

SOFTWARE

Open Access



Multioviz: an interactive platform for in silico perturbation and interrogation of gene regulatory networks

Helen Xie¹, Lorin Crawford^{2,3,4*}  and Ashley Mae Conard^{1,2,3*} 

[†]Lorin Crawford and Ashley Mae Conard have contributed equally.

*Correspondence:
lcrawford@microsoft.com;
ashleyconard@microsoft.com

¹ Department of Computer Science, Brown University, Providence, RI, USA

² Center for Computational Molecular Biology, Brown University, Providence RI, USA

³ Microsoft Research, Cambridge, MA, USA

⁴ Department of Biostatistics, Brown University, Providence, RI, USA

Abstract

In this paper, we aim to build a platform that will help bridge the gap between high-dimensional computation and wet-lab experimentation by allowing users to interrogate genomic signatures at multiple molecular levels and identify best next actionable steps for downstream decision making. We introduce `MULTIOVIZ`: a publicly accessible R package and web application platform to easily perform in silico hypothesis testing of generated gene regulatory networks. We demonstrate the utility of `MULTIOVIZ` by conducting an end-to-end analysis in a statistical genetics application focused on measuring the effect of in silico perturbations of complex trait architecture. By using a real dataset from the Wellcome Trust Centre for Human Genetics, we both recapitulate previous findings and propose hypotheses about the genes involved in the percentage of immune CD8+ cells found in heterogeneous stocks of mice. Source code for the `MULTIOVIZ` R package is available at <https://github.com/lcrawlab/multio-viz> and an interactive version of the platform is available at <https://multioviz.ccv.brown.edu/>.

Introduction

Phenotypic architecture is often driven by a collection of biological processes that occur through dynamic interactions across various molecular levels, including single nucleotide polymorphisms (SNPs), genes, and proteins [1]. Gene regulatory networks (GRNs) are directed graphs that effectively allow for the visualization of interactions between these components that constitute cellular pathways and signaling cascades [2]. Each node in a GRN represents a molecular variable such as a SNP or a gene, with each edge representing the interaction between two nodes. By simultaneously characterizing phenotypes at multiple genomic levels and modeling their interactions as a GRN, practitioners and data analysts can identify significant molecular variables for follow-up studies (e.g., through knockout experiments) [3].

Unfortunately, finding cost effective ways to investigate how a set of perturbations on a GRN will drive changes within a phenotype remains a challenge—especially as sequencing technologies continue to advance and, with this new depth, the space of potential



biomarkers continues to grow. This motivates using *in silico* approaches to explore initial hypotheses and to identify actionable candidates for downstream tasks. To date, many computational methods have been developed for this purpose in high-dimensional multi-omics datasets [4, 5]. These platforms often leverage variable importance measures, such as *p*-values and posterior inclusion probabilities (PIP), to infer GRNs [6, 7]. However, despite their usefulness, the software accompanying these algorithms usually require a fair amount of coding expertise to run them—thus, posing a challenge particularly for non-computational users [8]. Furthermore, the outputs of these methods are usually just lists of potential biomarker candidates which are not always easily amenable for determining the best next experimental action. To that end, there is a need for an accessible interactive platform that leverages statistical variable selection methods and subsequently enables non-computational researchers to efficiently test biological hypotheses *in silico*, prior to spending time and money in the wet-lab.

To meet this need in the field, we developed *MultiOviz*: a web-based platform and R package for *in silico* exploration and assessment of GRNs. While many GRN platforms have been developed, a majority do not allow for perturbation analyses where a user is able to impose modifications onto a network (i.e., the addition or subtraction of a node or edge) and invoke a statistical reanalysis to learn how a phenotype might change with new sets of molecular interactions [9–13]. More notably, existing platforms that do indeed have the capability to incorporate perturbation analyses, often do not offer a user-friendly interactive environment for efficiently visualizing changes to GRNs [14, 15]. The key contribution of *MultiOviz* is that it enables *in silico* perturbation experiments within an easy-to-use interface that includes the following three main features (Table 1). First, it allows users to couple summary statistics from a computational analysis (e.g., *p*-values or PIPs) along with a set of biological annotations (e.g., SNPs within the boundary of a gene) to visualize multi-level genomic relationships in the form of a GRN. Second, it allows users to perturb these learned networks and investigate the associated ramifications on a phenotype of interest. Lastly, *MultiOviz* integrates various variable

Table 1 *MultiOviz* combines the features of existing platforms to present a unified platform for gene regulatory network (GRN) based *in silico* hypothesis testing and perturbation analyses. Comparable platforms listed include: OpenXR [9], *vissE.cloud* [10], MONGKIE [11], MiBiOmics [12], GeNeCK [13], *scTenifoldKnk* [14], and GenYsis [15]

	MultiOviz	OpenXGR	<i>vissE.cloud</i>	MONGKIE	MiBiOmics	GeNeCK	<i>scTenifoldKnk</i>	GenYsis
Perturbation analysis	✓						✓	✓
Flexible inputs	✓	✓	✓			✓		
Multi-omic integration	✓	✓	✓	✓	✓			
Variable selection	✓	✓	✓	✓			✓	✓
Multi-level analysis	✓	✓	✓	✓	✓			
Interactive platform	✓	✓	✓	✓	✓	✓		
Total # of features	6	5	5	4	3	2	2	2

selection methods to give users a wide choice of statistical approaches that they can use to generate relevant multi-level genomic signatures for their analyses.

Overall, *Multioviz* provides an intuitive approach to *in silico* hypothesis testing, even for individuals with less computational and coding experience. Here, a user starts by inputting molecular data along with an associated phenotype to graphically visualize the relationships between significant variables. The user can then “knock out” a node in the GRN and rerun the statistical variable selection step to observe the effect of the perturbation. As a general illustration of our proposed platform, we will demonstrate how to perform perturbation analyses with the *Multioviz* online platform using “Biologically Annotated Neural Networks” (BANNs) which are a class of feedforward Bayesian machine learning models that integrate known biological relationships to perform association mapping on multiple molecular levels simultaneously [7]. The rest of the paper is organized as follows. In the next section, we describe the methodological and engineering details behind the three main features of *Multioviz*. Next, we demonstrate how to perform perturbation analyses using *Multioviz* with real quantitative traits assayed in a heterogeneous stock of mice from Wellcome Trust Centre for Human Genetics [16]. We also compare the GRN outputs produced by *Multioviz* to those generated by comparable platforms during a perturbation analysis on the same real data. Finally, we close with a discussion and a look towards future research directions. We believe that the *Multioviz* platform and its application are a step towards providing practitioners the ability to perform true human-in-the-loop assessment of the biological processes driving complex phenotypes and diseases.

Materials and methods

The *Multioviz* platform allows the user to (i) intuitively visualize gene regulatory networks (GRNs) from multi-omics data (ii) perform *in silico* hypothesis testing through perturbing those GRNs and uncovering the effect the phenotypic architecture and (iii) allows these features to be leveraged with virtually any variable selection method. The general end-to-end workflow of *Multioviz* is intended to be intuitive and straightforward to all users regardless of coding experience (Fig. 1). To begin, a user first inputs individual level data or summary statistics derived from a multi-omic dataset (Fig. 1a, b). In this paper, for the second step, we will demonstrate *Multioviz* using BANNs to perform variable selection on these input data. After statistically significant variables are identified, *Multioviz* outputs a GRN where the nodes correspond to genomic units (e.g., SNPs or genes) and the edges between nodes symbolize that there is some functional relationship that connects them. The BANNs method produces a posterior inclusion probability (PIP) for each molecular variable. These PIP scores lie on the unit interval and provide a prioritization score for each genomic variable in the data—with values closer to 1 indicating greater statistical significance [17]. In the *Multioviz* user-interface, these PIP values are displayed in different colors for nodes and edges, respectfully, which provides an interpretable view of important molecular variables. Insignificant SNPs appear yellow, progressing to more red for those that are significant. Similarly, insignificant genes appear as light blue and then progress to dark blue for more significant genes (Fig. 1c). In the third step of the *Multioviz* workflow, the user then has the flexibility to perturb any part of the GRN within the interface (e.g., by

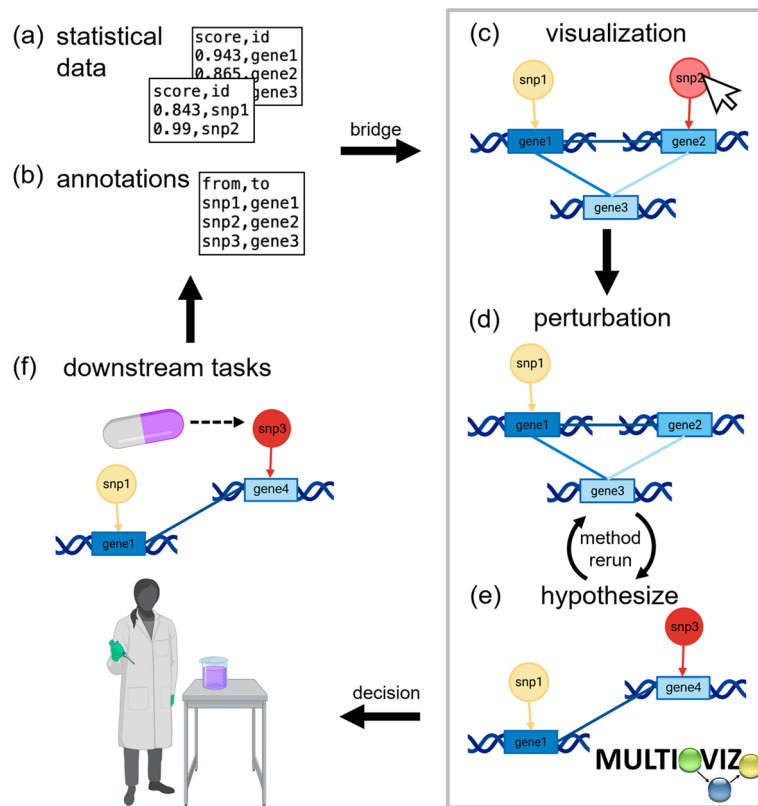


Fig. 1 Schematic overview of running an end-to-end computational analysis with the MultiOviz platform. **a**, **b** The user uploads their own individual-level data or summary statistics derived from an omics study. **c** Input data are visualized as gene regulatory networks (GRNs). Here, darker node colors denote greater statistical significance for a genomic variable. The mapping within and between molecular levels are given via edges which share the same color as the out-degree node. **d** MultiOviz allows users to visualize and perturb GRNs from a prioritized list of significant molecular variables. **e** Through the perturbation feature, users can explore their generated GRN and delete nodes, then rerun statistical analyses to produce a new GRN. Subsequent rerunning of the variable selection method regenerates data to be visualized as an updated GRN. **f** Human-in-the-loop perturbation analyses provide better informed *in silico* hypotheses to be tested and validated in the wet-lab

adding or removing variable nodes from the graph) to investigate *in silico* hypotheses (Fig. 1d). The user can subsequently click a button to rerun the statistical analyses (e.g., BANNs in web application or another variable selection method in the MultiOviz R package), and observe the newly visualized GRN (Fig. 1e). The human-in-the-loop perturbation analyses provided by MultiOviz will hopefully lead to better informed hypotheses to be tested and validated in the wet-lab for downstream tasks. In this section, we describe the GRN visualization, perturbation, and R package features in more detail.

Interpretable visualization of gene regulatory networks

The first step of the MultiOviz workflow is to visualize molecular variables in the context of a GRN (Fig. 2). The minimum required input is a file with two columns: (i) `id` which lists the molecular variables of interest and (ii) `score` which provides an associated summary statistic for each. MultiOviz directly visualizes these data as a GRN.

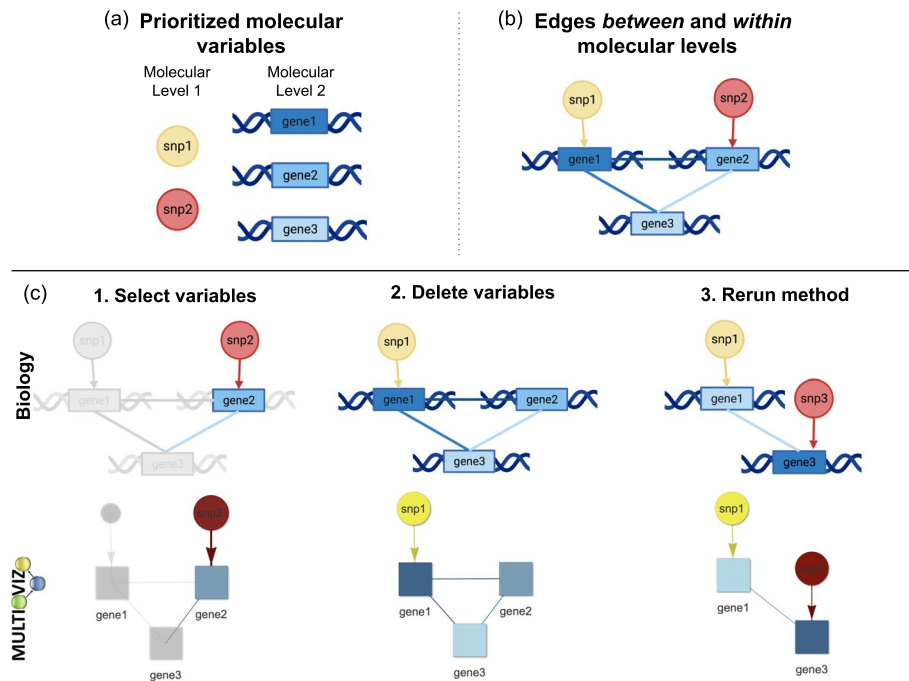


Fig. 2 Components of gene regulatory networks (GRNs) and a schematic overview for performing perturbation analyses in Multioviz. **a** Visualization of the components making up a GRN. Here, variables (e.g., SNPs and genes) are represented as nodes. The shape of these nodes are different depending on the molecular level that they represent and the color scheme describes the significance level of each variable according to a statistical model (e.g., *p*-value or PIP). Insignificant SNPs are more yellow and more statistically important SNPs are depicted in red. Similarly, insignificant genes are light blue, while significant genes are dark blue. **b** Directed edges are used to map nodes between molecular levels (e.g., SNPs reside in the boundaries of genes) [18]. Since we do not assume to have access to temporal information, interactions between variables on the same molecular level are represented by undirected edges. There are three classes of edges between and within molecular levels: no connectivity, sparse connectivity, and complete connectivity. Here, the first molecular level has no connectivity since there are no direct interactions between SNPs. However, the second molecular level has complete connectivity because there is an interaction between all genes. The nodes and edges together form a visual representation of a GRN. **c** To emulate perturbation analyses, Multioviz allows users to select molecular variables (i.e., nodes), delete them, and rerun the statistical analysis to generate a new GRN. To perform this type of analysis within in the Multioviz interface, users simply highlight the variable of interest by (1) clicking on the node and selecting “Edit”, (2) clicking on “Delete selected”, and then (3) clicking “Rerun” under the “Perturb” left drop-down menu. An overview of the Multioviz interface can be found in Fig. 3

The shape of each node in the GRN corresponds to the molecular level (e.g., a SNP versus a gene) and the node color represents the importance of each variable (e.g., PIP or *p*-value) (Fig. 2a). Overall, variables with greater degrees of significance are plotted with darker colors. For the purposes of demonstrating Multioviz functionality, we illustrate multi-omic data with SNPs and genes as the two molecular levels. Multioviz color-codes the first (SNPs) and second molecular levels (genes) as yellow to orange circles and light to dark blue rectangles, respectively. Because Multioviz leverages biological annotations that allow for the inference of biological hierarchies, there are directed edges between molecular levels. For example, given that SNPs can occur within the boundary of a gene and affect its function, we represent the interaction between the two levels as directed arrows when moving from the SNP to gene level [18]. Edges within the same molecular level are undirected, since we do not assume to have information

on temporality. To summarize, for a given GRN, there are three types of edge connectivities: (1) no connectivity between the nodes belonging to one molecular level because those variables do not interact biologically, (2) sparse connectivity, or (3) complete connectivity. In Fig. 2b, we show no connections on the SNP-level and complete connectivity between genes. For a given GRN, the user can then select a subset of molecular variables to perturb (i.e., add or delete) and rerun the method to identify which variables are significant in this new biological context (Fig. 2c).

When working directly in the MultiOviz interface, users can click the “Visualization” left-hand side drop-down to upload their own file of statistical importance scores for variables up to two molecular levels (Fig. 3a). This file should be a two-dimensional matrix with the column names labeled as “score” and “id”, respectively. Once these inputs are uploaded, the user can click “Run” to construct and view a corresponding GRN (Fig. 3b). Note that clicking on a specific node will highlight the variable itself along with its connected neighbors. If desired, the user can also upload their own biological annotations to define *a priori* relationships and generate sparse edges between genomic levels (e.g., SNPs-to-genes or genes-to-pathways). To further explore the output of the GRN,

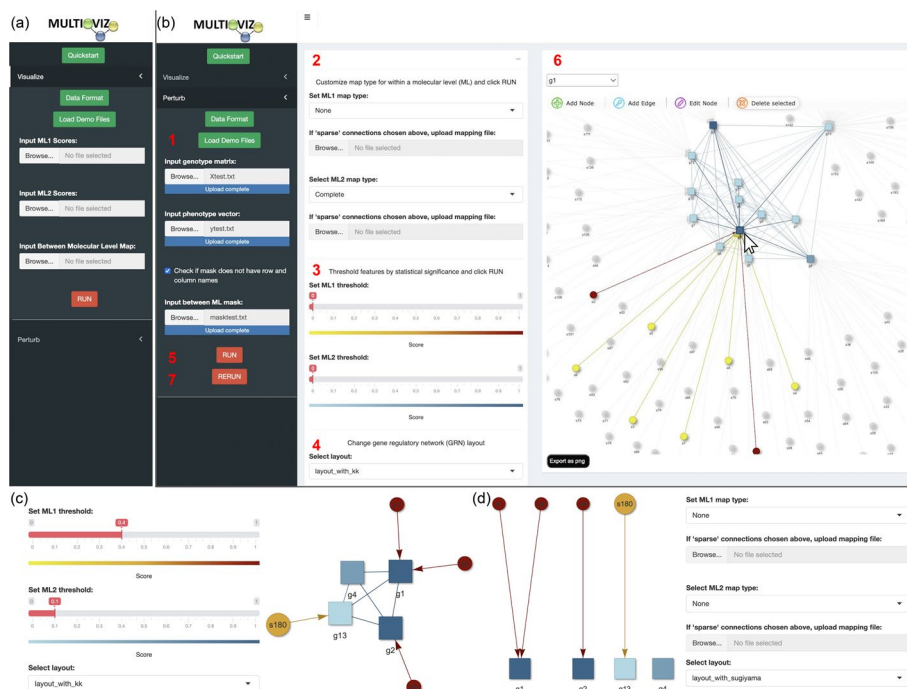


Fig. 3 Overview of the MultiOviz online platform user interface. **a** Users can select the visualization drop down to upload pre-generated variable ranks and annotation maps between molecular variables. **b** Under the perturbation drop down, the user can follow the outlined steps numbered in red. In step (1), the user will upload the required genotype matrix, phenotype vector, and biological annotation map (often given as a binary matrix where a “1” means that a variable belongs to a given group). In steps (2) through (4), the user can customize the type of method being used for association mapping, the threshold used to determine variable significance, and the layout for their gene regulatory network. In step (5) run BANNs (or some other variable selection approach) and subsequently generate the GRN. In step (6), the user can perform a perturbation analysis where they can select and delete variables of interest. Finally in step (7), the user will rerun the method to test in silico hypotheses. **c** MultiOviz allows the user to adjust significance thresholds for each molecular level. **d** The user is able to specify the degree of mapping within each molecular level, thereby changing and/or modifying the GRN layout

MultiOviz offers the user the flexibility to set importance thresholds for each molecular level and filter out variables with low significance (Fig. 3c). This can be particularly important when generating GRNs from large datasets with many variables. Finally, users can manually modify the GRN after it is generated to create layouts that are most digestible (Fig. 3d).

Facilitated perturbation analyses for in silico experimentation

The flexible functionality of MultiOviz allows for the in silico testing of hypotheses where the nodes and edges of a learned GRN can be perturbed to observe the influence of different molecular variables onto a phenotype of interest. These changes are run as altered inputs in a variable selection model that runs in the background of the software which then generates a new set of significant molecular variables that are then visualized as another interactive GRN (see Fig. 3). In the MultiOviz web application, the statistical method that is implemented is BANNs [7]. Like many linear and nonlinear models, BANNs requires a genotype matrix and a phenotype vector as input. Consider a biological study with N observations (e.g., the number of individuals, cells, tissues) that have been phenotyped for some response $\mathbf{y} = (y_1, \dots, y_N)$. Assume that the i -th sample has been genotyped, sequenced, or profiled for J variables (x_{i1}, \dots, x_{ij}) (e.g., gene expression, single nucleotide polymorphisms, proteomics). Collectively, all variables across all samples can be collected in an $N \times J$ matrix \mathbf{X} . In the MultiOviz interface, users can click the “Perturb” drop down menu and upload their data $\mathcal{D} = \{\mathbf{X}, \mathbf{y}\}$ along with a set of biological annotations encoded as a $J \times G$ binary mask matrix \mathbf{M} where G denotes the number of groups on the second molecular level. In this case, we have J SNPs that are grouped into G genes. To run BANNs (or a similar variable selection approach) and generate a GRN, the user should click “Run” once their data are uploaded. Once the user has a clear understanding of the GRN, in silico hypothesis testing can be carried out by clicking on a variable of interest, selecting “Edit” to delete the node and its connected edges, and then clicking “Rerun” to rerun BANNs and generate new sets of interactive GRN. Overall, this human-in-the-loop process facilitates the efficient testing of any number of hypotheses.

Flexible integration of statistical and machine learning methods

Part of the contribution of MultiOviz is that it is also available as a standalone R package. This enables users with coding experience to have more control over the statistical methodologies that run in the background of the software. This is important when there are unique theoretical considerations that need to be made for different types of omic data before performing perturbation experiments. Regardless of the method used, we recommend that the model have the ability to variable selection or regularization to ensure that the resulting GRN is reasonably sized (i.e., reducing an initial high-dimensional set of variables to a small number worthy of follow-up). Implementing MultiOviz within a developer script only requires two inputs once the package is installed: (i) a file of molecular variables and their associated scores, and (ii) a set of biological annotations.

Results

Demonstration of `Multioviz` on real data

To demonstrate the utility of `Multioviz`, we apply the software to real genetic data from a heterogeneous stock of mice collected by the Wellcome Trust Centre of Human Genetics (<http://mtweb.cs.ucl.ac.uk/mus/www/mouse/index.shtml>) [16]. The genotypes from this study were downloaded directly using the `BGLR` R package [20]. This study contains $N = 1,814$ heterogeneous stock of mice from 85 families (all descending from eight inbred progenitor strains) and 131 quantitative traits that are classified into 6 broad categories including behavior, diabetes, asthma, immunology, haematology, and biochemistry. Phenotypic measurements for these mice can be found freely available online to download (details can be found at <http://mtweb.cs.ucl.ac.uk/mus/www/mouse/HS/index.shtml>). In this study, we focus on modeling the percentage of CD8+ cells in these mice as our \mathbf{y} vector. For preprocessing, we corrected this trait for sex, age, body weight, season, and year [16]. The \mathbf{X} matrix that we input into `Multioviz` contains single nucleotide polymorphisms (SNPs) as variable, each of which are encoded as $\{0, 1, 2\}$ copies of a reference allele at each locus. For mice with missing genotypes, we imputed values by the mean genotype of that SNP in their corresponding family. Only polymorphic SNPs with minor allele frequency above 5% were kept for the analyses. This left a total of $J = 10,227$ SNPs that were available for all mice. Lastly, to create biological annotation file \mathbf{M} , we used the Mouse Genome Informatics database (<http://www.informatics.jax.org>) [21] to map SNPs to the closest neighboring gene(s). Unannotated SNPs located within the same genomic region were labeled as being within the “intergenic region” between two genes. Altogether, a total of $G = 2,616$ annotations were analyzed.

We input these files into `Multioviz` where we assumed that significant SNPs and genes would produce PIPs greater than or equal to 0.5—this is also known as the median probability model threshold in Bayesian statistics [19]. When viewing the corresponding GRN produced by the software, this resulted in 15 associated SNPs variables and 19 enriched genes (Fig. 4). Notably, we observed the SNP CEL-17_31069801 and gene *h1b156* on chromosome 17 as both being significant (PIPs = 1). As corroborating evidence, the genomic region where these molecular variables reside has been reported to contain highly significant SNPs that contribute to non-additive variation for CD8+ T-cells [16]. To investigate this region further, we perturbed the GRN in `Multioviz` by deleting CEL-17_31069801 and observed the emergence of CEL-17_31214920 as being important which also maps to the *h1b156* gene (PIP = 1). Two new gene-level variables that also became enriched upon perturbation and are both associated with CD8+ T-cell differentiation are *Anapc1* (PIP = 0.726) and *Pard3* (PIP = 0.998). *Anapc1* functions in the metaphase-to-anaphase transition in the cell cycle and has been associated with poor prognosis in T-cell acute lymphoblastic leukemia [22]. *Pard3* directs polarized cell growth and asymmetric cell division [23]. The asymmetric division of T-cells has been uncovered as a potential means by which effector and memory T cells are differentiated during immune responses [24]. Overall, we show here that `Multioviz` has the potential to enable users to generate new testable hypotheses in silico through its perturbation framework. These results suggest a honed set of molecular variables to explore in investigating mechanisms underlying the percentage of CD8+ T cells in heterogeneous mice.

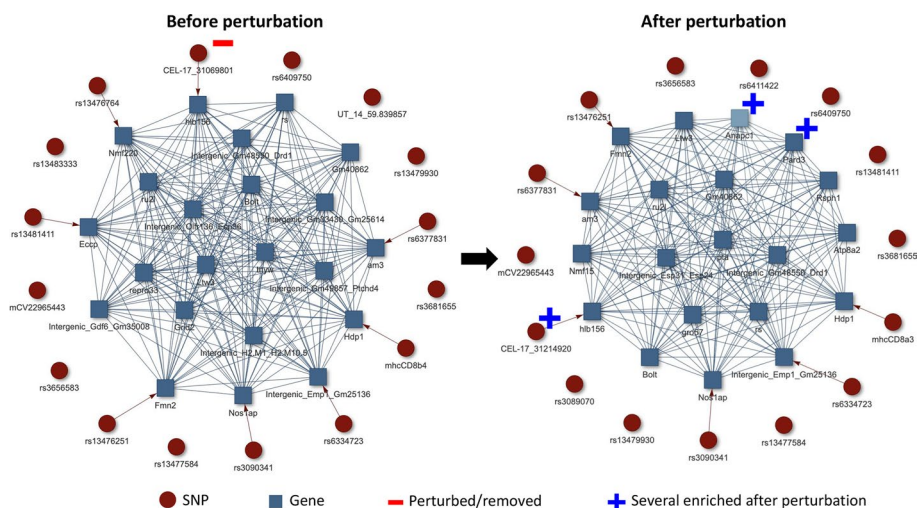


Fig. 4 Demonstration of an in silico analysis using Multioviz on a heterogeneous stock of mice dataset from the Wellcome Trust Centre of Human Genetics. We applied Multioviz to visualize a GRN with associated SNPs and enriched genes driving the architecture of CD8+ cell percentage. To generate this GRN, we set the following parameters in the Multioviz software: (i) Molecular Level 1 (ML1) Map Type = None; (ii) Molecular Level 2 (ML2) Map Type = Complete; (iii) ML1 Threshold = 0.5 which follows the median probability model [19]; (iv) ML2 Threshold = 0.5; and (v) GRN Layout = "layout_with_kk". In this figure, SNP-level variables in ML1 (red circles) map to gene-level variables in ML2 (blue squares). Upon deleting the CEL-17_31069801 SNP and rerunning Multioviz, we observe a new association with the SNP CEL-17_31214920 and a new enrichment of several genes, including *Anapc1* and *Pard3*. These are depicted in the perturbed GRN on the right. Note that both the deleted and newly enriched SNPs map to the *hlb156* gene

Comparing platforms during a perturbation analysis

To comprehensively assess Multioviz's performance during an in silico perturbation analysis, we compared Multioviz with two comparable platforms, OpenXGR [9] and visSE.cloud [10] (Table 1), which leverage Gene Set Enrichment Analysis (GSEA) [27] to visualize significant SNPs and genes that belong to subnetworks and enriched pathways (Table 1). For this platform comparison, we again utilized the heterogeneous stock of mice dataset from the Wellcome Trust Centre of Human Genetics [21]. While all platforms similarly aim to infer GRNs from high-dimensional multi-omics data, there are several differences, predominantly in data preprocessing, ease-of-use, and interface functionality.

Similar to Multioviz, both OpenXGR and visSE.cloud take precomputed statistics for the molecular level (e.g., genes) of interest. However, neither OpenXGR and visSE.cloud accommodate information about intergenic regions and, despite their potential significant regulatory influence, statistics corresponding to these features must be removed before these platform analyses can proceed [28]. While OpenXGR and visSE.cloud provide statistical confidence scores for molecular variables, these scores are presented in the form of lists and graphs (rather than being integrated into the output networks), which makes interactive variable selection less user-friendly. In the context of OpenXGR, only the gene table with functional descriptions and statistical significance scores is interactive, not the GRN itself. This can be limiting in settings where the goal for users is to interpret the GRNs. For visSE.cloud, various visuals for multi-scale analyses including GRNs and gene set enrichment exist. However, functional

connection between these scales is unclear, making it challenging to identify gene set affiliations and discern genes within specific subnetworks. Perhaps the most noticeable limitation in functionality for the `OpenXGR` and `visseE.cloud` platforms is that they lack integration support for a wide range of statistical models, a key component that is available for method developers to integrate in the `MultiOviz` R package. Further, while `MultiOviz` enables GRN generation that incorporates both genes and SNPs simultaneously, mirroring biological networks, `OpenXGR` and `visseE.cloud` only perform single molecular level GRN construction (e.g., creating only a SNP GRN or gene GRN, but not both). Currently, `OpenXGR` is restricted to human genomes, with plans to include compatibility with mouse data in future iterations of the platform [9]. Lastly, neither `OpenXGR` and `visseE.cloud` support direct *in silico* perturbation analysis.

Given the platform restrictions for both `OpenXGR` [9] and `visseE.cloud` [10], we needed to implement a few additional human-in-the-loop steps to their workflows in order to compare their performance with `MultiOviz`. The GSEA implementation in `visseE.cloud` requires that all G genes be input as a ranked list (in ascending order), a corresponding z -score statistic, or as p -values (P_1, \dots, P_g). `OpenXGR`, on the other hand, only accepts p -values as input. Thus, in order to re-analyze the same percentage of CD8+ cells phenotype, we used the `minSNP` procedure [25, 26]. Here, we ran a univariate linear model for each SNP individually, and attributed a p -value for each gene by using the SNP with the lowest p -value in that gene's region. This produced a list of genes with p -values from which we could determine a set of statistically significant genes. To ensure compatibility between platforms, we then filtered out any genomic features labeled as "intergenic" regions. Next, where applicable, we leveraged the Mouse Genome Informatics (MGI) database [29] to convert the mouse gene names to their corresponding human gene names to ensure compatibility with `OpenXGR`. With these paired gene and statistical inputs prepared, we were able to proceed with running perturbation analyses for all three platforms.

To implement a perturbation analysis in `OpenXGR` and `visseE.cloud`, we carry out the following steps. First, we run each platform with the paired gene and statistical measure inputs derived from the full mouse dataset. Second, we manually remove a statistically significant feature. Then, third, we rerun each platform without the significant feature to imitate an *in silico* knock-out (Fig. 5). It is worth noting that, for this particular dataset, neither of the competing platforms were able to generate GRNs using their default settings. Consequently, we had to manually adjust each of their hyper-parameters to investigate relationships between genes and pathways connect to CD8+ cell percentage. In `OpenXGR`, this meant setting the "functional interaction" to the lowest value of "medium confidence"; while, for `visseE.cloud`, we needed to fix the overlap threshold for gene set similarity measurement to the minimum value of 0.1. Notably, `MultiOviz` simplifies the process of identifying the degree of meaningful functional interactions by incorporating a toggle directly into its platform interface. In Fig. 5, we display `MultiOviz` with a selected edge threshold of 0.1 to ensure fair comparisons with the other platforms.

Each of the platforms we compare generates a slightly different type of visual GRN. `MultiOviz` provides an interactive GRN of SNPs and genes with their associated significance scores that enables users to explore and interact with the network dynamically

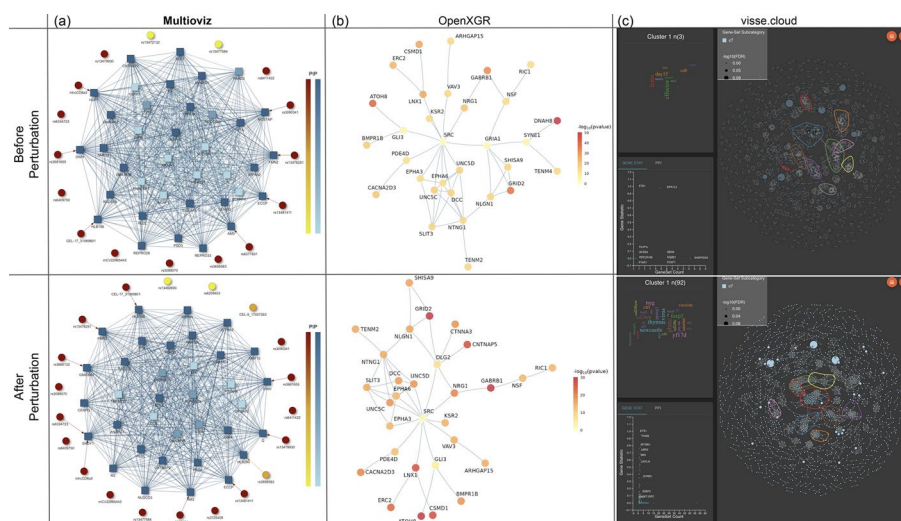


Fig. 5 Comparison of gene regulatory network outputs from Multioviz, OpenXGR, and visse.cloud during a perturbation analysis. Leveraging the same CD8+ cell percentage in the mice dataset from the Wellcome Trust Centre of Human Genetics (similar to Fig. 4) [21], we set out to compare these platforms. To ensure compatibility between platforms, we first preprocess data inputs by removing intergenic regions, and converting from mice gene names to human gene nomenclature where applicable. Then we proceed with perturbation analysis. **a** Perturbation analysis using Multioviz. The top panel shows an inferred GRN using Multioviz. The bottom perturbed GRN is generated by removing the most significant gene, *FMN2* and clicking “Rerun” on the platform. The total runtime for this analysis was approximately 10 min. **b** Similar perturbation analysis using OpenXGR [9]. The platform’s “Subnetwork Analyzer for Genes” (SAG) requires a list of genes and associated *p*-value statistics. To achieve this, we ran a series of univariate linear regressions for each SNP and determined the list of significant genes using the minSNP approach [25, 26]. This step was then followed by removing intergenic regions and converting from mice gene names to human gene nomenclature where possible. In OpenXGR, nodes represent genes in the inferred GRN, with darker colors indicating more significant genes. OpenXGR lacks in silico perturbation functionality. Thus to replicate the Multioviz pipeline, we manually remove the most significant gene *DNAH8* ($P = 3.25 \times 10^{-60}$) from the dataset. We then rerun the OpenXGR pipeline to obtain the new GRN. Runtime for this analysis was approximately 30 min. **c** Similar perturbation analysis using visse.cloud [10]. The visse.cloud platform uses “Gene Set Enrichment Analysis” (GSEA) [27] which requires a list of genes and their paired summary statistics. To achieve this, we again use the minSNP approach, removed intergenic regions, and where applicable, converted from mice gene names to human gene nomenclature. Given that the gene set network that visse.cloud outputs does not directly show which genes are most significant, we again perturbed *DNAH8* as we did in OpenXGR, resulting in the shown perturbed gene set network. The total runtime was approximately 40 min

(Fig. 5a). Conversely, OpenXGR outputs a static image of a gene level GRN, with a scrollable table of gene names and associated statistical measures below it (Fig. 5b). While the visse.cloud interface offers a wider range of genomic analyses, it does not directly link how the gene-level statistics correspond to gene set enrichment results. Instead, clusters of gene names and their associated statistics are displayed separately in a “Gene Stat” plot (Fig. 5c), while networks of connected pathways from GSEA are displayed in a different panel. Performing in silico perturbations also results in variations of detailed images from all three platforms. Due to us needing to include additional human-in-the-loop steps to overcome a lack of perturbation functionality for OpenXGR and visse.cloud, the total needed to time to run our in silico analysis took approximately 30 min for OpenXGR and approximately 40 min for visse.cloud. This compared to only needing 10 min to run an entire workflow for Multioviz. Overall, these comparison highlights promise of Multioviz to accelerate key steps in in silico perturbation

workflows. The `Multioviz` platform interface requires less data preprocessing for inputs, more flexible functionality for real time investigation, and requires less end-to-end runtime for analysis.

Discussion

`Multioviz` is an interactive platform for *in silico* hypothesis testing with GRNs. Both the web platform and the `R` package allow users to easily explore interactions between variables in omics datasets through clear visualizations and by enabling them to perform perturbation analyses. It is well known how valuable it can be to perform *in silico* knock-out or knock-down experiments to determine the best next actionable steps, prior to performing follow-up *in-vivo* and *in-vitro* experiments [30]. The `Multioviz` platform is in service of this goal.

Our real data results with the heterogeneous stock of mice dataset from the Wellcome Trust Centre for Human Genetics [16] serve as an illustrative example of how `Multioviz` can be used to identify a small set of candidate molecular variables that could be implicated in CD8+ T-cell differentiation. Our platform comparison results highlight the value of `Multioviz` by being a comprehensive *in silico* GRN perturbation platform, equipped with multi-level interactivity to ensure that practitioners can intuitively and efficiently test multiple hypotheses prior to spending energy and resources in the wet-lab. Further, our `R` package enables method developers to integrate other statistical models for biological experts to interact with subsequent GRN results. As part of future work, we want to extend `Multioviz` to integrate a wider array of mathematical and machine learning models into the web platform, as well as allow for integrating more than just two molecular levels for analysis.

Overall, we envision platforms like `Multioviz` being used for applications such as early drug development where the goal is often not only to identify potential druggable targets for disease pathways but also test the effects of drugs *in silico* prior to moving them into a biological system. Further, platforms including `Multioviz` could serve as a powerful means through which clinicians can generate and explore GRNs on a patient level and, as such, prescribe treatments and dosages tailored to each patient. The `Multioviz` platform is freely available, thereby providing researchers with an accessible way to analyze punitive molecular mechanisms underlying various traits across a wide array of biological levels.

Acknowledgements

This research was conducted using computational resources and services at the Center for Computation and Visualization (CCV) at Brown University. We would like to thank Brown's Computational Biology Core for hosting `Multioviz` online and making it free for users. Components of Fig. 1 were created with www.BioRender.com.

Author Contributions

HX, LC, and AMC conceived the study. LC and AMC supervised the project and provided resources. HX and AMC developed the software and performed the analyses. All authors wrote and revised the manuscript.

Funding

This research was supported by a David & Lucile Packard Fellowship for Science and Engineering awarded to L. Crawford. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of any of the funders.

Availability of data and materials

Project name: `Multioviz`. Project home page: <https://multioviz.ccv.brown.edu/> (user-facing website); <https://github.com/lcrawlab/multio-viz> (code); `docker pull ashleymaeconard/multioviz` (DockerHub image). Operating system(s): Platform independent. Programming language: `R`. Other requirements: To launch the `Multioviz` server locally, `R` version 4.1.2 or higher is required; DockerHub image is available; using `Multioviz` on the website has no

requirements except a web browser. License: GNU General Public License. Any restrictions to use by non-academics: No restrictions. `MultiOviz` is currently available as both a free web application and an R package. The web platform is hosted through the Center for Computation and Visualization at Brown University and can be accessed at <https://multioviz.ccv.brown.edu/>. The R package and other source code are also freely available at <https://github.com/crawlab/multioviz>. This GitHub repository also provides additional robust details on the platform and extensive tutorials on how to run analyses in the README. The heterogeneous stock of mice dataset from the Wellcome Trust Centre for Human Genetics can be found at <http://mtweb.cs.ucl.ac.uk/mus/www/mouse/index.shtml>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interest

The authors declare no competing interests.

Received: 11 October 2023 Accepted: 23 May 2024

Published online: 30 July 2024

References

1. Syvänen AC. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet.* 2001;2(12):930–42.
2. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004;5(2):101–13.
3. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet.* 2003;34(2):166–76.
4. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol.* 2008;9(10):770–80.
5. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17(6):333–51.
6. Bansal M, Belcastro V, Ambesi-Impiombato A, Di Bernardo D. How to infer gene networks from expression profiles. *Mol Syst Biol.* 2007;3(1):78.
7. Demetci P, Cheng W, Darnell G, Zhou X, Ramachandran S, Crawford L. Multi-scale inference of genetic trait architecture using biologically annotated neural networks. *PLoS Genet.* 2021;17(8): e1009754.
8. Yanai I, Chmielnicki E. Computational biologists: moving to the driver's seat. *Genome Biol.* 2017;18(1):223. <https://doi.org/10.1186/s13059-017-1357-1>.
9. Bao C, Wang S, Jiang L, Fang Z, Zou K, Lin J, et al. OpenXGR: a web-server update for genomic summary data interpretation. *Nucleic Acids Res.* 2023;51:W387.
10. Mohamed A, Bhuvana DD, Lee S, Liu N, Tan CW, Davis MJ. *vissE*: cloud: a webserver to visualise higher order molecular phenotypes from enrichment analysis. *Nucleic Acids Research.* 2023;51:W593.
11. Jang Y, Yu N, Seo J, Kim S, Lee S. MONGKIE: an integrated tool for network analysis and visualization for multi-omics data. *Biol Direct.* 2016;11(1):1–9.
12. Zoppi J, Guillaume JF, Neunlist M, Chaffron S. MiBiOmics: an interactive web application for multi-omics data exploration and integration. *BMC Bioinform.* 2021;22(1):1–14.
13. Zhang M, Li Q, Yu D, Yao B, Guo W, Xie Y, et al. GeNeCK: a web server for gene network construction and visualization. *BMC Bioinform.* 2019;20(1):1–7.
14. Osorio D, Zhong Y, Li G, Xu Q, Yang Y, Tian Y, et al. scTenifoldKnk: an efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. *Patterns.* 2022;3(3): 100434.
15. Garg A, Xenarios I, Mendoza L, DeMicheli G. An efficient method for dynamic analysis of gene regulatory networks and in silico gene perturbation experiments. In: Annual international conference on research in computational molecular biology. Springer; 2007:62–76.
16. Valdar W, Solberg LC, Gauguier D, Burnett S, Klenerman P, Cookson WO, et al. Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet.* 2006;38(8):879–87.
17. Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. *J Am Stat Assoc.* 1988;83(404):1023–32.
18. Brookes AJ. The essence of SNPs. *Gene.* 1999;234(2):177–86.
19. Barbieri MM, Berger JO. Optimal predictive model selection. *Ann Stat.* 2004;32(3):870–97. <https://doi.org/10.1214/00905360400000238>.
20. Perez P, de los Campos G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics.* 2014;198(2):483–95.
21. Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE. Mouse genome database. *Nucleic Acids Res.* 2019;47(D1):D801–6.
22. Fattizzo B, Rosa J, Giannotta JA, Baldini L, Fracchiolla NS. The physiopathology of T-cell acute lymphoblastic leukemia: focus on molecular aspects. *Front Oncol.* 2020;10:273.
23. Ludford-Menting MJ, Oliaro J, Sacirbegovic F, Cheah ETY, Pedersen N, Thomas SJ, et al. A network of PDZ-containing proteins regulates T cell polarity and morphology during migration and immunological synapse formation. *Immunity.* 2005;22(6):737–48.

24. Chang JT, Palanivel VR, Kinjyo I, Schambach F, Intlekofer AM, Banerjee A, et al. Asymmetric T lymphocyte division in the initiation of adaptive immune responses. *Science*. 2007;315(5819):1687–91.
25. Torkamani A, Topol EJ, Schork NJ. Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics*. 2008;92(5):265–72.
26. Hu Y, Deng L, Zhang J, Fang X, Mei P, Cao X, et al. A pooling genome-wide association study combining a pathway analysis for typical sporadic Parkinson's disease in the Han population of Chinese mainland. *Mol Neurobiol*. 2016;53:4302–18.
27. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102(43):15545–50.
28. Nelson CE, Hersh BM, Carroll SB. The regulatory content of intergenic DNA shapes genome architecture. *Genome Biol*. 2004;5:1–15.
29. (MGI) MGI. Mouse-Human Homology Relationships; 2024. Accessed: April 2024. Available from: http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt.
30. Moradi M, Golmohammadi R, Najafi A, Moghaddam MM, Fasihi-Ramandi M, Mirnejad R. A contemporary review on the important role of in silico approaches for managing different aspects of COVID-19 crisis. *Inf Med Unlock*. 2022;28: 100862.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.