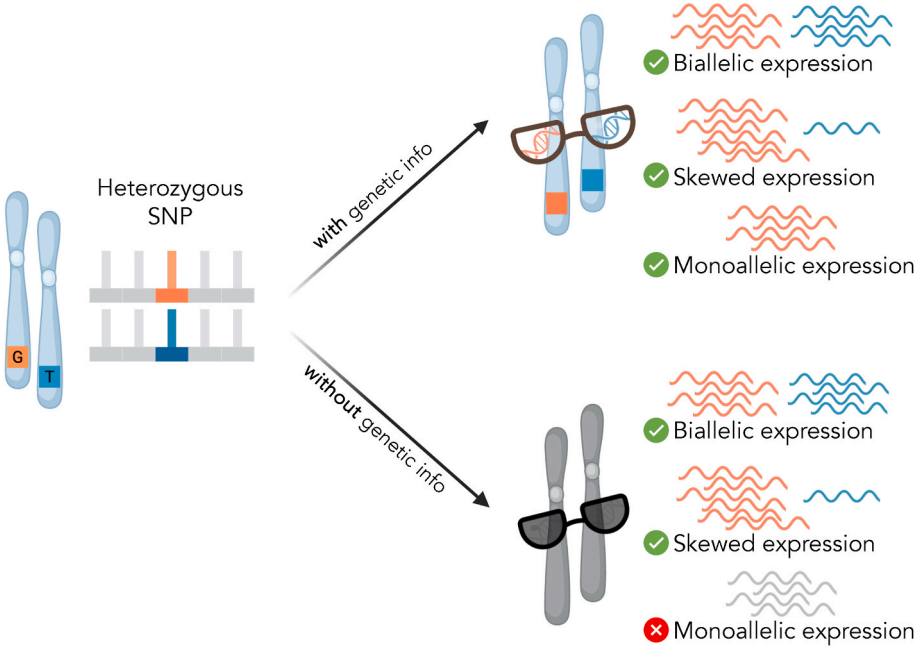


Fig. 2. ASE event detection capabilities with and without genetic information workflow. The possibility of identifying heterozygous sites (e.g., through DNA-seq or genotyping) is crucial for identifying monoallelic expression. Image partially created with biorender.com.

compare the pipelines, we evaluated the input requirements and preprocessing steps (as these sometimes need to be performed independently before running the analysis). In addition, we assessed how the results are presented, for example, by evaluating the possibility of producing graphical outputs, which is important for interpreting the results. To aid the reader and enhance clarity, a table of key terms was also included (Table 1).

We selected 26 pipelines for the evaluation. Of these, 15 are implemented in R (AllelicImbalance, BLMRM, Airpart, MBASED, SEESAW, stratAS, BSCET, ASEP, SCALE, DAESC, RASQUAL, spASE, GeneiASE, ANEVA-DOT, QuASAR, and EAGLE), six are Python packages (scDALI, BayesASE, scBASE, BYASE, AlleleSeq2, and LORALS), one relies on Python combined with other tools for visualization (BrewerIX), two are implemented in C/C++ (ASEQ, aScan), and one in Julia (IDP-ASE). We tested each pipeline using datasets provided by the authors and a bulk RNA-seq test dataset created from the Gene Expression Omnibus (GEO) dataset GSE110377 [28]. The main features of each tool are listed in Table 2. Moreover, to guide users in the selection the appropriate tool for their needs, we created a decision flowchart (Fig. 3).

At the time of writing, none of these pipelines (discussed based on the input data types below) can handle a full end-to-end automated analysis starting from the raw data (i.e., FASTQ or BAM files) to the results, including statistical testing and graphical visualization of results. A complete list of features characterizing these tools is provided in Supplementary Table 1.

3. Bulk RNA-seq analysis software

3.1. Tools requiring both DNA-seq and RNA-seq data

The optimal setting for bulk ASE analysis requires the availability of both genetic and transcriptomic data to enable the detection of pure monoallelic expression (Fig. 2).

When both data types are available, BayesASE, AllelicImbalance, aScan, ASEQ, BLMRM, and ANEVA-DOT can be used[29–35].

BayesASE is a flexible framework implemented in Python to estimate allelic imbalances that supports comparisons between different conditions [29]. The input requirements of BayesASE are FASTQ files (i.e., raw RNA-seq data containing details of unaligned reads), a reference genome, DNA-seq VCF files, and a design file for condition comparison. A guide for input preprocessing and data analysis is available on the GitHub page of the tool (see Supplementary Table 1).

The output produced by BayesASE consists of a text file with allelic ratios and the results of a Bayesian test comparing the analyzed conditions. While BayesASE requires limited preprocessing steps, in our opinion, its major drawbacks are the need to perform genotype phasing independently and the lack of a graphical representation of the results, which would make them easier to interpret.

AllelicImbalance is another tool for allelic imbalance analysis of RNA-seq data that enables multi-sample comparisons [35]. This is a well-maintained R Bioconductor package that provides instructions for input-data preparation, starting from an aligned RNA-seq BAM file and a set of relevant SNPs. Statistical testing applied downstream can be either a binomial or chi-squared test. AllelicImbalance also provides functions for visualizing the results to enhance comparisons among samples.

aScan [30] is implemented in C++ for ASE analysis at the individual level and uses aligned RNA-seq BAM files and a matched individual DNA-seq VCF file as inputs. RNA-seq alignment and DNA-seq variant calling must be performed independently prior to analysis (e.g., using STAR for RNA-seq and combining BWA with GATK for DNA-seq) [36–38]. Haplotype phasing is not required because aScan provides a module for haplotype reconstruction based on the ASE itself. Two additional inputs required by this tool are a reference GTF file and transcript-to-gene map file. The results includes a tab-separated file containing the gene (or transcript) name, number of heterozygous sites harbored by the gene (or transcript), number of reads supporting the high- and low-expression haplotypes, allele-specific p-value, and FDR. During our testing, we observed that the installation required bioinformatics proficiency and that the aScan's Docker image was deprecated. Furthermore, the tool does not provide progress status or updates during the analysis, and similar to BayesASE, it does not generate graphical output.

ASEQ was implemented in C++ [32] and includes three modes: "pileup" for the execution of optimized SNPs pileup analysis, "genotype" for SNPs genotyping, and "ASE", which can be executed using gene- or transcript-based modes. Each mode can be run easily from the command line using filters and options that can be tuned by the user. ASE mode requires a list of VCF files containing heterozygous SNPs, one file for each sample in the analysis, and RNA-seq BAM files as inputs. The final output consists of four file types: sample-specific files with pileup computation, a list of genes characterized by ASE and the number of samples used for the calculation, a file with ASE statistics, and a file containing the number of informative SNPs harbored by each analyzed gene.

ASEQ requires minimal data manipulation and preprocessing, and we found that it is easy to install and use. Overall, the results are very detailed, but there are no visual outputs for BayesASE and aScan.

Finally, we considered two additional tools implemented in R (BLMRM and ANEVA/ANEVA-DOT) that focused mainly on ASE tests.

BLMRM [33] enables the simultaneous ASE estimation of genes and variants using a Bayesian logistic mixed regression model. As input, it requires a table containing gene ID, SNP ID, and, for each sample replicate, maternal allelic counts and total counts (maternal + paternal), which can be automatically created by preprocessing the RNA-seq and DNA-seq data. The outputs consist of a table indicating genes exhibiting significant ASE, and a corresponding SNP-based table. However, no graphical output is available.

ANEVA and its extension, ANEVA-DOT [34], allow for the identification of genes subject to the effects of heterozygous genetic variants that significantly influence their expression. Using allelic expression data, ANEVA estimated V^G, which is the expected gene dosage variance due to inter-individual genetic differences within a population. ANEVA-DOT tests the null hypothesis that allelic imbalance in a sample is consistent with general population dosage variation (V^G). As input, ANEVA-DOT requires sample allelic counts (which can be obtained using phASER [39], as indicated in the documentation) and a vector with genetic standard deviation (square root of V^G scores). The outputs comprise a table with computed statistics and a scatter plot highlighting dosage-outlier genes.

3.2. Tools requiring RNA-seq only

As previously mentioned, when only RNA-seq data are available, ASE

Table 1
Terminology Table. This table summarizes the key terms necessary to understand this review.

Term	Explanation
ASE	Allele-Specific Expression (i.e. a gene's/SNP's/region's imbalanced expression toward one allele).
ASCA	Allele-Specific Chromatin Accessibility
ASB	Allele-Specific Binding
XCI	X chromosome imprinting
Haplotype phasing	The process through which, given the genotype of a sample, each allele is assigned to one of the two haplotypes.
Incomplete penetrance	Phenomena for which carriers of known disease-causing variants are phenotype negative
Variable expressivity	Phenotypic variability observed among individuals with the same genotype influencing the trait/disease of interest
BAM	Binary Alignment Map (file format)
ChIP-seq	Chromatin Immunoprecipitation Sequencing
ATAC-seq	Assay for Transposase-Accessible Chromatin Sequencing
VCF	Variant Call Format (file format)

Table 2
Allele-specific expression methods software. Summary of key information for each tool discussed in this review. Specifically, it details whether each method accepts bulk or single-cell sequencing data as input, whether autonomous preprocessing is required, and whether haplotype phasing information is needed.

Method	Data type(s)	Analysis	Phasing info required	Preprocessing required
MBASED	RNA-seq	1-sample or 2-sample (same individual) ASE	No	Yes
ASEQ	RNA-seq	ASE from matched RNA and DNA	No	Yes
GeneiASE	RNA-seq	Static or between-condition ASE in single individuals	No	Yes
RASQUAL	RNA-seq/ChIP-seq/DNaseI-seq	cis-QTLs association mapping	Yes	Yes
stratAS	RNA-seq/ATAC-seq/ChIP-seq	Allelic Imbalance between conditions or cell-states (eg. tumor vs. normal)	Yes	Yes
SCALE	scRNA-seq	ASE from single-cell RNAseq	Not explicetely required	Yes
scBASE	scRNA-seq	ASE from single-cell RNAseq	Yes	Yes
ASEP	RNA-seq	ASE or differential ASE among conitions in population	Not mandatory (pseudo-phasing)	Yes
BYASE	RNA-seq	ASE in polyploid organisms	Yes	Yes
BayesASE	RNA-seq	ASE between conditions	Not explicetely required	No
AllelicImbalance	RNA-seq	Detect, manage and visualize ASE	Yes	Yes
BSCET	RNA-seq	Cell-type specific ASE from bulk RNA	No	Yes
aScan	RNA-seq	ASE at individual level	No	No
scDALI	scRNA-seq/scATAC-seq	Differential Allelic Imbalance in single-cell	Yes	Yes
Airpart	scRNA-seq (or spatially/time resolved bulk RNA-seq)	Differential Allelic Imbalance between group of cells or samples	No	Yes
BrewerIX	RNA-seq/scRNA-seq	Imprining and XCI	No	No
DAESC	scRNA-seq	Differential ASE between multiple individuals	No	Yes
SEESAW	RNA-seq	Allelic Imbalance at the isoform level	Yes	Yes
spASE	spatially resolved gene expression data	Spatial ASE patterns cell-type aware	Yes	Yes
QuASAR	RNA-seq (but also other high-throughpyt seq)	Quantitative allele-specific analysis of reads from high throughput sequencing	Not mandatory	Yes
AlleleSeq2	RNA-seq/ATAC-seq/ChIP-seq	ASE, ASB and ASCA	Yes	No
BLMRM	RNA-seq	ASE genes and ASE variation within genes	Yes	Yes
ANEVA-DOT	RNA-seq	ASE test for extreme imbalance	Yes	Yes
EAGLE	RNA-seq	GxE interaction using ASE	No	Yes
IDP-ASE	long read + short read data (hybrid-seq)	Haplotyping and ASE quantification combining short-read and long-read sequencing technologies	No	Yes
LORALS	long read data	ASE in long read sequencing	Yes	Yes

analysis can be performed with the limitation of being blind to pure monoallelic expression. Several tools have been developed for RNA-seq-based ASE analysis.

BSCET [40] allows cell-type-specific ASE analysis from bulk RNA-seq by inferring cell composition from an external scRNA-seq dataset of the same tissue, and does not require haplotype information (i.e., phasing).

BSCET was implemented in R, and both inputs (bulk and scRNA-seq data) must be in the "ExpressionSet" format. BSCET involves two steps: i) estimation of bulk composition based on scRNA-seq reference, and ii) allelic imbalance inference. Although the concept of enabling access to single-cell dimensions starting from bulk RNA-seq data is attractive, this framework requires an appropriate external scRNA-seq dataset that is not always available. Visualization options are lacking, and bioinformatics skills in R are required to use BSCET.

ASEP, implemented in R, enables cross-individual detection of gene-level ASE under one condition and differential ASE between two conditions, relying solely on RNA-seq data [41]. In the former setting, ASEP requires a table containing allelic counts obtained from RNA-seq, which may be complex to create for users without a bioinformatics background. The table must be structured with columns containing the gene name, individual identifier, genomic location or name of each transcribed SNP, reference allele read count, and total read count. If genotype data are available, they can be supplied as additional inputs so ASEP can infer the haplotype phase. Without the genetic data, a pseudo-phasing strategy based on read counts (similar to MBASED) is adopted (assigning alleles with a larger read count to the same haplotype).

In the differential ASE setting, matched RNA-seq data from the two conditions (pre- and post-treatment) are required. The input table must have the same format as above, with two additional columns specifying the conditions to which the read counts refer, and which of the two conditions should be considered as the reference.

Both analyses undergo similar processing steps: initial pseudo-phasing, a generalized linear mixed-effects model for ASE detection, and p-value estimation. The output generated by ASEP consists of a matrix with gene-based ASE estimation and corresponding p-values, yet without a visual representation of the results.

A framework similar to that proposed by ASEP is offered by GeneiASE [42] and MBASED [43], which are two additional R packages that enable the detection of ASE events in single individuals and/or comparative ASE analyses between conditions or time points.

For ASEP, the input data are a table with allelic counts, and neither haplotype nor genetic data are required to perform the analysis.

GeneiASE works on a single RNA-seq read count table with the addition of an extra column containing the gene ID (or feature ID) harboring each SNP. For between-condition comparative analysis, the allelic counts for each condition are provided in the same table.

We found that GeneiASE is easy to run, although it requires extensive preprocessing to appropriately format the input table. The output format is a gene-based table with columns containing multiple ASE-related statistical test results.

In contrast, MBASED relies on RangedSummarizedExperiment R objects for both the input and output formats. If haplotype information is not available prior to the analysis, MBASED provides the possibility of performing pseudo-phasing following the read-count-based approach adopted by ASEP. Notably, MBASED is a Bioconductor package that is easy to install. Input preparation requires additional steps, which are clearly explained in the tool vignette (see Supplementary Table 1), and we found MBASED's output to be informative and easily interpretable. The MBASED framework was also incorporated into IMPALA, a Snakemake pipeline that includes the automated preprocessing of input files required by MBASED [23,44]. IMPALA also allows for the integration of additional information concerning allelic methylation, CNVs, and somatic variants, which could contribute to the explanation of ASE.

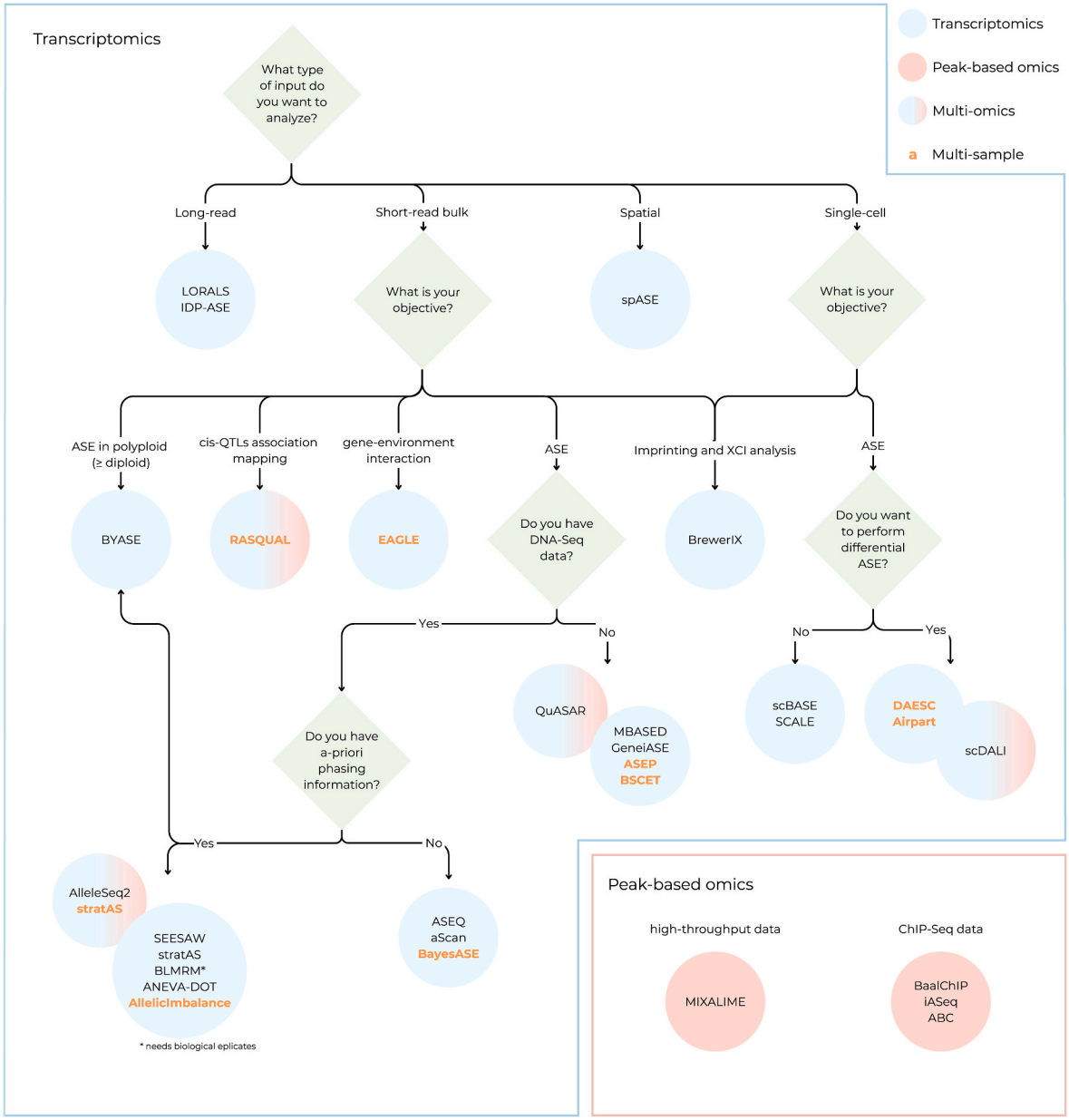


Fig. 3. ASE methods overview. This flowchart was designed to guide the reader in selecting the most appropriate method or tool based on the intended type of analysis. Each diamond shape represents a key question, while each round shape indicates one or more suitable tools. Methods highlighted in blue accept bulk or single-cell RNA-seq data as input, those in red are specialized tools for peak-based omics data analysis, and methods shaded in blue and red are compatible with either input type. Methods written in orange indicate the possibility of performing multi-sample or multi-individual analyses.

As for ASEP, the main drawback of MBASED (although overcome by IMPALA) and GeneiASE is the need to prepare read count tables prior to analysis, a potentially complex task for non-bioinformatic users. Nevertheless, this represents the only mandatory input because these tools do not require haplotype information. Visualization has not been implemented for all the tools discussed thus far.

QuASAR (implemented in R) allows joint genotyping and ASE detection from RNA-seq data [45]. For input preparation, QuASAR requires an aligned and filtered read file, a reference genome, and a BED file with variant information. The output consists of a table with allelic imbalance values (betas), standard errors, and p-values computed for each variant. ASE inference, based on a likelihood ratio test, is SNP specific. According to the authors, this approach can also be applied to other high-throughput sequencing datasets.

Although specifically tailored to the analysis of genomic imprinting and X-chromosome inactivation, BrewerIX [46] possesses several characteristics that make it stand out compared to other RNA-seq-based tools in terms of user friendliness.

BrewerIX, runnable as an app or command line tool, does not require any programming skills because the only required input is bulk (or single-cell) RNA-seq raw FASTQ files.

This tool provides a list of pre-compiled human and murine SNVs to check for biallelic signals without calling RNA-seq-based SNVs. Whenever genetic data are available, they can be used to create personalized sets of SNPs. The main output is a table complemented by two graphical panels in a dedicated user interface, featuring a gene summary and SNPs summary reporting allelic counts and monoallelic/biallelic classification. To our knowledge, this is the only available method for bulk/sc RNA-seq ASE analysis that automatically produces a graphical output summarizing the results of the analysis.

Another tool worth mentioning is EAGLE [47], which uses ASE from RNA-seq data to detect genetic and environmental interactions using a

binomial generalized linear mixed model. Considering the increasing use of RNA-seq in cohort studies, EAGLE represents a powerful and adaptable tool for studying the interplay between genetics, environment, and cellular networks in shaping gene expression [47]. When the genotype is unavailable, it can be determined from the RNA-seq data. However, at the time of writing, the documentation is outdated and the pipeline requires script amendments to successfully perform the installation. In addition, the test files are based on synthetic data and lack a straightforward guide for the preparation of real data.

4. Single-cell RNA-seq analysis software

ASE is highly tissue- and cell type-specific [5,6]. Notably, given its cell-type specificity, monoallelic expression is unlikely to be observed when analyzing bulk RNA-seq data, but it can be detected at the single-cell level. The inherent cell-specific variability inevitably hampers the detection of cell-specific signals, even in the presence of pure monoallelic expression phenomena, which often go undetected when bulk RNA-seq data are analyzed.

Therefore, tools specifically designed for single-cell allelic imbalance detection, and differential analysis are currently under development [48].

To the best of our knowledge, the current methods for single-sample single-cell ASE detection and classification include SCALE [49] and scBASE [50].

SCALE enables the genome-wide estimation of allele-specific transcriptional bursting kinetics and involves three steps: i) gene categorization into biallelic, monoallelic, or silent status through an empirical Bayesian framework; ii) allele-specific transcriptional kinetics estimation using a Poisson Beta hierarchical model, accounting for technical noise (such as dropouts or sequencing bias) and cell size (e.g., inferred by GAPDH gene expression); and iii) statistical testing [49]. SCALE processes allele-specific scRNA-seq read counts from homogeneous cell populations. The tool was implemented as an R package and is well-documented. Its output consists of a file containing the gene name, gene categorization (biallelic, monoallelic, or silent), burst (transcriptional activity pulses) frequency of each allele with related p-values, and proportion of cells expressing each allele.

Its R implementation makes SCALE straightforward to install and easy to use. The results are clear and easy to interpret owing to the automatically generated plots that aid in the interpretation of the results. Unfortunately, this method is mainly directed at SMART-seq single-cell data, that is, data from single cells isolated and sequenced in a bulk-like fashion. This choice ensured that proper coverage of many transcripts was achieved. However, this hampers its direct application in the widely used droplet-based and split-pool methods.

For scRNA-seq data, an alternative is scBASE, which classifies expression as paternal/maternal (i.e., monoallelic) or biallelic, with an estimation of the allelic proportions [50]. scBASE encompasses three main steps: i) counting allele-specific read alignments, ii) probabilistic classification of each gene as biallelic or monoallelic, and iii) optional partial pooling, considering the combined information across cells, to improve the individual cell-level estimation of allelic proportions.

Similar to SCALE, the required input is a table of allele-specific read counts, with scBASE requiring haplotype information. Prior to the main analysis, the files are converted to loom format, which was designed for the efficient management of large omics datasets. The final output contains the proportions of monoallelic (paternal/maternal) and biallelic cells, and the probabilities of cell-level allelic expression states for each gene (in loom format).

scBASE is implemented in Python3, and its installation can be challenging for non-bioinformatic users. An installation guide is provided in the GitHub tool repository; however, the tool has many dependencies and compliance requirements. No graphical output was produced.

If the purpose is differential ASE analysis using scRNA-seq data, Airpart [51] or DAESC [8] can be considered.

Airpart is a statistical framework for the detection of cell type-specific allelic imbalances, the identification of differential allelic imbalance patterns across samples, and the definition of allelic imbalance trends over time or space [51].

Airpart requires two matrices containing allelic counts of the effect and non-effect alleles alongside values concerning the variable under analysis, such as cell type annotation, spatial coordinates, or time points. Airpart is an R package, and its vignette provides all the necessary instructions to perform gene clustering based on allelic imbalance profiles, run ASE analysis across cell groups, and produce graphical outputs. Airpart is optimized to analyze long-read data, and while the tool also works on short reads, this is an important drawback in the single-cell context.

Similar to Airpart, DAESC supports both the single-sample ASE analysis and cross-sample comparisons [8]. Using latent variables and random effects, DAESC accounts for haplotype switching (implemented only in the full model, called DAESC-Mix) and the sample-repeat structure problem (multiple cells per individual) of single-cell data. Haplotype switching (i.e., swapping of reference and variant alleles) can result in two individuals characterized by an allelic imbalance of comparable magnitude but in opposite directions, resulting in erroneous ASE estimations. DAESC was developed in R and consists of two versions: the baseline beta-binomial model (DAESC-BB) and the full model (DAESC-Mix), which can be used for reasonably large datasets (e.g., $N \geq 20$ samples). The authors offer an example script for input preparation, starting from raw sequencing data, and a complete example of the workflow for conducting the analysis. Nevertheless, while DAESC is a versatile tool, both the preprocessing steps and the analysis itself require advanced bioinformatics skills. The final output is a matrix containing differential ASE effect sizes and p-values with no graphical visualization.

It is also worth mentioning spASE, a newly developed tool for allele-specific expression analysis of Slide-seq and Visium (spatial transcriptomic) data [52]. It relies on an underlying beta-binomial model to perform a cell type-aware spatial analysis of ASE. spASE was developed as an R package and integrates functions for the visualization of ASE events. Parametric and non-parametric analysis modes are offered, with the former being particularly suitable for hypothesis testing. Moreover, spASE provides the possibility of inspecting ASE events either spatially or within cell types (or both).

5. Multi-omics allele-specific analysis

Epigenetic factors are believed to play important roles in differential allelic expression. It has been demonstrated that both ASE and allele-specific chromatin accessibility (ASCA) can have an important joint influence on several disorders [53].

StratAS is a statistical framework for allele-specific analysis across different cell states and conditions [54], featuring population-scale tests for both ATAC- and RNA-seq allele-specific activity analyses.

Input preprocessing requires the use of several software programs (Eagle, WASP, and GATK ASEReadCounter) and the application of an R function available in the GitHub repository (see Supplementary Table 1). StratAS requires several inputs: a VCF file containing both phased genotype information and allelic counts, a sample file containing each sample's IDs and category/condition, a bed file with the physical regions of the features to be analyzed (i.e., peaks for ATAC-peaks or transcription start site (TSS) regions for RNA-seq), and a file with the parameters of the read count distribution. Optional flags can be added when using stratAS, corresponding to specific additional analyses such as a standard eQTL test. The result of the standard analysis is a table containing SNP-specific beta-binomial test statistics.

The need for user-autonomous preprocessing of raw files and independent installation of several software programs represent the main limitations of this framework. As in many other pipelines, data visualization was not implemented.

AlleleSeq2 is a framework for ASE, ASCA, and allele-specific binding (ASB) analyses that extends the capabilities of the previous version (AlleleSeq) [55,56]. It accepts RNA-seq, ChIP-seq, or ATAC-seq data but requires a VCF file containing DNA variants. The pipeline operates in two stages: (1) constructing a personal diploid genome and (2) detecting ASE (via RNA-seq), ASB (via ChIP-seq), or ASCA (via ATAC-seq) using a binomial test, followed by multiple test corrections. The main outputs are tables with allele-specific features.

scDALI (single-cell differential allelic imbalance) is a computational framework for analyzing allele-specific effects shared across cells or associated with specific cell states [57]. This tool does not necessarily require an *a priori* definition of the cell state, and the transition between states can be modelled either continuously or discretely.

Two main inputs are required: the allele-specific read count matrix and the cell-state representation that can be derived from the read counts (more details can be found in the documentation; Supplementary Table 1). The main output is a set of genomic regions characterized by cell-state-specific allelic imbalance, which can then be used for downstream analysis.

scDALI works with either scATAC-seq or scRNA-seq data. In our opinion, the need for input preprocessing makes this tool unsuitable for rapid analysis and for beginners. The tutorials covering the data preparation steps are available on the scDALI website (Supplementary Table 1). Graphical representation of results is not implemented, but the authors suggested plotting results independently following dimensionality reduction analyses, such as UMAP or tSNE.

RASQUAL is a statistical approach for the association mapping of cis-QTLs [58]. It is based on the integration of three main elements used to distinguish signals from noise: inter-individual differences, allele-specific signals, and potential bias introduced by sequencing technology. In the manuscript, the authors demonstrated how RASQUAL can generate "chromatin accessibility QTL" (caQTLs) if ATAC-seq data are available, and can highlight potentially imprinted regions. The required inputs are a read count table that must be converted to a specific binary file (using a script provided by the authors) and a VCF file with allelic counts and phased SNPs. The output is a table containing QTL mapping results for each SNP. Imprinting analysis can be performed to identify imprinted/randomly inactivated features, using the inputs described above.

Overall, we found that the installation was challenging because of unclear software requirements. Input files need extensive preprocessing, which requires upper-intermediate bioinformatics skills despite the clear explanations provided in the software documentation. The output is formatted as a table with no visualization features.

6. Allele-specific analysis on peak-based omics

Several tools have been developed to inspect allelic imbalances in peak-based omics. Although outside the main scope of this work, we list some example tools that could be used by those aiming to develop a multi-omic workflow, including these data types.

BaalChIP, iASeq, and ABC are widely used for ChIP-seq data analyses [59–61]. BaalChIP was designed to detect ASB events, with a particular focus on high-copy-number regions. In contrast, iASeq is tailored for joint analysis of multiple ChIP-seq datasets to improve the detection of ASB events. Finally, ABC allows for ASB analysis of the aligned ChIP-seq data for heterozygous SNVs.

MIXALIME offers a broader application for detecting allele-specific variants in high-throughput sequencing data, although it is highly specialized for peak-based input [62]. Examples include DNase-seq or ATAC-seq for the analysis of allele-specific chromatin accessibility and CAGE-seq for allele-specific promoter activity inspection.

7. SNP, gene or isoform level analysis

The ASE analysis results can be represented in different ways. Thus far, we have discussed methods that provide allele-specific expression as a function of transcribed SNP allele count. Occasionally, SNPs harbored by the same gene may provide discordant haplotype-specific allele counts owing to incomplete genomic data, limitations of sequencing technology, phasing errors, or inaccurate allele count predictions owing to low-complexity regions. This may complicate the classification of a given gene as mono or biallelic. Therefore, some methods include gene- or isoform-specific summaries or analyses in which the final classification is determined by considering all available evidence from the gene or isoform.

BYASE is a Python package for ASE analysis of diploid or polyploid organisms at the gene or isoform level [63] based on a Bayesian approach.

The inputs required for BYASE execution are a GFF annotation file, VCF file with input variants, and aligned BAM file. Both paired- and single-end RNA-seq can be used, with specific processing explained in the BYASE documentation in either case (see Supplementary Table 1). BYASE uses both phased and unphased variants to infer gene or isoform ASE status as precisely as possible. The entire analysis workflow was conducted using a graphical user interface that also plots the results. Despite this exciting idea, the software is poorly maintained, which complicates installation owing to version conflicts.

SEESAW was built on AirPart (as discussed above) for isoform-specific ASE analysis [64]. SEESAW includes several Bioconductor tools for allelic imbalance analysis at various resolutions. This versatile tool enables the user to choose between isoform-, gene-, or group-level (i.e., by TSS) analyses, with the latter favoring sensitivity over specificity and reducing inferential uncertainty compared with single-isoform-level analyses [64].

The inputs required by SEESAW include raw RNA-seq data, a reference genome (annotated for two haplotypes), and isoform information. The first step in SEESAW analysis is the creation of a diploid transcriptome using g2gtools [65]. Subsequently, allele-specific expression is quantified using Salmon [66], and statistical tests is performed using Swish [67]. Plotting and data exploration are performed in R, using the fishpond package [67].

SEESAW is not fully automated, as each step must be performed individually. However, the authors developed detailed tutorials. The output represents a complete snapshot of allelic imbalance at the isoform level for each sample.

Currently, the main limitation of SEESAW is the possibility of processing only isogenic samples.

8. Long reads

With the increasing adoption of third-generation sequencing technologies, many aspects of phasing, haplotype resolution, and isoform detection have changed rapidly. Cost and throughput remain the major challenges for the broad applicability of long-read sequencing. For instance, long-read nanopore sequencing analyses require dedicated GPUs to be efficiently performed, especially when calling DNA or RNA modifications (i.e., methylation) alongside sequence variants.

Nevertheless, as development progresses, long reads represent an exciting approach for analyzing allele-specific features [68].

We provide two example methods that were specifically developed for this purpose.

IDP-ASE is a method for haplotyping and ASE quantification at both gene and isoform levels by combining short- and long-read sequencing technologies [69]. It relies exclusively on RNA-seq data and leverages the strengths of next-generation sequencing (NGS) and third-generation sequencing (TGS) to overcome their specific limitations. For example, NGS provides less potential for haplotype phasing and covering multiple SNPs with a single read, whereas TGS is generally characterized by a lower throughput. IDP-ASE leverages the strengths of both NGS and TGS, requiring the input of SNPs and alignment files from hybrid-seq data.

LORALS is a Python-based software package designed for ASE and allele-specific transcript structure analysis, using long-read sequencing data [70]. It accepts inputs in the form of genome-aligned BAM files from long-read sequencing, and phased VCF files from whole-genome sequencing. By addressing the inherent bias in long-read sequencing, LORALS provides a tunable and versatile solution that can be applied to any long-read dataset.

9. Discussion

As discussed in this review, ASE analysis can be performed using a wide range of tools, each characterized by its advantages and drawbacks. A general trend observed in this study is that the tools developed for specific scenarios tend to be poorly maintained, making more broadly applicable workflows preferable.

Several tools have been developed to independently extract allele-specific read counts from raw RNA sequencing data[71–73], including WASP, PAC, and Ornaments. To date, WASP with the support of an external haplotype phaser has been the most commonly used method for this purpose. Similar to WASP, PAC (implemented as an automated Nextflow pipeline) aims to refine alignments and maximize the number of haplotype-resolved counts. Ornaments aim to improve WASP performance [74] using unphased heterozygous loci to estimate allele-specific transcript expression from RNA-seq.

Once allele-specific read counts are obtained, the methods and tools for statistical analysis can be categorized into two main groups: those relying on a Bayesian model (i.e., BYASE, SCALE, and BayesASE) and those adopting binomial tests (i.e., stratAS, scDALI, BrewerIX, BSCET, ASEP, GeneiASE, MBASED, ASEQ, DAESC, and RASQUAL) [8,29,32,40–43,46,49,57,58,63,75]. The advantages and limitations of both statistical approaches have been comprehensively discussed by Cleary et al. [27] and are beyond the scope of this study.

Several approaches have been developed to frame ASE results using gene- or isoform-specific (as opposed to SNP-based) methods. For example, ASEQ performs the first ASE test at the SNP level, and then exploits a thresholding approach to identify genes characterized by ASE. SEESAW relies on a different approach: computing allelic counts at the transcript level, aggregating them at the TSS level, and performing statistical testing on aggregated values.

Another important difference lies between tools that enable differential allele-specific expression analysis and those that support single-sample investigation. Options for differential analysis include i) comparison between isogenic samples under different conditions and ii) population-based tests controlling for haplotype switching and other potential population-based confounders.

Considering that single-cell approaches are rapidly becoming the norm, and that an exponentially increasing number of single-cell datasets are being produced, the choice of available pipelines tailored for this type of data is limited. In particular, we noticed that none of the pipelines supporting ASE analysis of scRNA-seq data were designed to work with high-throughput methods such as droplet- or split-pool-based scRNA-seq. This limits the use of single-cell allele-specific analyses to a relatively small proportion of datasets, excluding a significant share of the available single-cell datasets.

10. Limitations and challenges

Overall, the primary limitations of the described pipelines are twofold: i) the necessity to autonomously preprocess raw sequencing data before ASE analysis, which involves tasks such as alignment, filtering, phasing, and quality-checking raw data, as well as generating allelic count tables - tasks that require multiple tools; and ii) the absence of a multi-omics approach that can simultaneously handle RNA-seq and other data types (e.g., DNA-seq) in a fully automated manner, such as managing exome or genome sequencing data and performing automated phasing to support ASE analysis. Additionally, most of the described pipelines lack graphical outputs. Furthermore, we advocate for greater support for single-cell platforms, as the available tools often do not accommodate popular high-throughput single-cell RNA-seq technologies, thereby limiting their applicability. Overall, current tools tend to be task-specific rather than holistic, which hinders their reproducibility and scalability.

11. Future directions

Scientific workflow management systems, such as Snakemake [44] and Nextflow [31], support the creation of automated, scalable, and reproducible pipelines. These tools manage the execution of software written in almost any programming language, and are compatible with various software containers. In our compendium of methods, only IMPALA, which incorporates MBASED for ASE analysis, takes advantage of a workflow management system. In our opinion, this is a convenient method for handling a multistep workflow involving the installation and management of several software programs. The implementation of pipelines relying on workflow management systems would aid the user-friendliness, automation, and scalability of ASE analysis to large research cohorts and complex multi-omic workflows, simultaneously addressing several of the main drawbacks characterizing the existing pipelines (Fig. 4).

Another important aspect that should be addressed in ASE analysis tools in the near future is the development of an efficient visualization system. Only five of the 26 methods described here produced a thorough graphical output, with six performing at least one basic plot of the results. AllelicImbalance [35], thanks to the use of R libraries such as Gviz [76], offers the most complete visualization, along with BrewerIX [46]. In our opinion, most users would benefit from a graphical visualization of the results, both feature-specific and at a higher level. At the feature level, visualization options should provide details of the detected ASE events, allowing for a granular understanding of specific occurrences. By contrast, a broader view enables users to identify ASE patterns within a specific region, up to the chromosome, or across the entire genome or transcriptome.

Additionally, in multi-omics analyses, data from different omics can be overlaid to provide a more comprehensive understanding. Such multi-omic integration is desirable for future pipelines as it would contribute to advancing our knowledge of the effects and biological/clinical relevance of non-coding SNPs. Another interesting and viable option is to generate files that can serve as inputs for interactive genomic viewers such as IGV [77].

12. Concluding remarks

Despite numerous advances, a truly comprehensive pipeline for ASE analysis starting from raw data and automatically leading up to result visualization is still lacking. Most existing tools focus on isolated aspects of the workflow and lack end-to-end integration. In this respect, workflow management systems for automation and reproducibility are poorly utilized in ASE workflows. Scalability, graphical output production, multi-omic compatibility, and support for single-cell and spatial technologies are of fundamental importance for the next generation of ASE tools.

CRediT authorship contribution statement

Stefania Mattevi: Writing – review & editing, Writing – original draft, Methodology, Data curation, Conceptualization. **Francesco Mazzarotto:** Writing – review & editing, Writing – original draft, Conceptualization. **Paolo Martini:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization.

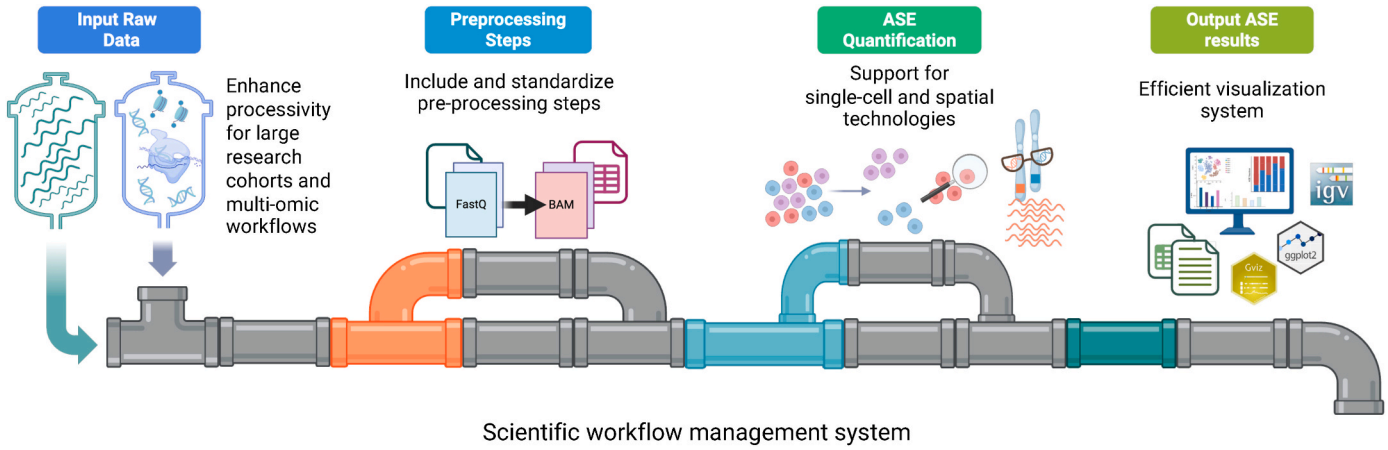


Fig. 4. Scientific workflow management system. Overview of the steps in the ASE analysis pipeline, with emphasis on areas requiring future development.

Ethics statement

Ethics statement is not necessary because we reviewed computational pipelines. We do not generate any data.

Funding

This work was supported by the European Molecular Biology Organization (EMBO), Germany, Scientific Exchange Grant Number 10653 to S.M.; by the Italian national project 'PRIN 2022' funded by the Italian Ministry of Education, University and Research, Italy [2022LJZRBY Project to P.M.]; and funding sources were not involved in the writing of the review or in the decision to submit the article for publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Stefania Mattevi: Writing–original draft, visualization. **Francesco Mazzarotto**: Writing, Review, and Editing. **Paolo Martini**: Writing, Supervision, Conceptualization.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2025.110890.

References

[1] D. Andergassen, C.P. Dotter, D. Wenzel, V. Sigl, P.C. Bammer, M. Muckenhuber, D. Mayer, T.M. Kulinski, H.-C. Theussl, J.M. Penninger, C. Bock, D.P. Barlow, F. M. Pauler, Q.J. Hudson, Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression, eLife 6 (2017), https://doi.org/10.7554/eLife.25125.

[2] C. Borel, P.G. Ferreira, F. Santoni, O. Delaneau, A. Fort, K.Y. Popadin, M. Garieri, E. Falconnet, P. Ribaux, M. Guipponi, I. Padioleau, P. Carninci, E.T. Dermitzakis, S. E. Antonarakis, Biased allelic expression in human primary fibroblast single cells, Am. J. Hum. Genet. 96 (2015) 70–80, https://doi.org/10.1016/j.ajhg.2014.12.001.

[3] A.R. Jeffries, D.A. Uwanogho, G. Cocks, L.W. Perfect, E. Dempster, J. Mill, J. Price, Erasure and reestablishment of random allelic expression imbalance after epigenetic reprogramming, RNA 22 (2016) 1620–1630, https://doi.org/10.1261/rna.058347.116.

[4] H.C. Naik, K. Hari, D. Chandel, S. Mandal, M.K. Jolly, S. Gayen, Semicoordinated allelic-bursting shape dynamic random monoallelic expression in pregastrulation embryos, iScience 24 (2021) 102954, https://doi.org/10.1016/j.isci.2021.102954.

[5] V. Savova, S. Vinogradova, D. Pruss, A.A. Gimelbrant, L.A. Weiss, Risk alleles of genes with monoallelic expression are enriched in gain-of-function variants and depleted in loss-of-function variants for neurodevelopmental disorders, Mol. Psychiatr. 22 (2017) 1785–1794, https://doi.org/10.1038/mp.2017.13.

[6] S.N. Kravitz, E. Ferris, M.I. Love, A. Thomas, A.R. Quinlan, C. Gregg, Random allelic expression in the adult human body, Cell Rep. 42 (2023) 111945, https://doi.org/10.1016/j.celrep.2022.111945.

[7] D.U. Kissiov, A. Ethell, S. Chen, N.K. Wolf, C. Zhang, S.M. Dang, Y. Jo, K. N. Madsen, I. Paranjpe, A.Y. Lee, B. Chim, S.A. Muljo, D.H. Raulet, Binary outcomes of enhancer activity underlie stable random monoallelic expression, eLife 11 (2022), https://doi.org/10.7554/eLife.74204.

[8] G. Qi, B.J. Strober, J.M. Popp, R. Keener, H. Ji, A. Battle, Single-cell allele-specific expression analysis reveals dynamic and cell-type-specific regulatory effects, Nat. Commun. 14 (2023) 6317, https://doi.org/10.1038/s41467-023-42016-9.

[9] E.A. Clayton, S. Khalid, D. Ban, L. Wang, I.K. Jordan, J.F. McDonald, Tumor suppressor genes and allele-specific expression: mechanisms and significance, Oncotarget 11 (2020) 462–479, https://doi.org/10.18632/oncotarget.27468.

[10] C.M. Bielski, M.T.A. Donoghue, M. Gadiya, A.J. Hanrahan, H.H. Won, M.T. Chang, P. Jonsson, A.V. Penson, A. Gorelick, C. Harris, A.M. Schram, A. Syed, A. Zehir, P. B. Chapman, D.M. Hyman, D.B. Solit, K. Shannon, S. Chandarlapaty, M.F. Berger, B.S. Taylor, Widespread selection for oncogenic mutant allele imbalance in cancer, Cancer Cell 34 (2018) 852–862.e4, https://doi.org/10.1016/j.ccell.2018.10.003.

[11] J. Luft, R.S. Young, A.M. Meynert, M.S. Taylor, Detecting Oncogenic Selection Through Biased Allele Retention in the Cancer Genome Atlas, 2020, https://doi.org/10.1101/2020.07.03.186593.

[12] Z. Zhang, F. van Dijk, N. de Klein, M.E. van Gijn, L.H. Franke, R.J. Sinke, M. A. Swertz, K.J. van der Velde, Feasibility of predicting allele specific expression from DNA sequencing using machine learning, Sci. Rep. 11 (2021) 10606, https://doi.org/10.1038/s41598-021-89904-y.

[13] M.P. Harwood, I. Alves, H. Edgington, M. Agbessi, V. Bruat, D. Soave, F.C. Lamaze, M.-J. Favé, P. Awadalla, Recombination affects allele-specific expression of deleterious variants in human populations, Sci. Adv. 8 (2022), https://doi.org/10.1126/sciadv.abl3819.

[14] D. van Beek, J. Verdonschot, K. Derks, H. Brunner, T.M. de Kok, I.C.W. Arts, S. Heymans, M. Kutmon, M. Adriaens, Allele-specific expression analysis for complex genetic phenotypes applied to a unique dilated cardiomyopathy cohort, Sci. Rep. 13 (2023) 564, https://doi.org/10.1038/s41598-023-27591-7.

[15] S.L. Farrow, S. Gokuladhas, W. Schierding, M. Pudjihartono, J.K. Perry, A. A. Cooper, J.M. O'Sullivan, Identification of 27 allele-specific regulatory variants in Parkinson's disease using a massively parallel reporter assay, npj Parkinson's Dis. 10 (2024) 44, https://doi.org/10.1038/s41531-024-00659-5.

[16] M. Langmyhr, S.P. Henriksen, C. Cappelletti, W.D.J. van de Berg, L. Pihlstrøm, M. Toft, Allele-specific expression of Parkinson's disease susceptibility genes in human brain, Sci. Rep. 11 (2021) 504, https://doi.org/10.1038/s41598-020-79990-9.

[17] R. Ghotbi, A. Gomez, L. Milani, G. Tybring, A.-C. Syvänen, L. Bertilsson, M. Ingelman-Sundberg, E. Aklillu, Allele-specific expression and gene methylation in the control of CYP1A2 mRNA level in human livers, Pharmacogenomics J. 9 (2009) 208–217, https://doi.org/10.1038/tpj.2009.4.

[18] L. Van Horebeek, A. Goris, Transcript-specific regulation in T-cells in multiple sclerosis susceptibility, Eur. J. Hum. Genet. 28 (2020) 849–850, https://doi.org/10.1038/s41431-020-0615-y.

[19] T. Johansson, J. Partanen, P. Saavalainen, HLA allele-specific expression: methods, disease associations, and relevance in hematopoietic stem cell transplantation, Front. Immunol. 13 (2022), https://doi.org/10.3389/fimmu.2022.1007425.

[20] I. Filip, A. Wang, O. Kravets, R. Orenbuch, J. Zhao, T.E. Perea-Chamblee, G. A. Manji, E. López de Maturana, N. Malats, K.P. Olive, R. Rabadan, Pervasiveness of HLA allele-specific expression loss across tumor types, Genome Med. 15 (2023) 8, https://doi.org/10.1186/s13073-023-01154-x.

[21] A. Saukkonen, H. Kilpinen, A. Hodgkinson, Highly accurate quantification of allelic gene expression for population and disease genetics, Genome Res. 32 (2022) 1565–1572, https://doi.org/10.1101/gr.276296.121.
[22] T.D. Nguyen, S. Markova, W. Liu, J.M. Gow, R.M. Baldwin, M. Habashian, M. V Relling, M.J. Ratain, D.L. Kroetz, Functional characterization of ABCC2 promoter polymorphisms and allele-specific expression, Pharmacogenomics J. 13 (2013) 396–402, https://doi.org/10.1038/tpj.2012.20.
[23] K. O'Neill, E. Pleasance, J. Fan, V. Akbari, G. Chang, K. Dixon, V. Csizmok, S. MacLennan, V. Porter, A. Galbraith, C.J. Grisdale, L. Culibrk, J.H. Dupuis, R. Corbett, J. Hopkins, R. Bowlby, P. Pandoh, D.E. Smailus, D. Cheng, T. Wong, C. Frey, Y. Shen, E. Lewis, L.F. Paulin, F.J. Sedlazeck, J.M.T. Nelson, E. Chuah, K. L. Mungall, R.A. Moore, R. Coope, A.J. Mungall, M.K. McConechy, L. M. Williamson, K.A. Schrader, S. Yip, M.A. Marra, J. Laskin, S.J.M. Jones, Long-read sequencing of an advanced cancer cohort resolves rearrangements, unravels haplotypes, and reveals methylation landscapes, Cell Genomics 4 (2024) 100674, https://doi.org/10.1016/j.xgen.2024.100674.
[24] Q. Zhou, Z. Wang, J. Li, W.-K. Sung, G. Li, MethHaplo: combining allele-specific DNA methylation and SNPs for haplotype region identification, BMC Bioinf. 21 (2020) 451, https://doi.org/10.1186/s12859-020-03798-7.
[25] J. Rosenski, A. Peretz, J. Magenheim, N. Loyfer, R. Shemer, B. Glaser, Y. Dor, T. Kaplan, Atlas of imprinted and allele-specific DNA methylation in the human body. https://doi.org/10.1101/2024.05.01.591988, 2024.
[26] B. Tycko, Allele-specific DNA methylation: beyond imprinting, Hum. Mol. Genet. 19 (2010) R210–R220, https://doi.org/10.1093/hmg/ddq376.
[27] S. Cleary, C. Seoighe, Perspectives on allele-specific expression, Annu. Rev. Biomed. Data Sci. 4 (2021) 101–122, https://doi.org/10.1146/annurev-biodatasci-021621-122219.
[28] S. Giulitti, M. Pellegrini, I. Zorzan, P. Martini, O. Gagliano, M. Mutarelli, M. J. Ziller, D. Cacchiarelli, C. Romualdi, N. Elvassore, G. Martello, Direct generation of human naive induced pluripotent stem cells from somatic cells in microfluidics, Nat. Cell Biol. 21 (2019) 275–286, https://doi.org/10.1038/s41556-018-0254-5.
[29] B.R. Miller, A.M. Morse, J.E. Borgert, Z. Liu, K. Sinclair, G. Gamble, F. Zou, J.R. B. Newman, L.G. León-Novelo, F. Marroni, L.M. McIntyre, Testcrosses are an Efficient Strategy for Identifying cis-regulatory Variation: Bayesian Analysis of Allele-specific Expression (Bayesase)., G3 (Bethesda), 2021, p. 11, https://doi.org/10.1093/g3journal/jkab096.
[30] F. Zambelli, M. Chiara, E. Ferrandi, P. Mandreoli, M.A. Tangaro, G. Pavesi, G. Pesole, aScan: a novel method for the study of Allele specific expression in single individuals, J. Mol. Biol. 433 (2021) 166829, https://doi.org/10.1016/j.jmb.2021.166829.
[31] P. Di Tommaso, M. Chatzou, E.W. Floden, P.P. Barja, E. Palumbo, C. Notredame, Nextflow enables reproducible computational workflows, Nat. Biotechnol. 35 (2017) 316–319, https://doi.org/10.1038/nbt.3820.
[32] A. Romanel, S. Lago, D. Prandi, A. Sboner, F. Demichelis, ASEQ: fast allele-specific studies from next-generation sequencing data, BMC Med. Genom. 8 (2015) 9, https://doi.org/10.1186/s12920-015-0084-2.
[33] J. Xie, T. Ji, M.A.R. Ferreira, Y. Li, B.N. Patel, R.M. Rivera, Modeling allele-specific expression at the gene and SNP levels simultaneously by a Bayesian logistic mixed regression model, BMC Bioinf. 20 (2019) 530, https://doi.org/10.1186/s12859-019-3141-6.
[34] P. Mohammadi, S.E. Castel, B.B. Cummings, J. Einson, C. Sousa, P. Hoffman, S. Donkervoort, Z. Jiang, P. Mohassel, A.R. Foley, H.E. Wheeler, H.K. Im, C. G. Bonnemann, D.G. MacArthur, T. Lappalainen, Genetic regulatory variation in populations informs transcriptome analysis in rare disease, Science 366 (1979) (2019) 351–356, https://doi.org/10.1126/science.aay0256.
[35] J.R. Gådin, F.M. van't Hooft, P. Eriksson, L. Folkersen, AllelicImbalance: an R/bioconductor package for detecting, managing, and visualizing allele expression imbalance data from RNA sequencing, BMC Bioinf. 16 (2015) 194, https://doi.org/10.1186/s12859-015-0620-2.
[36] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics 25 (2009) 1754–1760, https://doi.org/10.1093/bioinformatics/btp324.
[37] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21, https://doi.org/10.1093/bioinformatics/bts635.
[38] G.A. Van der Auwera, B.D. O'Connor, Genomics in the Cloud: Using Docker, GATK, and WDL in Terra, first ed., O'Reilly Media, Inc., 2020.
[39] S.E. Castel, P. Mohammadi, W.K. Chung, Y. Shen, T. Lappalainen, Rare variant phasing and haplotypic expression from RNA sequencing with phASER, Nat. Commun. 7 (2016) 12817, https://doi.org/10.1038/ncomms12817.
[40] J. Fan, X. Wang, R. Xiao, M. Li, Detecting cell-type-specific allelic expression imbalance by integrative analysis of bulk and single-cell RNA sequencing data, PLoS Genet. 17 (2021) e1009080, https://doi.org/10.1371/journal.pgen.1009080.
[41] J. Fan, J. Hu, C. Xue, H. Zhang, K. Susztak, M.P. Reilly, R. Xiao, M. Li, ASEP: Gene-based detection of allele-specific expression across individuals in a population by RNA sequencing, PLoS Genet. 16 (2020) e1008786, https://doi.org/10.1371/journal.pgen.1008786.
[42] D. Edsgärd, M.J. Iglesias, S.-J. Reilly, A. Hamsten, P. Tornvall, J. Odeberg, O. Emanuelsson, GeneiASE: detection of condition-dependent and static allele-specific expression from RNA-seq data without haplotype information, Sci. Rep. 6 (2016) 21134, https://doi.org/10.1038/srep21134.
[43] O. Mayba, H.N. Gilbert, J. Liu, P.M. Haverty, S. Jhunjhunwala, Z. Jiang, C. Watanabe, Z. Zhang, MBASED: allele-specific expression detection in cancer tissues and cell lines, Genome Biol. 15 (2014) 405, https://doi.org/10.1186/s13059-014-0405-3.
[44] F. Mölder, K.P. Jablonski, B. Letcher, M.B. Hall, C.H. Tomkins-Tinch, V. Sochat, J. Forster, S. Lee, S.O. Twardziok, A. Kanitz, A. Wilm, M. Holtgrewe, S. Rahmann, S. Nahnsen, J. Köster, Sustainable data analysis with snakemake, F1000Res 10 (2021) 33, https://doi.org/10.12688/f1000research.29032.2.
[45] C.T. Harvey, G.A. Moyerbrailean, G.O. Davis, X. Wen, F. Luca, R. Pique-Regi, QuASAR: quantitative allele-specific analysis of reads, Bioinformatics 31 (2015) 1235–1242, https://doi.org/10.1093/bioinformatics/btu802.
[46] P. Martini, G. Sales, L. Diamante, V. Perrera, C. Colantuono, S. Riccardo, D. Cacchiarelli, C. Romualdi, G. Martello, BrewerIX enables allelic expression analysis of imprinted and X-linked genes from bulk and single-cell transcriptomes, Commun. Biol. 5 (2022) 146, https://doi.org/10.1038/s42003-022-03087-4.
[47] D.A. Knowles, J.R. Davis, H. Edgington, A. Raj, M.-J. Favé, X. Zhu, J.B. Potash, M. M. Weissman, J. Shi, D.F. Levinson, P. Awadalla, S. Mostafavi, S.B. Montgomery, A. Battle, Allele-specific expression reveals interactions between genetic variation and environment, Nat. Methods 14 (2017) 699–702, https://doi.org/10.1038/nmeth.4298.
[48] G. Qi, A. Battle, Computational methods for allele-specific expression in single cells, Trends Genet. 40 (2024) 939–949, https://doi.org/10.1016/j.tig.2024.07.003.
[49] Y. Jiang, N.R. Zhang, M. Li, SCALE: modeling allele-specific gene expression by single-cell RNA sequencing, Genome Biol. 18 (2017) 74, https://doi.org/10.1186/s13059-017-1200-8.
[50] K. Choi, N. Raghupathy, G.A. Churchill, A Bayesian mixture model for the analysis of allelic expression in single cells, Nat. Commun. 10 (2019) 5188, https://doi.org/10.1038/s41467-019-13099-0.
[51] W. Mu, H. Sarkar, A. Srivastava, K. Choi, R. Patro, M.I. Love, Airpart: interpretable statistical models for analyzing allelic imbalance in single-cell datasets, Bioinformatics 38 (2022) 2773–2780, https://doi.org/10.1093/bioinformatics/btac212.
[52] L.S. Zou, D.M. Cable, I.A. Barrera-Lopez, T. Zhao, E. Murray, M.J. Aryee, F. Chen, R.A. Irizarry, Detection of allele-specific expression in spatial transcriptomics with spASE, Genome Biol. 25 (2024) 180, https://doi.org/10.1186/s13059-024-03317-4.
[53] S. Zhang, H. Zhang, Y. Zhou, M. Qiao, S. Zhao, A. Kozlova, J. Shi, A.R. Sanders, G. Wang, K. Luo, S. Sengupta, S. West, S. Qian, M. Streit, D. Avramopoulos, C. A. Cowan, M. Chen, Z.P. Pang, P. V Gejman, X. He, J. Duan, Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants, Science 369 (2020) 561–565, https://doi.org/10.1126/science.aay3983.
[54] A. Gusev, S. Spisak, A.P. Fay, H. Carol, K.C. Vavra, S. Signoretti, V. Tisza, M. Pomerantz, F. Abbasi, J.-H. Seo, T.K. Choueiri, K. Lawrenson, M.L. Freedman, Allelic Imbalance Reveals Widespread germline-somatic Regulatory Differences and Prioritizes Risk Loci in Renal Cell Carcinoma, 2019, https://doi.org/10.1101/631150.
[55] J. Rozowsky, A. Abyzov, J. Wang, P. Alves, D. Raha, A. Harmanci, J. Leng, R. Bjornson, Y. Kong, N. Kitabayashi, N. Bhardwaj, M. Rubin, M. Snyder, M. Gerstein, AlleleSeq: analysis of allele-specific expression and binding in a network framework, Mol. Syst. Biol. 7 (2011), https://doi.org/10.1038/msb.2011.54.
[56] J. Rozowsky, J. Gao, B. Borsari, Y.T. Yang, T. Galeev, G. Gürsoy, C.B. Epstein, K. Xiong, J. Xu, T. Li, J. Liu, K. Yu, A. Berthel, Z. Chen, F. Navarro, M.S. Sun, J. Wright, J. Chang, C.J.F. Cameron, N. Shoresh, E. Gaskell, J. Drenkow, J. Adrian, S. Aganezov, F. Aguet, G. Balderrama-Gutierrez, S. Banskota, G.B. Corona, S. Chee, S.B. Chhetri, G.C. Cortez Martins, C. Danyko, C.A. Davis, D. Farid, N.P. Farrell, I. Gabdank, Y. Gofin, D.U. Gorkin, M. Gu, V. Hecht, B.C. Hitz, R. Issner, Y. Jiang, M. Kirsche, X. Kong, B.R. Lam, S. Li, B. Li, X. Li, K.Z. Lin, R. Luo, M. Mackiewicz, R. Meng, J.E. Moore, J. Mudge, N. Nelson, C. Nusbaum, I. Popov, H.E. Pratt, Y. Qiu, S. Ramakrishnan, J. Raymond, L. Salichos, A. Scavelli, J.M. Schreiber, F. J. Sedlazeck, L.H. See, R.M. Sherman, X. Shi, M. Shi, C.A. Sloan, J.S. Strattan, Z. Tan, F.Y. Tanaka, A. Vlasova, J. Wang, J. Werner, B. Williams, M. Xu, C. Yan, L. Yu, C. Zaleski, J. Zhang, K. Ardlie, J.M. Cherry, E.M. Mendenhall, W.S. Noble, Z. Weng, M.E. Levine, A. Dobin, B. Wold, A. Mortazavi, B. Ren, J. Gillis, R. M. Myers, M.P. Snyder, J. Choudhary, A. Milosavljevic, M.C. Schatz, B.E. Bernstein, R. Guigó, T.R. Gingeras, M. Gerstein, The EN-TEx resource of multi-tissue personal epigenomes & variant-impact models, Cell 186 (2023) 1493–1511, https://doi.org/10.1016/j.cell.2023.02.018, e40.
[57] T. Heinen, S. Secchia, J.P. Reddington, B. Zhao, E.E.M. Furlong, O. Stegle, scDALI: modeling allelic heterogeneity in single cells reveals context-specific genetic regulation, Genome Biol. 23 (2022) 8, https://doi.org/10.1186/s13059-021-02593-8.
[58] N. Kumasaka, A.J. Knights, D.J. Gaffney, Fine-mapping cellular QTLs with RASQUAL and ATAC-seq, Nat. Genet. 48 (2016) 206–213, https://doi.org/10.1038/ng.3467.
[59] I. de Santiago, W. Liu, K. Yuan, M. O'Reilly, C.S.R. Chilamakuri, B.A.J. Ponder, K. B. Meyer, F. Markowetz, BaalChIP: bayesian analysis of allele-specific transcription factor binding in cancer genomes, Genome Biol. 18 (2017) 39, https://doi.org/10.1186/s13059-017-1165-7.
[60] Y. Wei, X. Li, Q. Wang, H. Ji, iASeq: integrative analysis of allele-specificity of protein-DNA interactions in multiple ChIP-seq datasets, BMC Genom. 13 (2012) 681, https://doi.org/10.1186/1471-2164-13-681.
[61] S.D. Bailey, C. Virtanen, B. Haibe-Kains, M. Lupien, ABC: a tool to identify SNVs causing allele-specific transcription factor binding from ChIP-seq experiments, Bioinformatics 31 (2015) 3057–3059, https://doi.org/10.1093/bioinformatics/btv321.
[62] A. Buyan, G. Meshcheryakov, V. Safronov, S. Abramov, A. Boytsov, V. Nozdrin, E. F. Baulin, S. Kolmykov, J. Vierstra, F. Kolpakov, V.J. Makeev, I.V. Kulakovskiy, Statistical framework for calling allelic imbalance in high-throughput sequencing

data, Nat. Commun. 16 (2025) 1739, https://doi.org/10.1038/s41467-024-55513-2.
[63] L. Dong, J. Wang, G. Wang, BYASE: a Python library for estimating gene and isoform level allele-specific expression, Bioinformatics 36 (2020) 4955–4956, https://doi.org/10.1093/bioinformatics/btaa636.
[64] E.Y. Wu, N.P. Singh, K. Choi, M. Zakeri, M. Vincent, G.A. Churchill, C.L. Ackert-Bicknell, R. Patro, M.I. Love, SEESAW: detecting isoform-level allelic imbalance accounting for inferential uncertainty, Genome Biol. 24 (2023) 165, https://doi.org/10.1186/s13059-023-03003-x.
[65] N. Raghupathy, A. Srivastava, M. Lloyd, g2gtools. https://github.com/churchill-lab/g2gtools, 2024.
[66] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression, Nat. Methods 14 (2017) 417–419, https://doi.org/10.1038/nmeth.4197.
[67] A. Zhu, A. Srivastava, J.G. Ibrahim, R. Patro, M.I. Love, Nonparametric expression analysis using inferential replicate counts, Nucleic Acids Res. 47 (2019) e105, https://doi.org/10.1093/nar/gkz622.
[68] E. Deng, Q. Shen, J. Zhang, Y. Fang, L. Chang, G. Luo, X. Fan, Systematic evaluation of single-cell RNA-seq analyses performance based on long-read sequencing platforms, J. Adv. Res. 71 (2025) 141–153, https://doi.org/10.1016/j.jare.2024.05.020.
[69] B. Deonovic, Y. Wang, J. Weirather, X.-J. Wang, K.F. Au, IDP-ASE: haplotyping and quantifying allele-specific expression at the gene and gene isoform level by hybrid sequencing, Nucleic Acids Res. 45 (2017) e32, https://doi.org/10.1093/nar/gkw1076.
[70] D.A. Glinos, G. Garborcauskas, P. Hoffman, N. Ehsan, L. Jiang, A. Gokden, X. Dai, F. Aguet, K.L. Brown, K. Garimella, T. Bowers, M. Costello, K. Ardlie, R. Jian, N. R. Tucker, P.T. Ellinor, E.D. Harrington, H. Tang, M. Snyder, S. Juul, P. Mohammadi, D.G. MacArthur, T. Lappalainen, B.B. Cummings, Transcriptome variation in human tissues revealed by long-read sequencing, Nature 608 (2022) 353–359, https://doi.org/10.1038/s41586-022-05035-y.
[71] B. van de Geijn, G. McVicker, Y. Gilad, J.K. Pritchard, WASP: allele-specific software for robust molecular quantitative trait locus discovery, Nat. Methods 12 (2015) 1061–1063, https://doi.org/10.1038/nmeth.3582.
[72] H. Younesy, T. Möller, A. Heravi-Moussavi, J.B. Cheng, J.F. Costello, M.C. Lorincz, M.M. Karimi, S.J.M. Jones, ALEA: a toolbox for allele-specific epigenomics analysis, Bioinformatics 30 (2014) 1172–1174, https://doi.org/10.1093/bioinformatics/btt744.
[73] C.A. Soderlund, W.M. Nelson, S.A. Goff, Allele workbench: transcriptome pipeline and interactive graphics for allele-specific expression, PLoS One 9 (2014) e115740, https://doi.org/10.1371/journal.pone.0115740.
[74] A. Adduri, S. Kim, Ornaments for efficient allele-specific expression estimation with bias correction, Am. J. Hum. Genet. 111 (2024) 1770–1781, https://doi.org/10.1016/j.ajhg.2024.06.014.
[75] A. Shetty, J.-H. Seo, C.A. Bell, E.P. O'Connor, M.M. Pomerantz, M.L. Freedman, A. Gusev, Allele-specific epigenetic activity in prostate cancer and normal prostate tissue implicates prostate cancer risk mechanisms, Am. J. Hum. Genet. 108 (2021) 2071–2085, https://doi.org/10.1016/j.ajhg.2021.09.008.
[76] F. Hahne, R. Ivanek, Visualizing Genomic Data Using Gviz and Bioconductor, 2016, pp. 335–351, https://doi.org/10.1007/978-1-4939-3578-9_16.
[77] H. Thorvaldsdottir, J.T. Robinson, J.P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration, Briefings Bioinf. 14 (2013) 178–192, https://doi.org/10.1093/bib/bbs017.