Allele-specific expression reveals interactions between genetic variation and environment

David A Knowles1,2, Joe R Davis1, Hilary Edgington3,4, Anil Raj1, Marie-Julie Favé3,5, Xiaowei Zhu6, James B Potash7, Myrna M Weissman8, Jianxin Shi9, Douglas F Levinson6, Philip Awadalla3–5, Sara Mostafavi10, Stephen B Montgomery1,11 & Alexis Battle12

Identifying interactions between genetics and the environment (GxE) remains challenging. We have developed EAGLE, a hierarchical Bayesian model for identifying GxE interactions based on associations between environmental variables and allele-specific expression. Combining whole-blood RNA-seq with extensive environmental annotations collected from 922 human individuals, we identified 35 GxE interactions, compared with only four using standard GxE interaction testing. EAGLE provides new opportunities for researchers to identify GxE interactions using functional genomic data.

Phenotypic variation results from the combined effect of environment and individual genetic background. Many environmental and behavioral influences have been shown to substantially affect human disease risk, and GxE interactions have been shown to be pervasive in model organisms1. However, the prevalence and importance of GxE interactions in human health is not well characterized, and identifying associations on a large scale in human populations has been both statistically and experimentally challenging2. Targeted experimental approaches are not always practical, and the detection of GxE interactions from genome-wide data faces such challenges as small genetic effect sizes for most complex traits and high multiple-hypothesis-testing burden.

In this study, we analyzed GxE interactions in the context of transcriptomic phenotypic traits—these traits can mediate disease risk, and the effects of genetic variation on gene expression are large enough for well-powered, reproducible, genome-wide detection of expression quantitative trait loci (eQTLs), even in modestly sized cohorts3,4. Gene expression can also reveal the impact of environmental factors5,6; and recently, in vitro immune stimulation has been used to detect hundreds of GxE interaction effects in human monocytes7 and dendritic cells8,9. The presence of extensive GxE interactions that affect the transcriptome, agnostic to the specific environment, is further supported by the presence of variance eQTLs10 and allele-specific expression (ASE)11 in monozygotic and dizygotic twins.

To improve power to discover GxE interactions, we developed environment-ase through generalized linear modeling (EAGLE), a novel method to test for GxE interactions using ASE (Supplementary Software and https://github.com/davidknowles/eagle). Intuitively, observing that allelic imbalance of a gene associates with a particular environmental factor suggests that there is a cis-regulatory effect whose impact on expression is modulated by that environment. For example, an environmentally responsive transcription factor binding to one allele better than the other allele (Fig. 1) would result in allelic imbalance of the target gene in that environmental context. By comparing two alleles within the same sample, ASE provides an ‘internally matched’ measure that inherently provides improved control for batch effects and other forms of confounding technical variation (Supplementary Fig. 1). EAGLE uses a binomial generalized linear mixed model (GLMM, Supplementary Fig. 1) and predicts the relative number of RNA-seq reads from each allele at exonic, heterozygous loci under different environmental conditions. EAGLE directly models allelic read counts, which we and others2,11 have found display extrabinomial variation. EAGLE estimates a per-locus overdispersion parameter (random effect variance) that accounts for both technical overdispersion (e.g., from PCR amplification) and extrinsic variation between individuals. Statistical power is shared across loci by learning a genome-wide prior on these variance parameters. We controlled for known cis-eQTL by including heterozygosity of the lead SNP as a covariate. EAGLE can also be used to identify associations between allelic imbalance and other factors other factors, such as genetic variants (Supplementary Fig. 2).

A naïve approach to associate an environmental factor with ASE is to calculate Spearman correlation with a standard definition of allelic imbalance, \(|y/n - 0.5|\), where \(y\) and \(n\) are the alternative and total counts, respectively. However, using a simulation study we have shown (Supplementary Note 2) that, by accounting for binomial sampling variance, EAGLE’s direct modeling of allelic read counts improved power (Supplementary Fig. 3) and reduced false positives (Supplementary Fig. 4). A binomial generalized linear model also failed to account for overdispersion, leading to overinflated

1Department of Genetics, Stanford University School of Medicine, Stanford, California, USA. 2Department of Radiology, Stanford University School of Medicine, Stanford, California, USA. 3Ontario Institute for Cancer Research, Toronto, Ontario, Canada. 4University of Toronto, Toronto, Ontario, Canada. 5CHU Sainte Justine, Montreal, Quebec, Canada. 6Department of Psychiatry, Stanford University School of Medicine, Stanford, California, USA. 7Department of Psychiatry, Columbia University and New York State Psychiatric Institute, New York, New York, USA. 8Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, USA. 9Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada. 10Department of Pathology, Stanford University School of Medicine, Stanford, California, USA. 11Department of Computer Science, Johns Hopkins University, Baltimore, Maryland, USA. Correspondence should be addressed to A.B. (sjbattle@cs.jhu.edu) or S.B.M. (smontgom@stanford.edu).

RECEIVED 15 August 2016; ACCEPTED 20 April 2017; PUBLISHED ONLINE 22 May 2017; DOI:10.1038/NMETH.4298
EAGLE associates allele-specific expression (ASE) with environmental covariates to detect GxE interactions. (a) Allelic imbalance can be driven by allele-specific binding of an environmentally responsive transcription factor (TF). (b) Relative to interaction QTL testing, using ASE increases power in the DGN cohort across 30 environmental variables. Interaction testing was performed on SNPs within 200 kb of each gene, and it was followed by Bonferroni correction. EAGLE provides an internally controlled test and integrates across the cis-regulatory landscape of a gene.

As a baseline, we mapped GxE interactions on total expression using a standard linear model interaction test (see Online Methods). EAGLE showed much greater power to detect GxE interactions than standard interaction QTL testing showed (Fig. 1b). In addition, using Bonferroni correction across the SNPs tested per gene (since there is no appropriate permutation strategy for interaction testing) followed by controlling the FDR at 10%, we found only four associations across all 30 environmental factors, compared with 35 discovered with EAGLE on the same set of tested genes.

We found enrichment of EAGLE associations in relevant pathways, transcription factor target sets, and trans-eQTL networks (Supplementary Notes 3–5 and Supplementary Fig. 9).

EAGLE detects GxE interactions missed by standard interaction QTL testing. (a) Blood-pressure medication modulates regulation of NPRL3, involved in fluid homeostasis. (b) Smoking interacts with regulation of IL10RA. N, no; Y, yes. (c–e) Using standard interaction QTL testing as a second phase within EAGLE hits, we detect rs685419 as a promising candidate variant for smoking’s association with IL10RA. The SNP lies 4 Mb from the TSS in a conserved region corresponding to an enhancer in CD14+ primary cells.

**Figure 1** | EAGLE associates allele-specific expression (ASE) with environmental covariates to detect GxE interactions. (a) Allelic imbalance can be driven by allele-specific binding of an environmentally responsive transcription factor (TF). (b) Relative to interaction QTL testing, using ASE increases power in the DGN cohort across 30 environmental variables. Interaction testing was performed on SNPs within 200 kb of each gene, and it was followed by Bonferroni correction. EAGLE provides an internally controlled test and integrates across the cis-regulatory landscape of a gene.

**Figure 2** | EAGLE detects GxE interactions missed by standard interaction QTL testing. (a) Blood-pressure medication modulates regulation of NPRL3, involved in fluid homeostasis. (b) Smoking interacts with regulation of IL10RA. N, no; Y, yes. (c–e) Using standard interaction QTL testing as a second phase within EAGLE hits, we detect rs685419 as a promising candidate variant for smoking’s association with IL10RA. The SNP lies 4 Mb from the TSS in a conserved region corresponding to an enhancer in CD14+ primary cells.
We investigated the validity of EAGLE associations by analyzing replication both within DGN and between independent studies. First, we split the DGN cohort into equal-sized discovery and replication sets while approximately matching sex and age. The proportion of EAGLE associations that replicate \((P < 0.05)\) increases with the stringency of the discovery \(P\)-value threshold, which is not the case for standard interaction QTL associations (Supplementary Fig. 10a).

Second, we checked for replication of EAGLE associations from DGN in 723 native French–Canadians from the CARTaGENE whole-blood cohort\(^{18,19}\). Despite differences in population, recruitment, and recording of environmental factors, we observed replication (Supplementary Fig. 10b); and the strongest pattern was observed for BMI, a measurement that has a quantitative definition and is thus likely to be consistent between the two studies. Ten EAGLE hits from DGN corresponded to environmental factors recorded in both cohorts. Of these, six replicated in CARTaGENE \((P < 0.05)\).

EAGLE’s improved power over standard interaction QTL testing may derive from multiple sources, including the controlled, within-individual nature of our ASE-based test (Supplementary Fig. 1) along with the direct modeling of read counts (Supplementary Fig. 3). Supported by a simulation study where we varied the level of confounding (Supplementary Note 6 and Supplementary Fig. 11), we hypothesize that confounders such as cell-type portion are a key reason why standard interaction QTL testing is underpowered. Further, EAGLE implicitly integrates over the entire cis-regulatory landscape of a gene rather than explicitly testing a specific candidate SNP and so reduces the multiple-hypothesis-testing burden and potentially captures the contribution of multiple regulatory variants.

Since EAGLE does not directly test individual candidate SNPs responsible for the association between environment and ASE, we applied a two-step procedure to find candidate variants driving GxE associations. In step one, EAGLE was used with a lenient FDR of 20% to give a shortlist of 57 GxE associations. In step two, we looked for candidate variants within 1 Mb of the transcription start site (TSS); we used meta-analysis to combine EAGLE with standard interaction testing (see Online Methods). SNPs with too few heterozygous individuals were not testable using EAGLE; for these SNPs we used standard interaction testing (see Online Methods). SNPs with too few dou-

\[ q = 1 \times 10^{-4} \]

| Immune stimulus | Proxy gene | Response gene | Replication in Fatlax et al. |
|-----------------|------------|---------------|-----------------------------|
| P = 0.01        | IPS31A     | ABCA7         |                             |

Figure 3 | EAGLE detects allele-specific effects of direct perturbations and environments measured by ‘proxy’ genes. (a) EAGLE recapitulates GxE interactions discovered using immune stimulation of monocytes in vitro\(^6\). We used genes differentially expressed under immune stimulation in vitro as proxies for the environment (stimulus). The genes detected by EAGLE as being modulated by these environmental proxies replicate in the in vitro data; i.e., they have detectable response QTLs. Network depicts all EAGLE predictions for each stimulus, with replicating interactions highlighted in yellow; each edge is annotated with the tested proxy gene for reference. (b) EAGLE detects allele-specific responses to treatment of rat livers with various toxicants. The strongest association for agonists of the PPAR\(\alpha\) receptor is a known target, Acot1. While total Acot1 expression is upregulated, we find that rats with the alternative C allele at exonic SNP Chr6:108042464 show no response. (c) Genes associated with PPAR\(\alpha\) by EAGLE show enrichment of relevant TF-binding motifs within 5 kb of the TSS.

Table 4. For example, we found that rs1538257, which is the top candidate variant to modulate BMI’s association with LGALS3 expression, is in LD \((R^2 = 0.55)\) with rs2274273, which is associated with LGALS3 protein levels \((P = 2 \times 10^{-188})\). Interestingly, LGALS3 has been shown to have a protective role in inflammation and diabetes induced by obesity in mice\(^{21}\).

We investigated the degree to which EAGLE analyses, conducted within a large cohort, recapitulate GxE interactions discovered in vitro.
The interplay of immune stimulation, gene expression, and genetics has been characterized in several recent in vitro studies\(^5\)\(^6\)\(^7\)\(^8\)\(^9\). We focused on Fairfax et al.\(^7\) because of its large sample size, genome-wide transcriptomic profiling, and choice of interferon-γ (IFN-γ) and lipopolysaccharide (LPS) immune stimulation, corresponding to viral and bacterial immune responses, respectively. Direct measurements of infection are not available for DGN, so we used the expression levels of differentially expressed genes for each stimulus as environmental ‘proxies’. We used 25, 16, and 26 genes for LPS at 2 h, LPS at 24 h, and IFN-γ, respectively, identified to have an absolute log-fold change greater than 4 in the Fairfax et al.\(^7\) data (The duration of IFN-γ was 24 h). We then applied EAGLE genome-wide to find associations between ASE and gene expression levels for each proxy gene. We excluded tests for interactions between proxy genes and allelic balance of genes on the same chromosome, since these associations could represent direct cis-regulation rather than interaction. At 10% FDR (accounting for testing multiple proxy genes per condition), we found 26, 6, and 14 GxE interactions across the proxy genes for LPS at 2 h, LPS at 24 h, and IFN-γ, respectively. Evaluating \(\gamma\)-statistics for the lead eQTL (Supplementary Note 7 and Supplementary Fig. 13), 11/26, 3/6, and 6/14 interactions replicated \((P < 10^{-4})\) for the three stimuli, respectively, in Fairfax et al.\(^7\) (Fig. 3a). We used random sets of nondifferentially expressed proxy genes to generate an empirical null distribution, which provided empirical \(P\) values for the observed replication rates of 0.048, 0.06, and 0.029, respectively, or 0.0017 for the overall replication frequency.

While we developed EAGLE in the context of an observational population-scale RNA-seq cohort, this method is equally applicable to direct perturbation experiments. We applied EAGLE to RNA-seq data from male Rattus norvegicus livers following oral or intravenous exposure to seven different classes of small molecules\(^22\). Since genotypes were unavailable, we called exonic SNPs from RNA-seq (see Online Methods). Despite moderate sample sizes (30 controls and 8–18 treated samples), we detected 442 associations (10% FDR) across the seven classes (Supplementary Fig. 14a). This power likely derives from controlled laboratory conditions, large effects of direct perturbations, and large haplotype blocks in the outbred rats used, where the exonic variant being tested will frequently cosegregate with the causal variant. EAGLE identified 117 associations (10% FDR) for agonists of PPAR\(\alpha\), a well-characterized transcription factor. Examples include the known targets C/EBP\(\gamma\) (Supplementary Note 8) and AcoT1. AcoT1 is significantly upregulated by PPAR\(\alpha\) (Supplementary Fig. 14b), but only for haplotypes with the reference allele at Chr6:108042464 (Fig. 3b). PPAR\(\alpha\)-associated genes showed enrichment of the binding motif for both PPAR\(\alpha\)/\(\gamma\) and the heterodimer with RXR around their TSS \((P < 0.05;\) Fig. 3c and Supplementary Note 9).

Out of 85 known targets of PPAR\(\alpha\)\(^23\)\(^24\) testable by EAGLE, 37 (44%), compared with 10% for other genes, hypergeometric \(P = 3 \times 10^{-7}\) showed evidence of allele-specific response (10% FDR; Supplementary Fig. 14c).

The associations detected by EAGLE indicate that common environmental risk factors—including substance use, exercise, and BMI—interact with individual genetic variation in regulation of gene expression. EAGLE provided a substantial increase in power over standard methods; yet the overall number of associations remained modest, indicating that GxE interaction effects on gene expression are not prevalent with large effect sizes compared with additive effects, or they are obscured by confounders. Additionally, there are allele-specific, cis-regulatory mechanisms other than genetic effects that could potentially explain some of the discovered associations—for example, epigenetic regulation of expression. As RNA-seq becomes increasingly prevalent in human cohort studies, EAGLE will be appropriate to obtain additional power to detect individual differences in response to diverse environmental conditions. More generally, EAGLE is a useful, extensible tool for understanding the combined effects of external stimuli, genetic variation, and cellular networks on regulation of gene expression.

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We would like to thank J. Leek for helpful comments and S. Kersten for providing the graphic from which the PPAR\(\alpha\) network figure was adapted. D.A.K. is supported by NIH U54CA169415. M.-J.F. is supported by a CIHR Neuroinflammation fellowship. P.A. is supported by the Ontario Ministry of Research and Innovation. A.B. and S.B.M. are supported by NIH R01MH101814 and NIH R01HG008150. A.B. is supported by the Searle Scholars Program, NIH R01GM101820, NIH R01GM109905-01, and NIH R01GM120167-010. S.B.M. is supported by the Edward Mallinckrodt Jr. Foundation.

**AUTHOR CONTRIBUTIONS**

D.A.K., S.B.M. and A.B. conceived the project and wrote the manuscript. D.A.K. and A.B. developed the method. D.A.K. implemented the software and performed the main analyses. J.R.D. and A.R. performed additional statistical analyses. X.Z., J.B.P., M.M.W., J.S., S.M. and D.F.L. gave input regarding the analyses. X.Z., J.B.P., M.M.W., J.S., S.M. and D.F.L. gave input regarding the methods.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
ONLINE METHODS

Interaction QTL testing. Total expression for the DGN cohort was quantified as previously described\(^1\), including controlling for known and latent confounders using HCP\(^2\). We quantitively normalize each gene to a standard normal distribution to remove outliers; and we perform standard interaction testing to find GxE interaction effects for the 8,795 genes testable using ASE. For a specific combination of SNP, gene and environment, consider the null model \(H_0\) and alternative model \(H_1\),

\[
H_0: t_i = \beta g_i + \beta e_i + \mu + e_i
\]

\[
H_1: t_i = \beta g_i + \beta e_i + \beta g \times e_i + \mu + e_i
\]

(1)

Where \(t_i\) is normalized total expression for individual \(i\); \(g_i\) is the genotype of the SNP encoded as \(0,1,2\); \(e_i\) is the environmental factor; \(\beta g\), \(\beta e\), and \(\beta g \times e\) are genetic, environment and interaction effect sizes, respectively; and \(\mu\) is an intercept. Under the null the likelihood ratio \(\max_\beta \beta P(t|\beta, H_0)/\max_\beta \beta P(t|\beta, H_1)\) is \(\chi^2\) distributed with one degree of freedom, which allows us to obtain a well-calibrated \(P\) value. We test all SNPs within 200 kb of the TSS (obtained from GENCODE, release 20). Since there is no appropriate permutation strategy for testing interaction terms\(^17\), we were constrained to using Bonferroni correction to obtain an approximate gene-level \(P\) value. The gene-level \(P\) values for a particular environment are then adjusted using the Benjamini–Hochberg procedure to control the FDR at a prespecified level.

Replication cohort. The replication cohort included 723 native French–Canadians, 346 men and 377 women from Montreal (\(n = 369\), Quebec (\(n = 221\)), and Saguenay (\(n = 133\)), all from the CARTiGENE cohort. Whole-blood samples from these individuals were used to perform genotyping on Illumina’s Omni2.5M array and RNA-seq using paired-end libraries on the Illumina HiSeq 2000 platform as previously described\(^19\). Heterozygous sites were filtered to include only exonic sites that have not been shown to exhibit mapping bias\(^25\). Read counts for both alleles were generated using a custom Perl script. Cis-eQTLs within 1 Mb of the TSS were called for 15,632 genes in a subset of the CARTiGENE cohort (\(n = 689\)) using the R package MatrixEQTL. EAGLE was then run on these data as for the DGN cohort.

Allele-specific expression quantification. Tophat2 (ref. 26; v2.1.0) with default settings was used to map reads to hg19 (for DGN) or rn5 (Ensembl RGSC3.4). Samtools\(^27\) mpileup (v1.3) was used to obtain reference and alternative allele counts at known common SNPs. For the rat data, genotype data are not available, so we determine which individuals are heterozygous at each exonic SNP by requiring (i) two reads mapping to both the reference and alternative allele; and (ii) that the alternate base observed in the RNA-seq reads matches the known allele.

EAGLE model. Existing approaches for calling allelic imbalance\(^28,29\) or leveraging allelic signal in molecular QTL mapping\(^12,30\) are unable to test for association between an environmental factor and allelic imbalance. We first present the EAGLE model itself and then motivate the various modeling choices. The null model \(H_0\) is

\[
\min(y_{is}, n_{is} - y_{is}) | \beta, \mu_s, e_{is} \sim \text{Binomial}[n_{is}, \sigma(\beta h_{is} + \mu_s + e_{is})]
\]

and the alternative model \(H_1\) is

\[
\min(y_{is}, n_{is} - y_{is}) | \beta, \mu_s, e_{is} \sim \text{Binomial}[n_{is}, \sigma(\beta h_{is} + \mu_s + e_{is})]
\]

where \(y_i\) is the alternative read count for individual \(i\) at locus \(s\); \(n_{is}\) is the total read count; \(\sigma(x) = 1/(1 + e^{-x})\) is the logistic function; \(h_{is}\) denotes whether the top cis-eQTL is heterozygous; \(\mu_s\) is an intercept term to take into account unexplained allelic imbalance unrelated to the environment (e.g., due to reference mapping bias\(^12,23\)); and \(e_{is}\) is \(\nu \sim N(0, \nu_j)\) is a per individual per locus random effect modeling overdispersion. This model can be derived by assuming the log expression of each allele is linear in the environment and SNP genotype (Supplementary Note 1). The variance itself is given an inverse gamma prior \(IG(a, b)\). We learn the hyperparameters \(a, b\) across all genes.

We expect that environmental effects on ASE are usually mediated by one or more causal cis-regulatory genetic variants, which would often be in linkage disequilibrium with the locus where ASE is measured. However, some responsive individuals may have different causal sites and therefore may exhibit opposite direction of allelic effect. EAGLE gains power by testing just a single association statistic per gene, rather than modeling each possible causal site and incurring a large multiple testing burden; but it therefore cannot assume a consistent direction of allelic effect across the cohort. Additionally, linkage disequilibrium may be weak, especially for more distal elements. The EAGLE model is applicable in settings where causal sites vary between individual and also handles unphased data. We model the absolute deviation from allelic balance by considering \(\min(y_{is}, n_{is} - y_{is})\) rather than the minor allele count \(y_{is}\) itself. This is analogous to using

\[
\frac{y_{is} - 1}{n_{is} - 2}
\]

as a quantitative measure of allelic imbalance but maintains the count nature of the data. We also experimented with introducing explicit auxiliary ‘flipping’ variables to provide implicit phasing, but we found this was susceptible to overfitting.

Accounting for cis-regulation. Standard cis-eQTL analysis allowed us to identify proximal genetic variants associated with the expression of each gene. These variants often explain a significant proportion of observed ASE. To account for this, we add a dependence on \(h_{is}\), an indicator of whether the top cis-eQTL for the gene containing locus \(s\) is heterozygous in individual \(i\). Additionally, in some cases one of the known cis-eQTLs could be the variant through which the environment influences the observed ASE, which we model by including an interaction term \(h_{is} e_{is}\) (Supplementary Note 10). We approximately integrate over the random effects \(e_{is}\) and per locus variance \(\nu_j\) using conjugate variational message passing\(^31\) while optimizing the coefficients \(\beta\) and hyperparameters \(a, b\) (Supplementary Note 11).

Parameter estimation and inference. Holding the overdispersion hyperparameters \(a, b\) fixed, we fit both the alternative and null models at each locus and use the variational lower bound as an approximation to the true marginal likelihood for each model, allowing us to calculate an approximate likelihood ratio. It is not obvious
that the usual asymptotic theory should hold here since (i) our data are not normally distributed, (ii) we only have an approximation of the true likelihood, and (iii) our model incorporates random effect terms. To investigate this we performed permutation experiments, using the conveniently valid strategy of separately permuting the individuals heterozygous or homozygous for the top cis-SNP. These experiments show that our approximate likelihood ratios do in fact follow the asymptotic $\chi^2$ distribution quite closely, while being slightly conservative (Supplementary Fig. 8). Therefore we choose to use the nominal likelihood ratio test $P$ values, avoiding having to run computationally expensive permutation analysis for every tested association.

Software. EAGLE was developed in C++ and R 3.1.2 using RcppEigen and is available as an R package at https://github.com/davidaknowles/eagle.

Data availability statement. Genotype, raw RNA-seq, quantified expression, covariates and environmental data for the DGN cohort are available by application through the NIMH Center for Collaborative Genomic Studies on Mental Disorders. Instructions for requesting access to data can be found at https://www.nimhgenetics.org/access_data_biomaterial.php, and inquiries should reference the Depression Genes and Networks study (D. Levinson, PI). Requests for replication cohort (CARTaGENE) data should be submitted to access@cartagene.qc.ca. For the rat toxicity study see SRA, SRP039021; GEO, GSE55347.

24. Mostafavi, S. et al. PLoS One 8, e68141 (2013).
25. Panousis, N.I., Gutierrez-Arcelus, M., Dermitzakis, E.T. & Lappalainen, T. Genome Biol. 15, 467 (2014).
26. Kim, D. et al. Genome Biol. 14, R36 (2013).
27. Li, H. et al. Bioinformatics 25, 2078–2079 (2009).
28. Harvey, C.T. et al. Bioinformatics 31, 1235–1242 (2015).
29. Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E. & Lappalainen, T. Genome Biol. 16, 195 (2015).
30. Kumasaka, N., Knights, A.J. & Gaffney, D.J. Nat. Genet. 48, 206–213 (2016).
31. Knowles, D.A. & Minka, T. Adv. Neural Inf. Process. Syst. 24, 1701–1709 (2011).