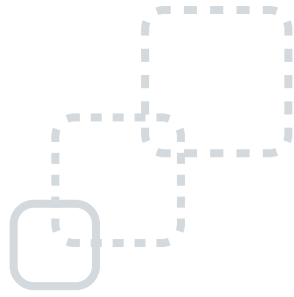


SevenBridges

# INDUSTRIAL-SCALE BIOINFORMATICS





# **INDUSTRIAL-SCALE BIOINFORMATICS**

## **DATA LOGISTICS AT PETABYTE SCALE**

**Logistical data challenges strain user experience and satisfaction. The Seven Bridges Platform can help you grow faster than 18,000 genomes a year.**

# **OFFER MEANINGFUL ANALYSIS, FAST GAINING INSIGHT FROM GENOMIC DATA IS A DEMANDING TASK. THE SEVEN BRIDGES PLATFORM SUPPORTS THE WORK OF BIOINFORMATICIANS.**

Data management infrastructure. no matter where it's stored 6  
Use and control your data with ease - be it in the cloud or on your cluster

Infinite computation and storage resources 6  
Flexible scaling

Turnkey world class compliance and security 7  
Encryption in transit and at rest, access controls, HIPAA, US and EU Safe Harbor compliant

Get started in just one day 7  
API for seamless integration

Use your own pipelines to analyze tens of thousands of samples 8  
SDK to support simple import of existing pipelines, a visual editor to help you develop new ones, and an API to scale analysis

Apply best practice pipelines to your data immediately 8  
More than 400 tools and 30 pipelines, ready-to-run

Interpret your variants 9  
Check your coverage quality with a heat map, then annotate, interpret, and filter your variants, and finally work on what remains with the collaborative genome browser

Case Study 10  
Cancer Genomics Research Laboratory at the National Cancer Institute

# DATA LOGISTICS AT PETABYTE SCALE

---

## **DATA MANAGEMENT INFRASTRUCTURE, NO MATTER WHERE IT'S STORED**

Use and control your data with ease be it in the cloud or on your cluster.

Between 2014 and 2018, production of new NGS data is estimated to exceed two exabytes. As sequencers become faster and less expensive to use, storage must get even smarter.

The Seven Bridges Platform simplifies your storage and data delivery needs completely or in conjunction with your existing infrastructure.

Store all your data with Seven Bridges for complete security and infinite scalability in our cloud-based data warehouse. Or, easily connect your storage systems to the Platform.

Either way, the Seven Bridges Platform will manage user permissions, metadata, run near-instant searches and link data to analyses as you run them.

## **INFINITE COMPUTATION AND STORAGE RESOURCES, AVAILABLE WHEN YOU NEED THEM**

Flexible Scaling

The Seven Bridges Platform is built for data management and analysis at scale.

Its flexibility allows you to instantly increase or decrease computational power and storage, as you need it. Even Illumina HiSeq X Ten users won't run into space limits or time restrictions for their data storage.

Seven Bridges has automated and optimized computation and storage resource allocation, allowing for efficiency and speed without the need for manual adjustment.

When you use the Seven Bridges Platform, data logistics are a solved problem.

## **TURNKEY WORLD CLASS COMPLIANCE AND SECURITY**

Encryption in transit and at rest,  
access controls, HIPAA, US and  
EU Safe Harbor compliant



See more at: [coalfire.com](https://coalfire.com)

## **GET STARTED IN JUST ONE DAY**

API for seamless integration

The Seven Bridges Platform is built with data security at its core. Data is encrypted at all times: during transfer, in storage, and during computation. Data volumes are encrypted using AES 256, and all file transfers and platform services communicate through encrypted SSL channels. With project level access controls, you control when a colleague on another continent sees your data - and when they don't.

As a result, Seven Bridges allows for compliance with HIPAA and both U.S. and E.U. Safe Harbor regulatory frameworks. Coalfire assessed the success of our HIPAA compliance.

When you use our storage, you also gain the service quality and reliability of Amazon Web Services (AWS). Not only can you do away with management of hard drives and worries over data scalability and integrity, but you also benefit from Amazon's management infrastructure. AWS manages its storage in alignment with the HIPAA, SOC, PCI DSS, and ISO 27001 standards. It also powers organizations such as NASA and the CIA.

The Seven Bridges Platform integrates easily with your project management system, LIMS, or your sequencer itself. Setting up your tools is faster than preparing your first library. Our API lets you build automated workflows to perform Quality Control checks, run analysis tools, and simplify the upload and metadata capture process. With the API, you can choose to perform analyses for your users and provide them with the results directly, or give them the freedom to automate their own analyses.

The Seven Bridges Platform also reduces data transfer logistics and time, since your data is hosted alongside programmable computational resources. You can invite new users directly into a project that already holds data, and use fine grained controls to limit their access to your other experimental data, and ability to execute new pipeline executions.

## OFFER MEANINGFUL ANALYSIS, FAST

---

### **USE YOUR OWN PIPELINES TO ANALYZE TENS OF THOUSANDS OF SAMPLES**

SDK to support simple import of existing pipelines and tools that you or colleagues have built, a visual editor to help you develop new ones, and an API to scale analysis

Our Docker based Software Development Kit (SDK) lets you easily wrap your preferred tools. This means that you can retain all capabilities of the command line, with the power of our Platform behind you.

You can put tools that you have developed yourself onto the Seven Bridges Platform, for on-demand access to our high computational power and large data repositories. The SDK lets you wrap your tools so that they can be executed on the Platform, with no need to reconfigure your existing command line applications to meet any proprietary format. And the tools are portable to a diverse range of infrastructures, should you want to run them on different platforms. Mix and match tools to build new pipelines, test them with all your data, and keep strong version control as you optimize the pipeline.

### **APPLY BEST-PRACTICE PIPELINES TO YOUR DATA IMMEDIATELY**

More than 400 tools and 30 pipelines, ready-to-run

The Seven Bridges Platform is designed to help you do your most common tasks in parallel for high throughput sequencing. Frequently used open source analysis pipelines such as BWA, GATK, and more are set up according to best practices. Our bioinformaticians optimize these pipelines for resource efficiency and speed; Intel asked for Seven Bridges Genomics' deep optimization expertise to help prepare key bioinformatics pipelines on its new Genomics Cluster solution.

Seven Bridges maintains and updates pipelines and tools so you can confidently access new versions. We also carefully version everything on the Platform, so that users know exactly what tool and version they used for every analysis. This gives you both the latest tools and perfect reproducibility at once.

When you want to customize one of the built in pipelines or tools, our visual editor gives you quick access to every parameter.

## INTERPRET YOUR VARIANTS

Check your coverage quality with a heat map, then annotate, interpret, and filter your variants, and finally work on what remains with the collaborative genome browser

As the number of samples you process per year crosses into tens of thousands, easily completing your analysis workflow is ever more important.

On the Seven Bridges Platform, you can quickly check coverage, annotate, interpret, and filter your data, all with simple to use interfaces. Then, you can use the fast collaborative genome browser to pool the problem solving knowledge of your entire team.

The screenshot shows a variant filtering interface. On the left, there are several filter sections: 'Amplicons that failed QC: 25' with a progress bar at 154/181; 'Keep SNVs that fit ALL of the following filters' with sub-filters for 'Quality by depth' (≥ 2), 'MQ rank sum' (≥ -12.5), and 'Mapping quality' (≥ 40); 'Keep Indels that fit ANY of the following filters' with 'Quality by depth' (≥ 2); and 'Keep all variants that fit ALL of the following filters' with 'Exclude variants in primer r...' set to 'True'. On the right, a 'Variant Table' is displayed with columns: Chr, Position, REF, ALT, DbsNP, COSMIC, QUAL, Gene\_Name, and Effect. The table shows 18 variants across chromosomes 2, 4, 5, 7, 9, 11, 13, and 18, with associated gene names like SMAD4, GNAQ, ATM, FIP1L1 PDGFRA, KDR RP11-53017.1, ERBB4, FLT3, and NPM1.

Simple Interface for Filtering Variants

The screenshot shows a collaborative genome browser interface for a region on Human g1k v37 (chromosome 12, coordinates 6,646,330-6,647,788). The interface includes several tracks: 'Bam Track - ERR009154\_1.fastq\_accepted\_hits.bam' showing read alignments; 'Gene track - UCSC\_hg19' showing gene models; 'Vcf Track - dbSNP137' showing variant annotations with their positions and reference/alternate alleles; and 'Reference Track - Human g1k v37' showing the reference genome sequence. The interface also features navigation controls like 'Update', 'Show Overview', and 'Collapse All Tracks'.

Collaborative Genome Browser

# CASE STUDY: CANCER GENOMICS RESEARCH LABORATORY AT THE NATIONAL CANCER INSTITUTE

---

## **AUTOMATED AND STREAMLINED PROCESSES SAVE MORE TIME WITH HIGHER THROUGHPUT**

The Cancer Genomics Research (CGR) Laboratory at the National Cancer Institute (NCI) uses the Seven Bridges Platform to streamline its data and analysis processes. As a result of this collaboration, the lab is able to process more samples simultaneously, and do its analytical work more quickly.

The CGR engages in sequencing and bioinformatics analysis for projects internal to the NCI and for external labs. In this capacity, they handle a broad spectrum of design, sequencing, and analysis project types. Instead of relying on a self built and supported system, they turned to the Seven Bridges Platform.

First the CGR utilizes a Seven Bridges plugin to automatically stream data from the sequencer directly to the cloud. By automating this routine task, Seven Bridges saved significant time for David Roberson, Senior Scientist at CGR:

**“The fact that the plugin will auto run means that the data will be uploaded before we even get to work the next day.”**

Whenever the CGR increases their sequencing throughput, the time savings are multiplied.

Due to the large number of projects and total data they handle, the CGR had a pressing need to accelerate virtually all aspects of their bioinformatics process. Seven Bridges Genomics worked with the CGR to accommodate this need.

The Seven Bridges bioinformatics team created an application that allows the CGR to visualize and perform statistics on their data in the language of their choice. This gives them an immediate view into their data, allowing them to interrogate across entire projects and make adjustments as they do additional sequencing.

We also worked with the CGR to develop a novel tool for visualization of coverage.



The CGR needed to quickly discover if any samples within a particular project needed re-sequencing or a different bench approach. We designed a dynamic utility to produce coverage analysis images so that they could quickly identify which samples didn't meet the coverage thresholds.

The CGR faces the particular challenge of running analyses for projects that span the spectrum of genomics applications. The lead CGR bioinformatician relies on the Seven Bridges bioinformatics and support teams for troubleshooting, development of new tools, and training on these. Through regular checkups and constant communication our team makes sure that the CGR is optimally equipped to complete their work quickly and efficiently. This personal connection has led to the development of new tools to increase the efficiency of the CGR.

Our work with the NCI influenced the many labs that converge on the CGR core facility. As the number of projects and volume of sequencing data increases, the Seven Bridges Platform helps the CGR team keep ahead of their users' storage and analytical needs.



**TEAM@SEVENBRIDGES.COM**  
**SEVENBRIDGES.COM**

**SevenBridges**