FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS & SCIENCES

METHODS FOR EXPRESSION QUANTITATIVE TRAIT LOCI (EQTLS) DETECTION USING SINGLE-CELL RNA-SEQUENCING DATA

By

ANDRES E. CANDIDO GETTINGS

A Thesis to be submitted to the
Department of Biological Science
in partial fulfillment of the
requirements for the degree of
Bachelor of Science in Computational Biology

The members of the Defense Committee:

Xian Mallory
Thesis Director

Hank Bass
Committee Member

Karen McGinnis
Committee Member

TABLE OF CONTENTS

TABLE OF FIGURES

Abstract

Expression Quantitative Trait Loci (eQTLs), are genomic loci associated with variations of downstream gene expression. The study of eQTLs provides insights into the genetic basis of complex traits and diseases by revealing how genetic variations influence the expression of genes, helping to bridge the gap between genotype and phenotype. Single-cell RNA-sequencing (scRNA-Seq) has emerged as a crucial technique for detecting eQTLs because it allows for the measurement of gene expression at a single-cell resolution, enabling the identification of cell-specific expression patterns and uncovering rare genetic variations, variations that occur to a small fraction of cells and thus might be missed in the traditional bulk RNA-seq. In this study, we will conduct a review of the methods used for detecting eQTLs using eQTLsingle, a unique eQTL analysis tool that uses single-cell RNA-sequencing data to measure gene expression and detect genetic variations at the single-cell level. In this thesis, we aim to enhance our understanding of how eQTLs can be detected at the single-cell level using scRNA-seq data and of how eQTL analysis tools such as eQTLsingle can accomplish doing so, as well as to address the inherent limitations of only utilizing scRNA-seq for detecting eQTLs and to find solutions to these limitations by using relevant literature.

Research Objectives

The primary objectives of this methodological and bibliographic review are to, firstly, enhance our understanding of how eQTLs can be detected at the single-cell resolution level using scRNA-seq data. Secondly, to review the methods used for detecting eQTLs in eQTLsingle, an eQTL analysis tool that uses single-cell RNA-sequencing data. Thirdly, to identify the inherent limitations of only utilizing scRNA-seq for detecting eQTLs. And finally, to identify possible solutions to these limitations by using relevant literature.

1. Introduction

Expression Quantitative Trait Loci (eQTLs) are genomic loci associated with variations of downstream gene expression. The study of eQTLs represents a pivotal intersection between genetics and gene expression, offering profound insights into the intricate mechanisms underlying complex traits and diseases [1,2]. Understanding how single nucleotide variations (SNVs) in these loci can lead to divergent outcomes in the expression levels of particular genes is instrumental in understanding the complex relationships between genotype and phenotype. As genetic associations with diseases and traits increasingly become a central focus of biomedical research, understanding the regulatory effects of genetic variants on gene expression is paramount. This understanding not only advances our comprehension of disease etiology but also holds the potential for the development of targeted therapeutic strategies [1-5].

The goal of eQTL analysis is to associate single nucleotide variants (SNVs) with gene expression profiles, which are typically measured in transcripts per million (TPM), counts per million (CPM), or other standardized metrics, in order to identify regions of the genome that are associated with the expression levels of particular genes, eQTLs (Fig.1). These eQTLs could then be potentially used for the development

of targeted therapeutic strategies for a variety of diseases and conditions such as type-1 diabetes [6], bipolar disorder [7], and more [1-5].

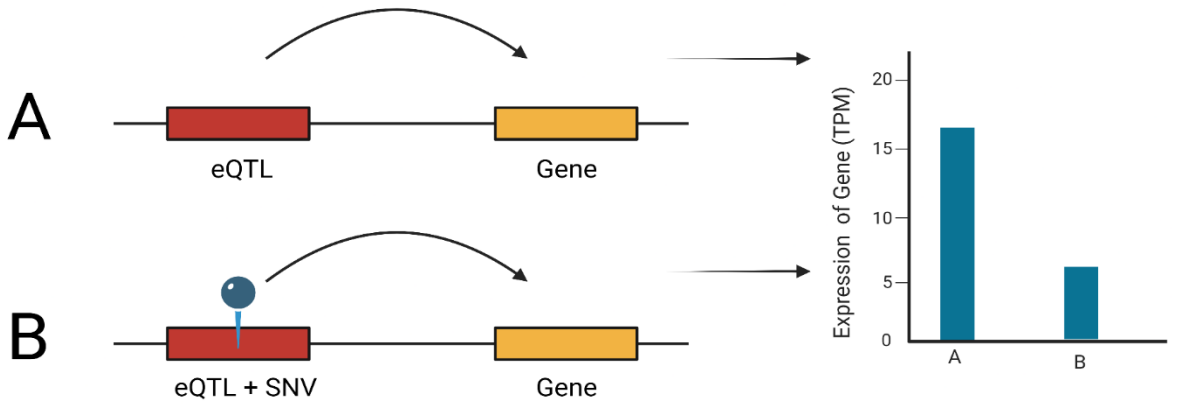

Fig.1

Fig. 1 - eQTL Detection

eQTLs can be categorized into two main classes based on their genomic locations relative to the genes they influence: Cis and Trans eQTLs.

In Cis-eQTLs, the "Cis" is short for "cis-acting," which means "on the same side" or “local.” In the context of eQTLs, cis-eQTLs are regions of the genome that are located near a specific gene, often within 1 million base pairs from the gene, and affect the expression of that gene (A, Fig.2) [5,8,9].

On the other hand, the "Trans" in trans-eQTLs is short for "trans-acting," which means "on the opposite side" or “distant.” In the context of eQTLs, trans-eQTLs are regions of the genome located at a different genomic location or farther away than 1 million base pairs from the gene they influence, meaning that they can affect both distant genes on the same chromosome and genes residing on different chromosomes (B, Fig.2). These eQTLs can affect the expression of one or more genes across the genome. Unlike cis-eQTLs, trans-eQTLs have a more indirect influence on gene expression and often act by regulating the activity of transcription factors or other regulatory elements that, in turn, affect the expression of multiple genes. These effects tend to have weaker influence on gene expression variance than those of Cis-eQTLs, which

consequently makes trans-eQTL identification much more challenging (C & D, Fig.2) [5,8,9]. Alternatively, we could also define some Trans-eQTLs as being both Cis and Trans-acting since they act on local genes which in turn act on other distant ones [5,9].

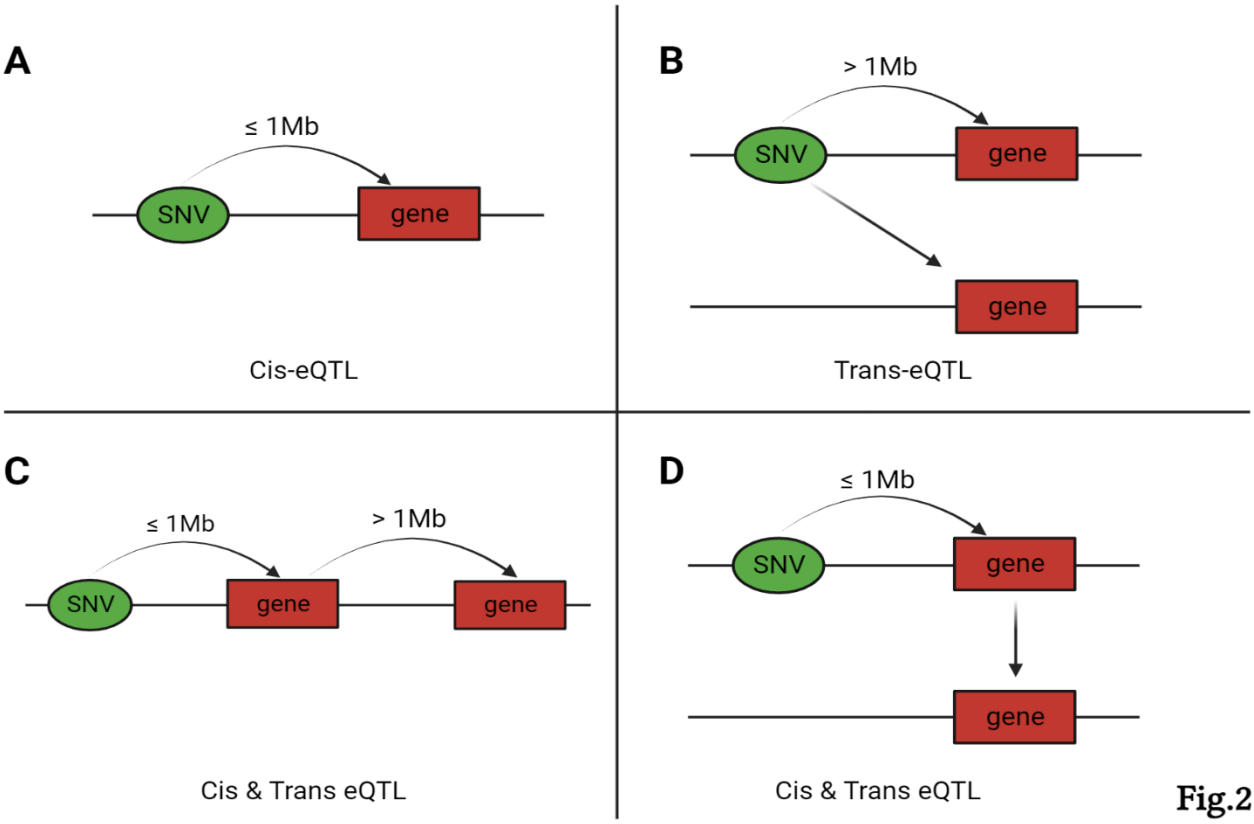

Fig. 2 - Cis and Trans-eQTLs

Additionally, it is important to understand that eQTLs are not the sole factor behind the control of the expression levels of genes, but rather, they are one of multiple factors that each explains for a fraction of the total level of gene expression, these other factors are commonly referred to as covariates in eQTL analysis [10,11].

2. Single-Cell Resolution in eQTL Studies

Traditionally, eQTL analysis has been conducted on bulk RNA sequencing data, where the gene expression profile of a sample is represented as the average gene expression of all cells in that particular sample (Fig. 3). While this approach has yielded valuable insights on the genetic influences on gene expression, it inherently disregards the cellular heterogeneity present within tissues. This limitation is particularly pronounced in complex tissues composed of diverse cell types, where distinct cellular subpopulations exhibit unique gene expression profiles. Consequently, the resulting averaged expression value can obscure vital regulatory associations specific to individual cell types or states [5,12].

This deficiency in traditional eQTL analysis necessitates the integration of single-cell RNA sequencing (scRNA-seq) technology. scRNA-seq enables the quantification of gene expression at the single-cell resolution. By profiling gene expression in individual cells from a sample, scRNA-Seq allows us to identify and group expression profiles of individual cells from each cell type present in a collected sample. Facilitating the discovery of cell-type-specific eQTLs that would be otherwise obscured when averaged across distinct cell populations (Fig. 3). This technology not only captures the intrinsic variability within cell populations but can also uncover rare cell subtypes and transient states that may be pivotal in the study of disease mechanisms [12-17].

Regardless of the inherent advantages of conducting eQTL analysis by leveraging single-cell resolution sequencing, the adoption and development of these technologies by the eQTL research community did not occur until fairly recently. In fact, by the end of 2021 only a dozen studies had been published conducting eQTL analysis using scRNA-seq data. One of these studies was responsible for the development of eQTLsingle, an eQTL analysis tool we will be reviewing in this paper [14].

Fig.3

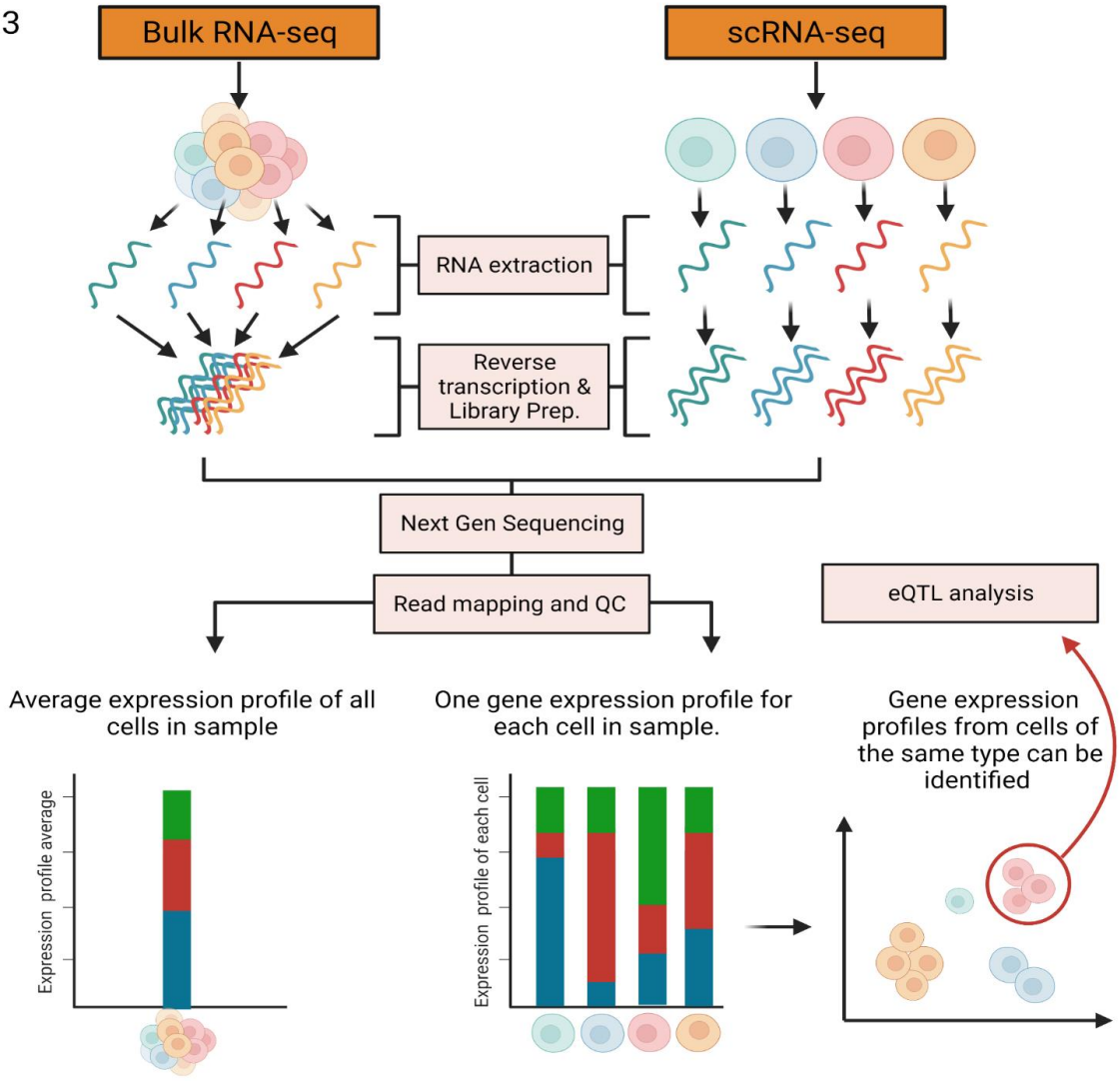

Fig. 3 - Bulk RNA-seq and scRNA-seq

3. eQTL Analysis Using scRNA-seq Data With eQTLsingle

3.1 eQTLsingle

eQTLsingle [18] is an eQTL analysis tool that allows us to detect eQTLs in cells of the same type by using scRNA-seq data. eQTLsingle is unique in the sense that it was designed to operate specifically on single-cell or single-nucleus RNA-seq data. This can be accomplished by detecting SNVs present in the scRNA-seq data in order to generate mutation profiles at the single-cell resolution. Finally, both of the generated

gene expression and mutation profiles are leveraged by eQTLsingle to detect cell-type-specific cis-eQTLs and trans-eQTLs in the exons, coding regions of the genome, from which the RNA transcripts represented in the gene expression profiles were generated. This approach provides high-quality genomic and expression data at the single-cell level in these exon regions.

3.2 Data Collection

Gene expression profiles can be obtained from a group of cells of the same type via isolating individual cells from a bulk sample and following the scRNA-seq workflow depicted in Fig.4 [19,20,21]. However, while eQTLsingle can conduct eQTL analysis with gene expression data collected using traditional protocols such as UMI-based technologies, the authors specified that eQTLsingle works better with scRNA-seq data obtained with full-length sequencing protocols such as SMART-seq [13,18,22].

3.3 Data Preprocessing

Once gene expression profiles from all cells in the sample are sequenced individually, we are left with raw sequencing reads in the form of FASTQ files. These reads are then aligned to a reference genome in order to map the location from which they originated, by using STAR [23], bowtie [24], or other RNA aligners. Quality control measures follow, filtering out low-quality cells by using the following metrics: (1) cells with less than a certain threshold of uniquely mapping reads are removed; (2) cells with percentage of uniquely mapped reads less than a certain threshold are removed. The threshold values used by the eQTLsingle authors as an example were (1) 1x10^6 and (2) 60% respectively [18].

The gene expression levels of each cell are then quantified using featureCounts [25]. The data is normalized to CPM (counts per million), which means, scaling up, or down, expression values proportionally as if the total number of counts was one million. This is done to facilitate comparisons between samples. In our case, we would use these CPM files to identify cells of the same type by comparing their gene expression profiles [14,18]. The final form of the gene expression profiles is a gene expression matrix (gene

x cell), rownames of this matrix are gene ids, colnames are cell Ids. The individual elements of this matrix represent the number of counts of a particular gene X in a particular cell Y (Fig.4).

Once we have identified our cell type of interest, we can use the specific raw reads from those cells to generate their mutation profiles via SNV calling. To accomplish this, the authors of eQTLsingle aligned the raw reads to single-nucleotide variant and indel reference genomes using GATK [26]. Which essentially allows them to identify the individual SNVs present in each cell, or at least the ones that can be identified with those references. The output of SNV calling is a mutation matrix (SNV x cells), rownames of this matrix are SNV ids, colnames are Cell ids. The individual elements of this matrix represent if a particular cell X has a particular SNV Y. These elements are divided into three groups: the reference group (non-mutated, no SNV), the alternate group (mutated, SNV present), and the insufficient information group (discarded). With the reference group denoted in the matrix as a 0, the alternate group denoted as a 1, the insufficient information group denoted as a -1 (Fig.4). The authors of eQTLsingle used bam-readcount [27] to generate the mutation matrix [18].

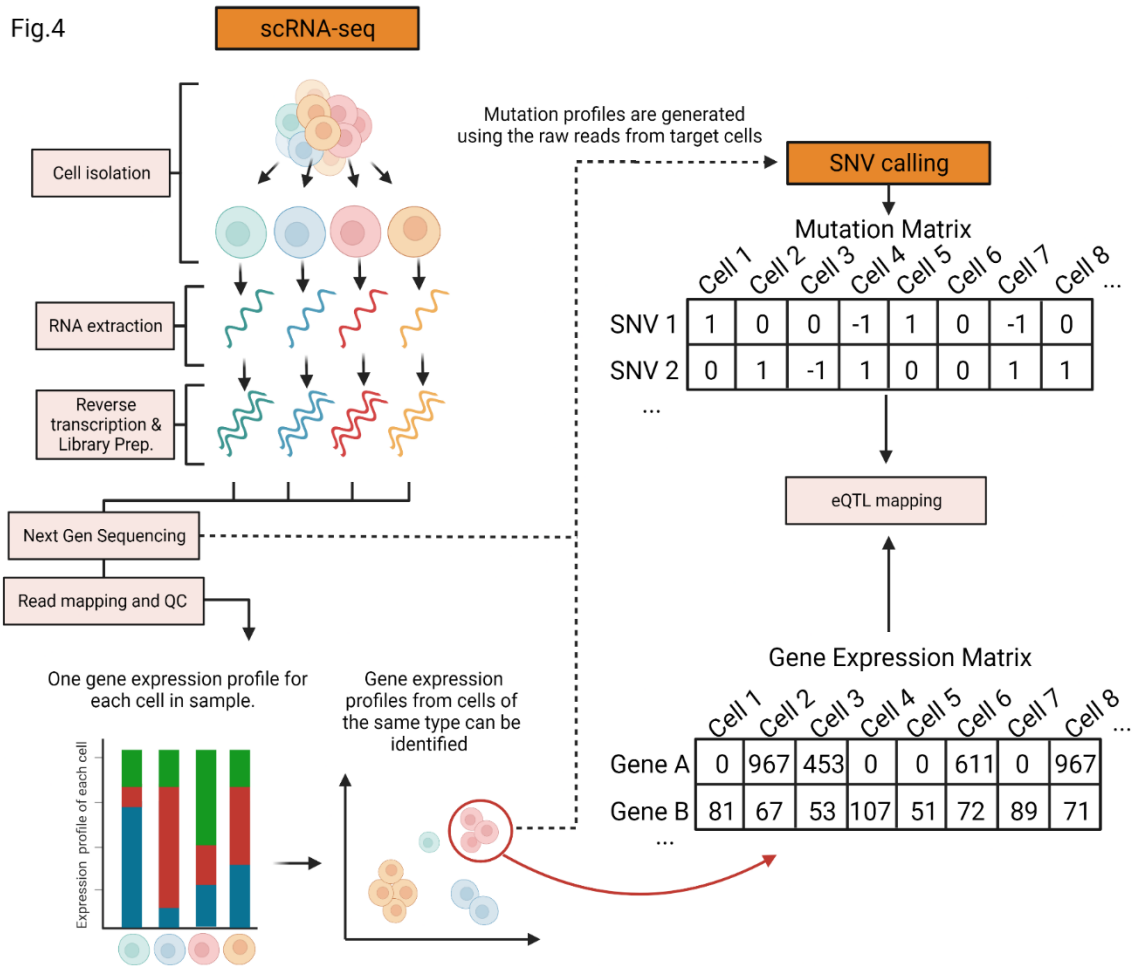

Fig. 4 - Data Collection and Preprocessing for eQTLsingle

3.4 eQTL mapping

In order to perform eQTL mapping, eQTLsingle goes through each detected SNV in the mutation matrix and looks for possible eQTL associations on all genes in the gene expression matrix. It can accomplish this by dividing the cells into the aforementioned reference and alternate groups of a particular SNV and gene pair (Fig. 5), it then builds a zero inflated negative binomial model (ZINB), for the gene expression level of a particular gene in each group of cells (Fig. 5). The ZINB model is particularly useful when dealing with scRNA-seq data since there is an over-representation of zero counts, hence the “zero inflated” part. In the context of eQTL analysis, this would mean that most genes have an expression level of zero (not

expressed) in each particular cell of any given cell type. The negative binomial model has been widely used for modeling count data, bulk RNA-seq. Therefore, by using a ZINB model, eQTLsingle can effectively model the gene expression of a particular gene in each group by using scRNA-seq data. Once the ZINB model has been generated for each group, eQTLsigle uses a likelihood ratio test (LRT) to compare the models of the two groups (Fig. 5). The genomic location of the SNV is identified as an eQTL for the target gene if the test finds a significant difference between the two models. This eQTL is then classified as a cis-eQTL if the distance between the SNV and the gene is less than, or equal to, one million base pairs (1Mb). Otherwise, if the distance is greater than 1Mb, or if the SNV and gene are on different chromosomes, the eQTL is classified as a trans-eQTL [18,28].

Fig.5

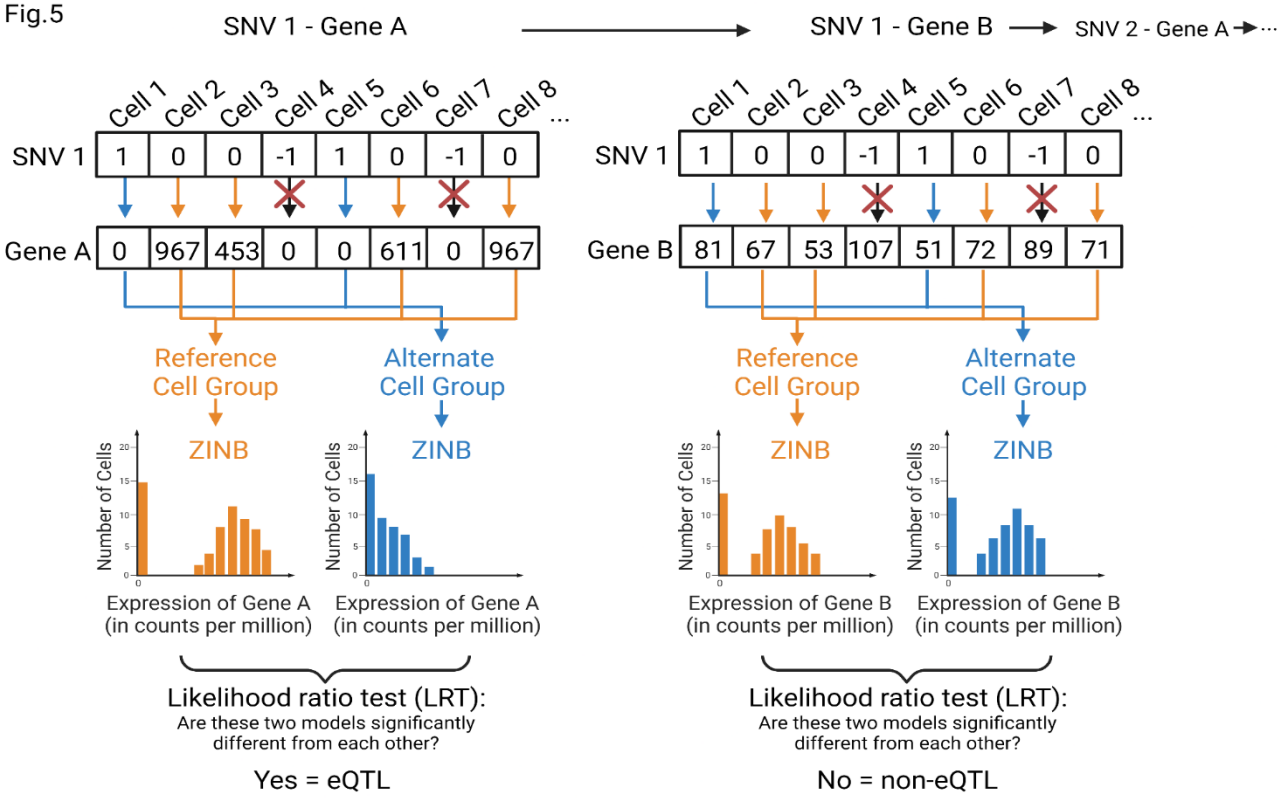

Fig. 5 - eQTL Mapping with eQTLsingle

4. Inherent limitations of eQTL Analysis Utilizing scRNA-seq Data Only

While eQTLsingle does allow us to detect cell-type-specific cis-eQTLs and trans-eQTLs, the fact that it generates both gene expression and mutation profiles from each cell by using only scRNA-seq data comes with some limitations. The raw reads we collect from RNA-seq data only have genetic information from the transcribed regions of the genome from which they originated, the exons [18].

According to the ENCODE Project, the encyclopedia of DNA elements, exons only represent around 2.94% of the entire human genome. This would mean that eQTL studies using single-cell RNA sequencing data only are not taking into account SNVs that could be present in around 97.06% of the genome, which consists of introns, non-protein-coding DNA sequences [3].

While introns do not encode proteins, they are integral to gene expression regulation, and therefore, are of great interest for eQTL studies. Additionally, by only using scRNA-seq data to generate the mutation profiles we could also miss SNVs present in lowly expressed genes that cannot be detected during the reverse transcription and cDNA amplification steps of scRNA-seq, this would mean that our effective coverage of the genome would be less than 2.94% at any given time. The ideal solution to both of these limitations would be for single-cell-type eQTL studies to sequence both the genome and transcriptome for each single-cell simultaneously [3,18].

5. Integrating scDNA-seq and scRNA-seq workflows to address limitations while preserving single-cell resolution for eQTL analysis.

By incorporating scDNA-seq into the previously established scRNA-Seq workflow used in eQTLsingle, we would be able to obtain both gene expression and mutation profiles from individuals cell simultaneously, while at the same time including the SNVs present in the previously absent intron regions of the genome and SNVs present in in lowly expressed exons into our final eQTL analysis (Fig. 6) [3,18,29-31].

We could start incorporating the scDNA-seq workflow after the cell isolation stage of the previously mentioned scRNA-Seq workflow (Fig.4). After individual cells are isolated from the bulk sample, cell lysis is induced, typically by using various detergents such as NP-40, Triton X-100, or Trizol. These detergents disrupt both the cellular and nuclear membranes while leaving DNA and RNA molecules unscathed. Following this step, poly(A) mRNA would need to be physically isolated from DNA, this could be accomplished by using a combination of oligo(dT) primers and magnetic beads or other methods, this scRNA-seq data would be used to generate our gene expression profiles needed for eQTL analysis. By isolating the mRNA, genomic DNA is left floating in a mixture of intra-cellular debris [29-38].

In order to separate the genomic DNA from the debris, DNA purification would follow. DNA purification can be achieved through various methods like ethanol precipitation or the use of reagents such as DNAzol [39]. Iterative cycles of DNA purification are performed until the desired purity level is attained, while preserving DNA integrity. The evaluation of purity and integrity often involves techniques such as gel electrophoresis or spectrophotometry. Once the desired level of purity is achieved, double-stranded DNA is denatured into single-stranded cDNA for subsequent processing. Similar to the scRNA-seq workflow (Fig. 4), cDNA would then be amplified, via polymerase chain reaction (PCR), fragmented, via mechanical or enzyme-based methods, and sequenced, via next generation sequencing (NGS) [40-45].

Following this, we would generate our mutation profiles by mapping the raw reads from the scDNA-seq to the single-nucleotide variant and indel reference genomes using GATK [26] and using bam-readcount [27] to generate the mutation matrix for eQTL analysis [18].

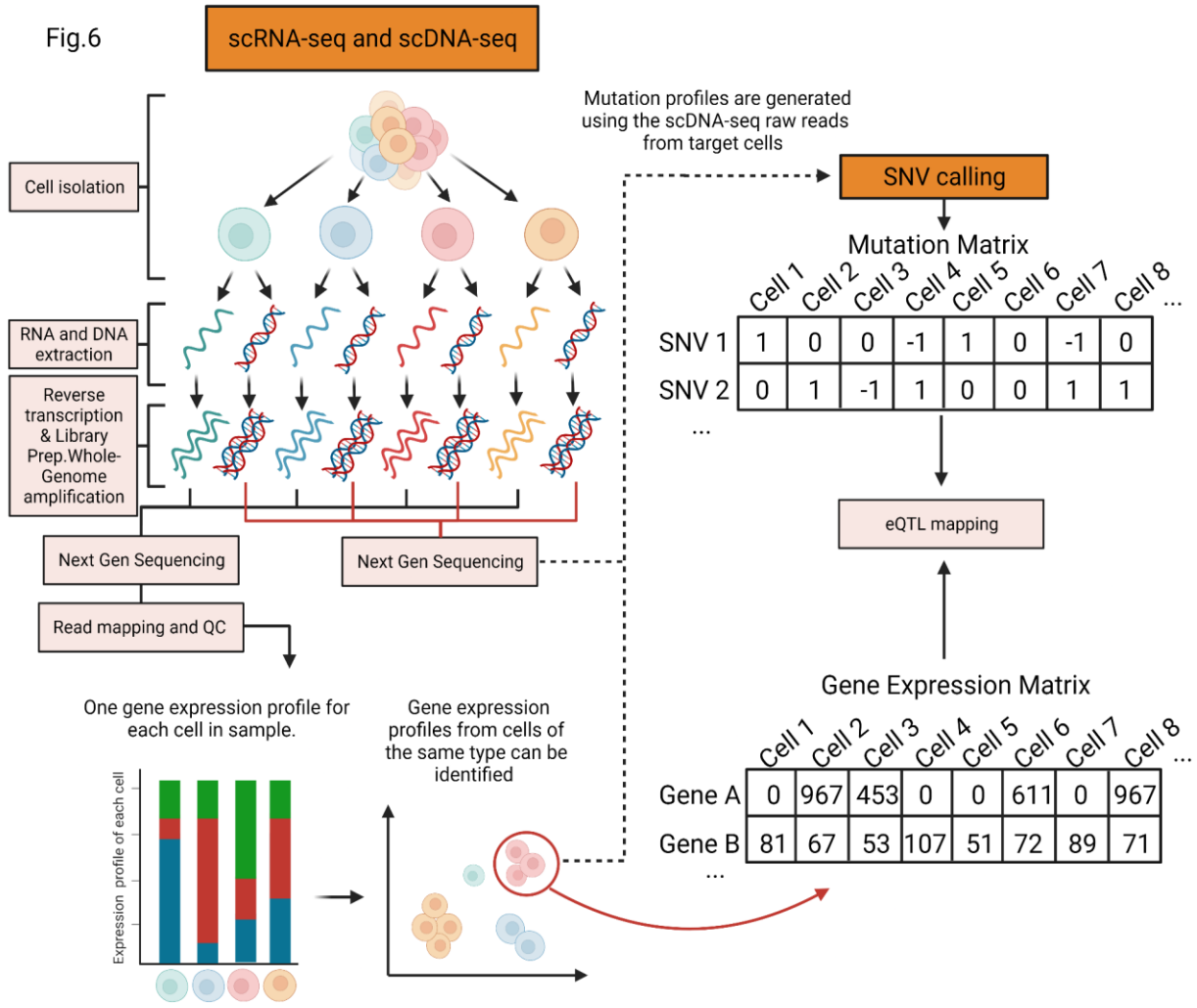

Fig. 6 - scRNA-seq and scDNA-seq for eQTL analysis

6. Conclusion and future research directions

In this paper, we explored how eQTLs can be detected at the single-cell-type resolution level using scRNA-seq data. We reviewed the methods used for detecting eQTLs in eQTLsingle, an eQTL analysis tool designed to exclusively use single-cell RNA-sequencing data to map eQTLs. We identified the inherent limitations of only utilizing scRNA-seq for detecting eQTLs. And finally, we identified possible solutions to these limitations by integrating scDNA-seq into the scRNA-seq workflow, which would allow us to detect eQTLs in both exon and intron regions of the genome by simultaneously sequencing complete gene expression and mutation profiles from single cells.

In light of these considerations, the next logical step when it comes to future research objectives would be to identify novel paired-sequencing methods and technologies available, meaning, methods and technologies that can simultaneously sequence both scDNA-seq and scRNA-seq data from single cells, and to review them in depth. Two notable candidates would be G&T-seq [29] and DRseq [30]. In the main eQTLsingle study, these paired-sequencing methods were mentioned to be under development and to have very poor data coverage, so poor in fact, that they were described as being the main motivation for the authors to develop eQTLsingle in the first place. However, it has been nearly three years since that, revisiting these methods could yield meaningful insights into the current state of the technologies. Possibly enabling researchers to have the tools needed to resolve the limitations of conducting eQTL analysis by only using scRNA-seq data and make novel and impactful discoveries that could ultimately increase our understanding the complex relationships between genotype and phenotype.

References:

1. GTEx Consortium et al. (2017). Genetic effects on gene expression across human tissues. Nature, 550(7675), 204-213. https://doi.org/10.1038/nature24277

2. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487, 330–337 (2012). https://doi.org/10.1038/nature11252

3. The ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74. https://doi.org/10.1038/nature11247

4. The FANTOM Consortium and the RIKEN PMI and CLST (DGT). (2014). A promoter-level mammalian expression atlas. Nature 507, 462–470. https://doi.org/10.1038/nature13182

5. Nica, A. C., & Dermitzakis, E. T. "Expression quantitative trait loci: present and future." Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, vol. 368, no. 1620, 2013, p. 20120362. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3682727/

6. Hakonarson, H., Grant, S., Bradfield, J. et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 448, 591–594 (2007). https://doi.org/10.1038/nature06010

7. WTCCC 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678 10.1038/nature05911 https://doi.org/10.1038/nature05911

8. Võsa, U., Claringbould, A., Westra, HJ. et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet 53, 1300–1310 (2021). https://doi.org/10.1038/s41588-021-00913-z

9. Shan, N., Wang, Z. & Hou, L. Identification of trans-eQTLs using mediation analysis with multiple mediators. BMC Bioinformatics 20 (Suppl 3), 126 (2019). https://doi.org/10.1186/s12859-019-2651-6

10. Wang, T., Liu, Y., Ruan, J. et al. (2021). A pipeline for RNA-seq based eQTL analysis with automated quality control procedures. BMC Bioinformatics 22 (Suppl 9), 403. https://doi.org/10.1186/s12859-021-04307-0

11. Andrey A. Shabalin, Matrix eQTL: ultra fast eQTL analysis via large matrix operations, Bioinformatics, Volume 28, Issue 10, May 2012, Pages 1353–1358, https://doi.org/10.1093/bioinformatics/bts163

12. Kulkarni, A., Anderson, A. G., Merullo, D. P., & Konopka, G. (2019). Beyond bulk: A review of single cell transcriptomics methodologies and applications. Current Opinion in Biotechnology, 58, 129-136. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6710112/

13. Haque, A., Engel, J., Teichmann, S. A., et al. (2017). A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Medicine, 9, 75. https://doi.org/10.1186/s13073-017-0467-4

14. Maria, M.; Pouyanfar, N.; Örd, T.; Kaikkonen, M.U. The Power of Single-Cell RNA Sequencing in eQTL Discovery. Genes 2022, 13, 502. https://doi.org/10.3390/genes13030502

15. Gross, A., et al. (2015). Technologies for Single-Cell Isolation. International Journal of Molecular Sciences, 16(8), 16897-919. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4581176/

16. Hu, P., Zhang, W., Xin, H., & Deng, G. (2016). Single Cell Isolation and Analysis. Frontiers in Cell and Developmental Biology, 4, 116. https://doi.org/10.3389/fcell.2016.00116

17. Mazutis, L., et al. (2013). Single-cell analysis and sorting using droplet-based microfluidics. Nature Protocols, 8(5), 870-891. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4128248/

18. Ma, Tianxing, et al. (2020). "Discovering Single-Cell eQTLs from scRNA-seq Data Only." Gene, https://doi.org/10.1101/2021.06.10.447906

19. Hu, P., Zhang, W., Xin, H., & Deng, G. (2016). Single Cell Isolation and Analysis. Frontiers in Cell and Developmental Biology, 4, 116. https://doi.org/10.3389/fcell.2016.00116

20. Mazutis, L., et al. (2013). Single-cell analysis and sorting using droplet-based microfluidics. Nature Protocols, 8(5), 870-891. https://10.1038/nprot.2013.046

21. uFluidix. "Single Cell Analysis in Microfluidics: Applications and Techniques." uFluidix, www.ufluidix.com/microfluidics-applications/single-cell-analysis/.

22. Picelli, S., Faridani, O., Björklund, Å. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9, 171–181 (2014). https://doi.org/10.1038/nprot.2014.006

23. Dobin, Alexander et al. "STAR: ultrafast universal RNA-seq aligner." Bioinformatics, vol. 29, no. 1, 2013, pp. 15-21, https://doi.org/10.1093/bioinformatics/bts635.

24. Langmead, B., Trapnell, C., Pop, M. *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009). https://doi.org/10.1186/gb-2009-10-3-r25

25. Liao, Yang, et al. "featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features." Bioinformatics, vol. 30, no. 7, 2014, pp. 923-930. Oxford University Press, https://doi.org/10.1093/bioinformatics/btt656.

26. G.A. Van der Auwera, M.O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K.v. Garimella, D. Altshuler, S. Gabriel, M.A. DePristo, From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline, Curr. Protoc. Bioinforma. (2013). https://doi.org/10.1002/0471250953.bi1110s43.

27. Chrisamiller, bam-readcount, (n.d.). https://github.com/genome/bam-readcount.

28. Zhun Miao, Ke Deng, Xiaowo Wang, Xuegong Zhang, DEsingle for detecting three types of differential expression in single-cell RNA-seq data, Bioinformatics, Volume 34, Issue 18, September 2018, Pages 3223–3224, https://doi.org/10.1093/bioinformatics/bty332

29. Macaulay, I., Haerty, W., Kumar, P. et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods 12, 519–522 (2015). https://doi.org/10.1038/nmeth.3370

30. Dey, S., Kester, L., Spanjaard, B. *et al.* Integrated genome and transcriptome sequencing of the same cell. *Nat Biotechnol* 33, 285–289 (2015). https://doi.org/10.1038/nbt.3129

31. Baysoy, A., Bai, Z., Satija, R. *et al.* The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* 24, 695–713 (2023). https://doi.org/10.1038/s41580-023-00615-w

32. Rio, D. C., Ares, M. Jr, Hannon, G. J., & Nilsen, T. W. (2010). Purification of RNA using TRIzol (TRI reagent). Cold Spring Harb Protoc, 2010(6), pdb.prot5439. https://doi.org/10.1101/pdb.prot5439

33. Sung, K., Khan, S. A., Nawaz, M. S., & Khan, A. A. (2003). A simple and efficient Triton X-100 boiling and chloroform extraction method of RNA isolation from Gram-positive and Gram-negative bacteria. FEMS Microbiology Letters, 229(1), 97–101. https://doi.org/10.1016/S0378-1097(03)00791-2

34. Le, A. V., Huang, D., Blick, T., Thompson, E. W., & Dobrovic, A. (2015). An optimised direct lysis method for gene expression studies on low cell numbers. Scientific Reports, 5, 12859. https://doi.org/10.1038/srep12859

35. Thermo Fisher Scientific. (n.d.). Detergents for Cell Lysis and Protein Extraction. https://www.thermofisher.com/us/en/home/life-science/protein-biology/protein-biology-learning-center/protein-biology-resource-library/pierce-protein-methods/detergents-cell-lysis-protein-extraction.html Date accessed: October 14th, 2023.

36. Green MR, Sambrook J. Isolation of Poly(A)+ Messenger RNA Using Magnetic Oligo(dT) Beads. Cold Spring Harb Protoc. 2019 Oct 1;2019(10). doi: 10.1101/pdb.prot101733. PMID: 31575797.

37. Tambe, A., & Pachter, L. (2019). Barcode identification for single cell genomics. BMC Bioinformatics, 20(1), 32. https://doi.org/10.1186/s12859-019-2612-0

38. Hong, M., Tao, S., Zhang, L., et al. (2020). RNA sequencing: New technologies and applications in cancer research. Journal of Hematology & Oncology, 13(1), 166. https://doi.org/10.1186/s13045-020-01005-x.

39. Thermo Fisher Scientific. (n.d.).Genomic DNA Extraction and Genomic DNA Isolation. https://www.thermofisher.com/us/en/home/life-science/dna-rna-purification-analysis/genomic-dna-extraction.html Date accessed: October 14th, 2023

40. Cooper, G.M. (2000). The Cell: A Molecular Approach (2nd edition). Sunderland (MA): Sinauer Associates. DNA Replication. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK9940/

41. Lee, P. Y., Costumbrado, J., Hsu, C. Y., & Kim, Y. H. (2012). Agarose gel electrophoresis for the separation of DNA fragments. Journal of Visualized Experiments, (62), 3923. https://pubmed.ncbi.nlm.nih.gov/22546956/

42. QIAGEN. (n.d.). Quantification of DNA. https://www.qiagen.com/us/knowledge-and-support/knowledge-hub/bench-guide/dna/analysing-dna/quantification-of-dna Date accessed: October 16th, 2023.

43. Liu, J., Huang, S., Sun, M. et al. (2012). An improved allele-specific PCR primer design method for SNP marker analysis and its application. Plant Methods 8, 34. https://doi.org/10.1186/1746-4811-8-34

44. Zhang, J., Yang, J., Zhang, L., & et al. (2020). A new SNP genotyping technology Target SNP-seq and its application in genetic analysis of cucumber varieties. Scientific Reports, 10, 5623. https://doi.org/10.1038/s41598-020-62518-6

45. New England Biolabs. (n.d.). Improving enzymatic DNA fragmentation for next-generation sequencing library construction. NEB. https://www.neb.com/en-us/tools-and-resources/feature-articles/improving-enzymatic-dna-fragmentation-for-next-generation-sequencing-library-construction Date accessed: October 15th, 2023.

Bibliographical Sketch:

Andres E. Candido Gettings is a dedicated and motivated student currently enrolled in the Computational Biology program at Florida State University, where he is actively pursuing a Bachelor of Science degree. In addition to his primary focus on Computational Biology, Andres has chosen to broaden his academic horizons by minoring in both Mathematics and Computer Science. This interdisciplinary approach reflects his commitment to acquiring a well-rounded and versatile set of skills. Always looking to grow intellectually and eager to learn new skills and to use new tools. Andres has had three years of computer programming experience using C++ and Python; He has conducted research on detecting eQTLs using scRNA-seq data only and eQTLsingle under the supervision of Dr. Xian Mallory for almost one year, he also learned to use R during this period. In terms of his future, Andres aims to leverage the skills he has acquired in computer science, mathematics, and biology to further his career into the field of genetics, either agricultural or bio-medical.

HONORS & AWARDS:

- Florida Bright Futures Academic Scholar

Issued by State of Florida · May 2019

- Orlando Health Scholarship

Issued by Orlando Health, Health Central Hospital · May 2019

- National Academy of Future Physicians and Medical Scientists Award of Excellence

Issued by National Academy of Future Physicians and Medical Scientists · Aug 2018