

Gene expression

ISoLDE: a data-driven statistical method for the inference of allelic imbalance in datasets with reciprocal crosses

Christelle Reynès^{1,2,*}, Guilhem Kister², Marine Rohmer³,
Tristan Bouschet¹, Annie Varrault¹, Emeric Dubois ³,
Stéphanie Rialle³, Laurent Journot ^{1,3} and Robert Sabatier^{1,2,*}

¹Institut de Génomique Fonctionnelle, IGF, Univ. Montpellier, CNRS, INSERM, 34094 Montpellier, France, ²Faculté de Pharmacie, Univ. Montpellier 34093 Montpellier, France and ³Montpellier GenomiX, MGX, Univ. Montpellier, CNRS, INSERM, 34094 Montpellier, France

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Received on December 14, 2018; revised on July 8, 2019; editorial decision on July 15, 2019; accepted on July 22, 2019

Abstract

Motivation: Allelic imbalance (AI), i.e. the unequal expression of the alleles of the same gene in a single cell, affects a subset of genes in diploid organisms. One prominent example of AI is parental genomic imprinting, which results in parent-of-origin-dependent, mono-allelic expression of a limited number of genes in metatherian and eutherian mammals and in angiosperms. Currently available methods for identifying AI rely on data modeling and come with the associated limitations.

Results: We have designed ISoLDE (Integrative Statistics of alleLe Dependent Expression), a novel nonparametric statistical method that takes into account both AI and the characteristics of RNA-seq data to infer allelic expression bias when at least two biological replicates are available for reciprocal crosses. ISoLDE learns the distribution of a specific test statistic from the data and calls genes ‘allelically imbalanced’, ‘bi-allelically expressed’ or ‘undetermined’. Depending on the number of replicates, predefined thresholds or permutations are used to make calls. We benchmarked ISoLDE against published methods, and showed that ISoLDE compared favorably with respect to sensitivity, specificity and robustness to the number of replicates. Using ISoLDE on different RNA-seq datasets generated from hybrid mouse tissues, we did not discover novel imprinted genes (IGs), confirming the most conservative estimations of IG number.

Availability and implementation: ISoLDE has been implemented as a Bioconductor package available at <http://bioconductor.org/packages/ISoLDE/>.

Contact: christelle.reynes@umontpellier.fr or robert.sabatier@umontpellier.fr

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

In diploid cells, genetic and epigenetic factors influence the relative expression levels of alleles. In the case of parental genomic imprinting, the preferred allele depends on its parental origin. Other forms of AI involve the random selection of one allele, allelic exclusion at the immunoglobulin and olfactory receptor loci or strain-of-origin-

dependent expression. The statistical methods used to infer AI from RNA-seq data all rely on data modeling. The inference of AI is then highly dependent on model accuracy as minor deviations of the experimental data from the model may lead to false positives or false negatives. A comprehensive survey of statistical methods used to infer AI is presented in the [Supplementary File, Section S2](#) and [Supplementary Table S1](#).

The importance of model accuracy is illustrated by the controversy over the number of genes undergoing parental genomic imprinting in mice and humans. A key feature of a bona fide parentally imprinted gene (IG) is the unequal expression of its two parental alleles that is demonstrated by parent-of-origin-dependent AI in RNA-seq data. Until 2010, the number of IGs in eutherian mammals was estimated to be approximately 100 (Babak *et al.*, 2008). In 2010, Gregg *et al.* suggested that 1300+ loci displayed parent-of-origin-dependent AI in the mouse embryonic and adult brain (Gregg *et al.*, 2010); this unexpected finding was highlighted in several comments (Keverne, 2010; Tollkuhn *et al.*, 2010; Wilkinson, 2010). In 2012, Babak *et al.* (DeVeale *et al.*, 2012) showed that ‘the vast majority of the novel reported imprinted loci are false positives explained by technical and biological variation of the experimental approach’. They showed that ‘allele-specific expression measured with RNA-seq is not accurately modeled with statistical methods that assume random independent sampling and that systematic error must be accounted for to enable accurate identification of imprinted expression’. The number of bona fide IGs was estimated to be approximately 150–200 (DeVeale *et al.*, 2012; Kelsey and Bartolomei, 2012). In 2015, Dulac’s group used a novel statistical method and found 160 IGs, including 41 new ones that displayed a weak to moderate parental bias (Perez *et al.*, 2015). This study suggested that a number of IGs remained to be identified in specific tissues or cell types. Similarly, Gregg’s group designed a new statistical method and identified 407 IGs, substantially more than the previous estimates, including 142 autosomal IGs that were not previously called ‘imprinted’ (Bonthuis *et al.*, 2015). Babak *et al.* re-analyzed published human and mouse datasets that covered many different organs and found only a handful of new IGs (Babak *et al.*, 2015).

To circumvent the limitations of the modeling approach, we designed Integrative Statistics of alleLe Dependent Expression (ISoLDE), a method that uses a nonparametric approach based on data resampling. We designed a specific nonparametric test statistic to take into account data specificities and make the best of biological replicates. ISoLDE identifies both biased and unbiased genes, leaving some genes as undetermined. It may be applied when at least two replicates of reciprocal crosses are available. We benchmarked ISoLDE against 7 other methods using 13 experimental datasets and 2 simulated datasets; we concluded that ISoLDE is more sensitive, specific and robust to the number of replicates than currently available methods. Using ISoLDE on different RNA-seq datasets generated from hybrid mouse tissues, we did not discover novel IGs, confirming the most conservative estimations of IG number.

2 Materials and methods

For the sake of simplicity, we describe how ISoLDE performs when AI depends on the parental origin of the alleles; the method works straightforwardly for AI with a different origin. A more detailed description of ISoLDE is available in the [Supplementary File, Section S1.1](#).

2.1 ISoLDE uses a robust and specific statistic

ISoLDE adapts usual statistics to RNA-seq data specificities. In a usual z-test, for example, the denominator accounts for the variability of the samples, assuming a binomial behavior, which underestimates real RNA-seq data variability. Classical variance is also inappropriate as many replicates are rarely available. To design a specific statistic, we chose the MAD, median absolute deviation (Hampel, 1974), to quantify the variability of the samples. MAD is the median of the absolute deviations from the sample’s median. For

a sample $X = (x_1, x_2, \dots, x_n)$, let Me be the median of X , then we denote $\text{MAD}(X) = \text{MAD}_i(x_i) = \text{median}_i(|x_i - \text{Me}|)$. Sequencing depth also has to be taken into account as more reads implies more reliable results. Hence, we divided the MAD by the median number of normalized reads in the sample. This variability estimation is a robust version of the coefficient of variation, using the MAD instead of the standard deviation and the median instead of the mean.

Let $n_{i,j,P}^g$ denote the normalized number of reads with paternal origin for gene g in the j th replicate ($j \in \{1, \dots, n_i\}$) of cross i ($i \in \{1, 2\}$) corresponding to one of the two reciprocal crosses, respectively $A \times B$ and $B \times A$) and $n_{i,j,M}^g$ the normalized number of reads with maternal origin. Then, for gene g , the statistic is expressed as

$$S_g = \frac{\pi_{M,1} - \pi_{P,2}}{\sqrt{\frac{\text{MAD}_{i1}(n_{i,j,M}^g)}{\text{median}_{i1}(n_{i,j,M}^g)} + \frac{\text{MAD}_{i2}(n_{i,j,P}^g)}{\text{median}_{i2}(n_{i,j,P}^g)}}$$

where

$$\pi_{M,1} = \frac{\sum_{j=1}^{n_1} n_{1,j,M}^g}{\sum_{j=1}^{n_1} n_{1,j,M}^g + \sum_{j=1}^{n_1} n_{1,j,P}^g}$$

is the proportion of normalized reads from maternal origin in the $A \times B$ cross and

$$\pi_{P,2} = \frac{\sum_{j=1}^{n_2} n_{2,j,P}^g}{\sum_{j=1}^{n_2} n_{2,j,M}^g + \sum_{j=1}^{n_2} n_{2,j,P}^g}$$

is the proportion of normalized reads from paternal origin in the $B \times A$ cross. If g is bi-allelically expressed, then $\pi_{M,1} \simeq \pi_{P,2} \simeq 0.5$ and S_g is close to 0. If g is parentally biased, then either $\pi_{M,1}$ is close to 1 and $\pi_{P,2}$ is close to 0 or the reverse. Hence, the numerator of S_g is close to either -1 (paternal expression) or $+1$ (maternal expression). As the denominator is always positive, the expected behavior of S_g is the following; bi-allelically expressed gene, S_g is close to 0; paternally expressed gene, S_g is strongly negative; maternally expressed gene, S_g is strongly positive.

2.2 ISoLDE calculates two different thresholds to detect bi-allelically expressed and allelically biased genes depending on the number of replicates

The next step was to specify thresholds defining both allelically biased and bi-allelically expressed genes while keeping the possibility that some genes remained undetermined with regards to statistical evidence. Two situations were considered depending on whether only two or more than two biological replicates per cross were available. ISoLDE was purposely not designed for experiments with no replicates because such experiments do not provide valuable biological information from statistical studies. In both cases, two different statistical tests, Q_1 and Q_2 , were performed; Q_1 was testing whether a given gene was allelically imbalanced, and Q_2 was testing whether it was bi-allelically expressed.

2.2.1 ISoLDE performed resampling when more than two biological replicates were available in both reciprocal crosses

When enough information was available, ISoLDE performed resampling. Some existing methods perform resampling to define a P -value threshold to compensate for approximate modeling of data behavior (Andergassen *et al.*, 2015; Babak *et al.*, 2015; Bonthuis *et al.*, 2015; DeVeale *et al.*, 2012; Rozowsky *et al.*, 2011; Zou *et al.*, 2014). ISoLDE performed resampling to learn the behavior of the S_g statistic in a nonparametric way, with no reference to a predefined distribution law.

For each gene and each biological replicate, the observed reads were randomly assigned ‘maternal’ or ‘paternal’ depending on the current test:

- For Q_1 , ISoLDE generated the distribution of S_g under the null hypothesis, i.e. bi-allelic expression and $\pi_{M,1} = \pi_{P,2}$, by equally assigning the reads to maternal and paternal origins using the distribution of proportions observed between replicates within one cross. Within a given cross, differences between maternal and paternal origins were expected to only account for biological noise.
- For Q_2 , ISoLDE generated the distribution of S_g under the null hypothesis, i.e. AI and $\pi_{M,1} \neq \pi_{P,2}$, by randomly choosing a parental bias ratio between 0.6 and 1 and randomly assigning the reads to maternal and paternal origins according to this ratio. We chose 0.6 as the minimum bias value because it is the lowest threshold value considered a significant bias so far (Andergassen et al., 2015; Smith et al., 2013; Tran et al., 2014; Wang and Clark, 2014).

Resampling was performed 10 000 times in order to obtain the distribution of S_g under the null hypothesis of Q_1 or Q_2 for each gene. Then, empirical P -values were estimated following the method described by Dudoit and van der Laan (2008) to obtain the null distribution of null shift and scale-transformed test statistics (Dudoit et al., 2004; Pollard and van der Laan, 2004; van der Laan et al., 2004) followed by Benjamini–Hochberg adjustment. Finally, the two P -values were used to call genes ‘allelically imbalanced’ (‘AI’), ‘bi-allelically expressed’ (‘BA’) or ‘undetermined’ (‘UN’).

As this method was stochastic, a low number of genes that were in-between allelically imbalanced and bi-allelic expression may have different calls after different runs. Hence, the algorithm was run three times by default in order to retrieve consensus results. When discrepancies were observed among the runs, genes were called ‘UN’.

To help the user detect potentially interesting genes among undetermined genes, we appended a consistency flag to undetermined genes that exhibited a consistent bias direction, i.e. genes having the same parental bias in all the replicates of each cross. We called ‘UN’ and appended a significance flag to genes called ‘AI’ but displaying at least one inconsistency in the bias direction among all the replicates of each cross.

2.2.2 ISoLDE relied on two predefined thresholds when only two biological replicates in at least one cross were available

In such experimental situations, there was not enough information to obtain reliable distributions of S_g under the null hypotheses for Q_1 and Q_2 . We then selected two thresholds for the S_g statistic based on the analysis of 10 different experimental datasets comprising two or more replicates for each cross:

- five datasets from Bouschet et al. (2017) including the two *in vivo* datasets used in the Results section and three *in vitro* experiments containing only two biological replicates;
- the dataset from Hasin-Brumshtein et al. (2014) comprising two replicates of mouse adipose tissue from the reciprocal crosses of C57BL/6J and DBA/2J strains;
- the dataset from Babak et al. (2008) comprising four replicates of E9.5 mouse embryos from reciprocal crosses of CAST/EiJ and C57BL/6J strains;
- three datasets from Lorenc et al. (2014) comprising three to six replicates of three mouse tissues [vomeronasal organ (VNO), hypothalamus (HYP) and liver (LIV)] from reciprocal crosses between WSB and PWD strains.

The majority of these datasets had few replicates, and we excluded other datasets with many replicates to avoid biasing the thresholds toward values that were sensible only when many replicates were available. Moreover, the selected datasets included very different biological conditions and sequencing depth to cover a range of experimental situations.

We used the resampling version of ISoLDE when datasets had more than two replicates. For two-replicate datasets, we applied the Storer–Kim test to each gene, or its normal approximation when appropriate. To prevent false positives, we set a Storer–Kim test P -value threshold using mock reciprocal crosses (DeVeale et al., 2012) with a false discovery rate (FDR) threshold set at 0.05. We applied two additional filters based on biological assumptions. The first filter selected only AI candidates whose parental bias was the same for both replicates. The second filter selected those genes for which the Storer–Kim test was significant for most replicates taken independently. We used again the mock reciprocal crosses approach to set a threshold on significance agreement.

Then, the statistic S_g was computed for all genes of a dataset. A first threshold was defined as the fifth percentile of the absolute value of the statistic distribution among allelically imbalanced candidates identified in the previous step. Genes whose statistic absolute value was higher than this threshold were considered allelically imbalanced. Finally, a second threshold was defined as the 99th percentile of the absolute value of the statistic distribution among genes not identified as allelically imbalanced candidates in the previous step. Genes whose statistic absolute value was lower than this threshold were considered bi-allelically expressed. Genes whose statistic absolute value was between the two thresholds were called ‘undetermined’ (‘UN’).

This method was applied to the 10 different datasets mentioned above. The two selected thresholds were the median values of the thresholds obtained among the 10 datasets: genes with a S_g absolute value lower than 0.7133 were called ‘bi-allelically expressed’ (‘BA’) and genes with a S_g absolute value higher than 1.7712 were called ‘allelically imbalanced’ (‘AI’).

2.3 Filtering method

Filtering is an important step when gene-by-gene statistical testing is performed. It directly impacts the number of positives in the final result. ISoLDE performed a two-step filtering:

- threshold definition: for each parental origin, the maximum number of reads was computed for genes with at least 66% of their replicates as zero counts in the raw data. The 95th percentile of the distribution of those maximum values was considered a threshold value below which the gene had close to no expression (e.g. threshold filter was 2.95 for the P0 cortex and 2.79 for the E13.5 cortex samples).
- only genes with all replicates for at least one parental origin above the threshold were considered expressed and kept for the subsequent analyses.

3 Results

3.1 Comparison of the resampling-based and predefined thresholds versions of ISoLDE

3.1.1 The resampling-based version of ISoLDE provides detailed and easy to process results

We first surveyed 18 mouse strains to determine the number of genes whose AI is searchable in the F1 progeny of any cross, i.e. the number of genes that have at least one exonic SNP for the cross under

study (see [Supplementary File, Section S1.4](#) and [Supplementary Fig. S1](#) for details). We selected the C57BL/6J and JF1 inbred strains whose genomes diverge at more than 12 million SNPs ([Takada et al., 2013](#)) and enable the interrogation of 21 256 genes out of the 22 569 that were surveyed (94.2%). These strains were reciprocally crossed, and three to four biological replicates of each cross were obtained. We ran the resampling-based version of ISoLDE on the RNA-seq data generated in our laboratory from E13.5 and P0 murine cortices ([Bouschet et al., 2017](#)). Read counts were normalized using the RLE method as implemented in the edgeR package ([Robinson et al., 2010](#)).

Each gene was projected according to its S_g denominator value (ordinate) and numerator value (abscissa), which corresponds to the allelic bias ([Fig. 1A](#) and [Supplementary Fig. S2A](#)). The ordinate scales are not the same in [Figure 1A](#) and [Supplementary Figure S2A](#) due to the different number of replicates in both experiments; three replicates for each cross at E13.5 and three and four replicates for the crosses at P0.

A total of 37 genes displayed significant parent-of-origin dependent expression at E13.5 ([Supplementary Fig. S2A](#)) or P0 ([Fig. 1A](#)), all of which have already been identified in previous AI studies. The allelic expression of nine genes was confirmed by other methods ([Bouschet et al., 2017](#)).

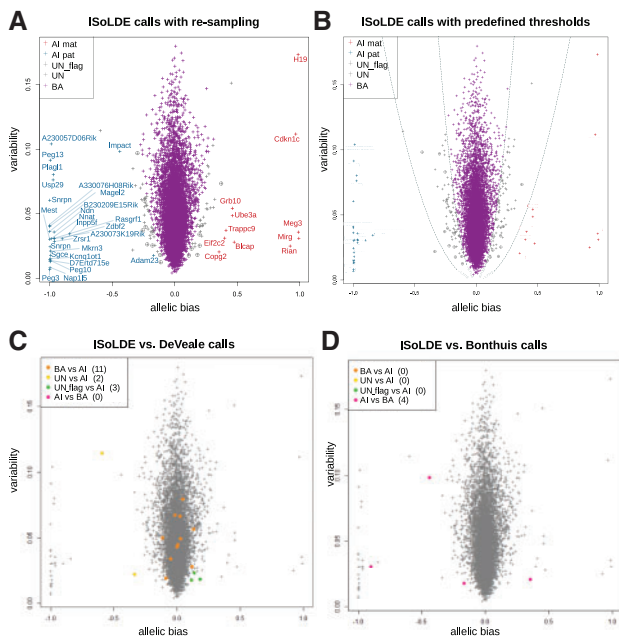


Fig. 1. ISoLDE calls on the P0 cortex dataset and comparison with the methods by DeVeale et al. and Bonthuis et al. (A) Output of the resampling version of ISoLDE. For each gene, the variability (denominator value of the S_g statistic) was plotted against the allelic bias (numerator value of the S_g statistic). Violet crosses correspond to bi-allelically expressed ('BA') genes. Red and blue crosses correspond to genes called maternally and paternally imbalanced ('AI mat' and 'AI pat', respectively). Gray crosses correspond to undetermined ('UN') genes. Gray circled crosses correspond to flagged genes (consistency or significance flag, 'UN_flag'). (B) Output of the ISoLDE version using predefined thresholds. Crosses were colored as in (A). Dashed lines represent the empirical threshold values. (C) Discordant ISoLDE (resampling version) and DeVeale calls. The P0 cortex dataset was analyzed using the method by [DeVeale et al. \(2012\)](#). Genes that were called 'AI' by only one of the two methods are represented as colored circles on the ISoLDE output. See the figure for the color code; the left call is the ISoLDE call. Numbers in brackets are the numbers of genes in each category. (D) Same as C using the method by [Bonthuis et al. \(2015\)](#)

As expected, lower allelic bias, i.e. lower absolute value of S_g numerator, required lower variability to result in an 'AI' call. The minimal bias that resulted in an 'AI' call was 0.35 for the E13.5 condition ([Supplementary Fig. S2A](#)). Because P0 samples included one additional replicate, ISoLDE was able to call genes 'AI' with a bias as low as 0.2 for the P0 condition ([Fig. 1A](#)). This illustrates the improved ability of ISoLDE to call genes 'AI' with only one additional replicate.

As the final decision is made on individual permutations at the gene level, genes with the same S_g value may be called differently. However, ISoLDE called only few genes differently from their neighbors; for example, at E13.5 ([Supplementary Fig. S2A](#)), *Igf2* was located between the paternally biased *Impact* and *Slc38a4* genes and was called 'UN' because of the high variability of the paternal read numbers (range, 25–795; MAD = 294) while the maternal reads were less variable (range, 13–64; MAD = 10.5). ISoLDE eliminated such ambiguous genes, and the flagging process highlighted those situations.

3.1.2 The predefined thresholds version of ISoLDE gave results comparable to the resampling-based version

The predefined thresholds and the resampling-based versions of ISoLDE generally agreed, with some noteworthy exceptions illustrating the benefits of the resampling version ([Fig. 1B](#) and [Supplementary Fig. S2B](#)). *Impact* and *Adam23* at P0, and *Impact* and *Asb4* at E13.5 were called 'UN_flag' using the predefined thresholds version due to low bias values and/or high variability ([Fig. 1A](#) and [Supplementary Fig. S2A](#)) whereas they were called 'AI' using the resampling version. *Impact*, *Adam23* and *Asb4* were all previously shown to undergo parental genomic imprinting, which is in favor of the resampling results. On the other hand, *Dlk1*, *Igf2*, *C130071C03Rik* and *Ltv1* were called 'AI' at E13.5 using the predefined thresholds version, whereas they were called 'UN_flag' by the resampling version. *Dlk1* and *Igf2* are bona fide IGs, whereas *C130071C03Rik* and *Ltv1* were not previously reported as imprinted. Very little is known about *C130071C03Rik*, and there is no published data supporting its parental genomic imprinting. Interestingly, *Ltv1* is located between *Plagl1*, an established IG, and *Phactr2*, which was also called 'AI' at E13.5 ([Supplementary Fig. S2A and B](#)) and in the placenta ([Wang et al., 2011](#)). In conclusion, both versions of ISoLDE called 'AI' or flagged known IGs. The more equivocal gene, *C130071C03Rik*, which had few reads and displayed high variability ([Supplementary Fig. S3](#)), was called 'AI' using the predefined thresholds version and flagged using the resampling version, illustrating that the resampling version performed better with the available information, especially for lowly expressed genes.

3.2 Benchmarking of the ISoLDE method

We used simulated datasets to compare ISoLDE specificity and sensitivity to those of other methods. We generated two simulated datasets (see [Supplementary File, Section S1.2](#) for details). The first dataset consisted of 10 882 genes, including 59 imprinted ones, with three replicates in each cross. The second dataset consisted of 11 889 genes, including 68 imprinted ones, with five replicates in each cross. We also used experimental datasets to further compare ISoLDE and other methods ([Table 1](#)). We occasionally did not succeed to match gene IDs between the published article and the corresponding data submitted to GEO. When we did not solve such discrepancies by manual curation and/or by contacting the authors, we discarded the corresponding genes from the comparisons. As the

Table 1. Datasets and methods used for benchmarking of ISoLDE

Reference dataset name	Replicate number in each cross	Statistical method used in the original publication
Bouschet et al. (2017)		
P0 cortex	4-3	ISoLDE
E13.5 cortex	3-3	
Hasin-Brumshtein et al. (2014)	2-2	Beta-binomial (DESeq)
Babak et al. (2008)	4-4	Binomial + bias direction agreement between replicates
Lorenc et al. (2014)		
LIV, HYP	6-3	χ^2 on the sums over replicates + mock reciprocal cross
VNO	4-3	
Perez et al. (2015)		
P8, P60	12-12	Generalized linear model + Bayesian parameters estimation
Bonthuis et al. (2015)		
DRN, ARN	9-9	Generalized linear model + mock reciprocal cross
Liver, muscle	9-9	

actual status of genes was unknown for experimental datasets, we attempted to make a decision on the discordant calls by observing data distribution of individual genes and by comparing the calls to the lists of published IGs.

We organized the benchmarking according to the modeling used by the published methods. We first compared ISoLDE to methods based on Poisson, binomial and beta-binomial modeling and representing a variety of significance testing. For the simulated datasets, we chose the method by [DeVeale et al. \(2012\)](#) that accounted for the most classical approaches based on binomial modeling, χ^2 test and mock reciprocal crosses to correct for false positives and filter for consistency in the bias direction. This method has been used in several other reports ([Babak et al., 2015](#); [Lorenc et al., 2014](#)). We also tested the method by [Pirinen et al. \(2015\)](#), which explored an alternative modeling based on beta-binomial and Bayesian estimation. For the experimental datasets, we used the method by [DeVeale et al.](#) as well as those by [Babak et al. \(2008\)](#) and [Lorenc et al. \(2014\)](#). We also tested the method by [Perez et al.](#), which used Poisson-based modeling and Bayesian estimation.

We then benchmarked ISoLDE against two methods based on negative-binomial modeling and GLM. We tested the method by [Bonthuis et al. \(2015\)](#) on simulated and experimental datasets, and the method by [Hasin-Brumshtein et al. \(2014\)](#) on the corresponding experimental dataset.

Finally, we tested the behavior of ISoLDE with respect to replicate number, and compared it to that of the methods by [Perez et al. \(2015\)](#) and [Bonthuis et al. \(2015\)](#).

When appropriate, we optimized the thresholds used in the compared methods as recommended by authors, independently for each used dataset.

3.2.1 ISoLDE is more conservative than methods based on Poisson, binomial and beta-binomial distributions

3.2.1.1 Comparison of ISoLDE, DeVeale and Pirinen methods on simulated datasets. Table 2 summarizes the results of the analysis of the two simulated datasets by the different methods. The results for

Table 2. Comparison of ISoLDE and other methods on simulated datasets

Simulated Dataset	Method	Se.	Sp.	Calls	True calls				
					'AI'	'BA'			
Three-replicate dataset	ISoLDE	94.9	99.4	'AI'	46	0			
				'UN_flag' consistency	10	66			
				'UN_flag' significance	0	0			
				'UN'	0	40			
				'BA'	3	10 717			
				'AI'	31	21			
	DeVeale	52.5	99.8	'BA'	28	10 802			
		Bonthuis	28.8	100	'AI'	17	0		
					'BA'	42	10 823		
	Pirinen	79.7	98.8	'AI'	47	126			
				'UN'	4	1347			
				'BA'	8	9350			
Five-replicate dataset	ISoLDE	94.1	100.0	'AI'	59	0			
				'UN_flag' consistency	4	0			
				'UN_flag' significance	1	2			
				'UN'	0	6			
				'BA'	4	11 813			
				'AI'	31	27			
	DeVeale	45.6	99.8	'BA'	37	11 494			
				Bonthuis	64.7	100	'AI'	44	0
							'BA'	24	11 821
	Pirinen	77.9	98.5	'AI'	53	172			
				'UN'	8	1694			
				'BA'	7	9955***			

Note: The number of genes in each category, the sensitivity (Se.) and the specificity (Sp.) values for each method are indicated

two additional datasets with 7 and 10 replicates are displayed in [Supplementary Table S2](#).

ISoLDE resulted in five possible calls; allelic imbalance ('AI'), undetermined call with consistency flag ('UN_flag'), undetermined call with significance flag ('UN_flag'), unflagged undetermined call ('UN') and bi-allelically expressed ('BA'). To compute sensitivity and specificity, we considered 'AI' and 'UN_flag' as 'positive' calls and 'UN' and 'BA' as 'negative' calls. The analysis of the simulated three-replicate dataset by ISoLDE resulted in 99.4% sensitivity and a 99% specificity; we obtained only three genes called 'BA' whereas they were imprinted. Moreover, the 10 remaining IGs were undetermined and flagged for consistency. On the other hand, 66 bi-allelically expressed genes were flagged for consistency and 40 other ones were called undetermined. We concluded that, in this application, flags did not provide specific information.

The analysis of the five-replicate dataset by ISoLDE resulted in 94.1% sensitivity and 100% specificity. Only 13 genes were considered as undetermined, 7 of which were flagged, 5 true IGs and only 2 true BA genes. In this case, the flags were informative. ROC curves and other performance indices can be found in [Supplementary Figure S4](#).

Analysis of the same datasets with the DeVeale method resulted in 52.5% sensitivity (28 false negatives) and 99.8% specificity (21 false positives) on the three-replicate dataset, and 45.6% sensitivity (37 false negatives) and 99.8% specificity (27 false positives) on the five-replicate one.

The Pirinen method provided six calls that we grouped in the following way to compare with ISoLDE calls:

- genes called ‘moderately’ or ‘strongly’ biased by the Pirinen method were considered ‘AI’;
- genes called ‘heterogeneous’ across replicates by the Pirinen method were considered ‘UN’;
- genes called ‘unbiased’ by the Pirinen method were considered ‘BA’.

As for ISoLDE, we considered ‘undetermined’ calls as ‘negative’ calls to compute sensitivity and specificity. This method achieved 79.7% sensitivity and 98.7% specificity. The undetermined genes comprised 8 IGs and 1694 BA genes.

This study on simulated data showed that ISoLDE compared favorably to published methods based on binomial modeling/ χ^2 test or beta-binomial modeling/Bayesian estimate.

3.2.1.2 Comparison of ISoLDE and DeVeale method on data from P0 and E13.5 cortex. DeVeale method was also applied to the P0 and E13.5 murine cortex datasets (Figs 1C and 2, Supplementary Figs S5 and S6). One gene called ‘AI’ by the method of DeVeale in the E13.5 dataset has been filtered out by ISoLDE in the P0 and E13.5 samples due to very low expression (see Section 2 for more details about filtering). Similarly, three and two genes called ‘AI’ by ISoLDE in the P0 and E13.5 datasets, respectively, were filtered out by the method of DeVeale. The other genes called ‘AI’ by ISoLDE in both datasets were also called ‘AI’ by the method of DeVeale. Sixteen and nine genes were called ‘AI’ by the method of DeVeale alone in the P0 and E13.5 datasets, respectively (Figs 1C and 2 and Table 3). Among these genes, 3 and 5, respectively, were called ‘UN’ and flagged by ISoLDE; 2 and 1, respectively, were called ‘UN’; 11 and 3, respectively, were called ‘BA’. The corresponding normalized read counts mostly supported ISoLDE’s calls (Supplementary Figs S5 and S6) but for a few exceptions; *Slc38a4* displayed a high variability and low read counts (Supplementary Fig. S5) and was called ‘UN’ by ISoLDE; visual inspection of the data (Supplementary Fig. S5) suggested AI.

We concluded that ISoLDE was more conservative and flagged genes whose expression was close to AI.

3.2.1.3 Comparison of ISoLDE and Babak method on Babak’s dataset. The dataset from Babak *et al.* (2008) contains four replicates of E9.5 mouse embryos from reciprocal crosses between CAST/EiJ and C57BL/6J. Forty genes were called ‘AI’ using a binomial modeling combined with signal consistency filter and mock reciprocal crosses to assess the FDR (Babak *et al.*, 2008). The resampling version of ISoLDE confirmed 24 of these and called two additional ones (Fig. 2 and Supplementary Fig. S7 and Table 3). Among the 16 genes not called ‘AI’ by ISoLDE, 6 were filtered out due to low expression levels, 7 were called ‘UN’, including one flagged for consistency and 3 ‘BA’. Supplementary Figure S8 showed that six of the undetermined genes (*Rian*, *Asb4*, *Neurabin*, *Usp29*, *Peg1* and *Dlk1*) exhibited strain-biased expression as, depending on the gene, either the C57BL or the CAST allele was systematically overexpressed. *Air*, *Meg3* and *Zac1* were biased in only one of the crosses. Finally, *Zdbf2* exhibited inconsistency in bias direction within replicates. These observations were not in favor of AI of these 10 genes.

On the other hand, two genes (*AK090117* and *AK142799*), which were not detected by the Babak method, were called ‘AI’ by ISoLDE with a strong paternal bias (Supplementary Figs S7 and S9).

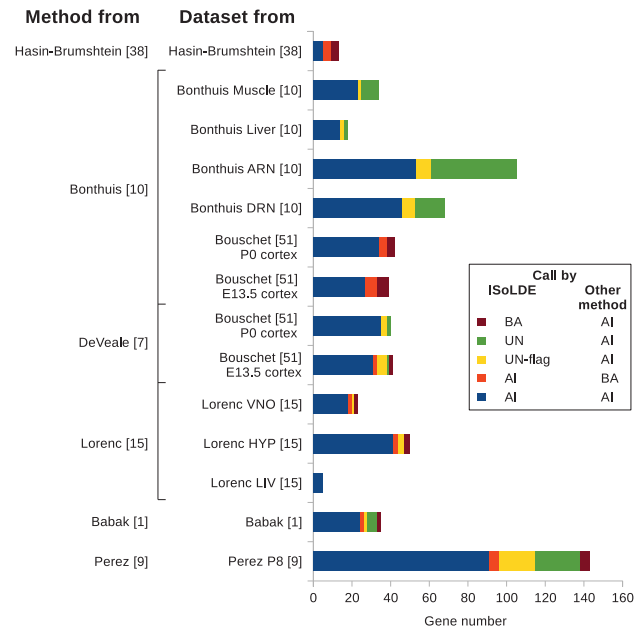


Fig. 2. Comparison of ISoLDE to other published methods on a series of experimental datasets. The indicated datasets were analyzed in parallel by ISoLDE and by one of the indicated published methods. For each method comparison, the genes called parentally biased (‘AI’) by either method were classified according to the calls of the two methods. ‘AI’, allelic imbalance. ‘BA’, bi-allelically expressed. ‘UN’, undetermined. ‘UN-flag’, undetermined and flagged by ISoLDE. The AI-AI pair in legend denotes genes called parentally biased by both ISoLDE and the compared method. The AI-BA and BA-AI pairs in legend denote genes called parentally biased by one method and bi-allelically expressed by the other. The UN-AI pair in legend, respectively UN-flag-AI, denotes genes whose parental bias could not be determined, respectively was flagged, by ISoLDE and was called parentally biased by the other method

A previous study supported imprinting of *AK142799* (Mould *et al.*, 2013). The marked difference between maternal and paternal normalized read counts of these two genes (Supplementary Fig. S9) suggested that the discrepancy between ISoLDE and the method of Babak might be due to inaccurate gene ID, as observed in other studies, rather than to methodological issues.

We concluded that ISoLDE was efficient at eliminating false positives.

3.2.1.4 Comparison of ISoLDE and Perez method on Perez’s datasets. As previously mentioned, Perez *et al.* proposed a method based on a complex generalized linear model assuming a Poisson distribution of variability (Perez *et al.*, 2015). This method uses MMSEQ to estimate the real expression level of each gene along with the variability. P8 and P60 cerebella of 24 mice derived from the reciprocal crosses between CAST/EiJ and C57BL/6J strains were RNA sequenced. As raw data were not available, we ran ISoLDE on MMSEQ average expression outputs (Supplementary Fig. S10).

In the P8 cerebella, 91 out of 143 genes called ‘AI’ by Perez were also called ‘AI’ by ISoLDE. Among the 52 discordant genes, 19 were flagged by ISoLDE including 15 significance flags (Fig. 2 and Supplementary Fig. S11, Table 3). Ten genes called AI by Perez were called bi-allelically expressed by ISoLDE (Fig. 2); the corresponding maternal and paternal numbers of reads did not support the calls by Perez method (Supplementary Fig. S12).

Five genes (*FR0149454*, *NM 009225.2-13*, *NM 011241.4-55*, *NM 013762.2-23* and *NM 027275.3-3*) were considered biased by

Table 3. Comparison of ISoLDE and other methods on experimental datasets

Dataset	Method	Calls	IsoLDE calls				
			'AI'	'UN_flag' (consist.)	'UN_flag' (signif.)	'UN'	'BA'
Perez P8	Perez	'AI'	91	4	15	23	10
		'BA'	5	0	141	2425	29 644
Babak	Babak	'AI'	24	1	1	5	3
		'BA'	2	1	2	8	628
Lorenc	Lorenc	'AI'	5	0	0	0	0
LIV		'BA'	0	3	1	8	7402
Lorenc	Lorenc	'AI'	41	3	0	0	5
HYP		'BA'	3	2	3	7	10 174
Lorenc	Lorenc	'AI'	18	1	0	0	0
VNO		'BA'	2	9	1	65	10 663
Bouschet	DeVeale	'AI'	35	3	0	2	11
P0 cortex		'BA'	0	13	4	52	10 685
Bouschet	Bonthuis	'AI'	34	0	0	0	0
P0 cortex		'BA'	4	16	4	54	10 696
Bouschet	DeVeale	'AI'	31	5	0	1	3
E13.5 cortex		'BA'	2	31	2	115	10 144
Bouschet	Bonthuis	'AI'	27	0	0	0	0
E13.5 cortex		'BA'	6	36	2	116	10 147
Hasin-Bru.	Hasin-Bru	'AI'	5	0	0	0	0
		'BA'	4	56	0	89	5690
Bonthuis	Bonthuis	'AI'	46	5	2	15	54
DRN		'BA'	0	0	0	83	12 496
Bonthuis	Bonthuis	'AI'	53	7	1	44	211
ARN		'BA'	0	0	0	85	12 547
Bonthuis	Bonthuis	'AI'	14	1	1	2	5
Liver		'BA'	0	0	1	36	10 776
Bonthuis	Bonthuis	'AI'	23	1	1	9	47
Muscle		'BA'	0	0	1	68	10 261

Note: The number of genes in each category is indicated

ISoLDE but not by the Perez method (Fig. 2); these discrepancies might be related to misidentification of gene IDs because the same transcript was sometimes linked to different IDs.

The comparison of ISoLDE and Perez methods on P60 data gave similar results (not shown).

These results showed that the sophisticated modeling of data by the Perez method may still lead to false positives. A possible reason for this observation was the high number of replicates, 12 per replicate; parametric modeling implies a rapid increase in confidence with sample size. The combination of this phenomenon with the potential variability underestimation due to modeling may lead to false positives.

3.2.1.5 Comparison of ISoLDE and Lorenc method on Lorenc's dataset. Three tissues were investigated: VNO, HYP and LIV (Lorenc et al., 2014). They were collected from reciprocal crosses of WSB and PWD strains with three to six biological replicates in each cross. The method of Lorenc used a χ^2 test, which involved the normal approximation of the binomial distribution, followed by signal consistency filter. Mock reciprocal crosses were used to control FDR in a way similar to the one used by DeVeale et al. (2012).

Lorenc et al. first identified significant SNPs and then annotated transcripts called 'AI' using *Mus musculus* GRCm38.68 (ENSEMBL) and named UCSC genes from the mm10 reference

annotation. As ISoLDE worked at the transcript/gene level, we first annotated SNPs using ANNOVAR and mm10 UCSC refFlat file and then summed all allele-specific sense read bases across a gene before running ISoLDE. As we could not use the annotation database of Lorenc, discrepancies between gene symbols might have occurred.

We analyzed all three datasets with ISoLDE (Supplementary Figs S13–S15). The vast majority of 'AI' calls were identical between ISoLDE and Lorenc method (Fig. 2 and Table 3). Among the seven genes called 'AI' in at least one tissue by ISoLDE only (Supplementary Figs S17 and S19), six were previously called 'AI' by others (Babak et al., 2008; Chiesa et al., 2012; Perez et al., 2015; Silva-Santiago et al., 2012; Tremblay et al., 1995; Wu et al., 2012; Xiao et al., 2006). Of note, *DQ267100* and *Snord64* are snoRNAs that might not be annotated in Lorenc's database. *DQ267100* is located in an intron of the maternally expressed *Rian* gene. *Snord64* is located in an intron of the paternally expressed *Snrpn* gene and overlaps *AK139082*, an unclassifiable mRNA from GenBank that was called paternal in HYP by Lorenc. Furthermore, *Kcnq1ot1*, a long noncoding RNA, might not be present in Lorenc's database. Hence, misannotation may be the source of most of the observed discrepancies.

Genes called 'AI' by the Lorenc method only displayed high variability with regards to bias (*Adam23*, *Bcl2l1*, *Ppp1r9a*, *Asb4*, *Igf2*, *Klhd10* and *Wars*), discrepancies in the bias direction depending on the replicate (*Ddc*, Supplementary Fig. S16) or allelic bias in only one cross (*Nnat*, Supplementary Fig. S18). Four of those genes were called 'UN' and flagged by ISoLDE with consistency across replicates in bias direction. ISoLDE called *Wars* 'BA' whereas the visual inspection of the read counts showed a weak and constant bias in favor of paternal reads (Supplementary Fig. S16).

In short, the inspection of the data challenged some of the calls made by the Lorenc method. The ability of ISoLDE to learn the variability of each gene from the data made it more efficient to remove genes that had low bias at the expense of the sensitivity for those genes displaying very weak bias.

In conclusion, the widely used modeling of RNA-seq data using binomial, beta-binomial or Poisson distributions resulted in underestimated variability and led to a high number of false positive 'AI' calls, which was the cause of the controversy over the number of IGs.

3.2.2 ISoLDE is more sensitive than methods based on negative binomial distribution

The negative binomial distribution was implemented in RNA-seq data analysis to correct the strong underestimation of variability in the more classical modeling using the binomial or Poisson distributions.

3.2.2.1 Comparison of ISoLDE and Bonthuis method on simulated datasets. The method by Bonthuis et al. (2015) is based on the edgeR package (Robinson et al., 2010; Zhou et al., 2014) and uses a negative binomial modeling combined with GLM for significance testing. The results obtained with this method on the three- and five-replicate datasets are given in Table 2; 28.8% sensitivity and 100% specificity on the three-replicate dataset; 64.7% sensitivity and 100% specificity on the five-replicate dataset. Compared to ISoLDE, the method by Bonthuis et al. displayed a marginally improved specificity (100% versus 99.39% or 99.98%) at the expense of sensitivity (28.8% versus 94.9%, three-replicate dataset; 64.7% versus 94.1%, five-replicate dataset).

3.2.2.2 Comparison of ISoLDE and Bonthuis method on data from P0 and E13.5 cortex. We applied the method by Bonthuis et al. to the E13.5 and P0 cortices datasets (Figs 1D and 2). In both datasets, all genes called ‘AI’ by the Bonthuis method were also called ‘AI’ by ISoLDE (Table 3). Four, respectively six, additional genes were called ‘AI’ by ISoLDE only in P0 (*Zdbf2*, *Copg2*, *Impact* and *Adam23*), respectively E13.5 (*Snrpn*, *Adam23*, *Phactr2*, *Slc38a4*, *Impact* and *Asb4*). Those genes have already been shown to be imprinted and their normalized read counts (Supplementary Figs S22 and S23) supported this conclusion. Hence, those results were consistent with the simulation results and confirmed the low sensitivity of the Bonthuis method. In our cortex datasets, the number of replicates was low (3–4 per cross), which caused edgeR and DESeq to be restrictive. In this context, ISoLDE was still able to make the most of the available data to highlight genes whose specific characteristics allow for positive decisions.

3.2.2.3 Comparison of ISoLDE and Hasin–Brumshtein method on Hasin–Brumshtein’s dataset. This dataset (Hasin–Brumshtein et al., 2014) comprised two replicate experiments of mouse adipose tissue from reciprocal crosses of C57BL/6J and DBA/2J strains. It used negative binomial modeling of data combined with GLM as implemented in the DESeq package (Anders and Huber, 2010). Because this dataset included two replicates only, we used the predefined thresholds version of ISoLDE.

Among the nine genes called ‘AI’ by ISoLDE (Supplementary Fig. S20), five genes (*Meg3*, *Zim1*, *Ndn*, *Trappc9* and *Sgce*) were also identified by Hasin–Brumshtein (Supplementary Fig. S20 and Fig. 2, Table 2). Four genes (*H13*, *Snrpn*, *Snurf* and *Hoxb9*) were called ‘AI’ by ISoLDE only. *H13*, *Snrpn* and *Snurf* are known IGs and inspection of maternal and paternal read counts of *Hoxb9* (Supplementary Fig. S21) supported the ‘AI’ call despite low read counts. ISoLDE called significantly more AI genes, with no obvious signs of false-positives.

3.2.3 ISoLDE is more robust to variations in replicate number

The comparison described above suggested that Perez method displayed overconfidence when the number of replicates increased. To test how ISoLDE behaved with respect to replicate number, we tested ISoLDE on the Bonthuis datasets that had nine biological replicates for each reciprocal cross in four different tissues: ARN, DRN, muscle and liver (Figs 2 and 3).

Surprisingly, this comparison revealed many probable false positives among the Bonthuis calls. Examples of discordant calls are shown in Figure 3C. *Ddc* was called ‘BA’ by ISoLDE and ‘AI’ by the Bonthuis method. Inspection of the read counts supported a strain bias; the average number of paternal reads was marginally lower than the average number of maternal reads, but the same number of samples displayed a maternal or a paternal bias. To confirm this, we ran ISoLDE in a strain bias mode; among the 132 genes called ‘AI’ by Bonthuis, 30, including *Ddc*, were called with a strain bias and no parental bias by ISoLDE (Supplementary Fig. S24). Other kinds of potential false positives were also identified. *Gm20715* was called ‘UN’ by ISoLDE; 11 replicates exhibited a maternal bias, and 7 replicates exhibited a paternal bias. Finally, *Cobl* was flagged for consistency by ISoLDE, and this conservative decision was reasonable given the low global difference between the maternal and paternal read counts and the observed variability among the replicates.

The comparison of ISoLDE with Perez or Bonthuis method confirmed the trend of parametric methods toward overfitting, especially when the number of replicates increased. In contrast, ISoLDE

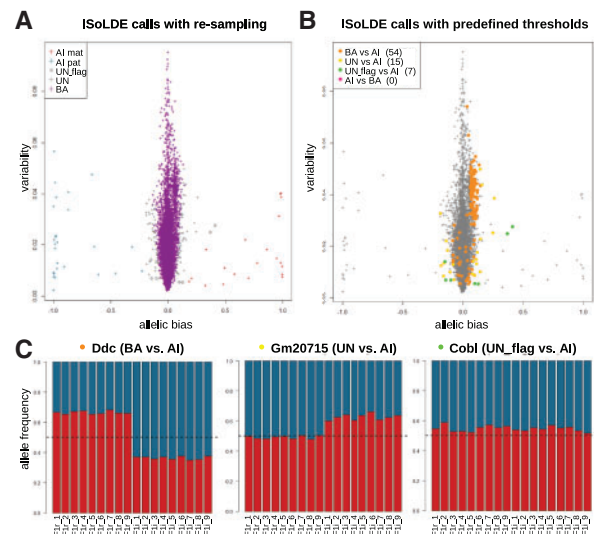


Fig. 3. Comparison of the ISoLDE and Bonthuis methods using the Bonthuis DRN dataset. (A) Output of the resampling version of ISoLDE on the Bonthuis dataset. For each gene, the variability (denominator value of SG statistics) is plotted against the allelic bias (numerator value of the S_g statistics). Violet crosses correspond to bi-allelically expressed genes (BA). Red and blue crosses correspond to genes called maternally and paternally biased (AI), respectively. Gray crosses label undetermined genes (UN). Gray circled crosses correspond to flagged genes (consistency or significance flag). (B) Discordant ISoLDE and Bonthuis calls. The Bonthuis dataset was analyzed using the resampling version of ISoLDE. Genes that were called AI by only one of the two methods are indicated by colored circles on the ISoLDE output. See the figure for color code (first call is ISoLDE call). Numbers in brackets are the number of genes in each category. (C) The fractions of maternal and paternal reads in each replicate sample of the Bonthuis dataset are displayed in red and blue, respectively. The black dashed line corresponds to an unbiased parental origin, i.e. a bi-allelic expression. A gene from each type of discordance in this dataset has been chosen. F1r and F1i in the replicate names refer to the reciprocal crosses

appeared to exhibit a more constant behavior whatever the number of replicates.

4 Discussion

To the best of our knowledge, statistical analysis of AI relied only on model-based methods. Classical distributions such as Poisson, binomial, beta-binomial and negative binomial have been used to model RNA-seq data. In contrast, ISoLDE relies on a nonparametric framework.

We designed a specific statistic to take into account both AI and the characteristics of the RNA-seq data, i.e. discrete values, variable sequencing depth and replicate variability. Its gene-by-gene distribution was learned directly from data through resampling. The comparison to published methods indicated that ISoLDE was more conservative than methods based on binomial, beta-binomial or Poisson distributions, whose intrinsic risk of underestimation of variability leads to false positives. ISoLDE also circumvents the limitations of methods based on negative binomial modeling, such as edgeR or DESeq. These methods have been specifically designed to avoid variability underestimation with few replicates and rapidly increase confidence with replicate number. Consequently, their sensitivity is adequate when many replicates are available but is limited for few replicates. On the other hand, their good specificity with few

replicates decreases with bigger datasets. The data-driven framework of ISoLDE seems to be more flexible and automatically adapts to the sample size.

Another important benefit of ISoLDE is its capacity to provide more informative results. Biological variability often prevents drawing definitive conclusions. The expression of some genes may be neither clearly allelically imbalanced nor clearly bi-allelic. ISoLDE classifies genes into three categories: allelically biased, bi-allelically expressed and undetermined. Moreover, ISoLDE flags two subsets of undetermined genes according to the significance or consistency in their bias. Finally, the statistics of all genes are illustrated in an informative, user-friendly graph so that the investigator's attention can be focused on genes having the most interesting characteristics depending on his/her needs. In the present study, we focused on gene level analyses but ISoLDE performs analysis at any level, gene, transcript or SNP.

Learning from the data is an advantage that comes with a limitation; it requires at least three replicates of both reciprocal crosses. To alleviate this limitation, we designed an option in ISoLDE that uses precomputed thresholds. Whenever possible, we recommend using the resampling option to compute gene-specific thresholds.

We applied ISoLDE to identify genes displaying parent-of-origin-dependent AI, which were therefore potentially imprinted. We did not find any new parentally biased gene in the cerebral cortex samples that we recently generated. We also found very few new parentally biased genes in other published RNA-seq data. Our conclusion is that the total number of IGs is likely close to the most conservative estimates, i.e. ~150–200. If new allelically biased genes are to be discovered, they should be searched for specific cell types, tissues and/or at precise developmental stages.

We designed ISoLDE with parent-of-origin-dependent AI in mind, but ISoLDE is able to detect AI whatever the origin of the bias. In the accompanying Bioconductor package, two options are available; parental and strain bias. More generally, the user can provide any two-level variable defining two conditions between which AI is to be compared.

ISoLDE is limited by the need for reciprocal crosses. This experimental design is becoming standard in model organisms but is obviously unworkable in human studies. We plan to develop a novel nonparametric method to analyze AI in human RNA-seq data in a future version of ISoLDE. This method will also be helpful to identify random AI, in particular in single-cell RNA-seq data.

Acknowledgements

We thank Dr Tomas Babak for advice on the analysis of the parent-of-origin-dependent gene expression using RNA-seq, and Drs Thomas Keane and David Adams (Sanger Institute) for providing the 129P2 genome sequence.

Funding

The authors acknowledge the financial support from CNRS, INSERM, Université de Montpellier and the Agence Nationale pour la Recherche (EpiNet, ANR-07-BLAN-52). MGX acknowledges the financial support from France Génomique National Infrastructure, funded as part of the 'Investissement d'Avenir' program managed by the Agence Nationale pour la Recherche (ANR-10-INBS-09).

Conflict of Interest: none declared.

References

- Andergassen,D. *et al.* (2015) Allelome.PRO, a pipeline to define allele-specific genomic features from high-throughput sequencing data. *Nucleic Acids Res.*, **43**, e146.
- Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
- Babak,T. *et al.* (2015) Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat. Genet.*, **47**, 544–549.
- Babak,T. *et al.* (2008) Global survey of genomic imprinting by transcriptome sequencing. *Curr. Biol.*, **18**, 1735–1741.
- Bonthuis,P.J. *et al.* (2015) Noncanonical genomic imprinting effects in offspring. *Cell Rep.*, **12**, 979–991.
- Bouschet,T. *et al.* (2017) In vitro corticogenesis from embryonic stem cells recapitulates the in vivo epigenetic control of imprinted gene expression. *Cereb. Cortex.*, **27**, 2418–2433.
- Chiesa,N. *et al.* (2012) The KCNQ1OT1 imprinting control region and non-coding RNA: new properties derived from the study of Beckwith–Wiedemann syndrome and Silver–Russell syndrome cases. *Hum. Mol. Genet.*, **21**, 10–25.
- DeVeale,B. *et al.* (2012) Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. *PLoS Genet.*, **8**, e1002600.
- Dudoit,S. *et al.* (2004) Multiple testing. Part I. Single-step procedures for control of general type I error rates. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article 13.
- Dudoit,S. and van der Laan,M.J. (2008) *Multiple Testing Procedures with Applications to Genomics*. Springer-Verlag, New York.
- Gregg,C. *et al.* (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*, **329**, 643–648.
- Hampel,F.R. (1974) The influence curve and its role in robust estimation. *J. Am. Stat. Assoc.*, **69**, 383–393.
- Hasin-Brumshtein,Y. *et al.* (2014) Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics*, **15**, 471.
- Kelsey,G. and Bartolomei,M.S. (2012) Imprinted genes ... and the number is? *PLoS Genet.*, **8**, e1002601.
- Keverne,E.B. (2010) Neuroscience: a mine of imprinted genes. *Nature*, **466**, 823–824.
- van der Laan,M.J. *et al.* (2004) Multiple testing. Part II. Step-down procedures for control of the family-wise error rate. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article 14.
- Lorenc,A. *et al.* (2014) Genetic differentiation of hypothalamus parentally biased transcripts in populations of the house mouse implicate the Prader–Willi syndrome imprinted region as a possible source of behavioral divergence. *Mol. Biol. Evol.*, **31**, 3240–3249.
- Mould,A.W. *et al.* (2013) Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation. *Epigenetics Chromatin*, **6**, 19.
- Perez,J.D. *et al.* (2015) Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *Elife*, **4**, e07860.
- Pirinen,M. *et al.* (2015) Assessing allele-specific expression across multiple tissues from RNA-seq read data. *Bioinformatics*, **31**, 2497–2504.
- Pollard,K.S. and van der Laan,M.J. (2004) Choice of a null distribution in resampling-based multiple testing. *J. Stat. Plan. Inference*, **125**, 85–100.
- Robinson,M.D. *et al.* (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
- Rozowsky,J. *et al.* (2011) AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Mol. Syst. Biol.*, **7**, 522.
- Silva-Santiago,S.C. *et al.* (2012) The human imprintome v1.0: over 120 imprinted genes in the human genome impose a major review on previous censuses. In: *2012 IEEE International Conference on Bioinformatics and Biomedicine*, pp. 1–5. Philadelphia, USA.
- Smith,R.M. *et al.* (2013) Whole transcriptome RNA-Seq allelic expression in human brain. *BMC Genomics*, **14**, 571.
- Takada,T. *et al.* (2013) The ancestor of extant Japanese fancy mice contributed to the mosaic genomes of classical inbred strains. *Genome Res.*, **23**, 1329–1338.
- Tollkuhn,J. *et al.* (2010) A custody battle for the mind: evidence for extensive imprinting in the brain. *Neuron*, **67**, 359–362.

- Tran,D.A. *et al.* (2014) Characterization of the imprinting signature of mouse embryo fibroblasts by RNA deep sequencing. *Nucleic Acids Res.*, **42**, 1772–1783.
- Tremblay,K.D. *et al.* (1995) A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. *Nat. Genet.*, **9**, 407–413.
- Wang,X. *et al.* (2011) A survey for novel imprinted genes in the mouse placenta by mRNA-seq. *Genetics*, **189**, 109–122.
- Wang,X. and Clark,A.G. (2014) Using next-generation RNA sequencing to identify imprinted genes. *Heredity*, **113**, 156–166.
- Wilkinson,L.S. (2010) Developmental biology. Which parental gene gets the upper hand? *Science*, **329**, 636–637.
- Wu,M.-Y. *et al.* (2012) An unexpected function of the Prader–Willi syndrome imprinting center in maternal imprinting in mice. *PLoS One*, **7**, e34348.
- Xiao,Y. *et al.* (2006) Characterization of three novel imprinted snoRNAs from mouse Irm gene. *Biochem. Biophys. Res. Commun.*, **340**, 1217–1223.
- Zhou,X. *et al.* (2014) Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Res.*, **42**, e91.
- Zou,F. *et al.* (2014) A novel statistical approach for jointly analyzing RNA-Seq data from F1 reciprocal crosses and inbred lines. *Genetics*, **197**, 389–399.