The BaseJumper™ research platform for single cell multi-omic analysis & interactive visualization

V. Weigman[1], V. Amin[1], C. Culler[1], I. Salas González[1], L. Coons[1], T. Tate[1], T. Laws[1], J.A.A. West[1]

[1]BioSkryb Genomics, Inc., Durham, NC, USA

BioSkryb GENOMICS · **ASHG** American Society of Human Genetics

Abstract

Translating the molecular alterations, especially those from a single cell, can be a powerful tool to understand underlying mechanisms of disease and somatic mosaicism. Along with managing the vast data volume from sequencing technologies, the complexity of these changes requires a multi-dimensional approach to first compute and identify the variance found within the cells of a study and then display these effects concomitantly. The platform was designed for researchers across computational skill levels, placing more energy on the ability ask direct questions of across the multiple forms within their dataset.

We present a secure, cloud-based bioinformatics and visualization platform that can be run from any laboratory computer on a standard web browser. Samples can be pulled from across repositories: local and cloud, to create analysis projects that can be wholly contained within an institution or shared across collaborators for distributed analysis considerations. Available pipelines provided within the platform enable simple (SNV/insertion/deletion) and complex (structural/copy number/highly polymorphic) variant detection in addition to cell identification, cell state and pathway status from gene expression. Pipelines have been benchmarked across single cells of NIST standards like NA12878 achieving a ~94% accuracy and ~99% precision for variant identification. Context for variation across cells is provided through annotation to several known repositories for genetic variation (gnomAD, ExAC, 1000 genomics) and clinical relevance (COSMIC, ClinVar). These can be directly queried for prevalence or in addition to other filtering strategies, where schema can be saved and toggled within and across research groups. This filtering across the whole genome dataset (hundreds of thousands to tens of millions of variants) occurs in just a few seconds.

Leveraging cellular or sample phenotypes, different methods can be directly applied to datasets to run association studies, and important genotypes presented, without the need for pulling extra compute.

Native to the platform is the ability to toggle these associative markers across transcript and genome-level of paired datasets. Output from one analysis can be used to guide visualization or narrow focus of another data type. Visualization applications can take datasets across genomic and transcriptomic results and can be manipulated based on expert interpretation of the researcher to maximize biological leverage. Compiling all of this within a browser maximizes accessibility of data so you can perform analyses anywhere inspiration takes you.
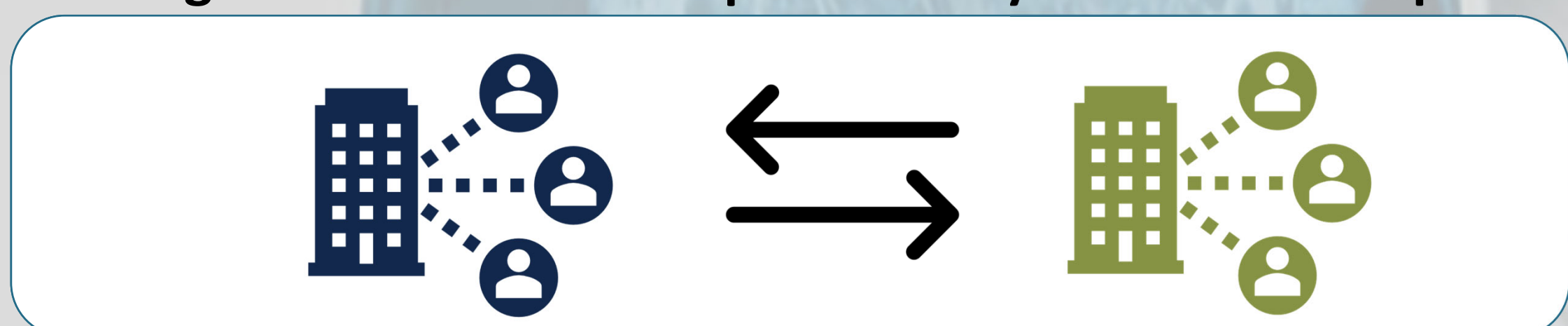
Access

BaseJumper protects your data using security best-practices in conjunction with technology integrated into our cloud (AWS) platform to provide isolation and safety for your genomic data.
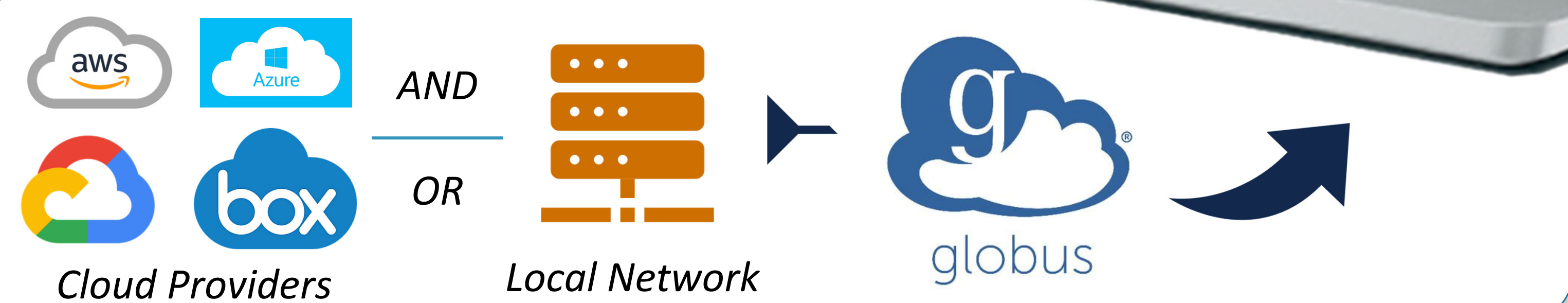
Log in from any device.

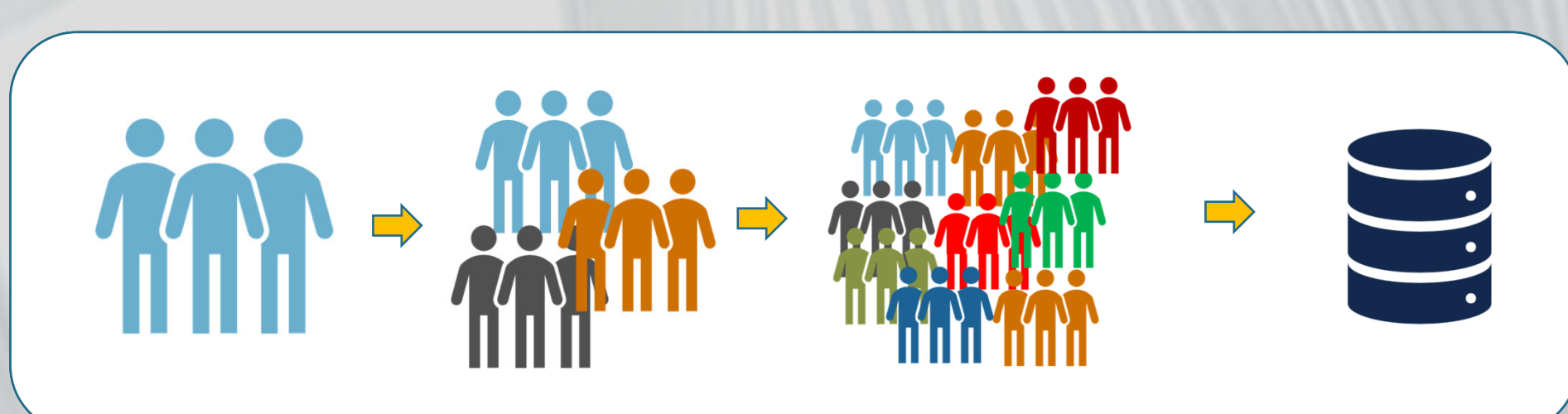**Access at Organization / Group / Individual Levels**

Group / Lab B — Researcher 1, Researcher 2, Researcher 3

Organization

Group / Lab A
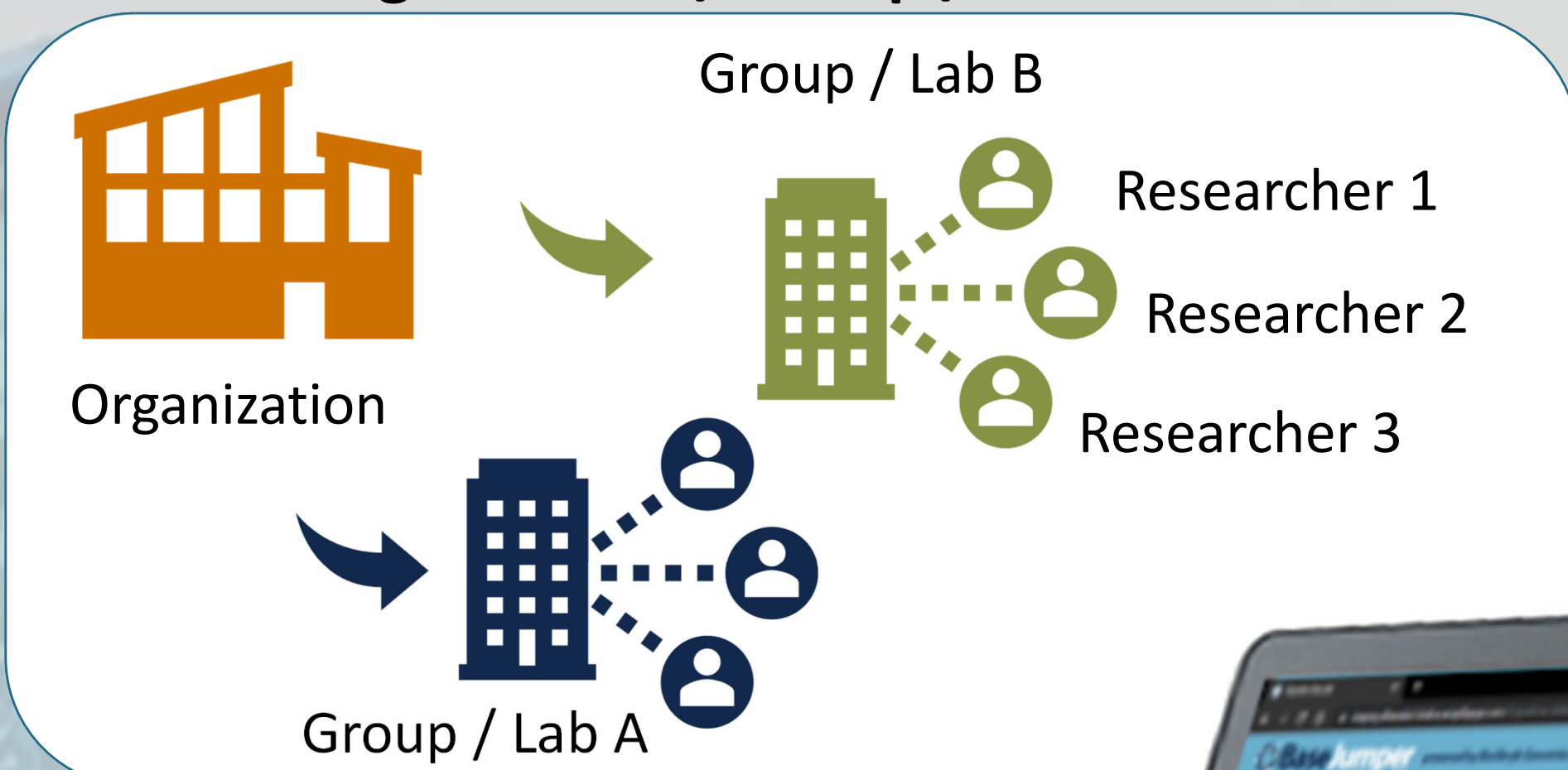
Sharing can be enabled for specific analyses across Groups

Build

Cloud Providers (aws, Azure, Google Cloud, box) **AND / OR** Local Network ▸ globus

Import sequencing data from any location through Globus

BaseSpace SEQUENCE HUB

Sync Illumina BaseSpace to your BaseJumper Organization

Grow Projects over time. Leverage one project space to incorporate latest results from the same study over time. Analysis version control ensures consistent processing and interpretation

Compute

Pipelines embedded on platform

DNA
- Whole Genome Alignment & Variants
- Whole Exome Alignment & Variants
- Copy Number Variation
- Structural Variation
- Repetitive Element Detection
- Tumor Mutational Burden
- Extensive Annotation (COSMIC, ClinVar, gnomAD, etc.)

RNA
- Gene and Isoform-level Expression
- Support for full-length and End-tagged workflows
- Cellular ID
- Gene Fusion Detection
- Variant Calling
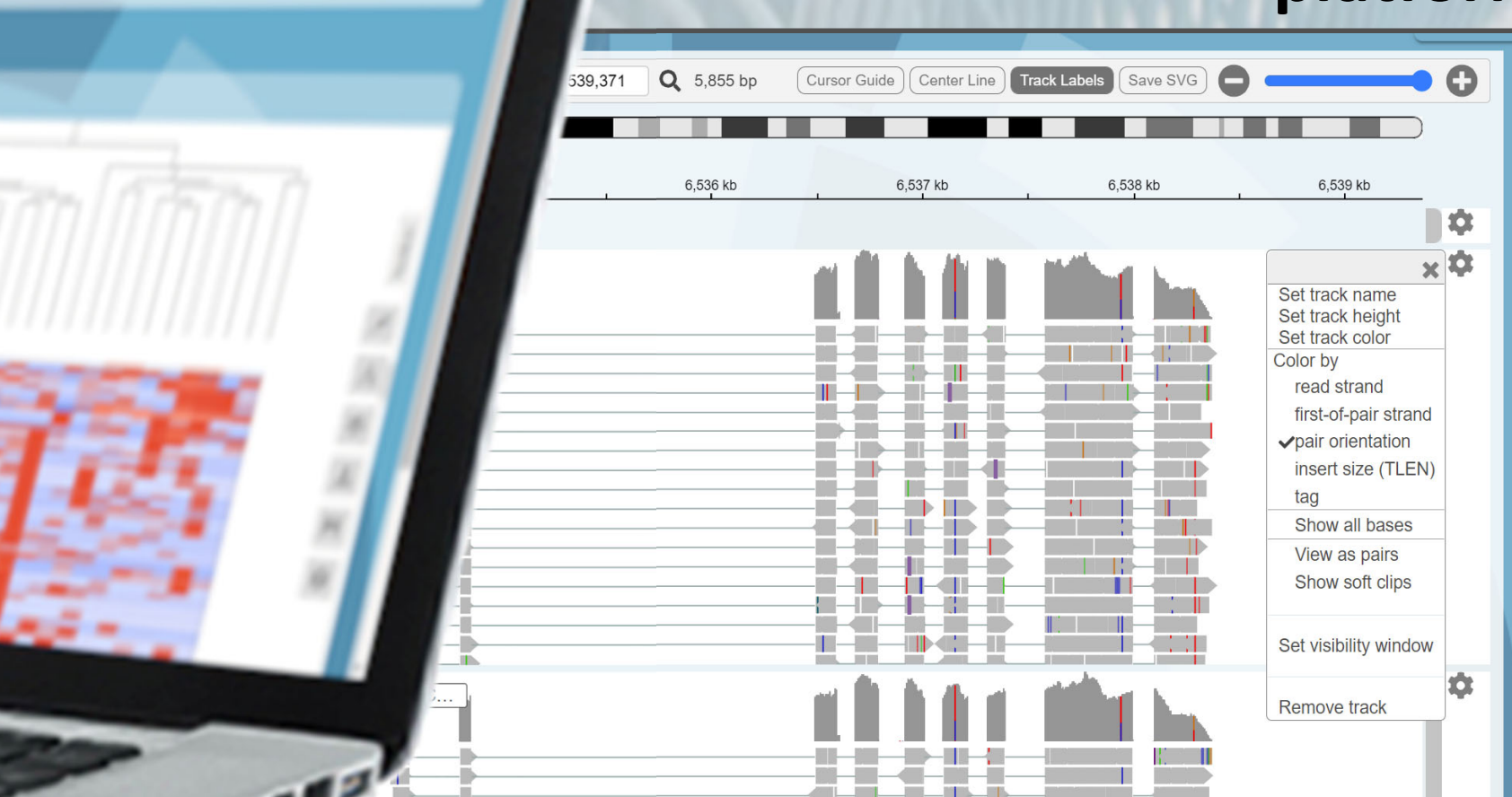- Differential Expression

Display

Aggregated pipeline QC
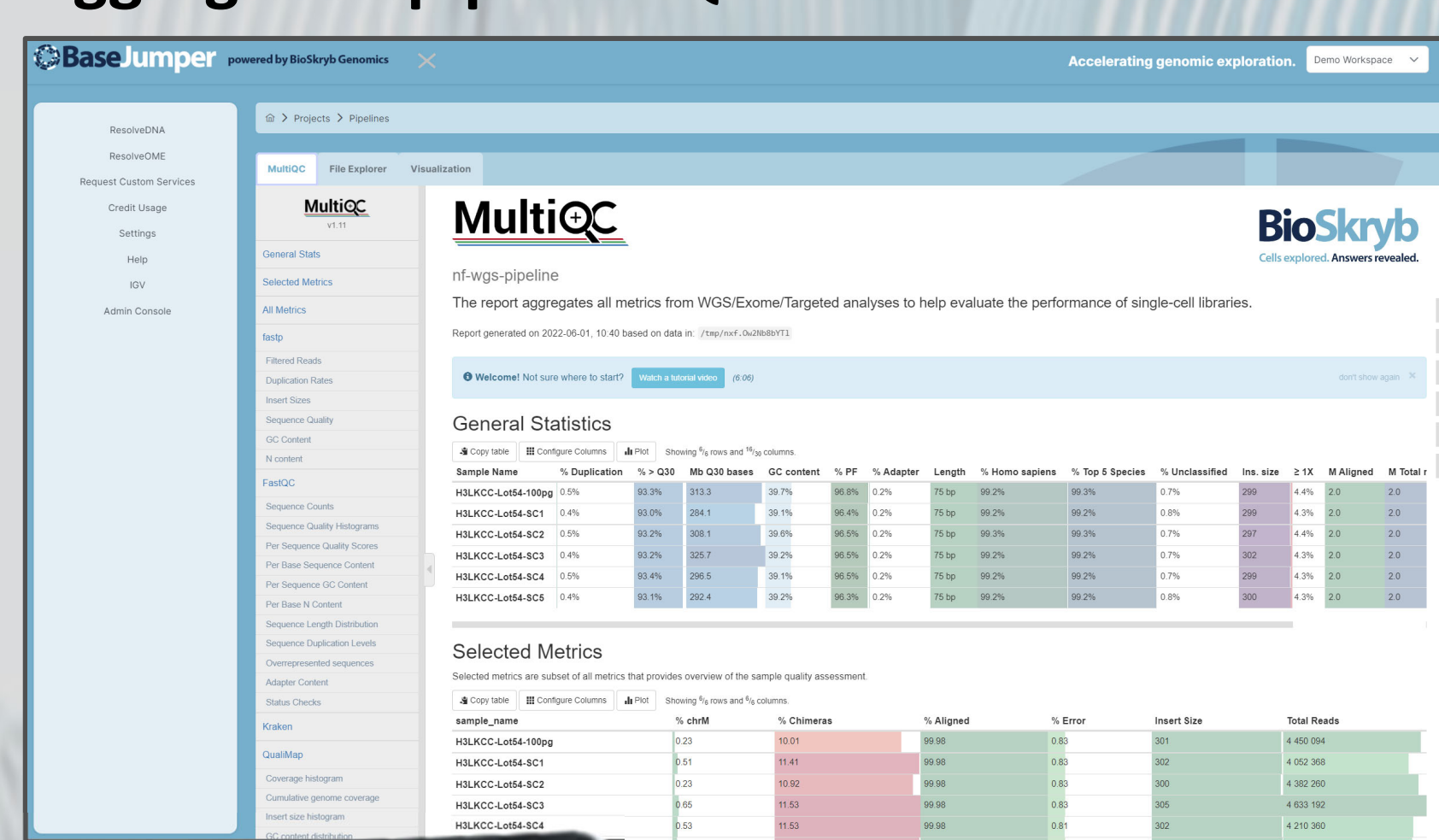
Multiple transcriptome layouts

Genome browsing within platform

Cell-by-cell copy number profiling with selectable, annotated regions

- *Rapid, dynamic and interactive figures and tables downloadable to device*
- *Exported biomarker lists in one application serve as inputs to another*
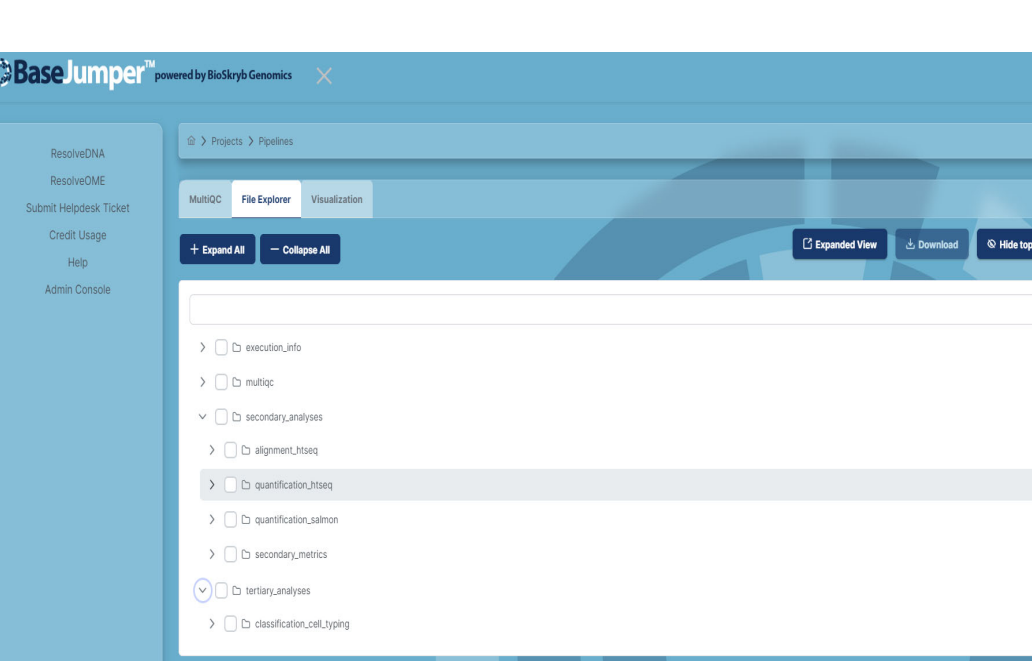- *Layouts / filters can be saved and reapplied to other projects*

Quickly filter through millions of variants in a project along with protein function prediction

Export

Push back to your Collection — globus

…or download back to device

ResolveOME™

BaseJumper™