



Communication

Technological Convergence: Highlighting the Power of CRISPR Single-Cell Perturbation Toolkit for Functional Interrogation of Enhancers

Reza Ghamsari ^{1,2} , Joseph Rosenbluh ³, A Vipin Menon ², Nigel H. Lovell ^{4,5}  and Hamid Alinejad-Rokny ^{1,6,*}

¹ BioMedical Machine Learning Lab (BML), The Graduate School of Biomedical Engineering, UNSW Sydney, Sydney, NSW 2052, Australia

² Epigenetics and Development Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3052, Australia; menon.v@wehi.edu.au

³ Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Melbourne, VIC 3800, Australia; sefi.rosenbluh@monash.edu

⁴ The Graduate School of Biomedical Engineering, UNSW Sydney, Sydney, NSW 2052, Australia

⁵ Tyree Institute of Health Engineering (IHealthE), UNSW Sydney, Sydney, NSW 2052, Australia

⁶ UNSW Data Science Hub, UNSW Sydney, Sydney, NSW 2052, Australia

* Correspondence: h.alinejad@unsw.edu.au; Tel.: +61-2-9385-3911

Simple Summary: Enhancers serve as logic gates of the regulatory mechanism of gene expression, and their malfunction is associated with numerous diseases. Therefore, the functional validation of enhancer elements is of great importance in genomics research. Recent technological advancements have enabled the perturbation of enhancers and the examination of their impact on the expression of nearby genes. Here, we review the progress made in experimental and computational methods, which have equipped researchers with a promising arsenal to uncover relationships between enhancers and phenotypes, providing mechanistic insights into diseases.

Abstract: Higher eukaryotic enhancers, as a major class of regulatory elements, play a crucial role in the regulation of gene expression. Over the last decade, the development of sequencing technologies has flooded researchers with transcriptome-phenotype data alongside emerging candidate regulatory elements. Since most methods can only provide hints about enhancer function, there have been attempts to develop experimental and computational approaches that can bridge the gap in the causal relationship between regulatory regions and phenotypes. The coupling of two state-of-the-art technologies, also referred to as *crisprQTL*, has emerged as a promising high-throughput toolkit for addressing this question. This review provides an overview of the importance of studying enhancers, the core molecular foundation of *crisprQTL*, and recent studies utilizing *crisprQTL* to interrogate enhancer-phenotype correlations. Additionally, we discuss computational methods currently employed for *crisprQTL* data analysis. We conclude by pointing out common challenges, making recommendations, and looking at future prospects, with the aim of providing researchers with an overview of *crisprQTL* as an important toolkit for studying enhancers.

Keywords: single-cell perturbation; enhancer; CRISPR; *crisprQTL*; epigenome



Citation: Ghamsari, R.; Rosenbluh, J.; Menon, A.V.; Lovell, N.H.; Alinejad-Rokny, H. Technological Convergence: Highlighting the Power of CRISPR Single-Cell Perturbation Toolkit for Functional Interrogation of Enhancers. *Cancers* **2023**, *15*, 3566. <https://doi.org/10.3390/cancers15143566>

Academic Editor: Shigenori Kadowaki

Received: 24 May 2023

Revised: 30 June 2023

Accepted: 3 July 2023

Published: 11 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The human genome contains about 19,000 protein-coding genes, which make up less than 1–2% of the genome. The expression of these genes is regulated by DNA elements that occupy the remaining part of our DNA, making up approximately 99% and known as ‘dark matter’. It is estimated that about 15 million transcription factor binding sites are located in more than 3 million regulatory DNA regions [1]. Understanding the mechanisms of gene expression regulation in cells is a necessity and is at the forefront of genomics research.

Non-coding regions play a crucial role in gene expression regulation, and it is estimated that more than 90% of disease- and trait-associated variants fall in these regions [2,3].

Our fascination with enhancers dates back to the 1980s [4]. Enhancers are non-coding sequences and a major class of cis-regulatory elements. They can regulate gene expression independent of their distances and orientation to the transcription starting site when bound by transcription factors [5]. Many causal non-coding loci, identified by genome-wide association studies (GWAS) that are associated with human diseases, are located in enhancer regions [6]. The contributions of single-nucleotide polymorphisms (SNPs) located in enhancers to several diseases, developmental disorders, tumorigenesis, and cancers, underscore the therapeutic potential of enhancers [7].

The Encyclopaedia of DNA Elements (ENCODE) project and the Roadmap Epigenomics Program have identified key features and characteristics of active enhancers, revealing a significant view of the global map of regulatory elements [8,9]. Despite a large body of work focused on enhancer prediction through biochemical and structural tools such as profiling of chromatin accessibility and histone marks, distinguishing between functional and non-functional enhancers remains one of the greatest challenges in the field. Besides the large distance between enhancers and their target genes, the cell-type-specific gene regulation makes predicting their target genes challenging. Advances in large-scale parallel perturbations of enhancers allow for analysis of enhancer–phenotype correlations and the proposal of causal relationships. Inspired by the research surrounding the expression-quantitative trait loci (eQTL), Gasperini et al. developed an experimental strategy, termed *crisprQTL*. In an analogous manner, they replaced individuals, variants, and tissue-level RNA sequencing in eQTL with cells, various combinations of sgRNAs, and single-cell RNA sequencing (scRNAseq) readouts in *crisprQTL*, respectively [10].

This review highlights the use of *crisprQTL* as a versatile toolkit for dissecting enhancer functions.

2. A High-Throughput Toolkit for Precision Epigenome Editing

The emergence of precision editing tools targeting the epigenome, without altering the genome, provides a versatile approach in interrogating causal relationships between the epigenome and transcriptome [11]. Different DNA-binding proteins such as zinc finger proteins (ZFPs), transcription-activator-like effectors (TALEs), and clustered regularly interspaced short palindromic repeats (CRISPR)/nuclease-deactivated Cas9 (dCas9), have been employed to target fused epigenetic modifiers to specific loci in the genome. Among these, dCas9 offers a robust adjustable DNA recognition mechanism without altering the underlying DNA sequence [12].

CRISPR was first discovered in the adaptive immune systems of bacteria and has advanced our understanding of gene function [13]. The CRISPR-Cas9 system consists of two main components: a Cas9 endonuclease and a guide RNA molecule, the latter directing the first component to a specific genomic location. Deactivated Cas9 coupled with various activators or repressors has been employed to investigate enhancer function. Compared to Cas9, dCas9 does not cut the DNA, but can be linked to chromatin modifiers such as Krüppel-associated box (KRAB) to induce chromatin changes [14,15]. The fusion of KRAB, a transcriptional repressor domain, with dCas9 equipped researchers with an incredibly powerful tool known as CRISPR interference (CRISPRi) [16]. A single-guide RNA (sgRNA) directs dCas9-KRAB to the genomic location, wherein KRAB induces a heterochromatin state of DNA, thereby promoting histone methylation and deacetylation. The precise epigenome editing capability of the dCas9-KRAB platform, enables the targeting of candidate enhancers in a highly specific manner [17]. A recent study has identified a novel putative enhancer located upstream of colony-stimulating factor 1 (CSF1), a highly expressed gene in triple-negative breast cancer cells. They used CRISPRi assays (dCas9-KRAB) followed by bulk RNA sequencing to validate the contribution of this enhancer to CSF1 expression [18]. Despite the capability of this precise perturbation in enhancing understanding of the correlation between enhancers and phenotypes, the need for a more

high-throughput toolkit becomes evident when investigating multiple perturbations within cell type and state-specific scope.

The introduction of scRNA-seq has significantly deepened the understanding of individual cells. Although single-cell technologies provide unprecedented descriptive information, they should be combined with other functional genomics tools to study causality of genes [19]. By combining CRISPRi and scRNA-seq toolkits, Perturb-seq and CRISP-seq have equipped scientists with new powerful tools to investigate and explore genomic element interactions. In these methods, sgRNA was not directly detectable via scRNA-seq due to the lack of a poly (A) tail [20–22]. However, by including the sgRNA at the 3' end of a polyadenylated mRNA transcript, CROP-seq made the direct sequencing of sgRNAs possible and further simplified the construction of sgRNA libraries [23].

One such approach, 'direct-capture Perturb-seq', implemented a hybridization-based target enrichment strategy, offered a flexible guide design and targeted sequencing in a cost-efficient way. The 5' and 3' platforms of this method were commercialized by 10× Genomics [24]. By integrating direct-capture Perturb-seq with CRISPR gene-tiling screen, Sc-Tiling was developed to investigate the functional domain of a gene at a sub-gene resolution [25]. In another study, expanded CRISPR-compatible cellular indexing of transcriptomes and epitopes by sequencing (ECCITE-seq) combined crisprQTL with cell surface antigen readouts to detect several modalities at a single-cell resolution. Given that protein detection is more sensitive than mRNA, ECCITE-seq, by detecting both modalities, can provide a more robust characterization of cellular phenotypes in response to perturbations. This can be particularly useful in studying heterogeneous cell populations in diseases like cancer [26]. Later, ECCITE-seq was applied to identify new regulatory genes involved in the differentiation mechanism of acute myeloid leukemia [27]. In another study, CRISPR–sciATAC was developed to capture changes in chromatin states following single-cell perturbation screens. This method can be applied to investigate the interplay between specific genetic variants and the chromatin accessibility profile, which can ultimately provide valuable insight into the underlying mechanisms of phenotypes and diseases [28]. Linking genotypes to the phenotype of individual cells has been achieved through the integration of CRISPR perturbations with image-based phenotyping in optical pooled screens. This approach facilitates the measurement of phenotypic features such as cellular morphology, cell–cell interactions, and the localization of molecules at single-cell resolution [29].

The use of crisprQTL to alter a large number of enhancers (pooled screens) in a large population of cells (Figure 1) serves as an efficient and high-throughput platform for the interrogation and validation of the function of enhancers in their native genomic contexts.

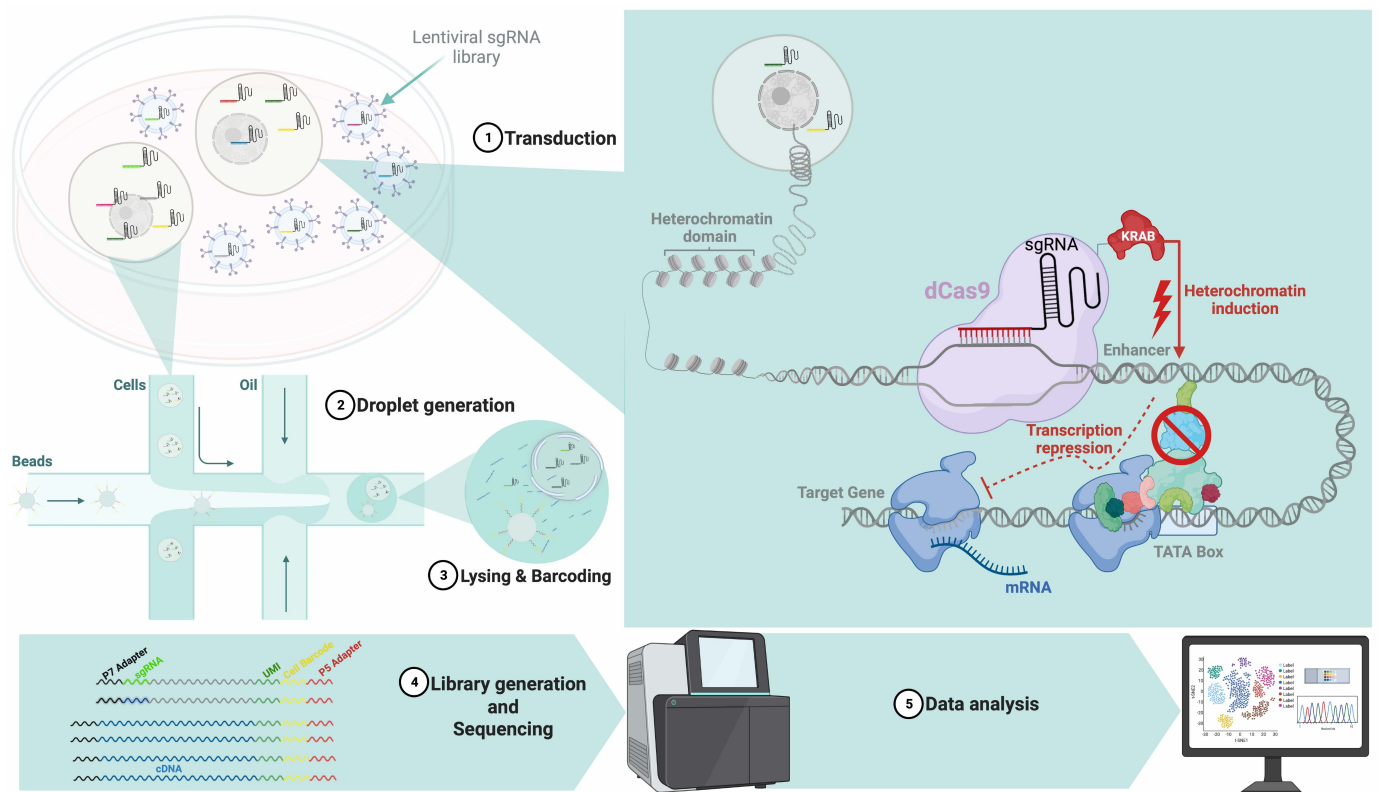


Figure 1. The overall workflow of the crisprQTL toolkit: (1) Transduction: lentiviral delivery of sgRNAs library designed to target candidate regions and the generation of stable cell lines at high multiplicity of infection (MOI). (2) Droplet generation: after screening and selection for the phenotype of interest, polyclonal cells go through droplet generation. (3) Lysing and barcoding: cell lysis is followed by reverse transcription and barcoding. (4) Library generation: pooling samples is followed by PCR to prepare the library ahead of sequencing. (5) Data analysis: data processing, mapping, and assigning sgRNAs to cells.

3. Enhancer Interrogation via crisprQTL

The pace of utilizing crisprQTL in functional genomics research has been astonishing, and so far, several research studies have been exploiting this toolkit to functionally explore enhancers. In one study, Mosaic-seq was employed to systematically perturb enhancers and to analyze their activity in both individual and combinational manners. Xie et al. showed that perturbing an enhancer has as significant an impact on gene expression, as targeting the gene itself. They also found that while only a few enhancers within each large cluster of enhancers, also known as super-enhancers, play a major role in repression, perturbing multiple weak enhancers can also change the expression of the target gene [30]. Later, they introduced Mosaic-seq2, which was optimized for the 10× Genomics platform, and also adopted polyadenylation from the CROP-seq method, as mentioned above [31].

In another study, this team optimized their method by increasing the number of utilized sgRNAs per enhancer and performing a gene-specific *p*-value correction. They used this method to identify the secondary target genes regulated by enhancers and to construct an enhancer-driven gene regulatory network. They showed that targeting an enhancer that regulates a transcription factor can result in downregulation of multiple secondary targets [32].

In a separate study conducted by Gasperini et al., crisprQTL mapping was utilized to investigate the presence of multiple sgRNAs within each individual cell. By avoiding an assumption about the target gene, crisprQTL avoids the limitations of gene-specific assays. They showed that introducing sgRNA at a high multiplicity of infection (even to ~28) within individual cells does not reduce the power of CRISPRi [10].

CROP-seq-based methods have certain drawbacks when it comes to genes with a low level of gene expression. Targeted Perturb-seq (TAP-seq) has increased the sensitivity of these approaches by amplifying only a limited number of genes of interest ahead of sequencing, thereby reducing sequencing requirements. TAP-seq not only increases the sensitivity of capturing low-level RNA transcripts but also increases the efficiency of retrieving the gRNA identity. This method also addresses the multiple hypothesis testing problems by focusing on a predefined set of genes and targeted readout of their relevant transcripts [33]. In another study, systematic targeting and inhibition of non-coding GWAS loci with single-cell sequencing (STING-seq) utilized crisprQTL to study gene-regulatory functions of multiple GWAS variants. This approach connects GWAS variants with their respective target genes, enabling high-throughput exploration of relationships between genetic variants and their functional outcomes, and holds a potential to discover therapeutic targets [34].

A recent study by Armendariz et al. employed crisprQTL to investigate the contribution of enhancers to developmental diseases by examining how they alter cell fate determination. They functionally interrogated 25 enhancers associated with congenital heart defects in stem cells, ahead of their differentiation to cardiomyocytes. They measured the transcriptome and found a subset of 16 enhancers where the perturbation caused a delay in the specification of human cardiomyocytes [35]. In another study, a multi-assay evaluation of GWAS target genes was carried out using functional CRISPR screens, High-throughput Chromosome Conformation Capture with Chromatin Immunoprecipitation (HiChIP), and crisprQTL to evaluate the relationship between candidate breast cancer-associated enhancers and their target genes [36].

Most recently, crisprQTL has been implemented in a comprehensive study, which focused on regions containing mutations associated with psychiatric disorders, perturbing almost one thousand enhancers in primary human astrocytes. They found almost one hundred active enhancers that contain disease-associated variants and additionally pinpointed 140 enhancer target genes linked to these disorders [37].

To date, most functional studies have focused on unraveling the mystery of gene regulation by perturbing enhancers at a DNA level. However, a recent study introduced Cas13 RNA Perturb-seq (CaRPool-seq) that can perform multiplexed combinatorial perturbations by targeting enhancer RNAs [38].

4. Computational and Statistical Toolkits

In parallel with the development of the experimental procedures of crisprQTL, there is an ongoing effort to develop computational methods to assist downstream analysis and improve the interpretation of generated data.

Large-scale functional studies have been conducted to evaluate the efficiency of the Cas system as a toolkit for functional genomics [39–42]. However, the design of sgRNA has been a major bottleneck across these studies, given its crucial role in determining the success of knockouts or knockdowns. To address this issue, several computational models, based on these functional studies, have been developed with the aim to identify design principles for highly effective sgRNA [43–46]. These models incorporate state-of-the-art machine learning [43,46] and deep learning [47,48] models to identify biological features that influence the sgRNA efficiency. This research has led to the development of high-performing libraries of sgRNA [46]. Both experimental and computational studies have shown that features related to the structural stability and flexibility of sgRNA, such as GC content and self-folding energy significantly contribute to the efficiency of the CRISPR system [43,44]. Tools are available online and via the command line for scoring highly effective sgRNA for any given library.

For the downstream analyses, a few public datasets exist for functional analysis of enhancers using crisprQTL. In two early reported datasets, different statistical methods were applied to analyze differential expression. In the first one, Xie et al. [30] used nonparametric test for independence, Chi-squared-like tests, which had limitations in dealing with

technical confounders. In the second one, Gasperini et al. [10] applied negative binomial general linear models (GLMs), which were not robust to the misspecification of the gene expression model. These two datasets were later reanalyzed with SCEPTRE and GLM-EIV. SCEPTRE implemented a form of conditional randomization test to address the miscalibration problem highlighted in the two aforementioned studies. By examining negative control sgRNAs, SCEPTRE provided a reassessment of the calibration of these methods. Upon applying SCEPTRE, a new set of enhancers (>40) was introduced to Gasperini's dataset, while a set of introduced enhancers (>20) in the original research was removed [49]. SCEPTRE demonstrates the sensitivity and specificity of crisprQTL and addresses the previously mentioned limitations. However, like other standard analysis methods, SCEPTRE imputes perturbation assignments onto cells by applying the thresholded regression method on the sgRNA counts and assumes away measured noise. In another attempt, a method known as GLM-based errors-in-variables (GLM-EIV), used a new class of measurement error models to select a threshold and estimate the probability of perturbation in each cell, as well as the effect size of it on differential gene expression. GLM-EIV was applied to both datasets and outperformed the thresholded regression method in settings with high background contamination, where an sgRNA incorrectly assigned to a cell [50].

In another computational approach, the single-cell model-based analysis of genome-wide CRISPR/Cas9 knockout (scMAGECK) utilized the negative binomial distribution, a generalized linear model, and expectation maximization approach to model perturbation. This approach introduced two different modules. (1) The first module can detect enhancers linked to only one single target gene, based on the robust rank aggregation algorithm. (2) The second module can assess the impact of perturbation on thousands of target genes based on linear regression [51].

In a different study, a computational pipeline was developed to remove the clonal cell expansion effect in crisprQTL experiments, leading to significant reduction of false discoveries. This pipeline leverages the combination of multiple sgRNAs in each cell, using them as a barcode to identify distinct clones [52]. In another separate study, GLiMMIRS was developed using negative binomial generalized linear models, accounting for sgRNA efficiency and several other covariates. Reanalysis of Gasperini's dataset [10] using this method, which investigated interactions between 3,808 enhancer pairs, showed that enhancers act in a multiplicative manner; however, the results provided no evidence for strong interactions between pairs of enhancers [53]. Most recently, SCREE has been introduced as a comprehensive workflow that facilitates crisprQTL data analysis. This flexible pipeline offers a platform for performing pre-processing and downstream analysis of RNA sequencing data, assay for transposase-accessible chromatin with sequencing (ATAC-seq) data, and multimodal 10x-based readout. It is implemented in open-source Python and R packages, and is accompanied by a tutorial [54].

5. Pitfalls and Recommendations

The experimental design of a pooled CRISPR screen was comprehensively described in a recent review [55]. Despite the great promise of crisprQTL, a plethora of variables may affect the experiment and should be considered prior to the experiment design and during the data analysis. The number of perturbations in each cell depends on the number of sgRNA stochastically introduced into that cell [33]. One of the primary stages of the experiment is designing optimal sgRNA. However, the design is limited by requirements of the protospacer adjacent motif (PAM), which restricts the number of target regions, including those involved in gene regulation. Another limitation is the trade-off between efficiency and specificity; sgRNAs with high specificity may have lower efficiency, whereas sgRNAs with high efficiency may have a higher likelihood of off-target effects. Computational prediction of off-target candidate regions can help overcome this limitation [56,57]. Recently, significant progress has been made in both experiment and computational methods to overcome these limitations. A recent study presented high-performance computational models for assessing the impact of sgRNA on different targets across Cas9 variants and

orthologs. The study generated high-throughput datasets for all Cas9 variants [58] and orthologs [59], and identified relevant biological features, such as GC content, melting temperature, and self-folding energy of spacer sequence, that influence sgRNA activity. The paper provides a web tool link for researchers to utilize variants and orthologs based on experiment needs. Although the current designs are highly efficient, further improvements can be achieved by integrating the structural features of both the spacer and scaffold of sgRNA. Previous research has suggested that altering the secondary structure of sgRNA through engineering sequences within tetra loop regions improves editing efficiency [60]. For assessing off-target effects, both heuristic and machine-learning-based methods have been developed, which could be incorporated to identify highly active and specific sgRNAs. However, they currently work only for the human genome and, to some extent, for mice. Assessments on major model organisms like *Drosophila*, zebrafish, and *C. elegans*, have yet to be achieved [56,61]. In the context of CRISPR multiplexing perturbations, the competition of several sgRNAs for endonucleases in a cell, known as retroactive effects, can change the efficacy of every sgRNA [62]. A focused validation, through targeting enhancers by an active Cas9 protein, can also help to balance speed with accuracy [35].

KRAB induces histone deacetylation, H3K9 methylation and a heterochromatin domain formation. During an investigation of enhancers located more than 1 kb distant from the transcription start site, it was observed that the KRAB effect does not spread toward promoter linearly along the DNA [33]. Moreover, due to three-dimensional interactions in the genome, KRAB can induce heterochromatin formation in a non-linear manner [17]. Therefore, the repression of the putative gene might be induced by KRAB rather than the enhancer itself. The extent of spreading of the KRAB effect within DNA depends on different factors, including time, and can range from 1 kb to 200 kb. Other factors, such as the level of activation, genomic locus, and neighboring elements like insulators, enhancers, and promoters can also influence this spreading [63]. Considering the time dependency of enhancer functions, a delay in the KRAB function could pose a technical challenge while sgRNAs are introduced at a fixed time point [35].

The interaction between neighboring cells can confound gene expression and should be taken into account [64]. Gene expression can be regulated by multiple enhancers, which may mask the effect of perturbation. Moreover, the activity of enhancers and their contribution to gene expression are heterogeneous across a cell population. It is also important to note that enhancers can compensate for one another, therefore sometimes targeting multiple weak enhancers among some super-enhancers is necessary to repress gene expression [30,65].

As discussed above, despite the high efficiency of CRISPRi for perturbation across a cell population, it is susceptible to incomplete inhibition. The variability in sgRNA targeting efficacy and dCas9-KRAB perturbation should be considered [10], and GLiMMIRS uses a statistical model to account for this covariate [53].

While a high ratio of a lentiviral transduction to the number of target cells, known as the multiplicity of infection (MOI), can improve the sensitivity and specificity of the screen [51], an excessively high MOI could also affect the physiological response of the cell. Therefore, it is important to consider the MOI tolerance threshold for different cell types, especially when most sgRNAs have an effect. Thus, it might be necessary to carefully optimize and calibrate the MOI tolerance. Multiple perturbations in a cell can either facilitate or impede the cell proliferation rate and trigger a clonal expansion, introducing a potential bias in sampling. By considering the clonal expansion artifacts during the experimental design and applying computational methods to filter out clonal cells, biases, and false discoveries can be minimized [52].

From a technical perspective, scRNA-seq data is intrinsically noisy data. Various studies have addressed data correction and noise reduction [66]. A low amount of start material can affect the success of screening, particularly for genes with low expression levels. Targeted sequencing or high-content read-outs can address this issue, although the latter is limited by sequencing costs [33]. Simultaneous profiling of gene expression

and other modalities within the same cell, such as multiome [67] and CITE-seq [26], offer potential solutions to overcome the drop-out limitation associated with the low abundance of transcripts.

Finally, to avoid unintended interferences and biases, it is crucial to consider and evaluate the influencing variables mentioned above prior to, during, and after the experimental process, throughout the data analysis.

6. Conclusions and Future Prospective

The crisprQTL toolkit has enabled a facile parallel perturbation of a large number of enhancers and the analysis of the resulting phenotype in each cellular compartment. We highlighted several recent research studies, and collectively, these studies have demonstrated the potential of crisprQTL in investigating the correlation between enhancers and their target genes. With an ever-growing generation of epigenomic data, such as enhancer-RNA expression, high-throughput chromosome conformation capture (Hi-C), ATAC-seq, and chromatin immunoprecipitation, followed by sequencing (ChIP-seq), introducing more candidate enhancers, the functional testing of enhancers will be indispensable. Astonishing progress has been made in single-cell multi-omics approaches, specifically the commercial multi-omics product 10x Multiome [67], which provides an unprecedented opportunity to profile the gene expression and chromatin accessibility within the same cell. Recruiting multiome instead of scRNA with the crisprQTL toolkit can multiply the power of enhancer interrogation. In the future, the incorporation of long-read sequencing and computational tools such as FLAMES (full-length alternative splicing quantification) [68] into the crisprQTL toolkit could also lead to significant advances in our understanding of enhancer roles in splicing and regulation of isoform usage. Other omics modalities could be integrated into the crisprQTL toolkit, along with advancements in genomics, to significantly deepen our understanding of the way enhancers regulate gene expression. In the near term, this will depend on the development of computational and statistical methods. We envision that crisprQTL, as a promising toolkit, can change the landscape of high-throughput functional epigenomics and play a pivotal role in the comprehensive analysis and interrogation of non-coding regions, which could ultimately pave the way to the discovery of novel therapeutic targets.

Author Contributions: R.G. designed the study. R.G. wrote the manuscript. R.G., H.A.-R., J.R. and N.H.L. edited the manuscript. A.V.M. wrote the sgRNA designing section. R.G. edited this section. R.G. generated the figure. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the UNSW Scientia Program Fellowship to HAR; and the Australian Research Council Discovery Early Career Researcher Award (DECRA), under grant DE220101210 to HAR.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank ChatGPT (GPT-4, OpenAI's large-scale language-generation model) for its contribution in meticulously reviewing and refining the grammar and punctuation of the final version of the manuscript (after acceptance). The assistance provided by this language model was invaluable in improving the overall quality of the written text.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chi, K.R. The dark side of the human genome. *Nature* **2016**, *538*, 275–277. [[CrossRef](#)] [[PubMed](#)]
2. Corradin, O.; Scacheri, P.C. Enhancer variants: Evaluating functions in common disease. *Genome Med.* **2014**, *6*, 85. [[CrossRef](#)] [[PubMed](#)]

3. Maurano, M.T.; Humbert, R.; Rynes, E.; Thurman, R.E.; Haugen, E.; Wang, H.; Reynolds, A.P.; Sandstrom, R.; Qu, H.; Brody, J.; et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **2012**, *337*, 1190–1195. [[CrossRef](#)] [[PubMed](#)]
4. Banerji, J.; Rusconi, S.; Schaffner, W. Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* **1981**, *27*, 299–308. [[CrossRef](#)]
5. Long, H.K.; Prescott, S.L.; Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **2016**, *167*, 1170–1187. [[CrossRef](#)]
6. Nasser, J.; Bergman, D.T.; Fulco, C.P.; Guckelberger, P.; Doughty, B.R.; Patwardhan, T.A.; Jones, T.R.; Nguyen, T.H.; Ulirsch, J.C.; Lekschas, F.; et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature* **2021**, *593*, 238–243. [[CrossRef](#)]
7. Nair, S.J.; Suter, T.; Wang, S.; Yang, L.; Yang, F.; Rosenfeld, M.G. Transcriptional enhancers at 40: Evolution of a viral DNA element to nuclear architectural structures. *Trends Genet.* **2022**, *38*, 1019–1047. [[CrossRef](#)]
8. Moore, J.E.; Purcaro, M.J.; Pratt, H.E.; Epstein, C.B.; Shores, N.; Adrian, J.; Adrian, J.; Kawli, T.; Davis, C.A.; Dobin, A.; et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **2020**, *583*, 699–710.
9. Roadmap Epigenomics, C.; Kundaje, A.; Meuleman, W.; Ernst, J.; Bilenky, M.; Yen, A.; Heravi-Moussavi, A.; Kheradpour, P.; Zhang, Z.; Wang, J.; et al. Integrative analysis of 111 reference human epigenomes. *Nature* **2015**, *518*, 317–330.
10. Gasperini, M.; Hill, A.J.; McFaline-Figueroa, J.L.; Martin, B.; Kim, S.; Zhang, M.D.; Jackson, D.; Leith, A.; Schreiber, J.; Noble, W.S.; et al. A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell* **2019**, *176*, 1516. [[CrossRef](#)]
11. O’geen, H.; Tomkova, M.; Combs, J.A.; Tilley, E.K.; Segal, D.J. Determinants of heritable gene silencing for KRAB-dCas9 + DNMT3 and Ezh2-dCas9 + DNMT3 hit-and-run epigenome editing. *Nucleic Acids Res.* **2022**, *50*, 3239–3253. [[CrossRef](#)]
12. Brocken, D.J.; Tark-Dame, M.; Dame, R. dCas9: A Versatile Tool for Epigenome Editing. *Curr. Issues Mol. Biol.* **2018**, *26*, 15–32. [[CrossRef](#)]
13. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D.A.; Horvath, P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007**, *315*, 1709–1712. [[CrossRef](#)]
14. Li, K.; Liu, Y.; Cao, H.; Zhang, Y.; Gu, Z.; Liu, X.; Yu, A.; Kaphle, P.; Dickerson, K.E.; Ni, M.; et al. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat. Commun.* **2020**, *11*, 485. [[CrossRef](#)]
15. Pulecio, J.; Verma, N.; Mejía-Ramírez, E.; Huangfu, D.; Raya, A. CRISPR/Cas9-Based Engineering of the Epigenome. *Cell Stem Cell* **2017**, *21*, 431–447. [[CrossRef](#)] [[PubMed](#)]
16. Alerasool, N.; Segal, D.; Lee, H.; Taipale, M. An efficient KRAB domain for CRISPRi applications in human cells. *Nat. Methods* **2020**, *17*, 1093–1096. [[CrossRef](#)] [[PubMed](#)]
17. Thakore, P.I.; D’ippolito, A.M.; Song, L.; Safi, A.; Shivakumar, N.K.; Kabadi, A.M.; Reddy, T.E.; Crawford, G.E.; Gersbach, C.A. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **2015**, *12*, 1143–1149. [[CrossRef](#)]
18. Lewis, M.W.; Wisniewska, K.; King, C.M.; Li, S.; Coffey, A.; Kelly, M.R.; Regner, M.J.; Franco, H.L. Enhancer RNA Transcription Is Essential for a Novel CSF1 Enhancer in Triple-Negative Breast Cancer. *Cancers* **2022**, *14*, 1852. [[CrossRef](#)] [[PubMed](#)]
19. Lee, S.; Kim, J.; Park, J.-E. Single-Cell Toolkits Opening a New Era for Cell Engineering. *Mol. Cells* **2021**, *44*, 127–135. [[CrossRef](#)]
20. Dixit, A.; Parnas, O.; Li, B.; Chen, J.; Fulco, C.P.; Jerby-Arnon, L.; Marjanovic, N.D.; Dionne, D.; Burks, T.; Raychowdhury, R.; et al. Perturb-Seq: Dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **2016**, *167*, 1853–1866.e17. [[CrossRef](#)]
21. Jaitin, D.A.; Weiner, A.; Yofe, I.; Lara-Astiaso, D.; Keren-Shaul, H.; David, E.; Salame, T.M.; Tanay, A.; van Oudenaarden, A.; Amit, I. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **2016**, *167*, 1883–1896.e15. [[CrossRef](#)]
22. Adamson, B.; Norman, T.M.; Jost, M.; Cho, M.Y.; Nuñez, J.K.; Chen, Y.; Villalta, J.E.; Gilbert, L.A.; Horlbeck, M.A.; Hein, M.Y.; et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **2016**, *167*, 1867–1882.e21. [[CrossRef](#)]
23. Datlinger, P.; Rendeiro, A.F.; Schmidl, C.; Krausgruber, T.; Traxler, P.; Klughammer, J.; Schuster, L.C.; Kuchler, A.; Alpar, D.; Bock, C. Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **2017**, *14*, 297–301. [[CrossRef](#)]
24. Replogle, J.M.; Norman, T.M.; Xu, A.; Hussmann, J.A.; Chen, J.; Cogan, J.Z.; Meer, E.J.; Terry, J.M.; Riordan, D.P.; Srinivas, N.; et al. Combinatorial single-cell CRISPR screens by direct guide RNA capture and targeted sequencing. *Nat. Biotechnol.* **2020**, *38*, 954–961. [[CrossRef](#)] [[PubMed](#)]
25. Yang, L.; Chan, A.K.N.; Miyashita, K.; Delaney, C.D.; Wang, X.; Li, H.; Pokharel, S.P.; Li, S.; Li, M.; Xu, X.; et al. High-resolution characterization of gene function using single-cell CRISPR tiling screen. *Nat. Commun.* **2021**, *12*, 4063. [[CrossRef](#)]
26. Mimitou, E.P.; Cheng, A.; Montalbano, A.; Hao, S.; Stoeckius, M.; Legut, M.; Roush, T.; Herrera, A.; Papalexi, E.; Ouyang, Z.; et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **2019**, *16*, 409–412. [[CrossRef](#)] [[PubMed](#)]
27. Wang, E.; Zhou, H.; Nadorp, B.; Cayanan, G.; Chen, X.; Yeaton, A.H.; Nomikou, S.; Witkowski, M.T.; Narang, S.; Kloetgen, A.; et al. Surface antigen-guided CRISPR screens identify regulators of myeloid leukemia differentiation. *Cell Stem Cell* **2021**, *28*, 718–731.e6. [[CrossRef](#)] [[PubMed](#)]

28. Liscovitch-Brauer, N.; Montalbano, A.; Deng, J.; Méndez-Mancilla, A.; Wessels, H.-H.; Moss, N.G.; Kung, C.-Y.; Sookdeo, A.; Guo, X.; Geller, E.; et al. Profiling the genetic determinants of chromatin accessibility with scalable single-cell CRISPR screens. *Nat. Biotechnol.* **2021**, *39*, 1270–1277. [[CrossRef](#)]
29. Feldman, D.; Funk, L.; Le, A.; Carlson, R.J.; Leiken, M.D.; Tsai, F.; Soong, B.; Singh, A.; Blainey, P.C. Pooled genetic perturbation screens with image-based phenotypes. *Nat. Protoc.* **2022**, *17*, 476–512. [[CrossRef](#)]
30. Xie, S.; Duan, J.; Li, B.; Zhou, P.; Hon, G.C. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol. Cell* **2017**, *66*, 285–299.e5. [[CrossRef](#)]
31. Xie, S.; Hon, G.C. Experimental and Computational Approaches for Single-Cell Enhancer Perturbation Assay. *Methods Mol. Biol.* **2019**, *1935*, 203–221. [[CrossRef](#)]
32. Xie, S.; Armendariz, D.; Zhou, P.; Duan, J.; Hon, G.C. Global Analysis of Enhancer Targets Reveals Convergent Enhancer-Driven Regulatory Modules. *Cell Rep.* **2019**, *29*, 2570–2578.e5. [[CrossRef](#)]
33. Schraivogel, D.; Gschwind, A.R.; Milbank, J.H.; Leonce, D.R.; Jakob, P.; Mathur, L.; Korbel, J.O.; Merten, C.A.; Velten, L.; Steinmetz, L.M. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nat. Methods* **2020**, *17*, 629–635. [[CrossRef](#)] [[PubMed](#)]
34. Morris, J.A.; Caragine, C.; Daniloski, Z.; Domingo, J.; Barry, T.; Lu, L.; Davis, K.; Ziosi, M.; Glinos, D.A.; Hao, S.; et al. Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science* **2023**, *380*, eadh7699. [[CrossRef](#)]
35. Armendariz, D.A.; Goetsch, S.C.; Sundarajan, A.; Sivakumar, S.; Wang, Y.; Xie, S.; Munshi, N.V.; Hon, G.C. CHD-associated enhancers shape human cardiomyocyte lineage commitment. *eLife* **2023**, *12*, e86206. [[CrossRef](#)] [[PubMed](#)]
36. Tuano, N.K.; Beesley, J.; Manning, M.; Shi, W.; Perlaza-Jimenez, L.; Malaver-Ortega, L.F.; Paynter, J.M.; Black, D.; Civitarese, A.; McCue, K.; et al. CRISPR screens identify gene targets at breast cancer risk loci. *Genome Biol.* **2023**, *24*, 59. [[CrossRef](#)] [[PubMed](#)]
37. Green, N. *Investigating Gene Expression Regulation in the Brain and Its Role in Neurodevelopmental Disorders*; UNSW Sydney: Kensington, Australia, 2023. [[CrossRef](#)]
38. Wessels, H.-H.; Méndez-Mancilla, A.; Hao, Y.; Papalexi, E.; Mauck, W.M.; Lu, L.; Morris, J.A.; Mimitou, E.P.; Smibert, P.; Sanjana, N.E.; et al. Efficient combinatorial targeting of RNA transcripts in single cells with Cas13 RNA Perturb-seq. *Nat. Methods* **2022**, *20*, 86–94. [[CrossRef](#)]
39. Chidley, C.; Darnell, A.M.; Gaudio, B.L.; Lien, E.C.; Barbeau, A.M.; Vander Heiden, M.G.; Sorger, P.K. A CRISPRi/a screening platform to study cellular nutrient transport in diverse microenvironments. *bioRxiv* **2023**. [[CrossRef](#)]
40. Kampmann, M. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chem. Biol.* **2017**, *13*, 406–416. [[CrossRef](#)]
41. Gilbert, L.; Larson, M.H.; Morsut, L.; Liu, Z.; Brar, G.A.; Torres, S.E.; Stern-Ginossar, N.; Brandman, O.; Whitehead, E.H.; Doudna, J.A.; et al. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell* **2013**, *154*, 442–451. [[CrossRef](#)]
42. Qi, L.S.; Larson, M.H.; Gilbert, L.A.; Doudna, J.A.; Weissman, J.S.; Arkin, A.P.; Lim, W.A. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **2013**, *152*, 1173–1183. [[CrossRef](#)]
43. Menon, A.V.; Sohn, J.-I.; Nam, J.-W. CGD: Comprehensive guide designer for CRISPR-Cas systems. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 814–820. [[CrossRef](#)]
44. Doench, J.G.; Hartenian, E.; Graham, D.B.; Tothova, Z.; Hegde, M.; Smith, I.; Sullender, M.; Ebert, B.L.; Xavier, R.J.; Root, D.E. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **2014**, *32*, 1262–1267. [[CrossRef](#)] [[PubMed](#)]
45. Xu, H.; Xiao, T.; Chen, C.-H.; Li, W.; Meyer, C.A.; Wu, Q.; Wu, D.; Cong, L.; Zhang, F.; Liu, J.S.; et al. Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **2015**, *25*, 1147–1157. [[CrossRef](#)]
46. Doench, J.G.; Fusi, N.; Sullender, M.; Hegde, M.; Vaimberg, E.W.; Donovan, K.F.; Smith, I.; Tothova, Z.; Wilen, C.; Orchard, R.; et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **2016**, *34*, 184–191. [[CrossRef](#)] [[PubMed](#)]
47. Chuai, G.; Ma, H.; Yan, J.; Chen, M.; Hong, N.; Xue, D.; Zhou, C.; Zhu, C.; Chen, K.; Duan, B.; et al. DeepCRISPR: Optimized CRISPR guide RNA design by deep learning. *Genome Biol.* **2018**, *19*, 80. [[CrossRef](#)]
48. Kim, H.K.; Kim, Y.; Lee, S.; Min, S.; Bae, J.Y.; Choi, J.W.; Park, J.; Jung, D.; Yoon, S.; Kim, H.H. SpCas9 activity prediction by DeepSpCas9, a deep learning-based model with high generalization performance. *Sci. Adv.* **2019**, *5*, eaax9249. [[CrossRef](#)]
49. Barry, T.; Wang, X.; Morris, J.A.; Roeder, K.; Katsevich, E. SCEPTRE improves calibration and sensitivity in single-cell CRISPR screen analysis. *Genome Biol.* **2021**, *22*, 344. [[CrossRef](#)]
50. Barry, T.; Katsevich, E.; Roeder, K. Exponential family measurement error models for single-cell CRISPR screens. *arXiv* **2022**, arXiv:2201.01879.
51. Yang, L.; Zhu, Y.; Yu, H.; Cheng, X.; Chen, S.; Chu, Y.; Huang, H.; Zhang, J.; Li, W. scMAGECK links genotypes with multiple phenotypes in single-cell CRISPR screens. *Genome Biol.* **2020**, *21*, 19. [[CrossRef](#)] [[PubMed](#)]
52. Wang, Y.; Xie, S.; Armendariz, D.; Hon, G.C. Computational identification of clonal cells in single-cell CRISPR screens. *BMC Genom.* **2022**, *23*, 135. [[CrossRef](#)] [[PubMed](#)]
53. Zhou, J.L.; Guruvayurappan, K.; Chen, H.V.; Chen, A.R.; McVicker, G.P. Genome-wide analysis of CRISPR perturbations indicates that enhancers act multiplicatively and without epistatic-like interactions. *bioRxiv* **2023**. [[CrossRef](#)]
54. Wei, H.; Han, T.; Li, T.; Wu, Q.; Wang, C. SCREE: A comprehensive pipeline for single-cell multi-modal CRISPR screen data processing and analysis. *Briefings Bioinform.* **2023**, *24*, bbad123. [[CrossRef](#)]

55. Bock, C.; Datlinger, P.; Chardon, F.; Coelho, M.A.; Dong, M.B.; Lawson, K.A.; Lu, T.; Maroc, L.; Norman, T.M.; Song, B.; et al. High-content CRISPR screening. *Nat. Rev. Methods Primers* **2022**, *2*, 8. [[CrossRef](#)]
56. Bae, S.; Park, J.; Kim, J.S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **2014**, *30*, 1473–1475. [[CrossRef](#)]
57. Kim, D.; Kang, B.-C.; Kim, J.-S. Identifying genome-wide off-target sites of CRISPR RNA-guided nucleases and deaminases with Digenome-seq. *Nat. Protoc.* **2021**, *16*, 1170–1192. [[CrossRef](#)]
58. Kim, N.; Kim, H.K.; Lee, S.; Seo, J.H.; Choi, J.W.; Park, J.; Min, S.; Yoon, S.; Cho, S.-R.; Kim, H.H. Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat. Biotechnol.* **2020**, *38*, 1328–1336. [[CrossRef](#)]
59. Kim, N.; Choi, S.; Kim, S.; Song, M.; Seo, J.H.; Min, S.; Park, J.; Cho, S.-R.; Kim, H.H. Deep learning models to predict the editing efficiencies and outcomes of diverse base editors. *Nat. Biotechnol.* **2023**, 1–14. [[CrossRef](#)] [[PubMed](#)]
60. Dang, Y.; Jia, G.; Choi, J.; Ma, H.; Anaya, E.; Ye, C.; Shankar, P.; Wu, H. Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol.* **2015**, *16*, 280. [[CrossRef](#)] [[PubMed](#)]
61. Listgarten, J.; Weinstein, M.; Kleinstiver, B.P.; Sousa, A.A.; Joung, J.K.; Crawford, J.; Gao, K.; Hoang, L.; Elibol, M.; Doench, J.G.; et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat. Biomed. Eng.* **2018**, *2*, 38–47. [[CrossRef](#)]
62. McCarty, N.S.; Graham, A.E.; Studená, L.; Ledesma-Amaro, R. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat. Commun.* **2020**, *11*, 1281. [[CrossRef](#)]
63. Lensch, S.; Herschl, M.H.; Ludwig, C.H.; Sinha, J.; Hinks, M.M.; Mukund, A.; Fujimori, T.; Bintu, L. Dynamic spreading of chromatin-mediated gene silencing and reactivation between neighboring genes in single cells. *eLife* **2022**, *11*, e75115. [[CrossRef](#)]
64. Holding, A.N.; Cook, H.V.; Markowitz, F. Data generation and network reconstruction strategies for single cell transcriptomic profiles of CRISPR-mediated gene perturbations. *Biochim. Biophys. Acta (BBA) Gene Regul. Mech.* **2019**, *1863*, 194441. [[CrossRef](#)] [[PubMed](#)]
65. Danko, C.G.; Choate, L.A.; Marks, B.A.; Rice, E.J.; Wang, Z.; Chu, T.; Martins, A.L.; Dukler, N.; Coonrod, S.A.; Wojno, E.D.T.; et al. Dynamic evolution of regulatory element ensembles in primate CD4+ T cells. *Nat. Ecol. Evol.* **2018**, *2*, 537–548. [[CrossRef](#)]
66. Luecken, M.D.; Theis, F.J. Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol. Syst. Biol.* **2019**, *15*, e8746. [[CrossRef](#)] [[PubMed](#)]
67. Houston, A. 10x Genomics Multiome ATAC+ GEX; 2023. Available online: <https://www.protocols.io/view/10x-genomics-multiome-atac-gex-n2bvj8kqpgk5/v1> (accessed on 23 May 2023).
68. Tian, L.; Jabbari, J.S.; Thijssen, R.; Gouil, Q.; Amarasinghe, S.L.; Voogd, O.; Kariyawasam, H.; Du, M.R.M.; Schuster, J.; Wang, C.; et al. Comprehensive characterization of single-cell full-length isoforms in human and mouse with long-read sequencing. *Genome Biol.* **2021**, *22*, 310. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.