Bio
Pharma
Trend

BiopharmaTrend.com
A fresh viewpoint on drug discovery, pharma, and biotech

info@biopharmatrend.com

Reimagining Genomics Data Analysis: Interview with Paolo Di Tommaso from Seqera Labs

April 18, 2023 by Andrii Buvailo

With BioNTech’s recent $200 million merger with OncoC4’s cancer drug, the demand for personalised medicines to treat diseases such as cancer continues to grow. Much of this development stems from pharmaceutical biotech companies sharing their cutting-edge bioinformatics data on open-source platforms for others to use.

Here is our interview with Paolo Di Tommaso, CTO & Co-founder of Seqera Labs which has a global community of over 15,000 researchers, where we talk about this largely unreported trend. Its flagship product, Nextflow, facilitates this approach to ‘open science’ and is the de facto software of its kind, transforming the way the sector approaches sharing and analysing data. Seqera Labs works with some of the largest pharmaceutical, life sciences and genomics companies worldwide – including AstraZeneca, Janssen Pharmaceuticals, Oxford Nanopore.

Paolo explains how having access to the necessary infrastructure to analyse genomic datasets at scale is speeding up innovation at the earliest levels of scientific research and what this trend means in terms of the scientists behind them, working together.

Andrii: Can you tell us a little about yourself and your journey as a bioinformatician, which ultimately led to the founding of Seqera Labs?

Paolo: I have a master’s degree in computer science and have worked as a software engineer for several years in varying industry fields including telecommunications and banking, as well as non-profit organisations like the Agency for Food and Agriculture of the United Nations (FAO).

With the rise of the cloud and big data, the potential application into the life sciences sector was substantial and, whilst also wanting to make a real world impact in this industry, I decided to get a master's degree in bioinformatics in 2008.

In 2010, I began working as a research engineer at the Centre for Genomic Regulation, with the purpose of exploring methods to help streamline the use of cloud in bioinformatics benchmarks and data analysis. Whilst offering powerful technologies, the initial solutions were ultimately failures and no longer aligned with what I was researching.

Around 2013, I began focusing on new ideas for a different approach to the deployment of data pipelines at scale and, born out of my frustrations of previous experiences, I realized the way forward was to enable researchers to scale data analysis in a portable and reproducible manner across different computing infrastructure without breaking their typical workflow as a developer. Hence, Nextflow was created.

In the first few months, the technology was similar to other lab projects I was running at the time, however a community of users rapidly grew, providing feedback and asking for new features.

It was ultimately under these conditions that Seqera Labs was launched and is very much a success story owing to the popularity of open source and the open science approach. We have seen this first hand through our technology; when you share knowledge, you increase its value.

Andrii: What was the inspiration behind creating Seqera Labs, and how does the company's mission to simplify complex data analysis pipelines in the cloud contribute to the biotech industry?

Paolo: The inspiration behind creating Seqera Labs was born out of mine, and my fellow co-founder Evan Floden’s, realization that the Nextflow project was growing well beyond what a typical research lab could offer.

We had enquiries from users that could only be fulfilled by creating an organization, now Seqera Labs, taking over this project and maximizing it’s potential.

Alongside this, we realized that developing and sharing analysis pipelines – such as the digital tools used to diagnose cancer from a biopsy of a tumour - across multiple infrastructure was still a significant problem.

Often when researching, valuable time and effort is spent downloading such tools, as well as reels of research, and so we sought to find a more efficient solution, by bringing the workload to the data, not the other way around.

In the past, much of this research would occur on forums such as GitHub where scientists could suggest modifications to the way something was analyzed which shows how this desire for collaboration has always existed amongst bioinformaticians looking to share their insights on current healthcare challenges.

This became the driving reason for building Seqera Labs: to define a new approach to building and sharing data pipelines, making them accessible to the next generation of scientists globally. Now in its 10th year, Nextflow is the de facto platform used by a full spectrum of researchers from individuals to those working within sectors behemoths such as AstraZeneca on their own journeys into data collaboration and compliance at scale.

As the use of cloud computing continues to play a foundational role in bioinformatics research, there remains a need for cost-effective and scalable ways to store and process large volumes of sequencing data. Cloud-based data orchestration tools like the Nextflow Tower are gaining significant traction post-pandemic, especially amongst big pharma companies as well as smaller labs, because of their ability to offer generous cost reductions and easier collaboration. It is likely we will see how these relationships will continue to be leveraged through cloud platforms across the life sciences industry, and a proliferation of pipeline data grounded predominantly in medical research.

Andrii: With over 15,000 researchers in your global community, can you share any standout collaborations or projects that have emerged as a result of using Seqera Labs' technology for data analysis and sharing?

Paolo: Seqera Labs represents a paradigm shift for genomics analysis especially as it enables scientists to build and share medical analysis pipelines from a manageable central location run in the cloud. This makes the data analysis process more secure and efficient, and can be run at scale whilst empowering scientists to collaborate. As such, we’ve seen first-hand the platform being used at the forefront of discovery in spaces such as oncology, precision medicine, genomics and RNA sequencing.

The Covid-19 pandemic stands out as a great example of how our technology was pivotal for data analysis and sharing, being used to discover and track the Alpha, Delta and Omicron variants of the virus. In this way, scientists who were working on vaccines were able to chart the global spread of the disease by variant. Our contribution here was instrumental such vaccines, and highlights how important collaboration is for scientists reactively working on some of the most pressing global health issues.

Andrii: The recent BioNTech and OncoC4 merger highlights the increasing demand for personalized medicine. Can you discuss the role of Seqera Labs in shaping the future of personalized medicine through collaborative efforts in the industry?

Paolo: As demand for personalization of medicine grows, being able to build scalable and repeatable tools which point to potential cures based on an individual’s genomic data is a driving force of the current market. This type of research and treatment is hugely expensive, which is why saving both money and time by being able to access data shared by other scientists working in the same areas allow for more innovation to be made – and faster.

We’re starting to see this being used in healthcare systems across the world – one such project is the UK’s 100,000 Genomes which is running alongside the NHS to deliver and continually improve genomic testing to help doctors and clinicians diagnose, treat and prevent illness. Such investments emphasise the need for initiatives that focus on the delivery of live saving personalised therapies, and much of this relies on the environments in which these researchers can process pipeline data.

Seqera Labs provides the tools for researchers to diagnose these conditions but what it is also allowing scientists to do is to share their results. Once a big enough bank of these is built up, it can be analyzed for trends and data which can inform the way we treat - and prevent - these diseases.

Andrii: How has Seqera Labs' work with major pharmaceutical and life sciences companies like AstraZeneca, Janssen Pharmaceuticals, and Oxford Nanopore impacted the way they approach data sharing and analysis in their research?

Paolo: Through our own platform, we are seeing how collaboration between Big Pharma companies and small biotechs can speed up productivity and efficiency at the earliest levels of research. This is usually because small, more agile biotechs have far higher technological capabilities but lack the access to vast datasets Big Pharma has acquired. Being able to free up resources otherwise used to build large scale infrastructure is helping to accelerate the pace of discovery and bring new therapies and diagnostics to the market much faster.

Interestingly, we also see scientists from multinational Big Pharma companies sharing the ways in which they can diagnose certain illnesses and provide treatment, often contributing to wider data pools to provide research to those looking at therapeutics. Seqera Labs provides the vehicle for all of this to take place at a reduced cost to pharmaceutical and research organisations of all sizes which can reinvest the saved money into projects that will improve patient outcomes.

Andrii: In terms of infrastructure, what specific capabilities does Seqera Labs provide that enable researchers to analyze genomic datasets at scale and facilitate innovation in the earliest stages of scientific research?

Paolo: Seqera Labs has a range of capabilities for its users. Its flagship technology, Nextflow - which celebrated its 10-year anniversary early this year - is an opensource workflow orchestrator which simplifies writing and deploying data-intensive pipelines at scale on any infrastructure. It is currently being downloaded over 130,000 times every month.

Another component is open access to nf-core, an independent community which collects and curates Nextflow analysis pipelines. With more than 15,000 contributors, nf-core democratizes access to quality open source scientific data and offers over 120 high-quality, product-ready pipelines.

And finally, we have Tower, which is an intuitive, centralized command post that enables large-scale collaborative data analysis. Its users can quickly launch, manage, and monitor scalable Nextflow data analysis pipelines and compute environments on-premises or across the cloud provider of their choice.

Andrii: As the trend of scientists working together and sharing their data continues to grow, how do you see the landscape of biotech and genomics research evolving, and what role will Seqera Labs play in this transformation?

Paolo: As the trend of data sharing in genomics and biotech continues to grow, there is unmatched potential for our platform to help solve some of the biggest issues in healthcare. Our contribution to identifying various Covid variants during the pandemic was instrumental to helping to develop vaccines, owing to the importance of collaboration within these industries in speeding up how we react to such pressing issues. It’s likely that we will see more projects like the UK’s 100,000 Genomes rolled out in healthcare systems around the world - and that the scientists behind them will be working together.

Through Seqera Labs’ own platform, we are seeing how this increasing demand to build scalable and secure pipelines is freeing up resources otherwise used to build complicated infrastructure to helping to accelerate the pace of discovery. Our mission is to become a generationally defining company with de facto use for all scientific data projects, helping customers to boost productivity, reduce cost and complexity, and simplify regulatory compliance.