High Resolution Modeling of COVID-19 Disease Biology

Renée Deehan-Kenney, PhD
Vice President, Computational Biology, QuartzBio

The COVID-19 pandemic has created an unprecedented urgency for identifying and developing therapies. In critical times such as these, computational biology is a powerful discipline that can be leveraged to analyze complex data and generate valuable insights which can help accelerate both disease understanding and therapeutic development. Here, we offer a proof of concept that demonstrates the real-world utility of leveraging computational biology approaches to streamline translational and clinical timelines by generating, testing and validating hypotheses using available molecular data. The results of this proof of concept indicate that the A549 cell line model can be used to screen drugs for mechanistic compatibility with COVID-19 disease biology. We also provide background on the application of computational approaches in the pursuit of novel treatments in any therapeutic area.

Using Computational Biology to Accelerate COVID-19 Therapeutic Development

Modern therapeutic development prioritizes identifying the intersection between a drug's mechanism of action (MoA) and a dysregulated pathway that causes or exacerbates a disease condition. In a typical drug discovery and translational medicine scenario, there is often a compound or compounds of interest—or at least a potential target—that researchers are attempting to modulate. Computational methods can accelerate the process of evaluating the compound's MoA in various indications and/or patient populations by leveraging proprietary or publicly available pre-clinical or clinical data. This permits rapid hypothesis testing as well as ongoing analysis enrichment when new data are generated or become available, adding layers of confidence to critical drug development decision-making.

Figure 1. Using Mechanistic Modeling to Identify Target Pathways in COVID-19

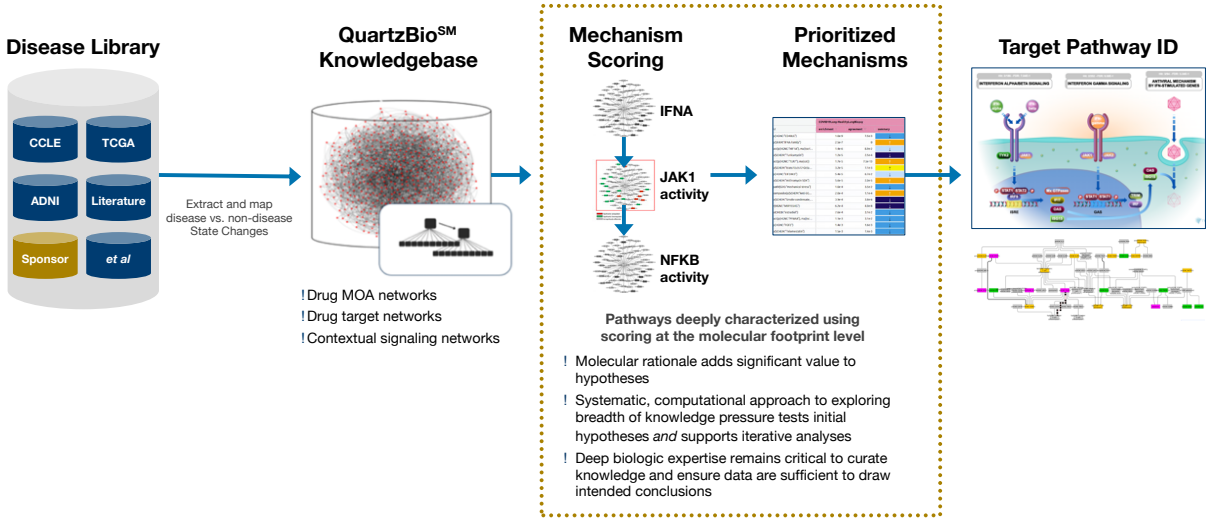

Illustration of target pathway courtesy of Reactome, a free, open-source, curated, and peer-reviewed pathway database.

*Current state as of June 2020

With COVID-19, the condition is already known, namely SARS-CoV-2 infection. In this case, the computational biology approach begins with characterizing the mechanisms that lead to severe disease by comparing samples from patients infected with COVID-19 and those without COVID-19. The differences between these samples represent dysregulated pathways which may also be drug target pathways. Analogous approaches can be taken with data captured from the drug of interest—or experiments that modulate the activity of a potential drug target—to compare the mechanisms dysregulated in disease to those that are influenced by drug target signaling.

Computational Biology in COVID-19: Proof of Concept

At QuartzBio, part of Precision for Medicine, we applied computational biology approaches to answering the following scientific questions:

1. Can we use existing pre-clinical and clinical data from COVID-19 or SARS-CoV-2 infection to investigate rational drug treatment pathways?
2. Can we identify particularly promising repurposing opportunities, without requiring significant upfront time and monetary investment in lab-based experimentation?

We approached this by first identifying and integrating COVID-19-relevant public data sets (see Table 1) as part of our QuartzBio team's COVID-19-Related Data Aggregation Initiative. Using these data, we proceeded with building a disease model, evaluating the viability and translatability of pre-clinical models, and exploring the potential of drugs to positively impact the target disease pathways.

- **Use Public Data Sets to Establish a Targeted Disease Model.** For this proof of concept, RNA-seq data was extracted from a publication by Cell (Blanco-Melo et al., 2020) for a statistical differential gene expression (DGE) analysis comparing four populations:
  1. Lung biopsies from two patients with COVID-19 compared to non-infected controls
  2. Normal human bronchial epithelial cells infected with SARS-CoV-2 compared to vehicle controls
  3. A549 lung carcinoma cells overexpressing ACE2 and infected with SARS-CoV-2 compared to vehicle controls
  4. SARS-CoV-2-infected A549 cells overexpressing ACE2 treated with JAK/STAT inhibitor ruloxitinib versus untreated
- **Map State Changes to a Knowledge Base.** After correcting for multiple testing, significant differentially expressed genes (DEGs) were selected for reverse causal inferencing (RCI). With RCI, measurable observations—in this case, the significant DEGs—are mapped onto representative nodes in our knowledge base. Statistical analysis of published and manually curated upstream controllers of DEGs enabled the systematic and rapid generation of inferences that describe potential modulators responsible for COVID-19 disease biology. For example, in this proof of concept, prior knowledge comprised those genes regulated by the interferon-alpha (IFNA) signaling family which were identified to be significantly enriched, suggesting that type I interferon signaling was activated in COVID-19 patients. As for IFNA signaling, inferences for over 2300 other mechanisms were made that represent mechanistic explanations for the data.

- **Perform Mechanistic QC.** QuartzBio's RCI Explorer enabled deep exploration of prioritized inferences through its visualization and analytic interface. The platform further enabled the team to perform RCI on other groups of interest—in this case cell line models—to determine their ability to recapitulate the human clinical infection. The output demonstrated that RCI is a sensible and rapid way to compare and contrast overlapping signaling mechanism. The output also allowed researchers to rapidly select appropriate preclinical models, thereby reducing the cost and time to clinic.
- **Identify Target Pathway.** We used the output of the RCI to identify and select target pathways and appropriate pre-clinical models that could be used to screen potential drugs for efficacy. In this case, we were able to demonstrate that the A549 cell line model reflected COVID-19 biology and could be used for screening drugs for mechanistic compatibility.

Table 1. Public Data Sets Utilized in COVID-19 Proof of Concept

Project Title	Accession / Link
Transcriptional response of human lung epithelial cells to SARS-CoV-2 infection	PRJNA615032
LY6E blocks coronavirus fusion and confers immune control of viral disease	PRJNA609134
Relative timing of type I interferon response and virus replication determines disease outcome during MERS-CoV infection	PRJNA545350
Transcriptomic analysis of MERS-CoV infected Calu-3 cell with or without AM580 treatment	PRJNA506733
Transcriptome profiling of influenza virus-infected human bronchial epithelial cells	PRJNA305099
Transcriptomic analysis of the Novel Middle East Respiratory Syndrome Coronavirus-MERS-CoV	PRJNA233943
Transcriptomic Analysis Of circRNAs-miRNAs-mRNAs upon Middle East Respiratory Syndrome Coronavirus -MERS-CoV- infection	PRJNA580021
SARS-CoV-Encoded Small RNAs Contribute to Infection-Associated Lung Pathology	PRJNA355238
E-MTAB-8871 - Transcriptomic analysis of immune response in healthy controls and COVID-19 cases using the NanoString Human Immunology Panel	E-MTAB-8871

Applying Computational Biology to Other Therapeutic Areas

There is no one-size-fits-all approach to modeling disease biology from high-throughput, multiomic data. Instead, selecting the right computational approach is dependent on the available data and the scientific questions being asked. With that said, computational biology modeling methods typically fall into two categories:

- Data-driven, where correlations in the data are explored in an unbiased way, such as unsupervised clustering
- Mechanistic, which relies on the use of prior knowledge to test a hypothesis, for example an analysis where particular biomarkers are selected based on their relevance to the disease or drug under study

Mechanistic modeling approaches can be extraordinarily valuable for both identifying and prioritizing target pathways and illuminating biological relationships around which evidence can be built. When starting with a disease in mind, the mechanistic modeling described in the proof of concept can be a powerful approach:

1. **Start with a Disease Library.** First, create a disease library containing mulitomics and clinical data from proprietary information or validated, curated public sources. Then, perform statistical analyses designed to identify significant state changes (e.g., relative changes in gene expression) between two groups of interest, for example those affected by the condition of interest and those not affected.
2. **Map the State Changes to a Knowledge Base.** Precision for Medicine's QuartzBio platform contains cause and effect relationships that have been curated from prior knowledge. Once the state changes have been mapped to the knowledge base, RCI analysis can be performed to predict what biological activity is likely to be modulated between the two groups of interest. This analysis will generate a prioritized list of mechanistic explanations for the data.
3. **Identify a Target Pathway.** The mechanistic output of RCI is then quality controlled and vetted by biological experts to identify significant drivers of disease pathogenesis.

It is worth noting that this same mechanistic approach can be taken with drug, rather than disease, data—as is more typical of a sponsor-led drug development program. In that scenario, the first step would be to build a high resolution MoA model and then evaluate the compound's MoA in various indications and/or patient populations by leveraging proprietary or publicly available pre-clinical or clinical data.

Conclusion

Recent advances in computational biology and translational informatics allow researchers to leverage data in the public domain to generate unique, deeply characterized biological insights to de-risk pre-clinical and clinical testing of potential drug candidates. Using these *in silico* approaches, we can expand our ability to test hypotheses and accelerate the identification of promising therapeutic targets. Moreover, we can confirm biological rationale throughout this process and advance targeting for any ongoing lab work.

In the context of a global pandemic, every day saved in getting a therapy approved and to patients can have significant impact on human life and the global economy. The urgent need to develop COVID-19 therapies has been met with broad-scale mobilization, as well as remarkable innovation in approaches to jumpstart SARS-CoV-2 drug development programs. We are cautiously optimistic that the breadth of engagement in such programs will support innovations that will streamline and accelerate drug development more broadly.

QuartzBio, part of Precision for Medicine, has been working on the leading edge of such innovation, including use of the computational biology approaches and workflows described in this article, as well as our other COVID-19 projects and services.

These innovations are quickly being extended to advance therapeutic development across other indications with rapid hypothesis generation and testing rooted in data-driven analyses and computationally developed mechanistic insights.