Genome Medicine

RESEARCH HIGHLIGHT

Single cell expression quantitative trait loci and complex traits

Enrico Petretto*

Abstract

The recently developed ability to quantify mRNA abundance and noise in single cells has allowed the effect of heritable variations on gene function to be re-evaluated. A recent study has shown that major sources of variation are masked when gene expression is averaged over many cells. Heritable variations that determine single-cell expression phenotypes may exert a regulatory function in specific cellular processes underlying disease. Masked effects on gene expression should therefore be modeled, not ignored.

Genetic regulation of gene expression

Understanding how and to what extent inter-individual genetic variation determines gene function in normal and pathological conditions can provide important insights into disease etiology. To this end, the rapid accumulation of large transcriptomic datasets across different tissues has prompted several population-based studies of gene expression variation [1]. In many of these studies, typical transcriptional analyses are carried out within or between whole tissue(s), with the aim of pinpointing gene expression signatures and/or (tissue-specific) genetic regulation of gene expression. Even at this level, context-dependent genetic regulation of gene expression has been shown to be important, and the underlying regulatory variants have more complex effects than previously anticipated [2]. For instance, characterizing different *cis*-regulatory mechanisms between tissues (such as opposite allelic effects) is important to understand the tissue-specific function exerted by disease-associated genetic variants.

The genetic variants that are associated with gene expression variation are commonly called expression quantitative trait loci (eQTLs). These can be mapped to the genome by modeling quantitative variation in gene expression and genetic variation (for example, single nucleotide polymorphisms (SNPs)) that have been assessed in the same population, family or segregating population. Essentially, mRNA levels can be treated as a quantitative phenotype and as such can be mapped to discrete genomic regions (genetic loci) that harbor DNA sequence variation affecting gene expression. In many cases, eQTL studies have provided direct insights into the complex regulatory mechanisms of gene expression - for instance, by allowing researchers to differentiate *cis* (or local) from *trans* (or distant) control of gene expression in a given tissue, experimental condition or developmental stage. Furthermore, eQTL analyses can be integrated with clinical genome-wide association studies (GWAS) to identify disease-associated variants [3,4]. Despite this recent, exciting progress in 'genetical genomics' (that is, eQTL studies), the growing number of single-cell transcriptomic analyses now prompts re-evaluation of our understanding of how heritable variations affect gene function in the cell.

Neglected single-cell differences and other hidden factors

Establishing a robust link between SNPs and gene expression variation is a non-trivial exercise when multiple cell types are jointly modeled. To aid this process, *ad hoc* methodological approaches that borrow information among tissues have been recently developed [5,6]. Nonetheless, emerging concepts such as single-cell transcriptomics have started changing our understanding of the genetic regulation of gene expression in individual cells, which can be hidden in ensemble-averaged experiments. In a recent study published in *Nature Biotechnology*, Holmes and colleagues [7] carried out single-cell quantification of gene expression for 92 genes in approximately 1,500 individual cells to disentangle the effect of gene variants on cell-to-cell variability, temporal dynamics or cell-cycle dependence in gene expression.

The authors looked at selected genes in fresh, naive B lymphocytes from three individuals and clearly showed how gene expression had much greater variability

*Correspondence: enrico.petretto@imperial.ac.uk
Medical Research Council (MRC) Clinical Sciences Centre, Faculty of Medicine, Imperial College London, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK

between cells within an individual than between individuals. This observation set the scene for a comprehensive investigation of the distributions of single-cell gene expression and the properties of gene expression noise in a larger population of cells. These analyses were focused on 92 genes affected by Wnt signaling (that can be chemically perturbed by a Wnt pathway agonist), of which 46 genes were also listed in the Catalog of Genome-Wide Association Studies, and resulted in four important outcomes.

First, perturbing the system with a Wnt pathway agonist exposed significant changes not only in whole-tissue gene expression but also in gene expression noise. Given the intrinsic stochastic nature of gene expression, it was expected that the number of mRNA copy numbers would vary from cell to cell, as previously shown in isogenic bacterial cell populations [8]. The single-cell transcriptomic analyses reported by Holmes and colleagues [7] highlight the large effect of fluctuations of mRNA copy numbers in HapMap lymphoblastoid cell lines, which has been mostly neglected and might influence eQTL detection in this system to a large extent.

Second, single-cell transcriptomic analysis allowed Holmes and colleagues to quantify both the noise from the regulation of transcription and the noise of RNA turnover, which therefore can be modeled independently. In keeping with previous observations [9], genes differed from each other primarily in terms of burst size (that is, the amount of RNA produced when the gene is switched on), resulting in an increased expression variance between cells that was greater than the expression mean. The expression 'Fano factor' (the gene expression variance divided by the mean) quantifies this phenomenon, and it represents another commonly neglected component that might be important in eQTL studies.

Third, when gene expression distributions were described in terms of heterogeneous cell subpopulations with respect to different stages of the cell cycle, Holmes and colleagues showed that the majority of genes analyzed had altered expression between G1 and early S phases. These apparent differences in cell cycle subpopulation proportions between samples represent another determinant of gene expression variation, which is expected to contribute significantly to gene regulation.

Finally, single-cell transcriptomics enabled the reliable quantification of the gene expression noise in the system. The latter can be considered as another source of variability, which can then be used to infer an expression network for each sample. Traditional gene co-expression networks assess gene-gene associations by correlating gene expression profiles across multiple samples. By contrast, in the *Nature Biotechnology* article, expression networks were built by correlating gene expression across multiple cells, which were profiled in the same lymphoblastoid cell line. For instance, one expression network built with approximately 200 cells from one of the lymphoblastoid cell lines revealed changes in cell-to-cell gene correlations in response to chemical perturbation of the Wnt signaling, which were not detectable at the level of whole-tissue expression. This approach allowed the authors to assess the extent to which the network connectivity of each gene varies in the system in response to other perturbations (for example, chemical, genetic), unmasking an additional factor that is potentially relevant for eQTL analysis.

Single-cell quantitative trait loci

After demonstrating (and quantifying) the important effect on gene function of a number of factors that reflect single-cell differences, Holmes and colleagues tested how each of these factors (alone or in combination) contributed to the detection of *cis*-eQTLs (that is, regulatory SNPs within 50 kb of the gene) [7]. This is an important question because integrated eQTL and clinical GWAS analyses are commonly employed to identify genes and pathways underlying disease, and eventually generate new hypotheses concerning diagnostic and prognostic biomarkers or potential therapeutic targets [10]. First, the eQTL associations detected at $-\log_{10}P = 3$ for whole-tissue gene expression (at both baseline and after chemical perturbation of the Wnt signaling) represented only a small fraction of the total number of eQTLs in the system (Figure 1). Overall, many more eQTL signals were detected for the other single-cell expression phenotypes tested. This highlights the extent to which different masked sources of variation (detailed above) can significantly affect the detection of *cis*-eQTLs in the system. Furthermore, it turns out that the complex spatio-temporal expression variability quantified by single-cell analysis ('single-cell expression') is more heritable than, or at least comparable to, gene expression levels averaged over many cells ('whole-tissue expression'), such that the authors of the study named this new class of associated genetic variants 'single-cell quantitative trait loci' (scQTLs) [7].

Notably, GWAS eQTL genes in particular demonstrated greater cell-cycle (G1 and early S phase) inter-individual variability compared with other genes and greater inter-individual variability of their network connectivities [7]. The implications of these results are two-fold: first, these studies urge caution in the interpretation of eQTL data published to date where only whole-tissue expression was considered; and second, they prompt a deeper evaluation (and accurate modeling) of these 'masked' sources of variation resulting from single-cell differences. It will be intriguing to extend these analyses to the study of more distant genetic control of gene expression at the single-cell level (that is, single-cell

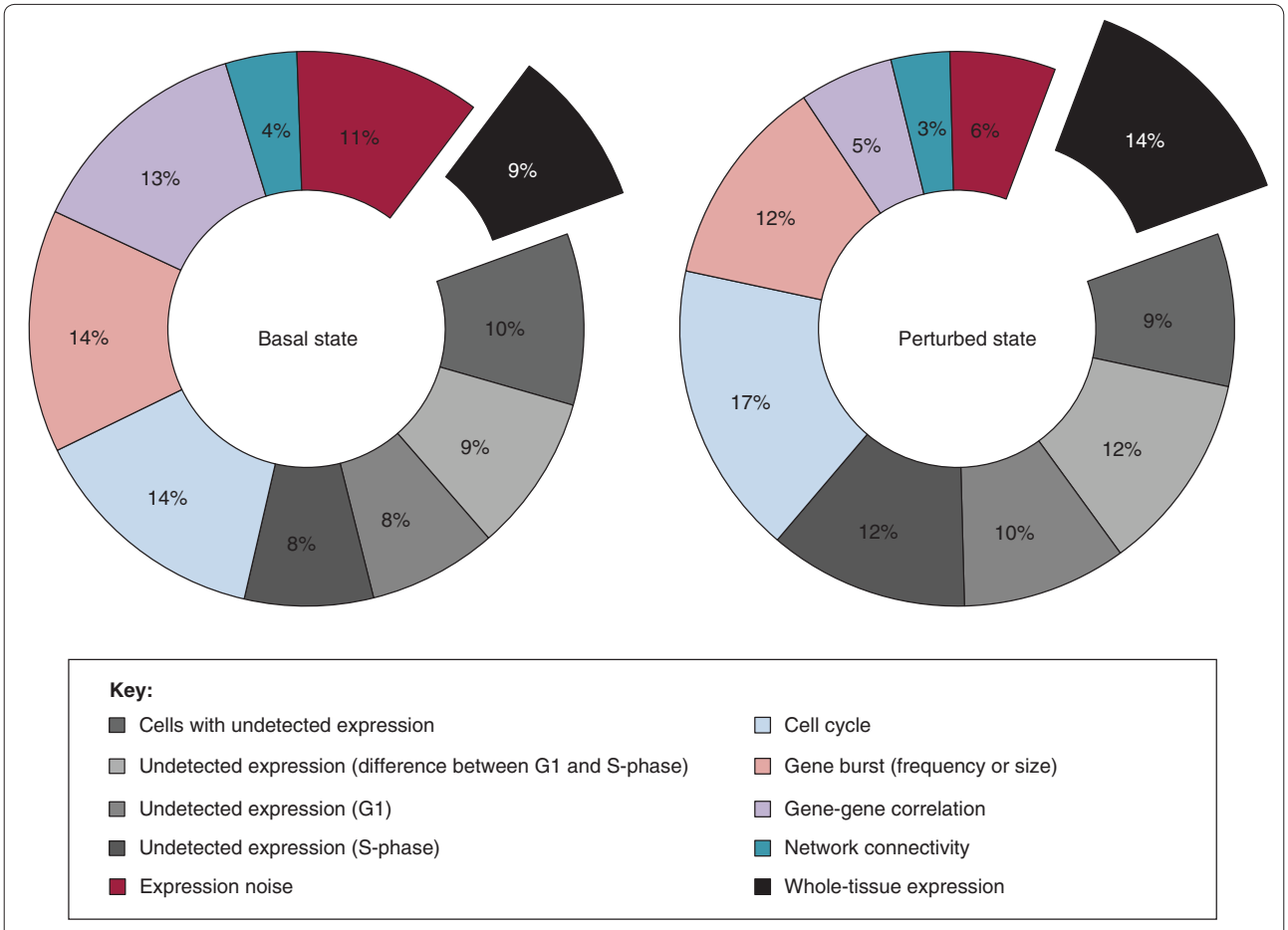

Figure 1. Distribution of single-cell quantitative trait loci detected at basal and perturbed states in HapMap lymphoblastoid cell lines derived from 15 unrelated individuals reported in [7]. The relative number of single-cell quantitative trait loci reported in Supplementary Table 1 from [7] is represented as a doughnut chart. Several different phenotypes derived from single-cell transcriptomic analysis were modeled as described in [7], and tested for association with single nucleotide polymorphisms within 50 kb of the gene. Beyond signals coming from cells with undetected expression (grey), a substantial number of single-cell quantitative trait loci associated with single-cell transcriptional variation due to cell cycle, gene burst, gene-gene correlation, network connectivity and expression noise were detected. The highlighted sector (black) denotes the relatively small contribution of whole-tissue expression quantitative trait loci, which were obtained using gene expression levels averaged over many cells.

trans-eQTLs) and to investigate the functional relevance of scQTLs on whole-body phenotypes in human and animal models. With the growing accessibility of single-cell technologies for transcriptomic studies, the time is right for a deep re-thinking of the key factors determining the observed complexity of gene expression and its regulation.

Abbreviations

eQTLs, expression quantitative trait loci; GWAS, genome-wide association study; scQTLs, single-cell quantitative trait loci; SNP, single nucleotide polymorphism.

Competing interests

The author declares that he has no competing interests.

Acknowledgments

EP is supported by the Medical Research Council UK and thanks Aida Moreno-Moral for proof-reading.

Published: 20 August 2013

References

1. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, Travers M, Potter S, Grundberg E, Small K, Hedman AK, Bataille V, Tzenova Bell J, Surdulescu G, Dimas AS, Ingle C, Nestle FO, di Meglio P, Min JL, Wilk A, Hammond CJ, Hassanali N, Yang TP, Montgomery SB, O'Rahilly S, Lindgren CM, Zondervan KT, Soranzo N, Barroso I, Durbin R, *et al.*: **The architecture of gene regulatory variation across multiple human tissues: the MuTHER study.** *PLoS Genet* 2011, **7:**e1002003.
2. Fu J, Wolfs MG, Deelen P, Westra HJ, Fehrmann RS, Te Meerman GJ, Buurman WA, Rensen SS, Groen HJ, Weersma RK, van den Berg LH, Veldink J, Ophoff RA, Snieder H, van Heel D, Jansen RC, Hofker MH, Wijmenga C, Franke L: **Unraveling the regulatory mechanisms underlying tissue-dependent genetic variation of gene expression.** *PLoS Genet* 2012, **8:**e1002431.
3. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ: **Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS.** *PLoS Genet* 2010, **6:**e1000888.
4. Min JL, Taylor JM, Richards JB, Watts T, Pettersson FH, Broxholme J, Ahmadi KR, Surdulescu GL, Lowy E, Gieger C, Newton-Cheh C, Perola M, Soranzo N, Surakka I, Lindgren CM, Ragoussis J, Morris AP, Cardon LR, Spector TD,

Zondervan KT: **The use of genome-wide eQTL associations in lymphoblastoid cell lines to identify novel genetic pathways involved in complex traits.** *PloS One* 2011, **6:**e22070.
5. Petretto E, Bottolo L, Langley SR, Heinig M, McDermott-Roe C, Sarwar R, Pravenec M, Hubner N, Aitman TJ, Cook SA, Richardson S: **New insights into the genetic control of gene expression using a Bayesian multi-tissue approach.** *PLoS Comput Biol* 2010, **6:**e1000737.
6. Flutre T, Wen X, Pritchard J, Stephens M: **A statistical framework for joint eQTL analysis in multiple tissues.** *PLoS Genet* 2013, **9:**e1003486.
7. Wills QF, Livak KJ, Tipping AJ, Enver T, Goldson AJ, Sexton DW, Holmes C: **Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments.** *Nat Biotechnol* 2013, **31**:748-752.
8. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS: **Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells.** *Science* 2010, **329:**533-538.
9. Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, Simpson ML, Weinberger LS: **Transcriptional burst frequency and burst size are equally modulated across the human genome.** *Proc Natl Acad Sci U S A* 2012, **109:**17454-17459.
10. Califano A, Butte AJ, Friend S, Ideker T, Schadt E: **Leveraging models of cell regulation and GWAS data in integrative network-based association studies.** *Nat Genet* 2012, **44:**841-847.

doi:10.1186/gm476
Cite this article as: Petretto E: **Single cell expression quantitative trait loci and complex traits.** *Genome Medicine* 2013, **5:**72.