

Quantum gene regulatory networks

James Cai (✉ jcai@tamu.edu)

Texas A&M University <https://orcid.org/0000-0002-8081-6725>

Cristhian Roman-Vicharra

Texas A&M University

Article

Keywords:

Posted Date: June 29th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1785614/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

1 Quantum gene regulatory networks

2

3 Cristhian Roman-Vicharra¹, James J. Cai^{1,2,*}

4 ¹Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843,

5 USA

6 ²Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843,

7 USA

8 *Correspondence: jcai@tamu.edu (JJC)

9 Abstract

10 In this work, we present a quantum circuit model for inferring gene regulatory networks (GRNs). The
11 model is based on the idea of using qubit-qubit entanglement to simulate interactions between genes.
12 We provide preliminary results that suggest our quantum GRN modeling method is competitive and
13 warrants further investigation. Specifically, we present the results derived from the single-cell
14 transcriptomic data of human cell lines, focusing on genes in involving innate immunity regulation. We
15 demonstrate that our quantum circuit model can be used to predict the presence or absence of
16 regulatory interactions between genes and estimate the strength and direction of the interactions,
17 setting the stage for further investigations on how quantum computing finds applications in data-driven
18 life sciences and, more importantly, to invite exploration of quantum algorithm design that takes
19 advantage of the single-cell data. The application of quantum computing on single-cell transcriptomic
20 data likewise contributes to a novel understanding of GRNs, given that the relationship between fully

21 interconnected genes can be approached more effectively by quantum modeling than by statistical
22 correlations.

23 Summary

24 Quantum computing holds the promise to achieve certain types of computation that would otherwise
25 be unachievable by classical computers. The advent in the development of quantum algorithms has
26 enabled a variety of applications in chemistry, finance, and cryptography. Here we introduce a
27 parameterized quantum circuit modeling method for constructing gene regulatory networks (GRNs)
28 using data from single-cell RNA sequencing (scRNA-seq). In the circuit, each qubit represents a gene, and
29 qubits are entangled to simulate the interaction between genes. The strength of interactions is
30 estimated using the rotation angle of controlled unitary gates between qubits after fitting the scRNA-seq
31 expression matrix data. We applied our quantum single-cell GRN (qscGRN) model to real scRNA-seq data
32 obtained from human lymphoblastoid cell lines and demonstrated its usage in recovering known and
33 detecting novel regulatory relationships between genes in the nuclear factor-kappa B (NF- κ B) signaling
34 pathway. Our quantum circuit model enables the modeling of vast feature space occupied by cells in
35 different transcriptionally activating states, simultaneously tracking activities of thousands of interacting
36 genes and constructing more realistic single-cell GRNs without relying on statistical correlation or
37 regression. We anticipate that quantum computing algorithms based on our circuit model will find more
38 applications in data-driven life sciences, paving the way for the future of predictive biology and precision
39 medicine.

40 I. Introduction

41 A gene regulatory network (GRN) defines the ensemble of regulatory relationships between genes in a
42 biological system. Inferring GRNs is a powerful approach for studying molecular mechanism of

43 transcriptional regulation and the function of genes in processes of cellular activities [1, 2]. A GRN is
44 often represented as a graph—which can be signed, directed, and weighted—representing the
45 relationships between transcription factors or regulators and their target genes whose expression level
46 is controlled. However, because gene regulation inside cells is difficult to observe, indirect
47 measurements of intracellular expression are often used as a proxy, and the statistical dependencies are
48 used to infer real regulatory relationships between genes. Thus, the power of different methods for GRN
49 inference depends on the types of computational algorithms and the resolution of the expression data
50 [3-5].

51 Single-cell technologies, which have recently been developed and improved, open up opportunities for
52 studying biology at unprecedented resolution and scale. Single-cell RNA sequencing (scRNA-seq) allows
53 us to measure gene expression in individual cells for thousands of cells in a single experiment [6]. The
54 GRN modeling can adopt scRNA-seq technology and leverage the unprecedented information from the
55 sheer number of cells to improve inference power [7]. The use of such data would allow us to learn
56 better and more detailed network models, which will also help us better understand the mechanics
57 behind cellular operations. A plethora of computational methods for inferring GRNs have been
58 developed for either population-level or single cell-level gene expression data. These methods apply
59 statistical approaches to identify likely regulatory relationships between genes based on their expression
60 patterns. These different methods are based on correlation and partial correlation [8], information
61 theory [7], regression [9, 10], Gaussian graphical model [11], Bayesian and Boolean networks [12-14],
62 and many others. Each method has its own set of assumptions and limitations, which are not always
63 stated explicitly.

64 In recent years, quantum computing has become an emerging technology and an intense field of
65 research constantly seeking applications [15]. Researchers have developed quantum algorithms with
66 applications in areas such as finance [16], cryptography [17], machine learning [18], drug discovery [19],

67 chemistry, and material science [20]. Theoretically, a speedup is expected in certain types of
68 computation using quantum algorithms versus classical algorithms because a quantum computer takes
69 advantage of superposition and entanglement phenomena during the computation [21, 22]. The most
70 iconic quantum algorithm is Shor's [23] for the factorization of large numbers, which can break the
71 Rivest-Shamir-Adleman encryption [24]. Due to the potential of quantum computing, the current
72 approach to scRNA-seq analysis and GRN inference may be rethought.

73 In this work, we introduce a quantum single-cell GRN (qscGRN) modeling method, which is based on a
74 parameterized quantum circuit and uses the quantum framework to recover biological GRNs from
75 scRNA-seq data. In the qscGRN model, a gene is represented using a qubit, and the structure is divided
76 into 2 types of layers: the encoder layer that translates the scRNA-seq data into a superposition state
77 and the regulation layers that entangle qubits and model gene-gene interactions in the quantum
78 framework. In contrast to the correlation-based inference methods, the qscGRN model maps the
79 binarized expression values onto a large vector space, known as Hilbert space, making full use of the cell
80 information in the scRNA-seq data. Thus, a large number of cells in the scRNA-seq data is important
81 because it improves the mapping of biological information in a superposition state. In addition,
82 parameterization in the qscGRN model allows the gene-gene relationships to be inferred all at once by
83 fitting the superposition state probabilities onto the distribution observed in the scRNA-seq data.

84 A quantum-classical framework for optimizing the qscGRN model is also introduced. The classical
85 component of our framework uses the Laplace smoothing [25] and the gradient descent algorithm [26]
86 to perform optimization by minimizing a Kullback-Leibler (KL) divergence [27] as a loss function. Finally,
87 we used the quantum-classical framework on a real scRNA-seq data set [28, 29] to show that gene-gene
88 interactions can be modeled using quantum computing, and the structure of such as previously
89 published GRN [30, 31] can be recovered from the parameter-optimized quantum circuit.

90 II. Methods

91 1. Quantum computing theory

92 We first introduce the basic, broad-audience background of quantum computing necessary for this
93 work. Classical computers manage information processing using bits for storage, computation, and
94 communication [32]. A bit is the unit of information being 0 or 1, also represented in Dirac notation [33]
95 as $|0\rangle = (1\ 0)^T$ or $|1\rangle = (0\ 1)^T$ respectively [34]. In quantum computing, a qubit is the unit of information
96 represented as $|\psi\rangle = (c_0\ c_1)^T = c_0 |0\rangle + c_1 |1\rangle$, where $|\psi\rangle$ is the quantum state in the superposition of $|0\rangle$
97 and $|1\rangle$ basis in a 1-dimensional Hilbert space, and c_0, c_1 are complex numbers. The state of a quantum
98 system is described by a unit vector in the Hilbert space; therefore, the square modulus sum $|c_0|^2 + |c_1|^2$
99 is equal to 1. In quantum mechanics, the measure of $|\psi\rangle$ results in 0 with a probability to be observed of
100 $|c_0|^2$, and 1 of $|c_1|^2$. Thus, the probability of measuring a basis is the squared modulus of the associated
101 complex number.

102 Single-qubit gates that are widely used include the *NOT* gate, Hadamard gate, Pauli gates X, Y and Z ,
103 phase shift gates, and parameterized rotation gates R_x, R_y and R_z . The Hadamard gate—represented as H
104 gate—is frequently used in various quantum algorithms and is defined as $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. The H gate maps the
105 basis state $|0\rangle$ to $H|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$ and $|1\rangle$ to $\frac{|0\rangle-|1\rangle}{\sqrt{2}}$, creating a superposition of the
106 basis states. The measurement of the quantum state $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ results in observing the basis state $|0\rangle$ with a
107 probability of 0.5 and $|1\rangle$ with 0.5. Furthermore, the rotation gate R_y is a single-qubit operation (FIGURE
108 1A) based on the exponentiation of the Pauli gate Y using a rotation parameter θ and is defined as

109
$$R_y(\theta) \equiv e^{\frac{-i\theta Y}{2}} = \cos\frac{\theta}{2}I - i\sin\frac{\theta}{2}Y = \begin{pmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix},$$

110 where Y is a Pauli operation defined as $\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ and I is the identity matrix. The R_y gate maps the basis
 111 state $|0\rangle$ to a superposition state $R_y(\theta)|0\rangle = \cos\frac{\theta}{2}|0\rangle + \sin\frac{\theta}{2}|1\rangle$.

112 In addition, the controlled gate is a 2-qubit gate—which applies an operation on a target qubit when the
 113 control qubit is in state $|1\rangle$. The operation is typically a single-qubit gate such as R_y gate. Thus, a
 114 controlled- R_y gate, represented as $c-R_y$ gate, is defined as

115
$$c - R_{y,1,0}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ 0 & 0 & \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{pmatrix},$$

116 where the control register is the qubit labeled q_1 and the target is q_0 (FIGURE 1B). In the case where the
 117 control register is the qubit labeled q_0 and the target is q_1 , the $c-R_y$ gate is defined as

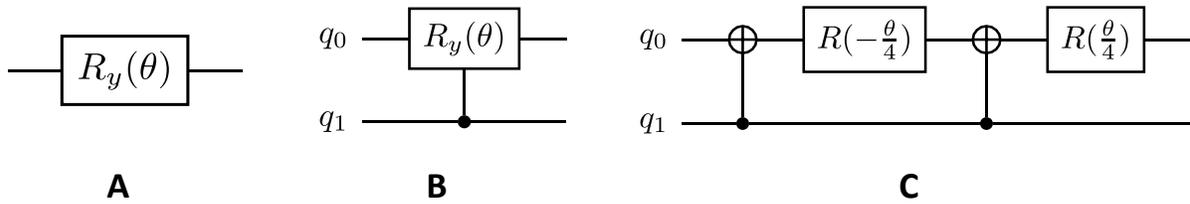
118
$$c - R_{y,0,1}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\frac{\theta}{2} & 0 & -\sin\frac{\theta}{2} \\ 0 & 0 & 1 & 0 \\ 0 & \sin\frac{\theta}{2} & 0 & \cos\frac{\theta}{2} \end{pmatrix}.$$

119 TABLE 1 shows the mapping of the basis states in a 2-dimension Hilbert space when using a $c-R_{y,1,0}(\theta)$
 120 gate. The basis states $|00\rangle$ and $|01\rangle$ have the first bit as 0, thus the operation R_y is not performed. On
 121 the other hand, the basis states $|10\rangle$ and $|11\rangle$ have the first bit as 1, thus the operation R_y is performed
 122 in the second bit.

Basis state $ x\rangle$	$c-R_{y,1,0}(\theta) x\rangle$
$ 00\rangle$	$ 00\rangle$
$ 01\rangle$	$ 01\rangle$
$ 10\rangle$	$\cos\frac{\theta}{2} 10\rangle + \sin\frac{\theta}{2} 11\rangle$
$ 11\rangle$	$-\sin\frac{\theta}{2} 10\rangle + \cos\frac{\theta}{2} 11\rangle$

123 **TABLE 1.** The controlled- $R_{y,1,0}(\theta)$ gate mapping of the basis state in a 2-dimension Hilbert space. The R_y
124 operation in the second bit is performed when the first bit is 1, and the second bit does not change
125 otherwise.

126 Generally, a controlled- U gate can be decomposed into single-qubit gates such as phase shift, c -NOT and
127 rotation gates, where U is a single qubit gate. Thus, the c - R_y gate is decomposable only into 2 single
128 rotation gates, represented as $R(c) = \begin{pmatrix} \cos c & -\sin c \\ \sin c & \cos c \end{pmatrix}$, and 2 c -NOT gates because there is no phase shift
129 operation. FIGURE 1C shows the decomposition of a c - $R_{y,1,0}$ gate into $R(\frac{\theta}{4})$, $R(-\frac{\theta}{4})$ gates and 2 c -NOT
130 gates, where the control register is q_1 , and target is q_0 . In other words, the effects of the rotation gates
131 sum up $\frac{\theta}{2}$ when the control qubit is 1 and cancel out each other otherwise.



132

133 **FIGURE 1.** Schematic representations of R_y gate, controlled- R_y gate, and its decomposition. (A) A
134 schematic representation of the R_y gate with a rotation parameter θ . In the representation, the input
135 register is on the left, and the output is on the right. (B) A schematic representation of the controlled- R_y
136 gate with a rotation parameter θ . In the representation, the control register is the qubit labeled q_1 , and
137 the target is labeled q_0 . The operation is performed on the target register when the control register q_1 is
138 1, and the rotation is not performed otherwise. (C) The schematic representation for the decomposition
139 of a controlled- R_y gate with a rotation parameter θ . The sequence consists of 2 rotation gates $R(\frac{\theta}{4})$,
140 $R(-\frac{\theta}{4})$ and 2 c -NOT gates. The operation in the target register is equal to an $R_y(\theta)$ gate when the control
141 register q_1 is 1, and no operation is performed otherwise.

142 2. The qscGRN model: a parameterized quantum circuit

143 In classical computation, a circuit is a model composed of a sequence of instructions (NOT, AND, OR
144 classical gates) that are not necessarily reversible. In the classical circuit, the input bits flow through the
145 sequence of instructions computing output bits for a certain task [35]. Similarly, a quantum circuit is a
146 model consisting of a sequence of quantum gates that perform operations on the qubits [36]. A
147 quantum circuit that is running an algorithm is usually initialized to $|0\rangle_n$, which means a string of n bits
148 of all zeros, and then put into a superposition state using $H^{\otimes n}$ transformation—which means an H gate
149 on each qubit—allowing all possible inputs to be tested [37]. Then, the register flows through a
150 sequence of quantum gates, and the output register is measured and decoded to interpret the result of
151 the algorithm.

152 Here, we introduce the quantum single-cell gene regulatory network (qscGRN) model, that is a quantum
153 circuit consisting of n qubits, and models a biological scGRN for n genes in the framework of quantum
154 computing. A qubit in the qscGRN model represents a gene in the biological scGRN. The sequence of
155 gates is grouped into 2 types of layers: The encoder layer L_{enc} consists of n R_y gates that translate
156 biological information (i.e., the state of gene activity or the frequency of genes actively expressed
157 among cells) onto a superposition state. In the L_{enc} layer, each qubit has a R_y gate (FIGURE 2A). The
158 regulation layer L_k consists of a sequence of $n-1$ $c-R_y$ gates that have the k th qubit as control and a
159 corresponding target such that the k th qubit is fully connected to other qubits (FIGURE 2B). In the L_k
160 layer, a $c-R_y$ gate—that has the k th qubit as control and the p th qubit as the target—models the
161 regulation relationship in the corresponding gene-gene pair. In particular, the parameter of a $c-R_y$ gate
162 quantifies the strength and determines the type of relationship between the k th and p th genes.

163 In FIGURE 2, we use the notation $\theta_{k,k}$ for the parameter of the R_y gate on the k th qubit in the L_{enc} layer
 164 and $\theta_{k,p}$ for the parameter on the c - R_y gate in the L_k layer that has the k th and p th qubits as control and
 165 target, respectively. Formally, the matrix representations for both layers are

$$166 \quad L_{enc} = R_y(\theta_{n-1,n-1}) \otimes \cdots \otimes R_y(\theta_{1,1}) \otimes R_y(\theta_{0,0}),$$

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

$$168 \quad L_k = \prod_{i=0, i \neq k}^{n-1} R_{y,n}(\theta_{k,i}) = R_{y,n}(\theta_{k,n-1}) \cdots R_{y,n}(\theta_{k,1}) R_{y,n}(\theta_{k,0}),$$

169 where $R_{y,n}(\theta_{k,i})$ denotes a c - R_y gate with the k th qubit as the control and i th qubit as the target in a n -
 170 qubit quantum circuit. Also, the computation of the matrix representation is not commutative, which
 171 means the order of the terms cannot be changed due to the operations needed are matrix
 172 multiplication and tensor product.

173 The qscGRN model is initialized to $|0\rangle_n$ state, and then put into a superposition state using a L_{enc} layer.
 174 The gene-gene interactions are then modeled using regulation layers L_0, L_1, \dots, L_{n-1} . Thus, the qscGRN
 175 model is a quantum circuit where each qubit is fully connected to every other qubit and has a total of n^2
 176 quantum gate parameters. Next, we construct the matrix representation θ of the qscGRN model using
 177 the collection of parameters $\theta_{k,p}$ on the quantum gates, where $0 \leq k, p < n$. Therefore, the matrix
 178 representation of the qscGRN model is denoted as

$$179 \quad \theta = \begin{bmatrix} \theta_{0,0} & \theta_{0,1} & \cdots & \theta_{0,n-1} \\ \theta_{1,0} & \theta_{1,1} & \cdots & \theta_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,0} & \theta_{n-1,1} & \cdots & \theta_{n-1,n-1} \end{bmatrix},$$

180 where the diagonal elements belong to the R_y gates in the L_{enc} layer, and the non-diagonal elements to
 181 the c - R_y gates in the regulation layers L_0, L_1, \dots, L_{n-1} .

182 The output register $|\psi_{out}\rangle$ of the n -qubit qscGRN model, encodes the gene-gene interactions in a
 183 superposition state according to the parameter θ and is formally defined as

$$184 \quad |\psi_{out}\rangle = \left(\prod_{k=0}^{n-1} L_k \right) L_{enc} |0\rangle_n = L_{n-1} \cdots L_1 L_0 L_{enc} |0\rangle_n.$$

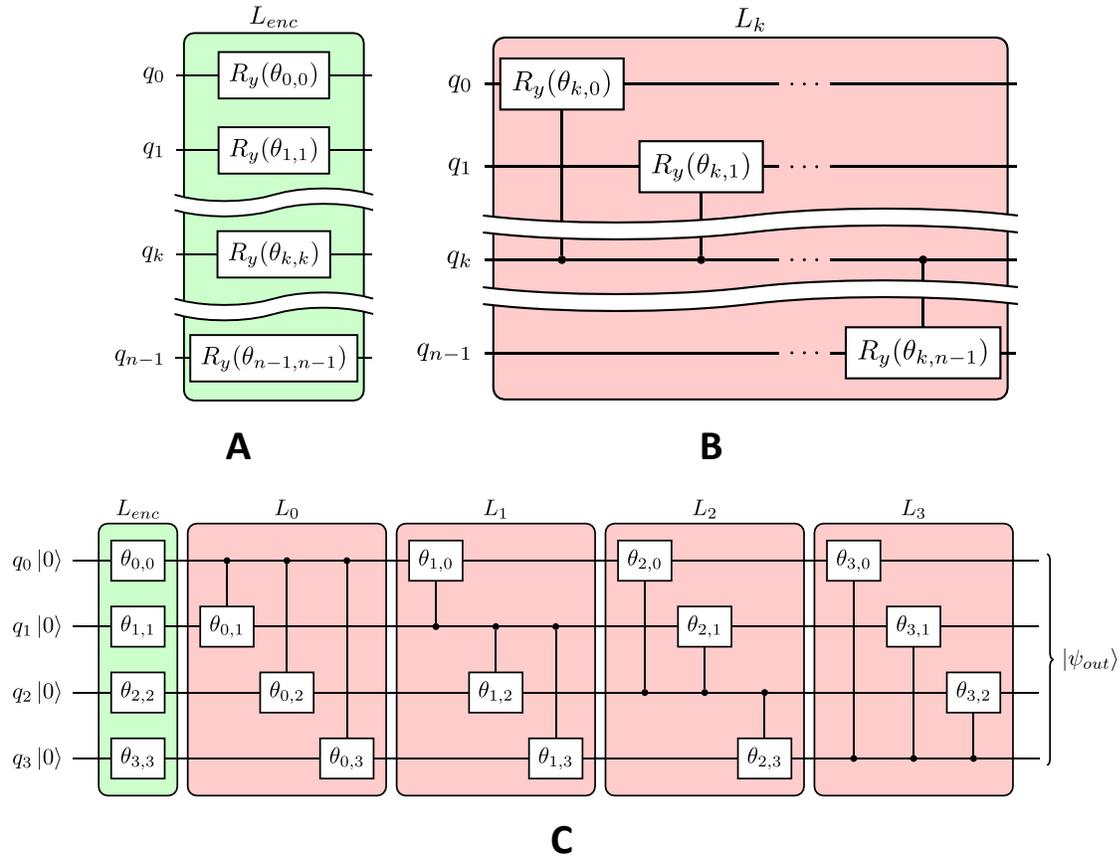
185 FIGURE 2C shows the schematic representation of a qscGRN model consisting of 4 qubits as an example
 186 for better interpretation of the equations. The quantum gate name is not shown for simplicity but only
 187 the corresponding parameter. According to the equations, the output register $|\psi_{out}\rangle$ in FIGURE 2C is
 188 defined as

$$189 \quad |\psi_{out}\rangle = L_3 L_2 L_1 L_0 L_{enc} |0\rangle_4,$$

190 where L_3, L_2, L_1 and L_0 have 3 parameters each, and L_{enc} has 4 parameters. Then, the matrix
 191 representation θ of the 4-qubit qscGRN model is denoted as

$$192 \quad \theta = \begin{bmatrix} \theta_{0,0} & \theta_{0,1} & \theta_{0,2} & \theta_{0,3} \\ \theta_{1,0} & \theta_{1,1} & \theta_{1,2} & \theta_{1,3} \\ \theta_{2,0} & \theta_{2,1} & \theta_{2,2} & \theta_{2,3} \\ \theta_{3,0} & \theta_{3,1} & \theta_{3,2} & \theta_{3,3} \end{bmatrix},$$

193 Finally, we understand the matrix θ as the adjacency matrix of the biological scGRN, which is a weighted
 194 directed fully connected network.



195

196 **FIGURE 2.** The quantum single-cell gene regulatory network (qscGRN) model. (A) The encoder layer L_{enc}
 197 for a qscGRN model that has n qubits, represented with a green background. The k th qubit has an R_y
 198 gate with a parameter $\theta_{k,k}$ —which is a diagonal element of the adjacency matrix in the qscGRN. (B) The
 199 regulation layer L_k for a qscGRN model that has n qubits, represented with a red background. The k th
 200 qubit uses a $c-R_y$ gate with a parameter $\theta_{k,p}$ —which is a non-diagonal element of the adjacency matrix
 201 in the qscGRN—to connect to the p th qubit, where $0 \leq p < n$ and $p \neq k$, thus the k th qubit is fully
 202 connected to other qubits. (C) A schematic representation of a qscGRN model that consists of 4 qubits.
 203 The L_{enc} layer consists of 4 R_y gates that put the input register into a superposition state. The L_k layer
 204 consists of $c-R_y$ gates that connect the k th qubit to the others. Thus, a 4-qubit qscGRN model uses an L_{enc}
 205 with 4 parameters and L_0, L_1, L_2 and L_3 with 3 parameters for each layer. The collection of parameters θ
 206 is the matrix representation of the 4-qubit qscGRN model.

207 3. Quantum-classical framework for optimization of the qscGRN model

208 **Gene selection and binarization:** The input data of the workflow is a transformed scRNA-seq expression
209 data matrix \mathbf{X} that has expression values for m cells. The transformation of the expression matrix can be
210 done using, for example, Pearson residuals [38]. Then, we select n genes from \mathbf{X} and binarize the
211 expression values. The binarization is achieved by applying an expression threshold of 0 to the
212 transformed expression matrix, which means that expression values greater than 0 are set to 1, and 0
213 otherwise. The outcome of the binarization is saved to \mathbf{X}^b , which is a matrix of dimension $n \times m$.

214 **Labeling and activation ratios (FIGURE 3A):** A label is assigned for each cell in \mathbf{X}^b , such that the label is a
215 string composed of the binarized expression of the n genes in a cell. In other words, a label is the
216 activation state of genes in a cell (colored in light blue). Then, we compute the percentage of
217 occurrences of each label in the m cells to obtain the observed distribution p^{obs} . The percentage of label
218 $|0\rangle_n$ in p^{obs} is set to 0, and the rest of the distribution is rescaled to sum to 1. The rationale for setting
219 the $|0\rangle_n$ probability to 0 is that cells with no expression levels are not informative in the quantum
220 framework due to dropout in the scRNA-seq experiment. Furthermore, the activation ratio act_k of the
221 k th gene is defined as the percentage of cells expressing that gene in \mathbf{X}^b (colored in light yellow). The n
222 genes in \mathbf{X}^b are ordered decreasingly by the activation ratio such that \mathbf{X}^b has n ordered rows labeled as
223 g_0, g_1, \dots, g_{n-1} .

224 **Initialization of the parameter θ in the qscGRN model (FIGURE 3B):** The parameters $\theta_{k,p}$ in the
225 regulation layers L_0, L_1, \dots, L_{n-1} are initialized to 0, where $0 \leq k, p < n$ and $k \neq p$. Besides, the parameters
226 $\theta_{k,k}$ in the encoder layer L_{enc} are initialized to $2 \cdot \sin^{-1} \sqrt{act_k}$ corresponding to the k th gene, where $0 \leq$
227 $k < n$. Therefore, the initial parameter θ is represented as

228
$$\boldsymbol{\theta}_{initial} = \begin{bmatrix} 2 \sin^{-1} \sqrt{act_0} & 0 & \dots & 0 \\ 0 & 2 \sin^{-1} \sqrt{act_1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 2 \sin^{-1} \sqrt{act_{n-1}} \end{bmatrix}.$$

229 The rationale for the formula $2 \sin^{-1} \sqrt{act_k}$ is that, independently on each qubit, the probability of
 230 observing 1 is the activation ratio of the corresponding gene after the L_{enc} layer.

231 **Optimization of the parameter $\boldsymbol{\theta}$ in the qscGRN model (FIGURE 3C):** We measure the output register
 232 $|\psi_{out}\rangle$ of the qscGRN model to obtain the output distribution p^{out} of observing the basis states. The
 233 probability of the state $|0\rangle_n$ in p^{out} is set to 0, and the rest of the distribution is rescaled to sum to 1.
 234 Then, Laplace smoothing is used to reshape p^{obs} and p^{out} to a different distribution \hat{p}^{obs} and \hat{p}^{out} ,
 235 respectively, thus handling the zero-probability problem when computing the loss function. The
 236 smoothed distribution is computed as

237
$$\hat{p}^i = \frac{\#ocu^i + \alpha}{m + 2^n \cdot \alpha},$$

238 where $i \in \{out, obs\}$ and α is the smoothing parameter which is typically 1. $\#ocu^i$ is the number of
 239 occurrences in the distribution p^i . In other words, $p^i = \frac{\#ocu^i}{m}$ is the original distribution.

240 The optimization of $\boldsymbol{\theta}$ is achieved by minimizing the loss function to a threshold value of $2^n \times 1e^{-6}$ using
 241 the gradient descent algorithm with a learning rate lr of 1. Otherwise, the optimization is performed for
 242 pre-defined iterations t . The loss and error function are defined as

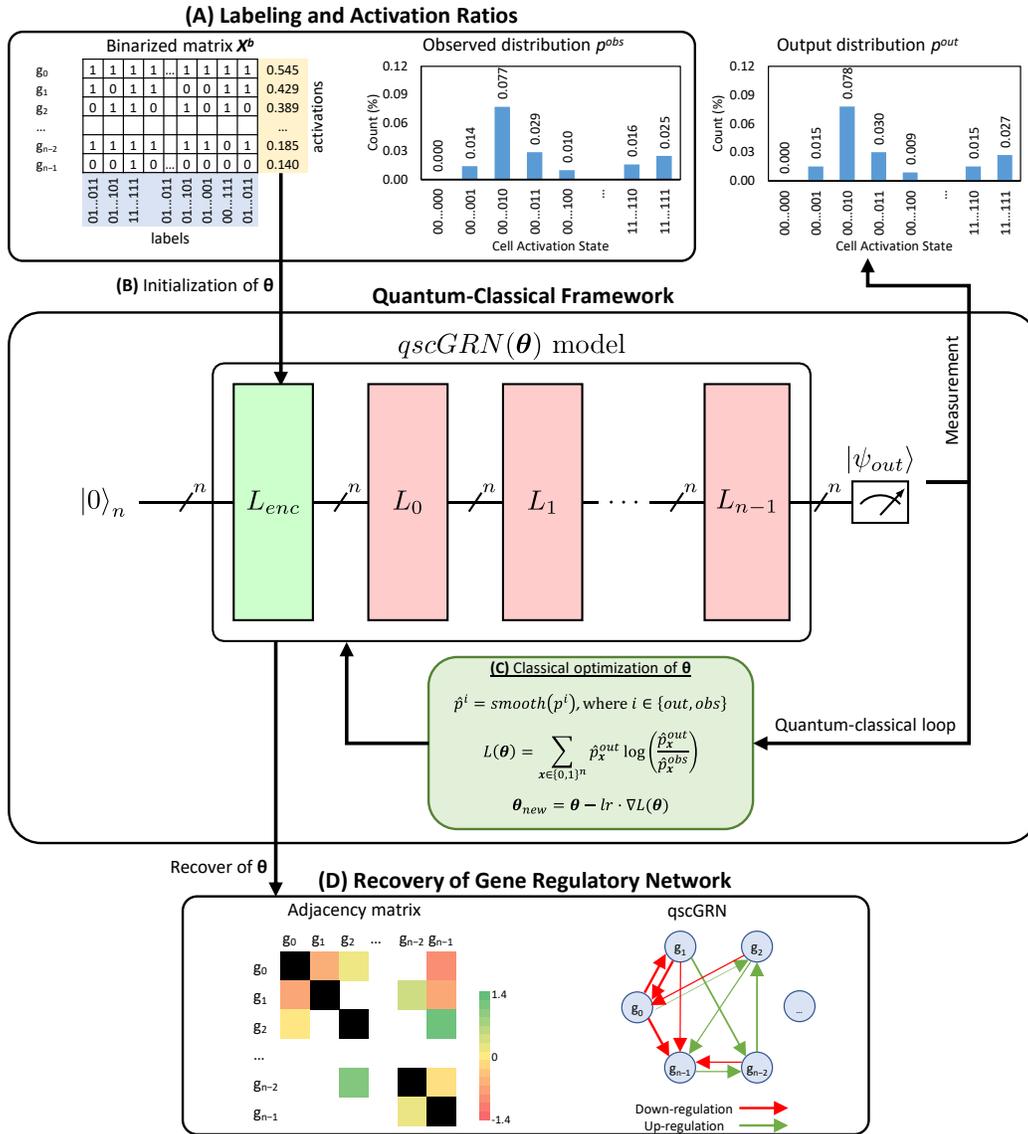
243
$$L(\boldsymbol{\theta}) = \sum_{x \in \{0,1\}^n} \hat{p}_x^{out} \log \left(\frac{\hat{p}_x^{out}}{\hat{p}_x^{obs}} \right),$$

244
$$E(\boldsymbol{\theta}) = \sum_{x \in \{0,1\}^n} (p_x^{out} - p_x^{obs})^2,$$

245 where p_x^i and \hat{p}_x^i are the probability of the state \mathbf{x} in the distributions, $i \in \{out, obs\}$.

246 The parameters $\theta_{k,k}$ in the L_{enc} layer are not trained during optimization according to the assumption
247 that these parameters encode a unique binarized scRNA-seq matrix into the quantum framework. Thus,
248 no training of the parameters $\theta_{k,k}$ implies that the L_{enc} layer encodes the same biological information
249 onto a superposition state in each iteration, making the optimized parameter θ meaningful from a
250 biological perspective.

251 **Recovery of Gene Regulatory Network (FIGURE 3D):** We use the values of parameter θ of qscGRN
252 model to construct the adjacency matrix of the biological scGRN, as described in the matrix
253 representation for qscGRN model step. In the adjacency matrix, parameters with an absolute value less
254 than 0.087 are set to 0 because no significant rotation is performed by the corresponding $c-R_y$ gate. The
255 heatmap representation of the adjacency matrix is shown on the left side of FIGURE 3D, where rows and
256 columns represent control and target genes, respectively, and strong regulation relationships are
257 highlighted. Finally, we construct the signed, directed, weighted network representation (right side of
258 FIGURE 3D) of the biological scGRN using the adjacency matrix.



259

260 **FIGURE 3.** The quantum-classical framework using the qscGRN model to infer the corresponding
 261 biological scGRN. The input matrix \mathbf{X}^b is the binarized scRNA-seq data with n genes selected. (A) Labels
 262 are assigned to each cell in \mathbf{X}^b such that a label is the binarized expression string for the n genes in a cell.
 263 The observed distribution p^{obs} is computed as the percentage of occurrences of each label. The
 264 percentage in the $|0\rangle_n$ label is set to 0, and the rest of p^{obs} is rescaled. The activation ratio of a gene is
 265 the percentage of cells expressing that gene on \mathbf{X}^b . (B) The parameters in the L_{enc} layer are initialized
 266 such that, independently, each qubit has a probability of observing 1 equal to the corresponding

267 activation ratio, and the parameters in L_0, L_1, \dots, L_{n-1} are initialized to zero. In the matrix representation
268 of θ , the diagonal values belong to L_{enc} layer, and the non-diagonal values to the regulation layers. **(C)**
269 The classical optimizer fits the smoothed distributions \hat{p}^{out} into \hat{p}^{obs} using a gradient descent-based
270 algorithm with the KL-divergence as the loss function and the mean square as the error metric. **(D)** The
271 matrix representation of the qscGRN model is used as the adjacency matrix of the biological scGRN. In
272 the adjacency matrix, parameters smaller than 0.087 are dropped (left side), assuming the interaction is
273 not strong enough to report. Thus, the network representation of the biological scGRN (right side) is
274 constructed using the remaining values in the adjacency matrix.

275 Data sets

276 **LCL data set:** The scRNA-seq data was generated from lymphoblastoid cell lines (LCLs), which are widely
277 used cell model systems derived from human primary B cells. Information about the experimental
278 handling and acquisition of data is provided in reference to our original study [28]. The data set has been
279 deposited in the Gene Expression Omnibus (GEO) database with the accession number GSE126321. For
280 this study, we merged this data set with another LCL scRNA-seq data set [29], for which the gene-
281 barcode matrix files were downloaded from the GEO database using the accession number GSE158275.
282 The raw data matrix was pre-processed using scGEAToolbox [39]. The processed matrix of 9,905 genes
283 and 28,208 lymphoblastoid cells was then transformed using the Pearson residuals normalization [38].
284 Then, expression values of six genes (IRF4, REL, PAX5, RELA, PRDM1, AICDA) of the NF- κ B signaling
285 pathway were extracted. The 6-gene expression matrix (6-by-28,208) was used as the input of the
286 qscGRN analysis of this study. The known regulatory relationships between genes were obtained from
287 the previously established B-cell differentiation circuit model [30, 31].

288 III. Results

289 Real data LCL

290 We used our quantum-classical framework to build the qscGRN model—a fully connected quantum
291 circuit—and compute the observed distribution p^{obs} using the data set described in **Methods**. The input
292 scRNA-seq expression matrix contained more than 28,000 cells belonging to the same cell type, the
293 lymphoblastoid cell. The p^{obs} distribution maps the population of the state in the scRNA-seq data into a
294 vector space (p^{obs} is represented in blue—FIGURE 4A). The qscGRN model schema for the data set is a 6-
295 qubit system and consists of an encoder layer and six regulation layers (FIGURE 4B). We measured the
296 output register of the qscGRN model to recover the output distribution p^{out} from the quantum
297 framework. To improve visualization, FIGURE 4A only shows the states with a probability greater than
298 0.01 due to the large size of the vector space.

299 Then, we optimized the parameter θ in the qscGRN model for 50,000 iterations by minimizing the loss
300 function $L(\theta)$ and using the smoothed distributions for p^{out} and p^{obs} . Therefore, the distribution \hat{p}^{out} is
301 fitted into \hat{p}^{obs} during the optimization, as shown in FIGURE 4A that visually proves the similarity of
302 both distributions after optimization. The p^{out} after optimization is represented in orange—FIGURE 4A.
303 The similarity is quantified using the loss function and error metrics that reached values of $8.03e-4$ and
304 $3.04e-5$, respectively.

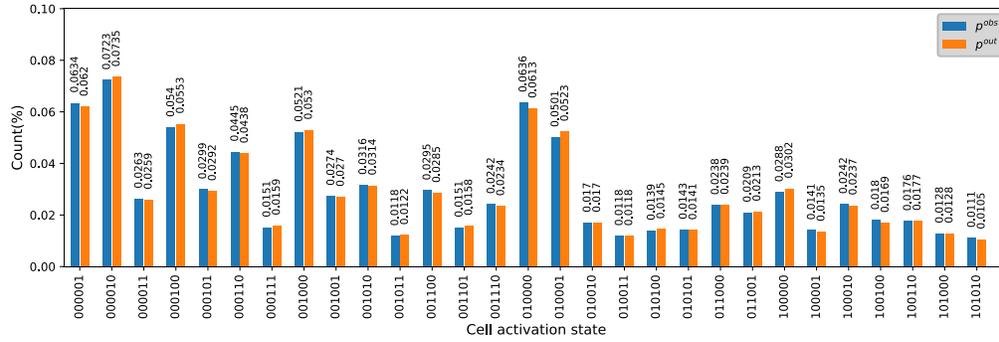
305 The value of the parameter θ after optimization retrieves an adjacency matrix (FIGURE 4C) that is used
306 to construct the biological scGRN. Then, we constructed a weighted, directed network from the
307 quantum framework using the non-diagonal elements of θ , as shown in FIGURE 4D. We compared the
308 sign of the element of each pair of genes with the corresponding regulatory effect in the previously
309 published network, i.e., the baseline GRN [30, 31]. The comparison results were measured using 3
310 metrics: accuracy, f1 score, and precision, to quantify the prediction performance of the classical-

311 quantum framework. The qscGRN model recovers gene-gene relationships with an accuracy, f1-score,
312 and precision of 0.63, 0.72, and 0.78, respectively.

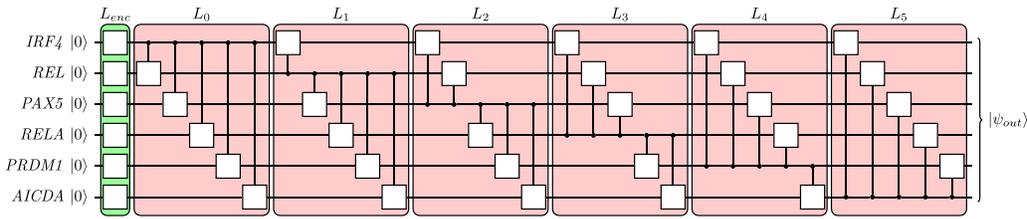
313 FIGURE 4E shows the evolution of parameters for 8 gene-gene pairs (a control-target pair) in the qscGRN
314 model during the optimization. These pairs are relationships recovered from the quantum framework
315 and are present in the NF- κ B network. Gene pairs correctly recovered are represented using a solid line,
316 and pairs incorrectly recovered in long-dash-dot lines. The 8 gene-gene pairs almost reach their
317 optimized value by 10,000 iterations. Specifically, pairs such as IRF4-AICDA, PAX5-AICDA, and PRDM1-
318 PAX5 that have reached their optimized value earlier than others are strongly supported by the
319 quantum framework. IRF4 is known to induce AICDA expression through an indirect mechanism in the
320 NF- κ B signaling cascade [40]. PAX5 is suggested to be a player in the B-lineage-specific control of AICDA
321 transcription in a previous study [41]. PRDM1 is a master regulator that represses PAX5 expression in B
322 cells [42].

323 FIGURE 4F shows 3 gene-gene pairs with regulatory relationships are present in the published baseline
324 GRN but are removed by our qscGRN estimation. The exclusion of these gene pairs suggests that the
325 regulatory relationships between them might be through indirect links. The baseline model failed to
326 distinguish between indirect and direct links. For example, one of the removed gene pairs is IRF4-
327 PRDM1 which has a parameter larger than 0.1 during the first 1,000 iterations. The parameter value
328 decreases gradually to close to zero after 5,000 iterations. The dropping suggests that IRF4-PRDM1's
329 regulatory relationship might be through a third-party modulator. Indeed, IRF4 is known to inhibit BCL6
330 expression, and because BCL6 can repress PRDM1 [43, 44], it has been formally speculated that the
331 effects of IRF4 on PRDM1 expression might have been mediated through inhibition of BCL6 expression
332 [45].

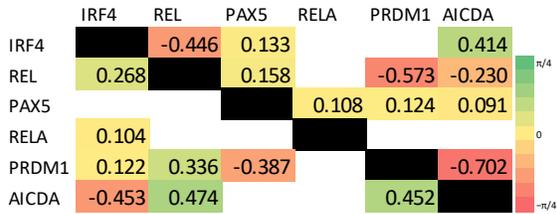
333 Ten novel regulatory relationships between genes, which are not in the published baseline GRN, were
334 predicted by the qscGRN model when the computation nearly reached their optimized value by 10,000
335 iterations (FIGURE 4G). We found that, for at least two of these newly discovered regulatory
336 relationships, there is experimental evidence supporting the involved gene pairs. The first gene pair is
337 IRF4-PAX5, for which our model predicted that PAX5 is positively regulated by IRF4. Indeed, in a previous
338 study of the phosphoinositide-3 kinase signaling in B cells [46], an experiment searching for transcription
339 factors that are activated by FOXO1 revealed that IRF4 is a potential candidate for PAX5 regulation. The
340 second gene pair is PRDM1-AICDA. PRDM1 has been found to be able to silence the expression of
341 AICDA, probably in a dose-dependent manner [47]. These results suggest that our qscGRN method
342 recovered regulatory relationships that were missed in the published baseline model.



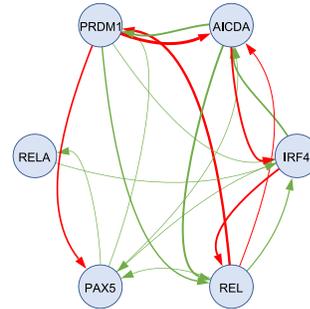
A



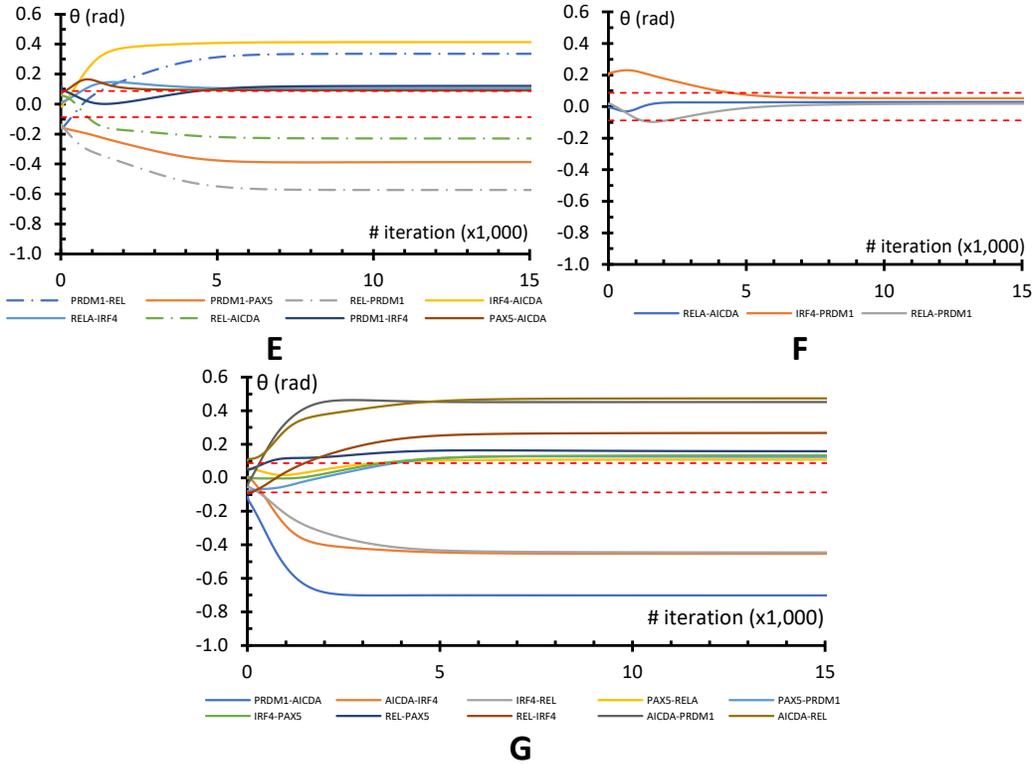
B



C



D



344

345 **FIGURE 4.** Application of the qscGRN modeling with real scRNA-seq data. (A) The observed and output
346 distributions (p^{obs} and p^{out}) colored in blue and orange, respectively. p^{obs} is computed using the
347 transformed scRNA-seq data. p^{out} is obtained by measuring the output register in the qscGRN model
348 when the optimization is completed. (B) The schematic for a 6-genes qscGRN model consists of a L_{enc}
349 layer and six regulation layers that map relationships between genes in the quantum framework. (C) The
350 adjacency matrix of the biological scGRN, in which parameters with an absolute value less than 0.087
351 are removed. The heatmap reveals the strength and direction of the interaction for a gene-gene
352 interaction. The diagonal elements are colored in black due to these parameters are not trained during
353 optimization. (D) A directed-weighted representation of the biological scGRN recovered from the
354 quantum circuit. Up-regulation and down-regulation relationships are colored in green and red,
355 respectively. The thickness of each edge is proportional to the absolute value of the corresponding
356 parameter in the adjacency matrix. (E) Evolution of parameters for gene-gene interaction that were

357 recovered from the quantum framework and are present in the NF- κ B network during the first 15,000
358 iterations. **(F)** Evolution of parameters for gene-gene interactions that were not recovered but are
359 present in the NF- κ B network during the first 15,000 iterations. **(G)** Evolution of parameters for gene-
360 gene interaction that are predicted by the quantum framework during the first 15,000 iterations.

361 IV. Discussion

362 Finding ways to apply quantum computing in biological research is an active research area [48-51].
363 Many questions in biology can benefit from quantum computing by exploring many possible parallel
364 computational paths, but identifying such questions remains challenging. Especially understanding how
365 to exploit quantum computers for progress in solving important biological questions is crucial. The latest
366 development of scRNA-seq has made it possible to gather the transcriptome information from tens of
367 thousands of individual cells in a high-throughput manner. These complex data sets with unprecedented
368 detail are driving the development of new computational and statistical tools that are revolutionizing
369 our understanding of cellular processes. However, quantum computation has not yet received enough
370 attention in the face of this single-cell big data revolution. As a consequence, we present our qscGRN
371 method for modeling interactions between genes to derive the quantum computing framework for
372 constructing GRNs. Below we discuss several aspects of application issues.

373 Mapping of scRNA-seq data in the quantum framework

374 Typically, a correlation-based method obtains enough information from a scRNA-seq data set to infer a
375 GRN with a large number of genes. The gene-gene interaction is calculated as a single value (summary
376 statistic) using the expression values of the available cells. On the other hand, our quantum approach for
377 GRN inference models a small number of genes due to the vector space size, which is equal to the
378 number of basis states, increases exponentially with the number of genes. In other words, the number
379 of cells in binarized scRNA-seq data might be mapped to a moderate number of basis states such that

380 each basis state is mapped from the biological data. For example, a 15-qubit qscGRN model requires
381 $2^{15}=32,768$ basis states; however, a scRNA-seq data set with roughly 32,768 cells would retrieve an
382 observed distribution without meaningful biological information mapped to the basis states. The latest
383 scRNA-seq technology has the capacity to allow the transcriptome of millions of cells to be measured. To
384 obtain enough cells, we may also integrate multiple scRNA-seq data sets from the same cell types or
385 similar biological sources and perform statistical correction to remove the batch effect. We can select
386 informative genes such as highly variable genes [52] in advance, then proceed with our quantum-
387 classical pipeline. Thus, reducing the burden of a large number of genes in the model while maintaining
388 the biological relevance.

389 Initialization values for the parameter θ in the qscGRN model

390 The initialization of the parameter θ determines the starting point in the landscape of the loss function,
391 thus indirectly setting the difficulty for the optimization due to barren plateaus—which is an issue when
392 optimizing a parameterized quantum circuit. In section II.3. Initialization of the parameter θ in the
393 qscGRN, an all-zeros approach is taken for the parameters in the $c-R_y$ gates. Additionally, 2 more
394 initialization approaches for the $c-R_y$ gates were tried using a random initialization with uniform and
395 normal distributions. The 3 methods initialize the parameter θ at 3 different positions in the landscape;
396 however, only the all-zeros approach defines the same point when running the workflow again. Thus,
397 the all-zeros approach would recover a biological scGRN consistently from the quantum framework
398 since the gradients are computed at the same starting point. Finally, the all-zeros approach recovered 8
399 gene-gene interactions with biological support, which is larger than the other approaches.

400 Advantages of the qscGRN over correlation- and regression-based GRNs

401 Correlation and regression-based methods are the most widely used methods for GRN inference due to
402 their computational efficiency. These methods typically compute a correlation or regression coefficient

403 for a gene pair using the total number of cells in the data. The issue with correlation and regression
404 methods is that they rely on summary statistics. The relationship between the two genes is measured
405 using a single value of the summary statistics: correlation or regression coefficient. Once computed, the
406 coefficient becomes independent of the total number of cells. Moreover, increasing the number of cells
407 would not substantially change the correlation and regression coefficients. Thus, information in the
408 scRNA-seq data is not fully used. The other issue is that the coefficient is computed only between the
409 two focal genes, regardless of the expression values of other genes in the same biological system. The
410 effect of not considering other genes in the computation can result in a biased coefficient, which does
411 not represent the true behavior of the interaction. There are methods such as partial correlation [8],
412 principal component regression [9], and LASSO [53] that may correct this. But the correcting effect is
413 limited given the all-to-all interactions cannot be modeled.

414 In contrast, our qscGRN method is based on the quantum framework, which uses the Hilbert space and
415 maps the binarized expression of each cell for n genes at once. The mapping of n genes to the Hilbert
416 space allows manipulating the superposition of basis states in an n -qubit system and fitting the output
417 distribution to the observed distribution in the binarized scRNA-seq data. Thus, biological information
418 passes through the quantum circuit and is encoded in a superposition state. In other words, a gene-gene
419 relationship is computed using information from an n gene biological system all at once in the quantum
420 framework, which is an improvement compared to correlation-based methods.

421 Acknowledgments

422 This work was supported by the DoD grant GW200026 for JJC.

423 Data availability

424 The scRNA-seq data analyzed during the current study is available in the NCBI GEO database with the
425 accession numbers GSE126321 and GSE158275. The processed data and the source code
426 implementation of the qscGRN package are provided in the GitHub repository at
427 <https://github.com/cailab-tamu/QuantumGRN/>. The repository also includes tutorials written in Python
428 language.

429 Author contributor

430 Conceptualization, JJC; methodology, CR and JJC; implementation of the software, CR; formal analysis,
431 CR and JJC; writing and editing, CR and JJC; supervision, JJC. All authors reviewed and contributed to the
432 manuscript.

433 Competing Interest

434 The authors declare no competing financial or non-financial interest.

435 References

- 436 1. Huynh-Thu, V.A. and G. Sanguinetti, *Gene Regulatory Network Inference: An Introductory*
437 *Survey*, in *Gene Regulatory Networks: Methods and Protocols*, G. Sanguinetti and V.A. Huynh-
438 Thu, Editors. 2019, Springer New York: New York, NY. p. 1-23.
- 439 2. Hecker, M., et al., *Gene regulatory network inference: data integration in dynamic models-a*
440 *review*. *Biosystems*, 2009. **96**(1): p. 86-103.
- 441 3. Chen, S. and J.C. Mar, *Evaluating methods of inferring gene regulatory networks highlights their*
442 *lack of performance for single cell gene expression data*. *BMC Bioinformatics*, 2018. **19**(1): p.
443 232.
- 444 4. Pratapa, A., et al., *Benchmarking algorithms for gene regulatory network inference from single-*
445 *cell transcriptomic data*. *Nat Methods*, 2020. **17**(2): p. 147-154.
- 446 5. Diaz, L.P.M. and M.P.H. Stumpf, *Gaining confidence in inferred networks*. *Sci Rep*, 2022. **12**(1): p.
447 2394.
- 448 6. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells*. *Nat Commun*,
449 2017. **8**: p. 14049.
- 450 7. Chan, T.E., M.P.H. Stumpf, and A.C. Babbie, *Gene Regulatory Network Inference from Single-Cell*
451 *Data Using Multivariate Information Measures*. *Cell Syst*, 2017. **5**(3): p. 251-267 e3.
- 452 8. Kim, S., *ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients*.
453 *Commun Stat Appl Methods*, 2015. **22**(6): p. 665-674.
- 454 9. Osorio, D., et al., *scTenifoldNet: A Machine Learning Workflow for Constructing and Comparing*
455 *Transcriptome-wide Gene Regulatory Networks from Single-Cell Data*. *Patterns (N Y)*, 2020. **1**(9):
456 p. 100139.
- 457 10. Huynh-Thu, V.A., et al., *Inferring regulatory networks from expression data using tree-based*
458 *methods*. *PLoS One*, 2010. **5**(9).

- 459 11. Kotiang, S. and A. Eslami, *A probabilistic graphical model for system-wide analysis of gene*
460 *regulatory networks*. *Bioinformatics*, 2020. **36**(10): p. 3192-3199.
- 461 12. Lahdesmaki, H., et al., *Relationships between probabilistic Boolean networks and dynamic*
462 *Bayesian networks as models of gene regulatory networks*. *Signal Processing*, 2006. **86**(4): p.
463 814-834.
- 464 13. Friedman, N., et al., *Using Bayesian networks to analyze expression data*. *J Comput Biol*, 2000.
465 **7**(3-4): p. 601-20.
- 466 14. Shmulevich, I., et al., *Probabilistic Boolean Networks: a rule-based uncertainty model for gene*
467 *regulatory networks*. *Bioinformatics*, 2002. **18**(2): p. 261-74.
- 468 15. Gyongyosi, L. and S. Imre, *A Survey on quantum computing technology*. *Computer Science*
469 *Review*, 2019. **31**: p. 51-71.
- 470 16. Egger, D.J., et al., *Quantum Computing for Finance: State-of-the-Art and Future Prospects*. *IEEE*
471 *Transactions on Quantum Engineering*, 2020. **1**: p. 1-24.
- 472 17. Fernández-Caramès, T.M. and P. Fraga-Lamas, *Towards Post-Quantum Blockchain: A Review on*
473 *Blockchain Cryptography Resistant to Quantum Computing Attacks*. *IEEE Access*, 2020. **8**: p.
474 21091-21116.
- 475 18. Ramezani, S.B., et al. *Machine Learning Algorithms in Quantum Computing: A Survey*. in *2020*
476 *International Joint Conference on Neural Networks (IJCNN)*. 2020.
- 477 19. Cao, Y., J. Romero, and A. Aspuru-Guzik, *Potential of quantum computing for drug discovery*.
478 *IBM Journal of Research and Development*, 2018. **62**(6): p. 6:1-6:20.
- 479 20. Bauer, B., et al., *Quantum Algorithms for Quantum Chemistry and Quantum Materials Science*.
480 *Chemical Reviews*, 2020. **120**(22): p. 12685-12717.
- 481 21. Bharti, K., et al., *Noisy intermediate-scale quantum algorithms*. *Reviews of Modern Physics*,
482 2022. **94**(1): p. 015004.

- 483 22. Huang, H.Y., et al., *Quantum advantage in learning from experiments*. Science, 2022. **376**(6598):
484 p. 1182-1186.
- 485 23. Shor, P.W., *Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a*
486 *Quantum Computer*. SIAM Review, 1999. **41**(2): p. 303-332.
- 487 24. Bhatia, V. and K.R. Ramkumar. *An Efficient Quantum Computing technique for cracking RSA*
488 *using Shor's Algorithm*. in *2020 IEEE 5th International Conference on Computing Communication*
489 *and Automation (ICCCA)*. 2020.
- 490 25. Peng, F., D. Schuurmans, and S. Wang, *Augmenting Naive Bayes Classifiers with Statistical*
491 *Language Models*. Information Retrieval, 2004. **7**(3): p. 317-345.
- 492 26. Ruder, S., *An overview of gradient descent optimization algorithms*. arXiv preprint
493 arXiv:1609.04747, 2016.
- 494 27. Kullback, S. and R.A. Leibler, *On Information and Sufficiency*. The Annals of Mathematical
495 Statistics, 1951. **22**(1): p. 79-86.
- 496 28. Osorio, D., et al., *Single-cell RNA sequencing of a European and an African lymphoblastoid cell*
497 *line*. Scientific Data, 2019. **6**(1): p. 112.
- 498 29. SoRelle, E.D., et al., *Single-cell RNA-seq reveals transcriptomic heterogeneity mediated by host-*
499 *pathogen dynamics in lymphoblastoid cell lines*. Elife, 2021. **10**.
- 500 30. Roy, K., et al., *A Regulatory Circuit Controlling the Dynamics of NFkappaB cRel Transitions B Cells*
501 *from Proliferation to Plasma Cell Differentiation*. Immunity, 2019. **50**(3): p. 616-628 e6.
- 502 31. Sciammas, R., et al., *An incoherent regulatory network architecture that orchestrates B cell*
503 *diversification in response to antigen signaling*. Mol Syst Biol, 2011. **7**: p. 495.
- 504 32. Meter, R.V. and M. Oskin, *Architectural implications of quantum computing technologies*. J.
505 Emerg. Technol. Comput. Syst., 2006. **2**(1): p. 31-63.

- 506 33. Dirac, P.A.M., *A new notation for quantum mechanics*. Mathematical Proceedings of the
507 Cambridge Philosophical Society, 1939. **35**(3): p. 416-418.
- 508 34. Rieffel, E. and W. Polak, *An introduction to quantum computing for non-physicists*. ACM Comput.
509 Surv., 2000. **32**(3): p. 300–335.
- 510 35. Lee, E.A., *The Past, Present and Future of Cyber-Physical Systems: A Focus on Models*. Sensors,
511 2015. **15**(3).
- 512 36. Vatan, F. and C. Williams, *Optimal quantum circuits for general two-qubit gates*. Physical Review
513 A, 2004. **69**(3): p. 032315.
- 514 37. Mosca, M., *Quantum algorithms*. arXiv preprint arXiv:0808.0369, 2008.
- 515 38. Lause, J., P. Berens, and D. Kobak, *Analytic Pearson residuals for normalization of single-cell
516 RNA-seq UMI data*. Genome Biol, 2021. **22**(1): p. 258.
- 517 39. Cai, J.J., *scGEAToolbox: a Matlab toolbox for single-cell RNA sequencing data analysis*.
518 Bioinformatics, 2019.
- 519 40. Sciammas, R., et al., *Graded Expression of Interferon Regulatory Factor-4 Coordinates Isotype
520 Switching with Plasma Cell Differentiation*. Immunity, 2006. **25**(2): p. 225-236.
- 521 41. Yadav, A., et al., *Identification of a ubiquitously active promoter of the murine activation-induced
522 cytidine deaminase (AICDA) gene*. Molecular Immunology, 2006. **43**(6): p. 529-541.
- 523 42. Boi, M., et al., *PRDM1/BLIMP1: a tumor suppressor gene in B and T cell lymphomas*. Leukemia &
524 Lymphoma, 2015. **56**(5): p. 1223-1228.
- 525 43. Tunyaplin, C., et al., *Direct repression of prdm1 by Bcl-6 inhibits plasmacytic differentiation*. J
526 Immunol, 2004. **173**(2): p. 1158-65.
- 527 44. Shapiro-Shelef, M. and K. Calame, *Regulation of plasma-cell development*. Nat Rev Immunol,
528 2005. **5**(3): p. 230-42.

- 529 45. Teng, Y., et al., *IRF4 negatively regulates proliferation of germinal center B cell-derived Burkitt's*
530 *lymphoma cell lines and induces differentiation toward plasma cells*. European Journal of Cell
531 Biology, 2007. **86**(10): p. 581-589.
- 532 46. Abdelrasoul, H., et al., *PI3K induces B-cell development and regulates B cell identity*. Sci Rep,
533 2018. **8**(1): p. 1327.
- 534 47. Nutt, S.L., et al., *The genetic network controlling plasma cell differentiation*. Semin Immunol,
535 2011. **23**(5): p. 341-9.
- 536 48. Outeiral, C., et al., *The prospects of quantum computing in computational molecular biology*.
537 WIREs Computational Molecular Science, 2020. **11**(1): p. e1481.
- 538 49. Marx, V., *Biology begins to tangle with quantum computing*. Nat Methods, 2021. **18**(7): p. 715-
539 719.
- 540 50. Emani, P.S., et al., *Quantum computing at the frontiers of biological sciences*. Nat Methods,
541 2021. **18**(7): p. 701-709.
- 542 51. Cheng, H.P., et al., *Application of Quantum Computing to Biochemical Systems: A Look to the*
543 *Future*. Front Chem, 2020. **8**: p. 587143.
- 544 52. Osorio, D., et al., *Single-Cell Expression Variability Implies Cell Function*. Cells, 2019. **9**(1).
- 545 53. Tibshirani, R., *Regression Shrinkage and Selection via the Lasso*. Journal of the Royal Statistical
546 Society. Series B (Methodological), 1996. **58**(1): p. 267-288.
- 547