

## ARTICLE OPEN



## Quantum gene regulatory networks

Cristhian Roman-Vicharra<sup>1,2</sup> and James J. Cai<sup>1,2</sup>✉

Immune regulation is a process that enables us to fight pathogens while preventing responses against 'harmless' substances using ignorance or the induction of tolerance.

'transcriptome,' mean the complete set of all the RNA molecules (called transcripts) expressed in some given entity, such as a cell, tissue, or organism

In this work, we present a quantum circuit model for inferring gene regulatory networks (GRNs) from single-cell transcriptomic data. The model employs qubit entanglement to simulate interactions between genes, resulting in competitive performance and promising potential for further exploration. We applied our quantum GRN modeling approach to single-cell transcriptomic data from human lymphoblastoid cells, focusing on a small set of genes involved in innate immunity regulation. Our quantum circuit model successfully predicted the presence and absence of regulatory interactions between genes, while also estimating the strength of these interactions. We argue that the application of quantum computing in biology has the potential to provide a better understanding of single-cell GRNs by more effectively approaching the relationship between fully interconnected genes compared to conventional statistical methods such as correlation and regression. Our results encourage further investigation into the creation of quantum algorithms that utilize single-cell data, paving the way for future research into the intersection of quantum computing and biology.

npj Quantum Information (2023)9:67 ; <https://doi.org/10.1038/s41534-023-00740-6>

## INTRODUCTION

A gene regulatory network (GRN) defines the ensemble of regulatory relationships between genes in a biological system. Inferring GRNs is a powerful approach for studying transcriptional regulation and the molecular basis of the regulatory mechanism, to understand the function of genes in processes of cellular activities<sup>1,2</sup>. A GRN is often represented as a graph—which can be signed, directed, and weighted—to depict relationships between transcription factors or regulators and their targets whose expression level is regulated. However, because the regulatory activity inside a cell is difficult to observe, measurements of static, intracellular gene expression are often used as a proxy, and the statistical dependencies are used to infer real regulatory relationships between genes.

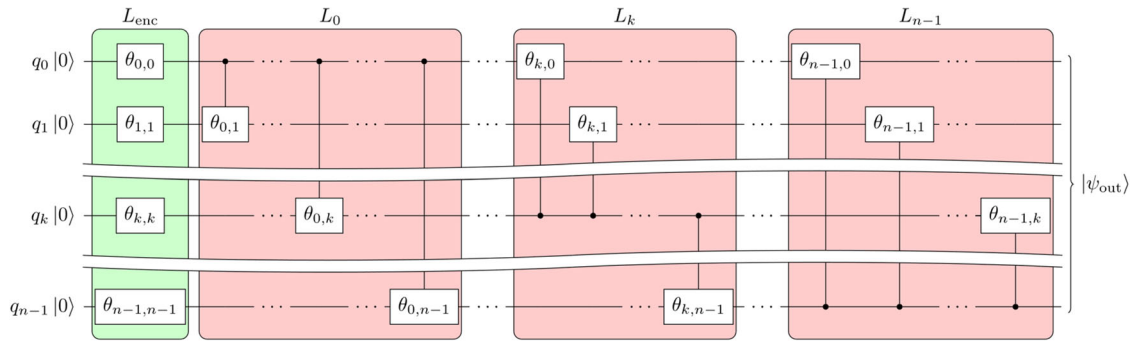
Single-cell technologies, which have recently been developed and improved, open up opportunities for studying biology at remarkable resolution and scale. Single-cell RNA sequencing (scRNA-seq), for example, allows us to measure the expression of thousands of genes in each of thousands of cells<sup>3</sup>. Computational methods for constructing GRNs can adopt scRNA-seq data and leverage the information from the sheer number of cells to improve the inference power<sup>4–6</sup>. Thus, the utilization of single-cell data can lead to the development of more detailed and precise network models, which will help us gain a better understanding of the molecular mechanisms involved in cellular activities.

Numerous computational methods have been developed for constructing GRNs. These methods use statistical approaches to detect dependencies between expression profiles of genes and establish potential regulatory relationships between genes. The typical strategies that have been employed broadly fall into several categories such as correlation, regression, information theory, Gaussian graphical model, and Bayesian and Boolean networks<sup>4–12</sup>. For a broader perspective on the topic, readers are referred to several review articles<sup>13–16</sup>. It is important to note that each method has its own set of assumptions and limitations that are not always explicitly stated<sup>17–19</sup>. More importantly, none of these conventional methods fully exploits simultaneous, inter-regulatory connections between all genes. There is still a need for a general and principled approach to model GRNs.

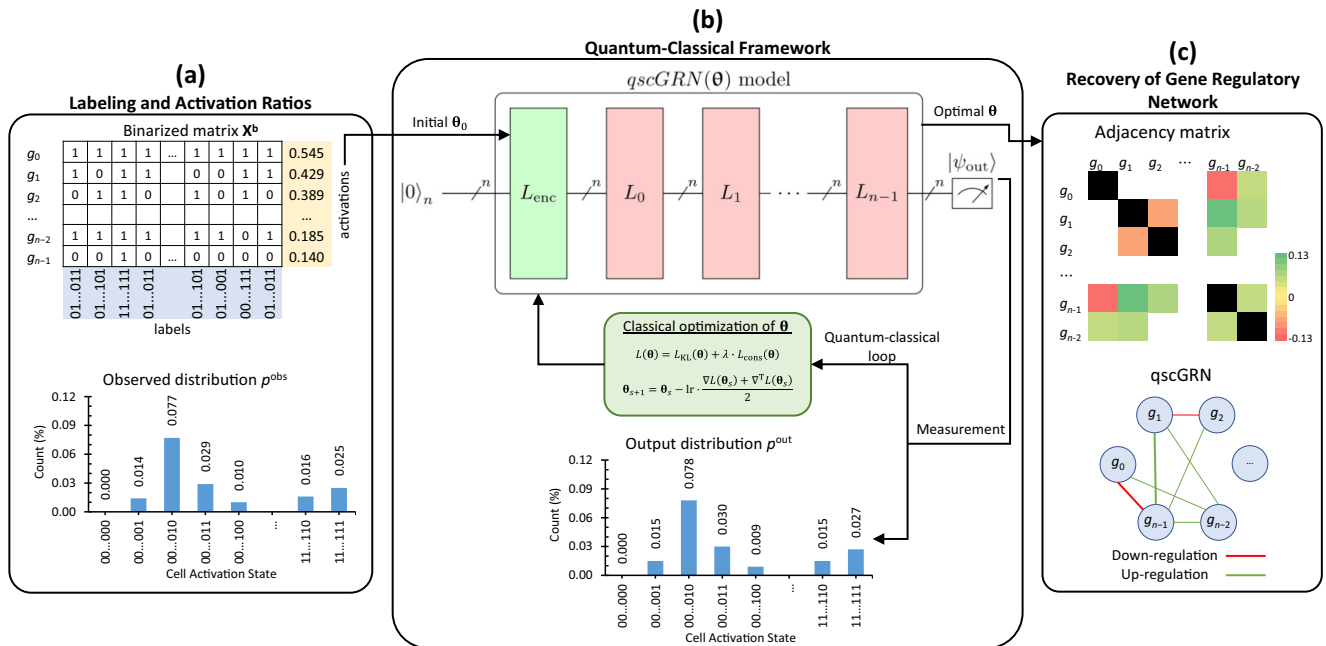
Quantum computing has become an emerging technology and an intense field of research constantly seeking applications<sup>20</sup>. Researchers have developed quantum algorithms with applications in areas such as finance, cryptography, machine learning, drug discovery, chemistry, and material science<sup>21–25</sup>. A theoretical speedup is expected in certain types of computation using quantum algorithms versus classical algorithms because a quantum computer takes advantage of superposition and entanglement phenomena during the computation<sup>26,27</sup>. Given the potential of quantum computing, conventional strategies for inferring GRNs might be expanded by taking advantage of the quantum computing framework.

In this work, we introduce a quantum single-cell GRN (qscGRN) modeling method, which is based on a parameterized quantum circuit and uses the quantum computing framework to infer biological GRNs from scRNA-seq data. In our qscGRN model, each gene is represented using a qubit, and the circuit structure is divided into two types of layers: the encoder layer that translates the scRNA-seq data into a superposition state, and the regulation layers that entangle qubits to model gene-gene interactions in the quantum framework. Our qscGRN model maps binarized gene expression values onto a large vector space, known as Hilbert space, making full use of the information in the individual cells. Thus, the signal from thousands of cells is leveraged to improve the mapping of regulatory relationships between genes. The parameterization of our qscGRN model allows gene-to-gene regulatory relationships to be inferred all at once by fitting the superposition state probabilities onto the distribution observed in the scRNA-seq data. We include a quantum-classical framework for optimizing the parameters of our qscGRN model for a given scRNA-seq data set. The classical component of our framework uses the Laplace smoothing<sup>28</sup> and the gradient descent algorithm<sup>29</sup> to perform optimization by minimizing a loss function based on Kullback-Leibler (KL) divergence<sup>30</sup>. We apply the quantum-classical framework to real scRNA-seq data sets<sup>31,32</sup> to show that gene regulatory relationships can be modeled using quantum computing, and the network recovered from the parameter-optimized quantum circuit is largely consistent with a previously published GRN<sup>33,34</sup>.

<sup>1</sup>Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA. <sup>2</sup>Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA. ✉email: [jcai@tamu.edu](mailto:jcai@tamu.edu)



**Fig. 1** The qscGRN model consisting of  $n$  qubits that models a biological GRN for  $n$  genes. The quantum circuit is composed of an encoder layer  $L_{\text{enc}}$  and regulation layers  $L_0, L_1, \dots, L_{n-1}$ .



**Fig. 2** The quantum-classical framework using the qscGRN model to infer the corresponding biological GRN. The input matrix  $X$  is binarized into matrix  $X^b$  and  $n$  genes are selected to be modelled. **a** Observed distribution  $p^{\text{obs}}$  and activation ratios  $\text{act}_k$  are computed using the labels and binarized values, respectively. The ratios  $\text{act}_k$  are used in the initial setup of the parameter  $\theta_0$ . **b** The qscGRN is trained to fit the output distribution  $p^{\text{out}}$  into  $p^{\text{obs}}$  by minimizing a loss function based on KL-divergence. **c** The adjacency and network representation of the biological GRN using the optimal parameter  $\theta$ .

## RESULTS

### The qscGRN model and its optimization framework

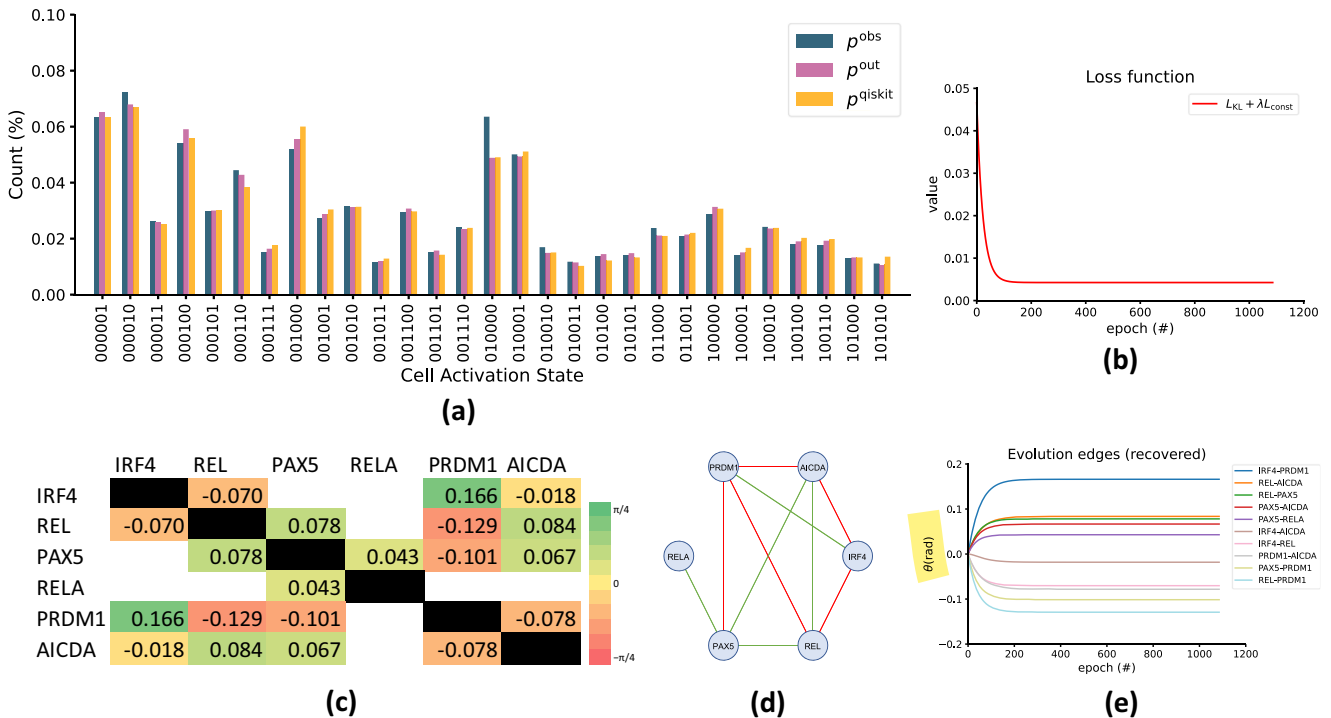
Our qscGRN model is a quantum circuit consisting of  $n$  qubits and models a biological GRN for  $n$  genes in the framework of quantum computation giving a **qubit-gene equivalence** (Fig. 1). A complete quantum-classical framework, which employs the qscGRN model to infer the corresponding biological GRN, is also introduced (Fig. 2). The methods section provides a detailed explanation of the model and its optimization framework.

### Applying qscGRN model to scRNA-seq data of lymphoblastoid cells

This section outlines the practical application of our qscGRN model in constructing a 6-gene GRN from real scRNA-seq data sets. The process began by feeding an input expression matrix, containing the expression values of 6 genes in over 28,000 lymphoblastoid cells, into the framework. The 6 genes, IRF4, REL, PAX5, RELA, PRDM1, and AICDA, are members of the NF- $\kappa$ B signaling pathway. The  $p^{\text{obs}}$  distribution was used to show the

frequencies of the  $2^6 = 64$  possible cell states mapped into a vector space. The  $p^{\text{obs}}$  is represented in blue in Fig. 3a, in which only the states with a probability greater than 0.01 are shown. The qscGRN model schema for the data set was a 6-qubit system and consisted of an encoder layer and six regulation layers. We measured the output register of the qscGRN model to recover the output distribution  $p^{\text{out}}$  from the quantum framework. Then, we optimized the parameter  $\theta$  in the qscGRN model using 1087 iterations to minimizing the loss function  $L(\theta)$ . The distribution  $p^{\text{out}}$  was fitted into  $p^{\text{obs}}$  during the optimization—smoothed distributions for  $p^{\text{out}}$  and  $p^{\text{obs}}$  show the similarity of the two distributions after optimization. The  $p^{\text{out}}$  after optimization is also represented in pink in Fig. 3a. The similarity is quantified using the loss function and error metrics that reached values of  $4.25e-3$  and  $3.21e-4$ , respectively (Fig. 3b). We validated the optimized parameter  $\theta$  running a quantum simulator that uses the Aer Simulator backend (colored in yellow in Fig. 3a).

The value of the parameter  $\theta$  after optimization retrieved an adjacency matrix (Fig. 3c), which was used to construct the biological GRN. Then, we constructed a weighted network from



**Fig. 3** Application of the qscGRN modeling with real scRNA-seq data from human lymphoblastoid cells. **a** The observed, output, and simulated frequency distributions ( $p^{obs}$ ,  $p^{out}$  and  $p^{qiskit}$ ) of cell activation states, colored in blue, pink and yellow respectively. **b** Loss function changes during training until optimization. **c** The adjacency matrix of the biological GRN. The heatmap shows the strength of gene-gene interactions. The diagonal elements are colored in black due to these parameters are not trained. **d** A weighted representation of the biological GRN recovered from the quantum circuit, where the thickness is proportional to the corresponding adjacency matrix. **e** Evolution of parameters in the qscGRN model recovered from the quantum framework during the optimization.

the quantum framework using the non-diagonal elements of  $\theta$ , as shown in Fig. 3d. We compared the sign of the element of each pair of genes with the corresponding regulatory effect in the previously published network, i.e., the baseline GRN<sup>33,34</sup>. Figure 3e shows the evolution of parameters for 10 regulator-target gene pairs in the qscGRN model during the optimization. These gene pairs are among the relationships recovered from the quantum framework, or present in the baseline NF- $\kappa$ B network<sup>33,34</sup>. Gene pairs IRF4-PRDM1, REL-AICDA, PAX5-PRDM1, REL-PRDM1 and PAX5-AICDA are correctly recovered, IRF4-AICDA incorrectly recovered, while IRF4-REL, REL-PAX5, PAX5-RELA and PRDM1-AICDA are predicted in our workflow. These recovered relationships are supported by previous studies, for example, PAX5 plays a role in the B-lineage-specific control of AICDA transcription as suggested by a previous study<sup>35</sup>. PRDM1 is a master regulator that represses PAX5 expression in B cells<sup>36</sup>. IRF4-PRDM1's regulatory relationship might be through a third-party modulator. Indeed, IRF4 is known to inhibit BCL6 expression, and because BCL6 can repress PRDM1<sup>refs. 37,38</sup>, it has been formally speculated that the effects of IRF4 on PRDM1 expression might have been mediated through inhibition of BCL6 expression<sup>39</sup>. Although, several relationships are correctly recovered, IRF4 is known to induce AICDA expression through an indirect mechanism in the NF- $\kappa$ B signaling cascade<sup>40</sup>, suggesting the inference power is still limited.

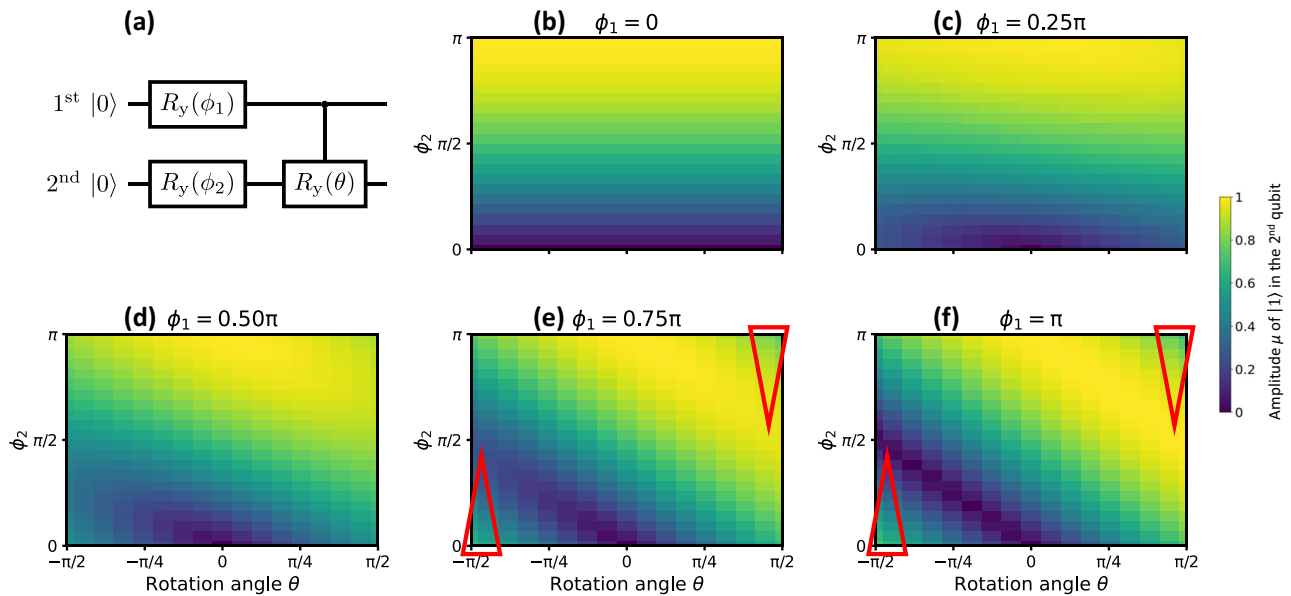
The qscGRN model predicted four regulatory relationships between genes that were not present in the published baseline GRN. These included the gene pair PRDM1 and AICDA, which may indeed interact as shown that PRDM1 can silence AICDA expression a dose-dependent manner<sup>41</sup>. These results indicate that our qscGRN method has the potential to uncover regulatory relationships that were previously missed in the baseline model.

## DISCUSSION

Finding ways to apply quantum computing in biological research is an active research area<sup>42–46</sup>. Many questions in biology can benefit from quantum computing by exploring many possible parallel computational paths, but identifying such questions remains challenging. Especially, understanding how to exploit quantum computers for progress in solving important biological questions is crucial. The latest development of scRNA-seq technology has made it possible to gather transcriptome information from tens of thousands of individual cells per assay in a high-throughput manner. These complex data sets with higher detail are driving the development of new computational and statistical tools that are revolutionizing our understanding of cellular processes. However, quantum computation has not yet received enough attention in the face of this single-cell big data revolution.

Here, we present our qscGRN method for modeling interactions between genes to derive the quantum computing framework for constructing GRNs. In the GRN inference, the interaction between two genes determines the level of production of the target gene based on the expression of a control gene, whether this interaction is promotion or repression. Similarly, the parameter in a c- $R_y$  gate indicates the degree of rotation of a target qubit based on the state of a control qubit. We took inspiration from the analogy between these two phenomena to design the quantum circuit in the quantum algorithm and used probability distribution to constrain the parameter of the circuit. Below we discuss three aspects of application issues.

1) Conventional correlation- or regression-based methods for GRN construction can handle a large number of genes because, for these methods, the gene-gene interaction is calculated as a single summary statistic from the expression profile of genes in measured cells. In contrast, our quantum approach for GRN inference can only model a small number of genes due to the



**Fig. 4** The impact of rotation angle  $\theta$  of a  $c\text{-}R_y$  gate on the amplitude,  $\mu$ , of the  $|1\rangle$  state of the target qubit in a quantum circuit. **a** The circuit configuration consists of a control qubit (the 1st qubit) and a target qubit (the 2nd qubit), each rotated by an  $R_y$  gate with angles  $\phi_1$  and  $\phi_2$ , respectively. **b–f** The heatmap panels display various combinations of  $\phi_1$  (0, 0.25 $\pi$ , 0.5 $\pi$ , 0.75 $\pi$  and 1.0 $\pi$ ) and  $\phi_2$  (0–1.0 $\pi$ ) settings. The heatmap colors indicate the amplitude of  $|1\rangle$  state of the target qubit with respect to  $\theta$ . In general, the amplitude of  $|1\rangle$  state of the target qubit increases or decreases monotonically with respect of  $\theta$ . In regions marked by red triangles in (**e** and **f**), the pattern is reversed.

vector space size—which is equal to the number of basis states—increases exponentially with the number of genes. In other words, cells in binarized scRNA-seq matrix may only be mapped to a moderate number of basis states such that each basis state is occupied by at least one cell. For example, a 15-qubit qscGRN model offers  $2^{15} = 32,768$  basis states, while a scRNA-seq data set with 20,000 cells can take at most 61% of activations states in the best case. Thus, our qscGRN model may retrieve an observed distribution with no biological information mapped to many basis states. Insufficient mapping may happen even though the latest scRNA-seq technology has the capacity to allow the transcriptome of millions of cells to be measured. To obtain enough cells, we can merge multiple scRNA-seq data sets as long as they are from the same cell types or similar biological sources and the batch effect can be corrected<sup>47</sup>. On the other hand, we can select most biologically informative genes such as highly variable genes<sup>48</sup> to be included in the analysis, reducing the burden of a large number of genes in the model while maintaining the biological relevance.

2) To simulate the regulatory relationship between two genes, we use a  $c\text{-}R_y$  gate to create a link between each pair of qubits in the regulation layers. The rotation angle of the  $c\text{-}R_y$  gate indicates the strength of interaction between the control gene and the target gene. The rotation angles of  $c\text{-}R_y$  gates are parameterized and mapped to the adjacency matrix after optimization to form the GRN. Throughout the paper, we assume that the rotation angle reflects the interaction strength—this is, the greater the angle, the stronger the interaction. However, we discovered that this is not always the case. We provide a simple example in Fig. 4 to illustrate the problem. Figure 4a shows the basic unit circuit, initialized in  $|00\rangle$  state, that consists of a control qubit (1st qubit, rotated using an  $R_y$  gate with an angle  $\phi_1$ ), a target qubit (2nd qubit, rotated using an  $R_y$  gate with an angle  $\phi_2$ ) and a  $c\text{-}R_y$  gate with rotation angle  $\theta$ . Figure 4b–f show the effect of rotation  $\theta$  in the  $c\text{-}R_y$  gate on the amplitude of  $|1\rangle$  of the 2nd qubit,  $\mu$ , under different settings with various combinations of  $\phi_1$  and  $\phi_2$ . When considering  $\mu$  as a function of  $\theta$ , we can see in most cases,  $\mu$  increases with increasing  $\theta$  or vice versa which is consistent with our assumption in gene regulation simulation. However, in some cases with specific combinations of  $\phi_1$  and  $\phi_2$  (as indicated with red triangles in Fig. 4e, f, the pattern is opposite— $\mu$  increases with

dalla figura sembra il contrario:  $\mu$  aumenta quando  $\theta$  che aumenta

decreasing  $\theta$  or vice versa. The opposite pattern becomes evident when  $\phi_1$  approaches  $\pi$ . We regard this phenomenon “boundary effect”, which may influence the interpretation of our modelling results. However, it should not have a great impact on our analysis. This is because: First, the boundary effect happens when the absolute value of rotation angle  $\theta$  of  $c\text{-}R_y$  gate approaches  $\pi/2$ . In our real-data study, as shown in Fig. 3e, we found the values of  $\theta$  for all genes are in the range between  $-\pi/4$  and  $\pi/4$ , in which the boundary effect is neglectable. Second, the boundary effect only happens in limited areas in the regions with specific combinations of states of control and target qubits. The phenomenon becomes pronounced when the rotation angle of  $R_y$  gate for the control qubit, i.e.  $\phi_1$ , is greater than 0.75 $\pi$ , which means the gene is being activated in more than 85% of cells. In the case of the rotation angle is  $\pi$ , the corresponding gene is always activated in all cells. In single-cell biology, a fully activated gene is most likely to happen for so called “house-keeping” genes, which are consistently expressed in a high level. These genes are essential for cell survival but are less likely to play any important regulatory role. Our previous study<sup>48</sup> provides evidence that highly variable genes such as those expressed in 50% of cells and inactivated in the other 50% of cells are most functionally important for any given cell type. Taken together, we acknowledge the potential impact of the boundary effect in our model but argue that, as the impact is likely to be limited, the interpretability of the  $c\text{-}R_y$  rotation angle as a measure of interacting strength remains largely intact.

Correlation- and regression-based are the most widely used methods for GRN inference, owing to in part their computational efficiency. These methods typically compute correlation or regression coefficients for gene pairs using the total number of cells in the data. The issue with these methods is that they deal with gene pairs across cells, not fully exploiting complex expression patterns by incorporating another degree of information. The relationship between any two genes is measured using a single value of summary statistics such as correlation or regression coefficient. Once computed, the coefficient becomes independent of the total number of cells. Increasing the number of cells would have little influence on correlation or regression coefficient. The other issue is that the coefficient is computed only between the two genes, regardless of the expression values of other genes in the same



**Table 1.** Mapping basis states using a  $c\text{-}R_y$  gate.

Basis state $ x\rangle$	$c\text{-}R_y(\theta)  x\rangle$
$ 00\rangle$	$ 00\rangle$
$ 01\rangle$	$ 01\rangle$
$ 10\rangle$	$\cos\frac{\theta}{2} 10\rangle + \sin\frac{\theta}{2} 11\rangle$
$ 11\rangle$	$-\sin\frac{\theta}{2} 10\rangle + \cos\frac{\theta}{2} 11\rangle$

biological system. Not considering other genes in the computation may result in a biased coefficient, which does not represent the true behavior of underlying interactions. There are methods such as partial correlation<sup>7</sup>, principal component regression<sup>5</sup>, and LASSO<sup>49</sup> that may correct this. But, the correcting effect is limited given that all-to-all interactions cannot be easily modeled.

## METHODS

The implementation of our package QuantumGRN is achieved using NumPy, Pandas, Matplotlib, iGraph and Qiskit—an open-source library for working with quantum computer simulators. Our package uses the Aer Simulator backend for a noisy circuit simulator. More details about code implementation and dataset can be found in data and code availability sections.

### Quantum computation theory

In this section, we introduce broad-audience background of quantum computation. In classical computation, a bit is the unit of information being  $|0\rangle$  or  $|1\rangle$  in Dirac notation, defined as  $(1\ 0)^T$  and  $(0\ 1)^T$  respectively<sup>50–52</sup>. In quantum computation, a qubit is the unit of information being  $|\psi\rangle = c_0|0\rangle + c_1|1\rangle$  in superposition, where  $|\psi\rangle$  is the quantum state,  $c_0$  and  $c_1$  are complex numbers, and  $|c_0|^2 + |c_1|^2 = 1$ . The measurement of  $|\psi\rangle$  results in 0 with a probability to be observed of  $|c_0|^2$  and 1 of  $|c_1|^2$ .

The Hadamard gate  $H$  is a single-qubit gate frequently used in quantum algorithms and is defined as  $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ , creating superpositions of the basis states (i.e.,  $H|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{|0\rangle + |1\rangle}{\sqrt{2}}$ ). Furthermore, the rotation gate  $R_y$  is also a single-qubit gate that uses a rotation parameter  $\theta$  and is defined as  $R_y(\theta) = \begin{pmatrix} \cos\theta/2 & -\sin\theta/2 \\ \sin\theta/2 & \cos\theta/2 \end{pmatrix}$ . In addition, a controlled gate is a 2-qubit gate that performs an operation on a target qubit when the control qubit is in state  $|1\rangle$ , where the operation is typically a single-qubit gate. For example, Table 1 shows the mapping of basis states when using a controlled- $R_y$  gate that has the first qubit as control and the second qubit as target. The  $R_y$  operation is performed in basis states  $|10\rangle$  and  $|11\rangle$  because the control qubit is 1, no operation is performed otherwise.

In classical computation, a circuit is a model composed of a sequence of gates (NOT, AND, OR operations) having input bits that flow through such a sequence eventually computing the output bits for a given task<sup>53</sup>. Similarly, a quantum circuit is a model consisting of a sequence of quantum gates that perform operations on input qubits<sup>54</sup>. In a quantum algorithm, the input qubits are usually initialized to  $|0\rangle_n$ , meaning a string of  $n$  bits of zeros. Then, the register flows through the sequence of gates, computing an output register that is measured and decoded to interpret the result.

### The qscGRN model: a parameterized quantum circuit

Here, we introduce the quantum single-cell gene regulatory network (qscGRN) model that is a quantum circuit consisting of  $n$  qubits and models a biological GRN for  $n$  genes in the framework of quantum computation giving a qubit-gene equivalence (Fig. 1, Algorithm 1). The sequence of gates is grouped into 2 types of

layers: The encoder layer  $L_{\text{enc}}$  consists of a  $R_y$  gate in each qubit and translates biological information (i.e., the frequency of gene actively expressed among cells) onto a superposition state. The regulation layer  $L_k$  consists of a sequence of  $c\text{-}R_y$  gates that have the  $k$ th qubit as control and a corresponding target such that the  $k$ th qubit is fully connected to other qubits. In the  $L_k$  layer, a  $c\text{-}R_y$  gate—that has the  $k$ th qubit as control and the  $p$ th qubit as the target—models the regulation interaction in the corresponding gene-gene pair. In particular, the parameter of the  $c\text{-}R_y$  gate quantifies the strength of the gene-gene interaction.

In this work, we used the notation  $\theta_{k,k}$  for the parameter of the  $R_y$  gate on the  $k$ th qubit in the  $L_{\text{enc}}$  layer, and  $\theta_{k,p}$  for the  $c\text{-}R_{y,n}$  gate with the  $k$ th qubit as control and the  $p$ th qubit as target, in the layer  $L_k$  of a  $n$ -qubit system. Thus, two layers were defined respectively as

$$L_{\text{enc}} = R_y(\theta_{n-1,n-1}) \otimes \cdots \otimes R_y(\theta_{1,1}) \otimes R_y(\theta_{0,0}), \quad (1)$$

where the  $\otimes$  operator is the tensor product, and

$$L_k = \prod_{i=0, i \neq k}^{n-1} c\text{-}R_{y,n}(\theta_{k,i}) = c\text{-}R_{y,n}(\theta_{k,n-1}) \cdots c\text{-}R_{y,n}(\theta_{k,1}) c\text{-}R_{y,n}(\theta_{k,0}). \quad (2)$$

The computation of  $L_{\text{enc}}$  and  $L_k$  is noncommutative due to the needed operations are matrix multiplication and tensor product.

The qscGRN model was initialized to  $|0\rangle_n$  and put into a superposition state using the  $L_{\text{enc}}$  layer. Then, the gene-gene interactions were modeled using regulation layers  $L_0, L_1, \dots, L_{n-1}$ . Thus, the qscGRN model is a quantum circuit that has  $n^2$  quantum gate parameters given by a matrix representation  $\theta$  for the set of parameters  $\theta_{k,p}$  in the quantum gates:

$$\theta = \begin{bmatrix} \theta_{0,0} & \theta_{0,1} & \cdots & \theta_{0,n-1} \\ \theta_{1,0} & \theta_{1,1} & \cdots & \theta_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,0} & \theta_{n-1,1} & \cdots & \theta_{n-1,n-1} \end{bmatrix}, \quad (3)$$

where the diagonal elements belong to the  $R_y$  gates in the  $L_{\text{enc}}$  layer, and the non-diagonal elements to the  $c\text{-}R_{y,n}$  gates in the regulation layers  $L_0, L_1, \dots, L_{n-1}$ . We recognized the matrix  $\theta$  as the adjacency matrix of the biological GRN.

Therefore, the output register  $|\psi_{\text{out}}\rangle$  of the qscGRN model encodes the gene-gene interactions in superposition as a function of the matrix  $\theta$  and was defined as

$$|\psi_{\text{out}}\rangle = \left( \prod_{k=0}^{n-1} L_k \right) L_{\text{enc}} |0\rangle_n = L_{n-1} \cdots L_1 L_0 L_{\text{enc}} |0\rangle_n. \quad (4)$$

### Algorithm 1. Construction of qscGRN model

**Require:** Number of qubits  $n$ , Parameter  $\theta$

- 1: Create  $n$ -qubit quantum circuit qscGRN
- 2: **for**  $k = 0, 1, \dots, n-1$  **do**
- 3:   Create and append  $R_y(\theta_{k,k})$  gate in qubit  $k$
- 4: **end for**
- 5: **for**  $k = 0, 1, \dots, n-1$  **do**
- 6:   **for**  $p = 0, 1, \dots, n-1$  **do**
- 7:     **if**  $k \neq p$  **then**
- 8:       Create and append  $c\text{-}R_{y,n}(\theta_{k,p})$  gate having  $k$  as control and  $p$  as target
- 9:     **end if**
- 10:   **end for**
- 11: **end for**
- 12: **return** qscGRN

### Quantum-classical framework for optimization of the qscGRN model

In this section, we introduce the complete quantum-classical framework using the qscGRN model to infer the corresponding biological GRN (Fig. 2).

**Gene selection.** The input data of the workflow was a scRNA-seq expression data matrix  $\mathbf{X}$  that contains expression values of  $N$  genes in  $m$  cells. The data matrix  $\mathbf{X}$  was normalized using Pearson residuals<sup>55</sup>. Then,  $n$  out of  $N$  genes were selected to be analyzed in the next step.

**Binarization.** The normalized expression matrix  $\mathbf{X}$  was binarized by applying the expression threshold of 0, which means that expression values greater than 0 are set to 1, and 0 otherwise<sup>56,57</sup>. The outcome of the binarization was saved to  $\mathbf{X}^b$ , which is a matrix of dimension  $n \times m$ .

**Labeling.** Labels were assigned for each cell in  $\mathbf{X}^b$ , such that the label is a string vector composed of the binarized expression of the  $n$  genes in a cell. Thus, a label is the activation state of a gene in the corresponding cells.

**Activation ratios.** Activation ratios were computed for each gene as the percentage of cells expressing that gene in  $\mathbf{X}^b$ . Then, the  $n$  rows in  $\mathbf{X}^b$  were ordered decreasingly by the activation ratio and were labeled as  $g_0, g_1, \dots, g_{n-1}$ .

**Observed distribution.** We computed the percentage of occurrences of each label within the  $m$  cells to obtain the observed distribution  $p^{\text{obs}}$ . The percentage of label  $|0\rangle_n$  in  $p^{\text{obs}}$  was set to 0, and the rest of the distribution was rescaled to sum to 1. The rationale for setting the  $|0\rangle_n$  probability to 0 is that only cells with expression values for at least one of  $n$  genes are informative. Sparsity is a common characteristic of scRNA-seq data because of the dropout event that occurs during the sequencing process. Figure 2a shows the steps just described in the workflow of the quantum-classical framework.

**Initialization of the parameter  $\theta$ .** The non-diagonal elements  $\theta_{k,p}$  corresponding to  $c\text{-}R_y$  gates were initialized to 0. The diagonal elements  $\theta_{k,k}$  corresponding to  $R_y$  gates were initialized to  $2 \cdot \sin^{-1} \sqrt{\text{act}_k}$ , where  $\text{act}_k$  is the activation ratio for the  $k$ th gene. The rationale for the formula is that, independently on each qubit, the probability of observing 1 is the activation ratio of the corresponding gene after the  $L_{\text{enc}}$  layer. Algorithm 2 illustrates the initial setup of the workflow.

#### Algorithm 2. Initial setup of the workflow

**Require:** Normalized scRNA-seq matrix  $\mathbf{X}$

- 1: Select  $n$  gene from  $\mathbf{X}$
- 2:  $\mathbf{X}^b$  = Binarized matrix  $\mathbf{X}$  for the selected  $n$  genes
- 3: Label each cell in  $\mathbf{X}^b$  of dimension  $n \times m$
- 4: **for**  $k = 0, 1, \dots, n-1$  **then**
- 5:  $\text{act}_k = \# \text{cells}_k / m$ , where  $\# \text{cells}_k$  is the number of cells expressing the  $k$  gene
- 6: **end for**
- 7: **for** each label  $\mathbf{x} \in \{0, 1\}^n$  **then**
- 8:  $p_{\mathbf{x}}^{\text{obs}} = \# \text{cells}_{\mathbf{x}} / m$ , where  $\# \text{cells}_{\mathbf{x}}$  is the number of cells having  $\mathbf{x}$  as label
- 9: **end for**
- 10: Rescale  $p^{\text{obs}}$
- 11:  $\theta$  = Create an  $n \times n$  matrix of all elements 0
- 12: **for**  $k = 0, 1, \dots, n-1$  **then**
- 13: **for**  $p = 0, 1, \dots, n-1$  **then**
- 14: **if**  $k = p$  **then**
- 15:  $\theta_{k,k} = 2 \cdot \sin^{-1} \sqrt{\text{act}_k}$
- 16: **end if**
- 17: **end for**
- 18: **end for**
- 19: **return** initial parameter  $\theta, p^{\text{obs}}$

**Measuring the output register of the qscGRN model.** We measured the output register  $|\psi_{\text{out}}\rangle$  to obtain the output distribution  $p^{\text{out}}$  of observing the basis states. The probability of the state  $|0\rangle_n$  in  $p^{\text{out}}$  was set to 0, and the rest of the distribution was rescaled to sum to 1.

**Smoothing  $p^{\text{obs}}$  and  $p^{\text{out}}$ .** Laplace smoothing was used to reshape  $p^{\text{obs}}$  and  $p^{\text{out}}$  to distributions  $\hat{p}^{\text{obs}}$  and  $\hat{p}^{\text{out}}$  respectively. These smoothed distributions were computed as  $\hat{p}^i = \frac{\# \text{occ}_i + a}{m + 2^n \cdot a}$ , where  $i \in \{\text{out}, \text{obs}\}$ ,  $a$  is the smoothing parameter being typically 1 and  $\# \text{occ}_i$  is the number of occurrences in the distribution  $p^i$ . In other words,  $p^i = \frac{\# \text{occ}_i}{m}$  is the original distribution.

A che serve?  
is a technique used to smooth count data, eliminating issues caused by certain values having 0 occurrences.

**Loss function.** The loss function consists of KL and constrain terms, named as  $L_{\text{KL}}$  and  $L_{\text{cons}}$ , were defined as

$$L_{\text{KL}}(\theta) = \sum_{\mathbf{x} \in \{0,1\}^n} \hat{p}_{\mathbf{x}}^{\text{out}} \log \left( \frac{\hat{p}_{\mathbf{x}}^{\text{out}}}{\hat{p}_{\mathbf{x}}^{\text{obs}}} \right), \quad (5)$$

$$L_{\text{cons}}(\theta) = \sum_{\theta_i \in \theta} \frac{1}{\left[ \theta_i^4 - \left( \frac{\pi}{2} \right)^4 \right]^2}, \quad (6)$$

where  $\theta$  is the parameter in the qscGRN model and  $\{0, 1\}^n$  is the  $n$  Cartesian power of the set  $\{0, 1\}$ . Thus, the loss function was defined as

$$L(\theta) = L_{\text{KL}}(\theta) + \lambda \cdot L_{\text{cons}}(\theta), \quad (7)$$

where  $\lambda$  is a dynamic coefficient that rescales  $L_{\text{cons}}$  to the same order of magnitude than  $L_{\text{KL}}$ . In summary, the  $L_{\text{KL}}$  term fits the output distribution  $p^{\text{out}}$  in the observed distribution  $p^{\text{obs}}$ . Meanwhile, the  $L_{\text{cons}}$  constraints any parameter in  $\theta$  to not get close to  $\pi/2$ .

**Optimization of the parameter  $\theta$ .** The optimization was achieved by minimizing iteratively the loss function to a threshold value of  $2^n \times 1e-4$  using a modified-gradient descent algorithm with a learning rate  $\text{lr}$  of 0.05. Otherwise, the optimization was performed for a pre-defined iterations  $t$ . Then, the parameter  $\theta$  in the iteration  $s+1$  was defined as

$$\theta_{s+1} = \theta_s - \text{lr} \cdot \frac{\nabla L(\theta_s) + \nabla^T L(\theta_s)}{2}, \quad (8)$$

where  $\nabla^T$  is the transpose of the gradient of loss function, allowing to keep the parameter  $\theta$  as a symmetric matrix. The diagonal parameters  $\theta_{k,k}$  were not trained during optimization under the assumption that these parameters encode the binarized scRNA-seq matrix given as an input to the quantum framework. Algorithm 3 illustrates the optimization of parameter  $\theta$  and Fig. 2b shows details of the optimization in the workflow. Our work is also integrated to Qiskit—an open source library for working with quantum computers—that simulates a noisy quantum circuit using Aer Simulator backend with default parameters.

#### Algorithm 3. Optimization of parameter $\theta$

**Require:** Initial parameter  $\theta_0, p^{\text{obs}}$

- 1:  $\hat{p}^{\text{obs}} = \text{smooth}(p^{\text{obs}})$
- 2: **for**  $s = 0, 1, \dots, t-1$  **then**
- 3: qscGRN = Constructed quantum circuit using  $\theta_s$
- 4: Measure output register and obtain  $p^{\text{out}}$
- 5:  $\hat{p}^{\text{out}} = \text{smooth}(p^{\text{out}})$
- 6:  $\text{loss} = L(\theta_s)$
- 7: **if**  $\text{loss} < \text{loss\_threshold}$  **then**
- 8: **return**  $\theta_s$
- 9: **end if**
- 10: Compute gradient  $\nabla L(\theta_s)$
- 11:  $\theta_{s+1} = \theta_s - \text{lr} \cdot \frac{\nabla L(\theta_s) + \nabla^T L(\theta_s)}{2}$
- 12: **end for**
- 13: **return**  $\theta_t$

**Recovery of gene regulatory network.** We removed non-diagonal parameters in  $\theta$  that had an absolute value less than  $\frac{\pi}{180 \cdot 2}$  because no significant rotation was performed by the corresponding  $c\text{-}R_y$  gate. Next, we used the remaining parameter in  $\theta$  to construct the adjacency matrix of the biological GRN, which is a weighted symmetric network. Figure 2c shows this last step in the workflow of the quantum-classical framework.

### Single-cell transcriptomic data

The scRNA-seq data used in this study was generated from lymphoblastoid cell lines (LCLs), which are widely used cell line systems derived from human primary B cells. The single cell sequencing libraries were prepared using the 10x Genomics platform. Information about the experimental procedure and the acquisition of sequence data is provided in reference to our original study<sup>31</sup>. The data set has been deposited to the Gene Expression Omnibus (GEO) database and can be accessed with accession number GSE126321. To increase the number of cells in this study, we merged our data set with another LCL scRNA-seq data set downloaded from the GEO database with accession number GSE158275 ref.<sup>32</sup>. The data matrices were pre-processed using scGEAToolbox<sup>58</sup> and combined to produce the final matrix, which contains expression counts of 9,905 genes of 28,208 cells. The matrix was then normalized using the Pearson residuals method<sup>55</sup>. Normalized expression values of six genes: IRF4, REL, PAX5, RELA, PRDM1, and AICDA in the NF- $\kappa$ B signaling pathway, were extracted. The 6-gene expression matrix with dimensions of  $6 \times 28,208$  was binarized and used as the input of our qscGRN analysis. The biological regulatory relationships between these genes, called the baseline model of GRN, were obtained from the previously established B-cell differentiation circuit model<sup>33,34</sup>.

### DATA AVAILABILITY

The scRNA-seq data analyzed in the current study is available in the NCBI GEO database under accession numbers GSE126321 and GSE158275.

### CODE AVAILABILITY

The processed data and the source code implementation of the qscGRN package are provided in the GitHub repository at <https://github.com/cailab-tamu/QuantumGRN/>. The repository also includes tutorials written in Python language.

Received: 22 June 2022; Accepted: 29 June 2023;

Published online: 13 July 2023

### REFERENCES

- Huynh-Thu, V. A. & Sanguinetti, G. Gene regulatory network inference: an introductory survey, in *Gene Regulatory Networks: Methods and Protocols*, Sanguinetti G. & Huynh-Thu V. A., Editors, Springer: New York, NY, 1–23 (2019).
- Hecker, M., Lambeck, S., Toepfer, S., van Someren, E. & Guthke, R. Gene regulatory network inference: data integration in dynamic models—a review. *Biosystems* **96**, 86–103 (2009).
- Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
- Chan, T. E., Stumpf, M. P. H. & Babbie, A. C. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell. Syst.* **5**, 251–267.e3 (2017).
- Osorio, D., Zhong, Y., Li, G., Huang, J. Z. & Cai, J. J. scTenifoldNet: a machine learning workflow for constructing and comparing transcriptome-wide gene regulatory networks from single-cell data. *Patterns* **1**, 100139 (2020).
- Yang, Y. et al. scTenifoldXct: a semi-supervised method for predicting cell-cell interactions and mapping cellular communication graphs. *Cell. Syst.* **14**, 302–311.e4 (2023).
- Kim, S. ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* **22**, 665–674 (2015).
- Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* **5**, e12776 (2010).
- Kotiang, S. & Eslami, A. A probabilistic graphical model for system-wide analysis of gene regulatory networks. *Bioinformatics* **36**, 3192–3199 (2020).
- Lahdesmaki, H., Hautaniemi, S., Shmulevich, I. & Yli-Harja, O. Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks. *Signal Process.* **86**, 814–834 (2006).
- Friedman, N., Linial, M., Nachman, I. & Pe'er, D. Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **7**, 601–620 (2000).
- Shmulevich, I., Dougherty, E. R., Kim, S. & Zhang, W. Probabilistic Boolean Networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* **18**, 261–274 (2002).
- Delgado, F. M. & Gomez-Vela, F. Computational methods for gene regulatory networks reconstruction and analysis: a review. *Artif. Intell. Med.* **95**, 133–145 (2019).
- Zhao, M., He, W., Tang, J., Zou, Q. & Guo, F. A comprehensive overview and critical evaluation of gene regulatory network inference technologies. *Brief. Bioinform.* **22**, 1–15 (2021).
- Blencowe, M. et al. Network modeling of single-cell omics data: challenges, opportunities, and progresses. *Emerg. Top. Life Sci.* **3**, 379–398 (2019).
- Cha, J. & Lee, I. Single-cell network biology for resolving cellular heterogeneity in human diseases. *Exp. Mol. Med.* **52**, 1798–1808 (2020).
- Chen, S. & Mar, J. C. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinform.* **19**, 232 (2018).
- Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **17**, 147–154 (2020).
- Diaz, L. P. M. & Stumpf, M. P. H. Gaining confidence in inferred networks. *Sci. Rep.* **12**, 2394 (2022).
- Gyongyosi, L. & Imre, S. A survey on quantum computing technology. *Comput. Sci. Rev.* **31**, 51–71 (2019).
- Egger, D. J. et al. Quantum computing for finance: state-of-the-art and future prospects. *IEEE Trans. Quantum Eng.* **1**, 1–24 (2020).
- Fernandez-Carames, T. M. & Fraga-Lamas, P. Towards post-quantum blockchain: a review on blockchain cryptography resistant to quantum computing attacks. *IEEE Access* **8**, 21091–21116 (2020).
- Cao, Y., Romero, J. & Aspuru-Guzik, A. Potential of quantum computing for drug discovery. *IBM J. Res. Dev.* **62**, 6:1–6:20 (2018).
- Ramezani, S. B., Sommers, A., Manchukonda, H. K., Rahimi, S. & Amirlatifi, A. Machine Learning Algorithms in Quantum Computing: A Survey. in *International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, 1–8, <https://doi.org/10.1109/IJCNN48605.2020.9207714>, (2020).
- Bauer, B., Bravyi, S., Motta, M. & Kim-Lic Chan, G. Quantum algorithms for quantum chemistry and quantum materials science. *Chem. Rev.* **120**, 12685–12717 (2020).
- Bharti, K. et al. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.* **94**, 015004 (2022).
- Huang, H. Y. et al. Quantum advantage in learning from experiments. *Science* **376**, 1182–1186 (2022).
- Peng, F. C., Schuurmans, D. & Wang, S. J. Augmenting naive Bayes classifiers with statistical language models. *Inf. Retr. J.* **7**, 317–345 (2004).
- Ruder, S. An overview of gradient descent optimization algorithms. Preprint at <https://arxiv.org/abs/1609.04747> (2016).
- Kullback, S. & Leibler, R. A. On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86 (1951).
- Osorio, D., Yu, X., Yu, P., Serpedin, E. & Cai, J. J. Single-cell RNA sequencing of a European and an African lymphoblastoid cell line. *Sci. Data* **6**, 112 (2019).
- SoRelle, E. D. et al. Single-cell RNA-seq reveals transcriptomic heterogeneity mediated by host-pathogen dynamics in lymphoblastoid cell lines. *eLife* **10**, e62586 (2021).
- Roy, K. et al. A regulatory circuit controlling the dynamics of NF $\kappa$ B cRel transitions B cells from proliferation to plasma cell differentiation. *Immunity* **50**, 616–628.e6 (2019).
- Sciammas, R. et al. An incoherent regulatory network architecture that orchestrates B cell diversification in response to antigen signaling. *Mol. Syst. Biol.* **7**, 495 (2011).
- Yadav, A. et al. Identification of a ubiquitously active promoter of the murine activation-induced cytidine deaminase (AICDA) gene. *Mol. Immunol.* **43**, 529–541 (2006).
- Boi, M., Zucca, E., Inghirami, G. & Bertoni, F. PRDM1/BLIMP1: a tumor suppressor gene in B and T cell lymphomas. *Leuk. Lymphoma* **56**, 1223–1228 (2015).
- Tunaypin, C. et al. Direct repression of prdm1 by Bcl-6 inhibits plasmacytic differentiation. *J. Immunol.* **173**, 1158–1165 (2004).

38. Shapiro-Shelef, M. & Calame, K. Regulation of plasma-cell development. *Nat. Rev. Immunol.* **5**, 230–242 (2005).
39. Teng, Y. et al. IRF4 negatively regulates proliferation of germinal center B cell-derived Burkitt's lymphoma cell lines and induces differentiation toward plasma cells. *Eur. J. Cell. Biol.* **86**, 581–589 (2007).
40. Sciammas, R. et al. Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation. *Immunity* **25**, 225–236 (2006).
41. Nutt, S. L., Taubenheim, N., Hasbold, J., Corcoran, L. M. & Hodgkin, P. D. The genetic network controlling plasma cell differentiation. *Semin. Immunol.* **23**, 341–349 (2011).
42. Outeiral, C. et al. The prospects of quantum computing in computational molecular biology. *WIREs Comput. Mol. Sci.* **11**, e1481 (2021).
43. Marx, V. Biology begins to tangle with quantum computing. *Nat. Methods* **18**, 715–719 (2021).
44. Emani, P. S. et al. Quantum computing at the frontiers of biological sciences. *Nat. Methods* **18**, 701–709 (2021).
45. Cheng, H. P., Deumens, E., Freericks, J. K., Li, C. & Sanders, B. A. Application of quantum computing to biochemical systems: a look to the future. *Front. Chem.* **8**, 587143 (2020).
46. Woolnough, A. P., Hollenberg, L. C. L., Cassey, P. & Prowse, T. A. A. Quantum computing: a new paradigm for ecology. *Trends Ecol. Evol.* **3142**, 9 (2023).
47. Tran, H. T. N. et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* **21**, 12 (2020).
48. Osorio, D. et al. Single-cell expression variability implies cell function. *Cells* **9**, 14 (2019).
49. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **58**, 267–288 (1996).
50. Dirac, P. A. M. A new notation for quantum mechanics. *Math. Proc. Camb. Philos. Soc.* **35**, 416–418 (1939).
51. Meter, R. V. & Oskin, M. Architectural implications of quantum computing technologies. *ACM J. Emerg. Technol. Comput. Syst.* **2**, 31–63 (2006).
52. Rieffel, E. & Polak, W. An introduction to quantum computing for non-physicists. *ACM Comput. Surv.* **32**, 300–335 (2000).
53. Lee, E. A. The past, present and future of cyber-physical systems: a focus on models. *Sensors* **15**, 4837–4869 (2015).
54. Vatan, F. & Williams, C. Optimal quantum circuits for general two-qubit gates. *Phys. Rev. A* **69**, 032315 (2004).
55. Lause, J., Berens, P. & Kobak, D. Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data. *Genome Biol.* **22**, 258 (2021).
56. Qiu, P. Embracing the dropouts in single-cell RNA-seq analysis. *Nat. Commun.* **11**, 1169 (2020).
57. Bouland, G. A., Mahfouz, A. & Reinders, M. J. T. Consequences and opportunities arising due to sparser single-cell RNA-seq datasets. *Genome Biol.* **24**, 86 (2023).
58. Cai, J. J. scGEAToolbox: a Matlab toolbox for single-cell RNA sequencing data analysis. *Bioinformatics* **36**, 1948–1949 (2019).

## ACKNOWLEDGEMENTS

The identification of the boundary effect in the  $c-R_y$  gate impact analysis is credited to the anonymous reviewer for whom we express our gratitude. This work was supported by the DoD grant GW200026 for J.J.C.

## AUTHOR CONTRIBUTIONS

Conceptualization, J.J.C.; methodology, C.R. and J.J.C.; implementation of the software, C.R.; formal analysis, C.R. and J.J.C.; writing and editing, C.R. and J.J.C.; supervision, J.J.C. All authors reviewed and contributed to the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to James J. Cai.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons

Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023