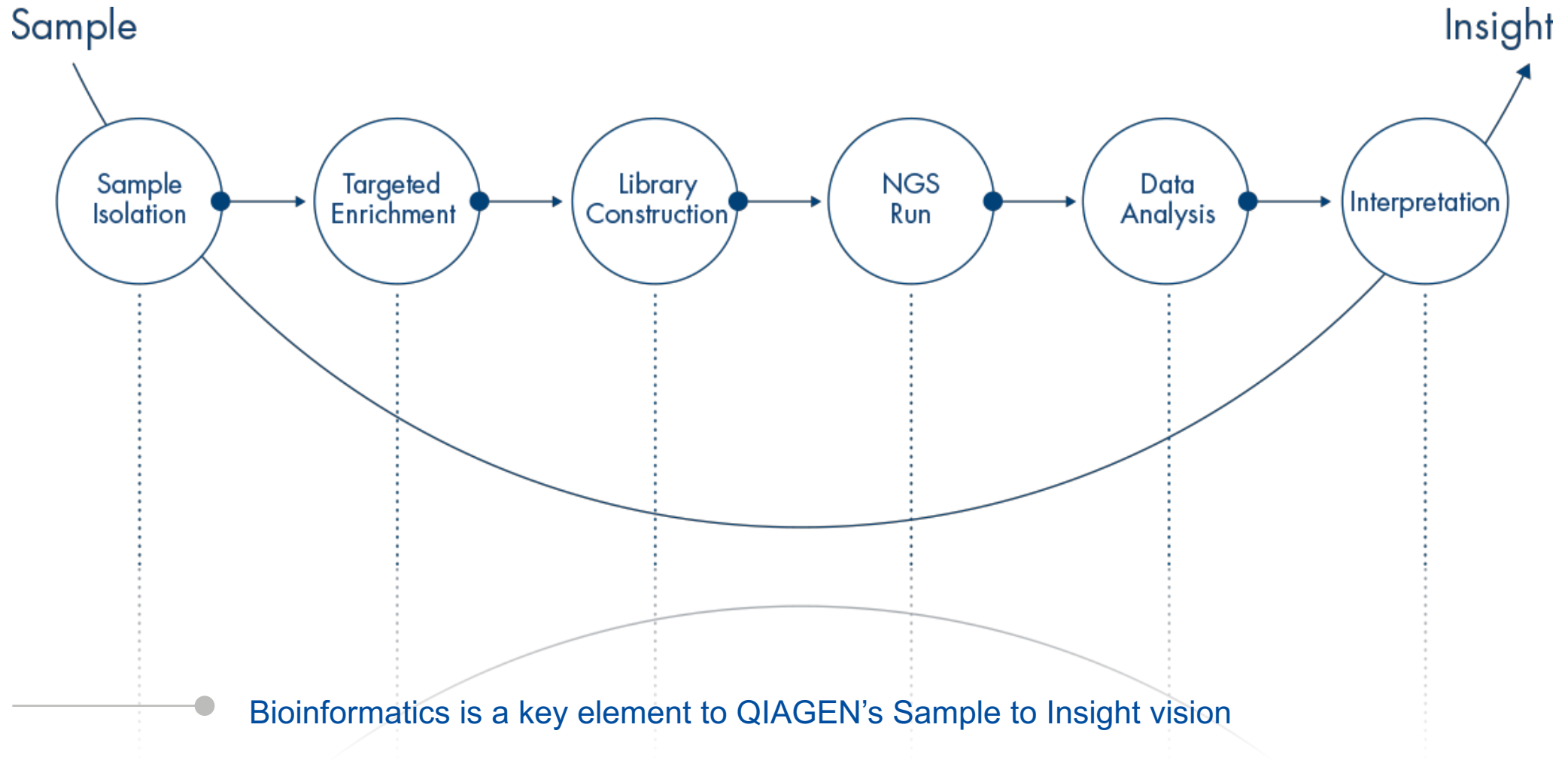


OmicSoft Introduction and Roadmap

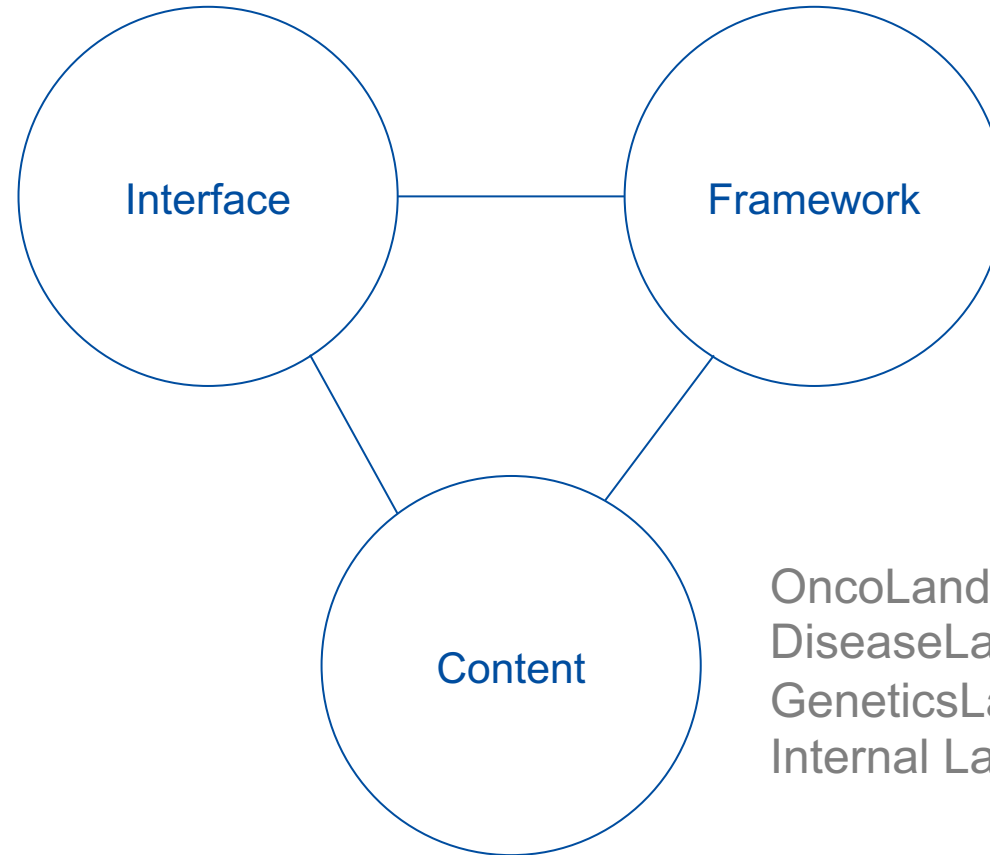
Joseph Pearson, Ph.D., OmicSoft Global Product Manager

QIAGEN vision - Sample to Insight solutions



OmicSoft Array Suite

Array Studio
Oshell/APIs
Land Explorer



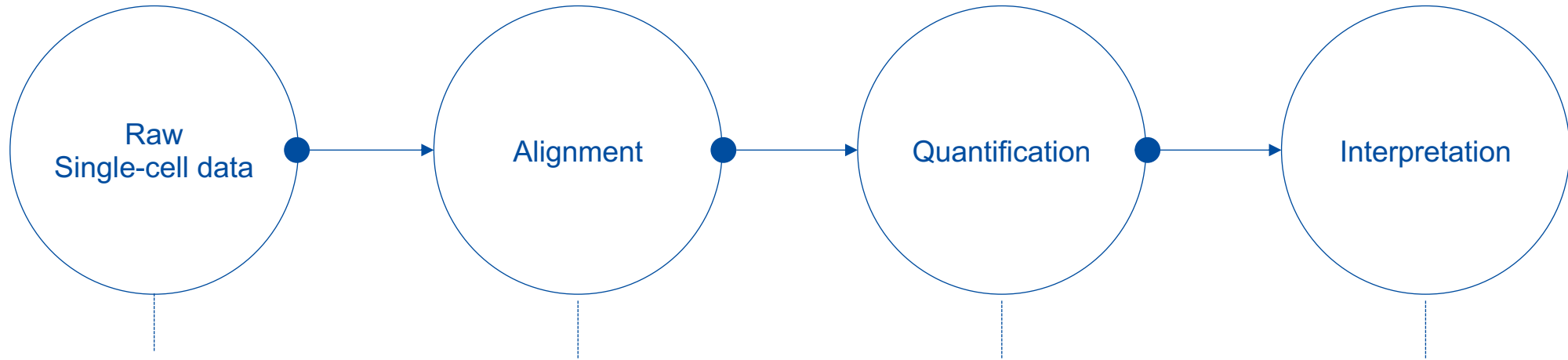
Land Technology
Project Management
Omics Analysis
Cloud Integration

OncoLand
DiseaseLand
GeneticsLand
Internal Lands

Agenda

- Interface
- Framework
- Content
- On the horizon

Single-cell RNA-seq analysis and project management



- Raw data QC
- Barcode/UMI extraction
- Cell / read filtering
- 10x Preprocessing pipeline

- Barcoded alignment
- Post-alignment QC

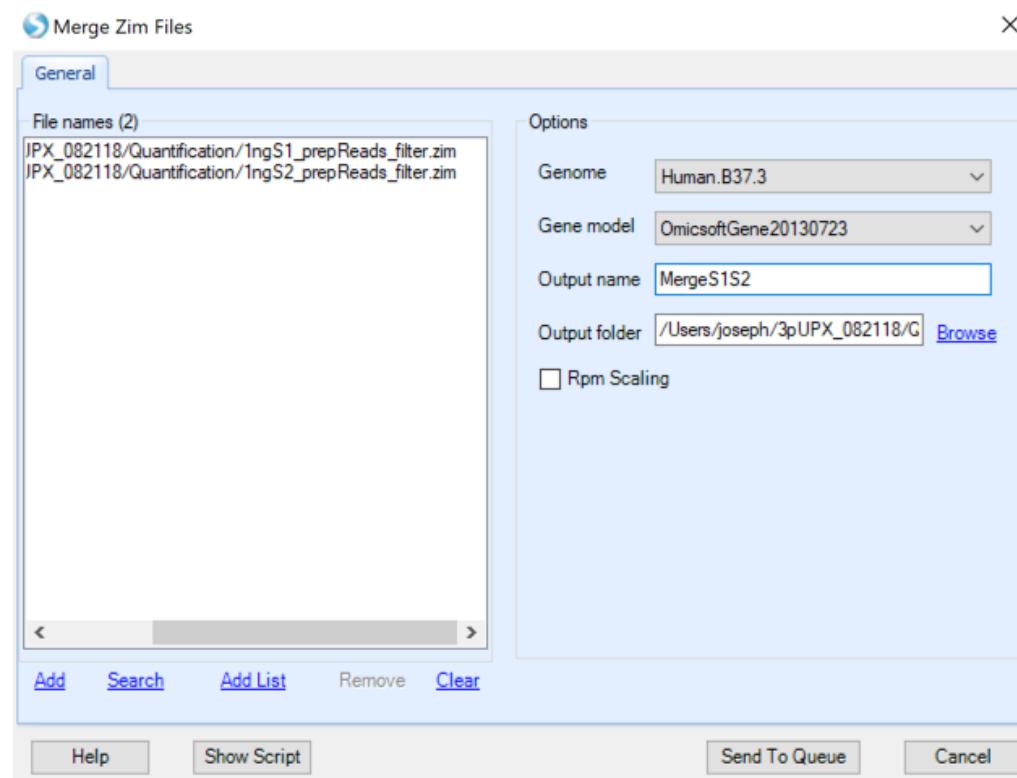
- Barcoded quantification
- Memory-efficient matrix storage
- Import results from external pipelines

- Expression scatter / violin plots
- tSNE, Seurat integration
- Overlay expression on clustering
- Send results to IPA
- Share projects

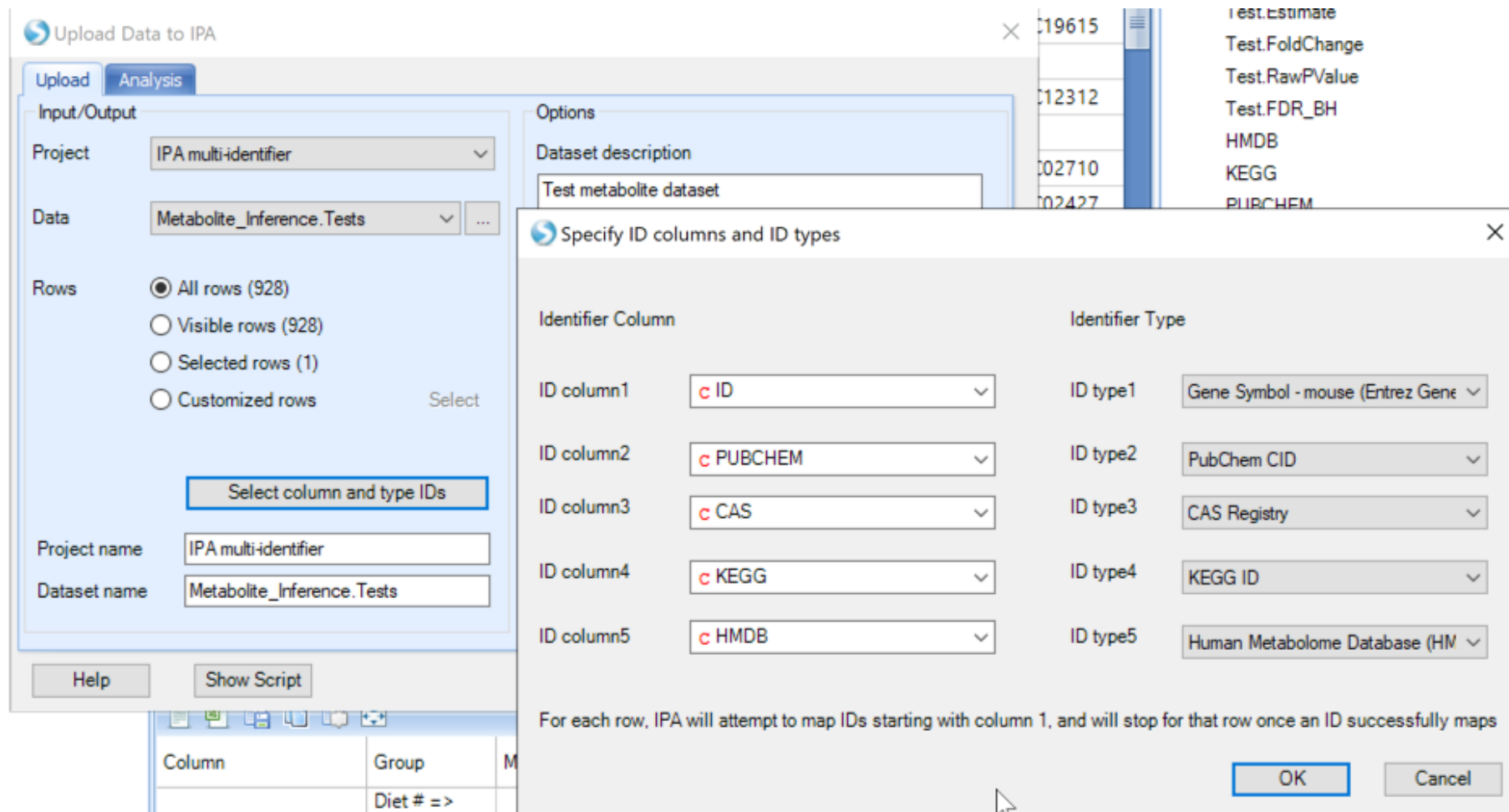
Improved handling of large expression matrices

Analysis performed external pipeline, import into Array Suite for visualization and project management

- Merge “Zero Inflated Matrices” from multiple samples into a single matrix
- Quantification and RPM-scaling from multiple samples is much more memory-efficient



Ingenuity Pathway Analysis (IPA) integration – Upload multiple identifiers for molecules



The screenshot displays the 'Upload Data to IPA' application window. The 'Analysis' tab is active, showing the 'Input/Output' section with 'Project' set to 'IPA multi-identifier' and 'Data' set to 'Metabolite_Inference.Tests'. The 'Rows' section has 'All rows (928)' selected. A 'Specify ID columns and ID types' dialog box is open, allowing for the configuration of identifier columns and their corresponding types.

Specify ID columns and ID types

Identifier Column	Identifier Type
ID column1: c ID	ID type1: Gene Symbol - mouse (Entrez Gene)
ID column2: c PUBCHEM	ID type2: PubChem CID
ID column3: c CAS	ID type3: CAS Registry
ID column4: c KEGG	ID type4: KEGG ID
ID column5: c HMDB	ID type5: Human Metabolome Database (HM)

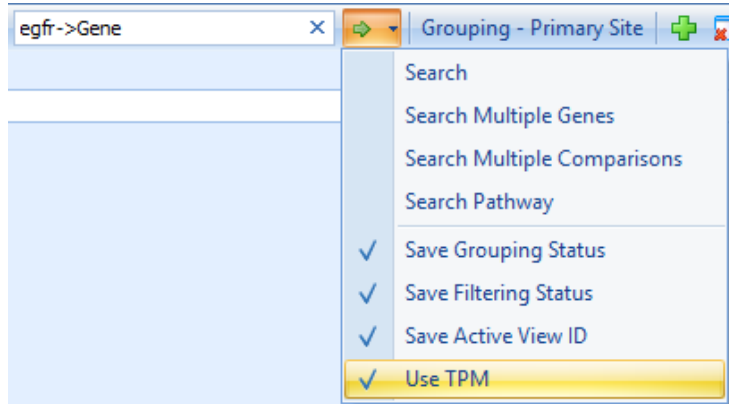
For each row, IPA will attempt to map IDs starting with column 1, and will stop for that row once an ID successfully maps

Buttons: OK, Cancel

HGMD 2018 mutation classifier: Available as add-on for all Lands (add to Land.cfg2)

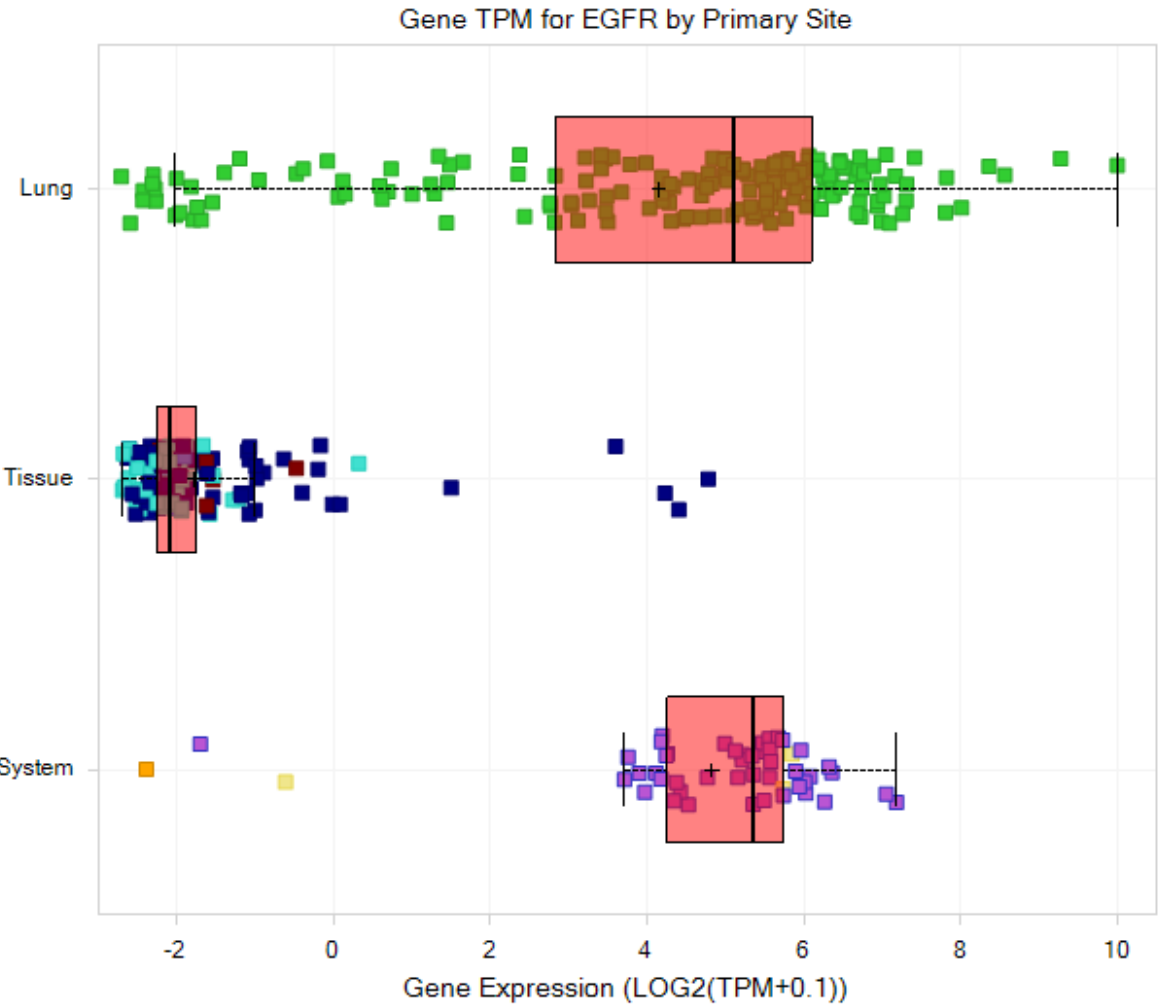


TPM scaling in all Land RNA-seq data

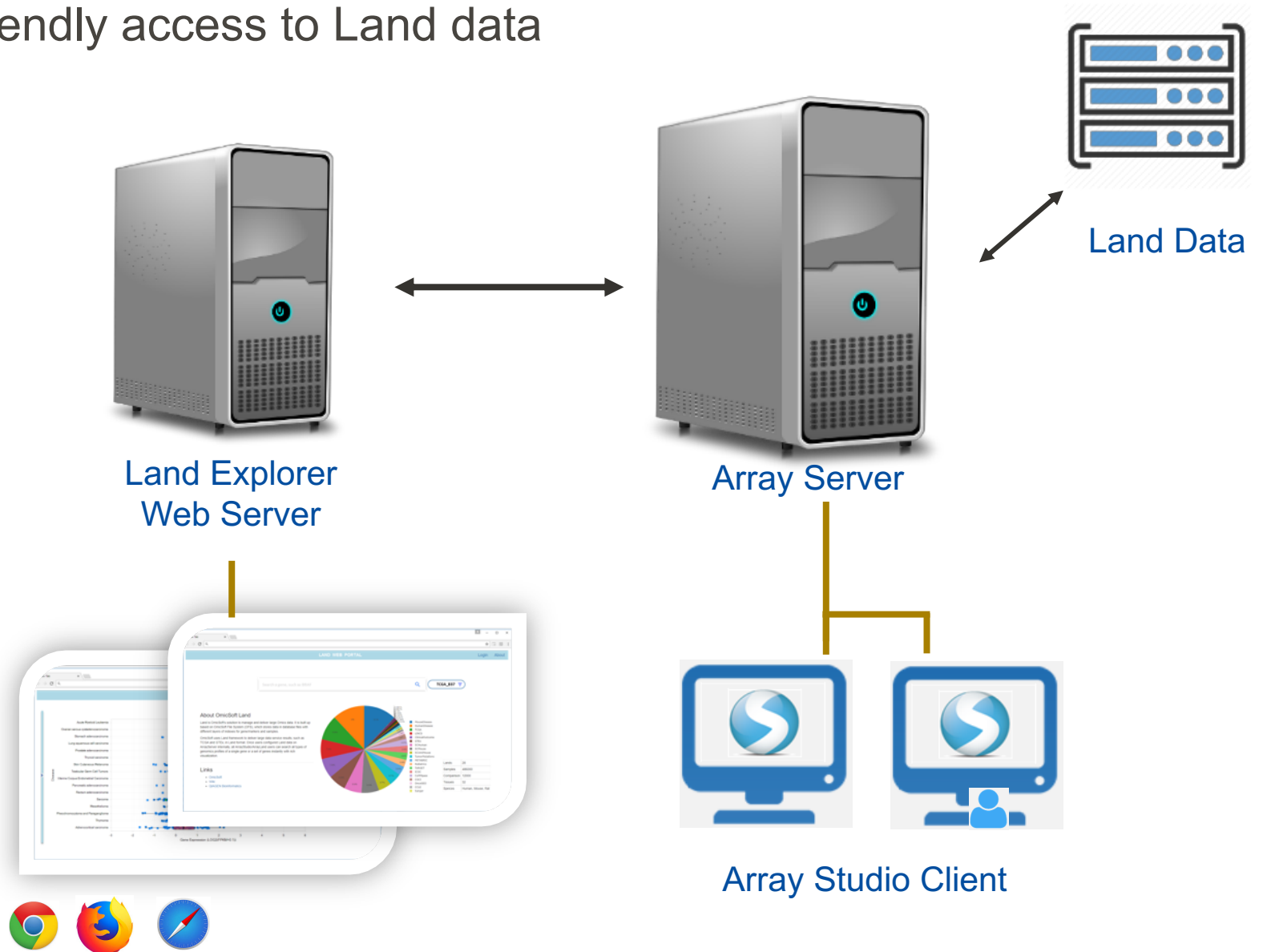


Primary Site
Haematopoietic And Lymphoid Tissue

Central Nervous System



Land Explorer: Biologist-friendly access to Land data



Comparison Explorer: Cross-Land exploration

Apply

Search...

Land Name ▲

Check All Check None Invert

Hematology_B37

HumanDisease_E

LINCS_B37

MetastaticCancer

MouseDisease_B

Pediatrics_B37

RatDisease_B6

TCGA_B37

TCGA_B38

TherapeuticArea ▼

Case.TissueCa... ▼

Case.Tissue ▼

Case.Disease... ▲

Check All Check None Invert

allergy

Comparisons by Case.DiseaseCategory

ComparisonID	ComparisonContrast	LandName
<input type="text"/>	DiseaseState <input type="button" value="▼"/> <input type="button" value="x"/>	<input type="text"/>
GSE99253.GPL6244.te...	DiseaseState => Parkinson's disease (PD) vs normal control	HumanDisease_B37
GSE74517.GPL17021....	DiseaseState => Parkinson's disease (PD) vs normal control	MouseDisease_B38
GSE93695.GPL17117.t...	DiseaseState => Parkinson's disease (PD) vs normal control	RatDisease_B6

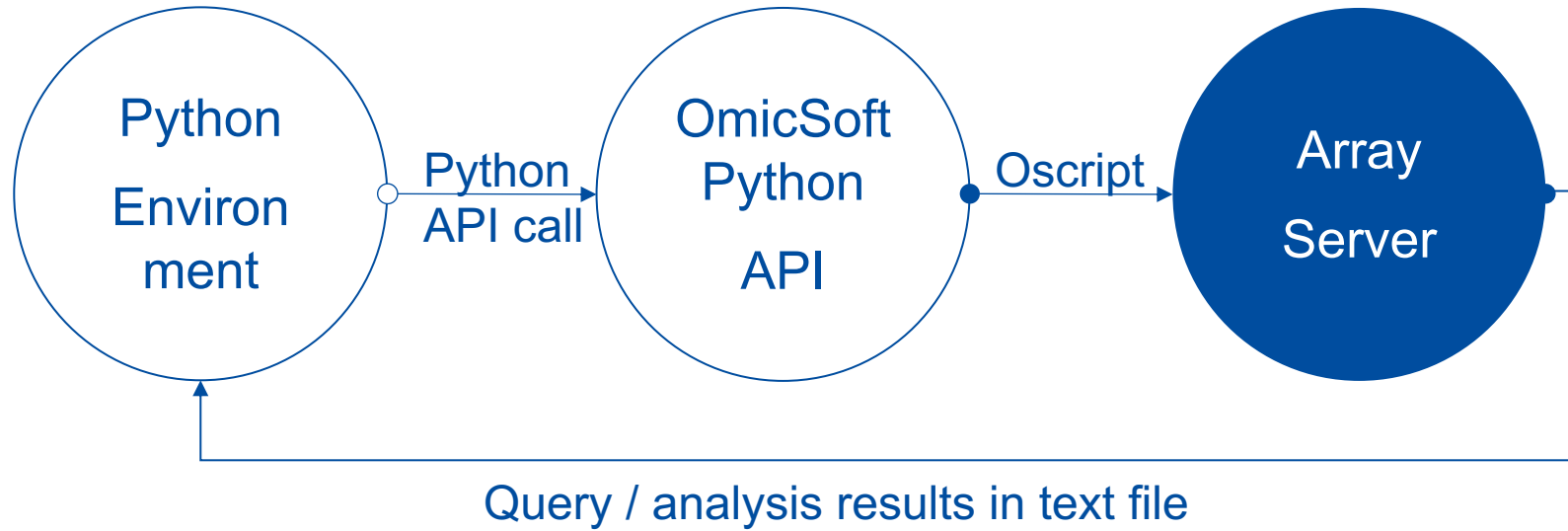
Comparison Category

Therapeutic Area

TCGA_B37	HumanDisease_B37
MouseDisease_B38	RatDisease_B6

■ DiseaseLand
 ■ OncoLand

OmicSoft's Python API allows OmicSoft analysis within the Python environment



- 1 Standard APIs
- 2 Custom API
- 3 Automated workflows

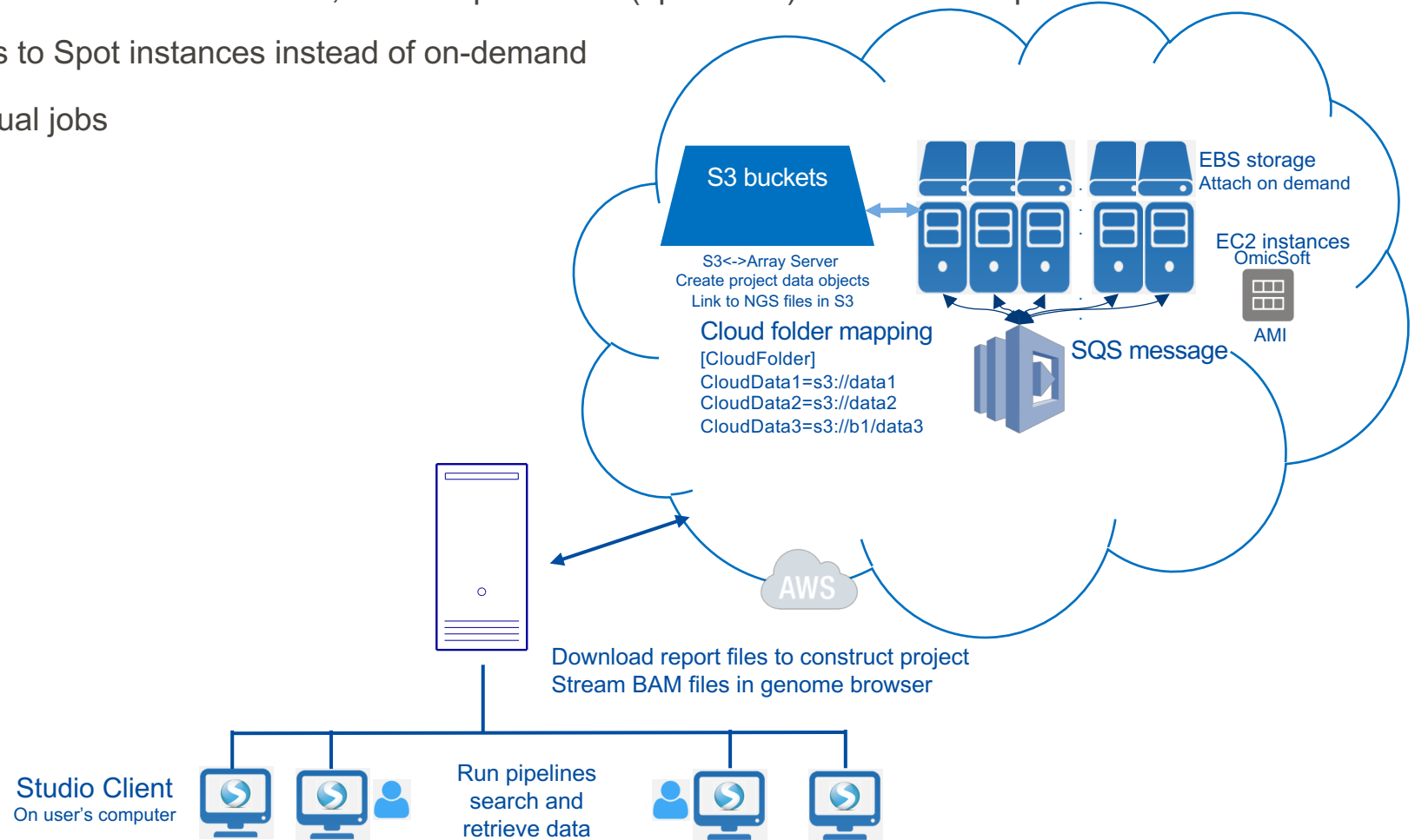
Agenda

- Interface
- Framework
- Content
- On the horizon

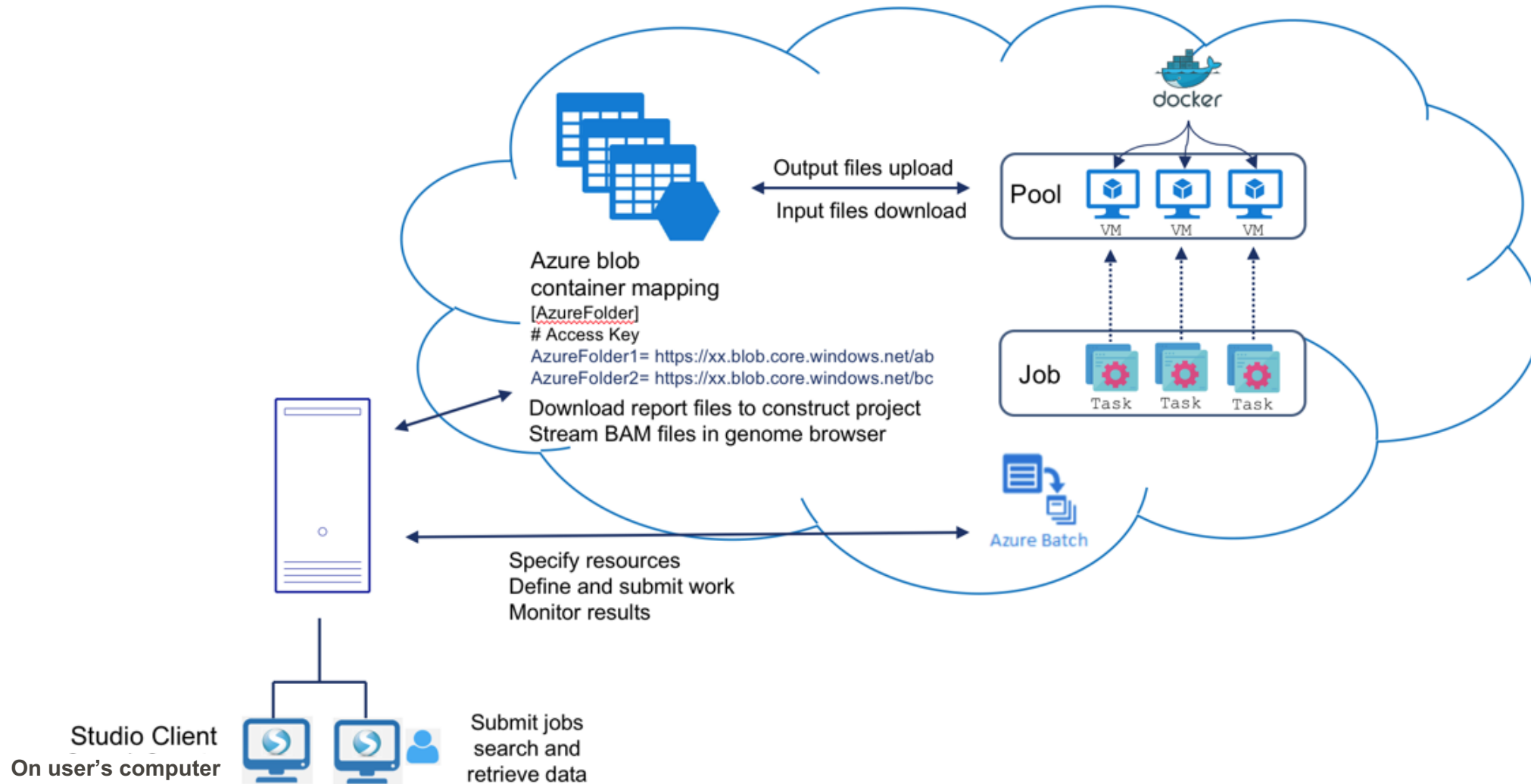
AWS Spot instance support

AWS Spot instances allow users to run jobs on unused EC2 instances, at a steep discount (up to 90%) to on-demand prices

- Array Server can launch cloud analysis jobs to Spot instances instead of on-demand
- Specify for all AWS-enabled jobs, or individual jobs



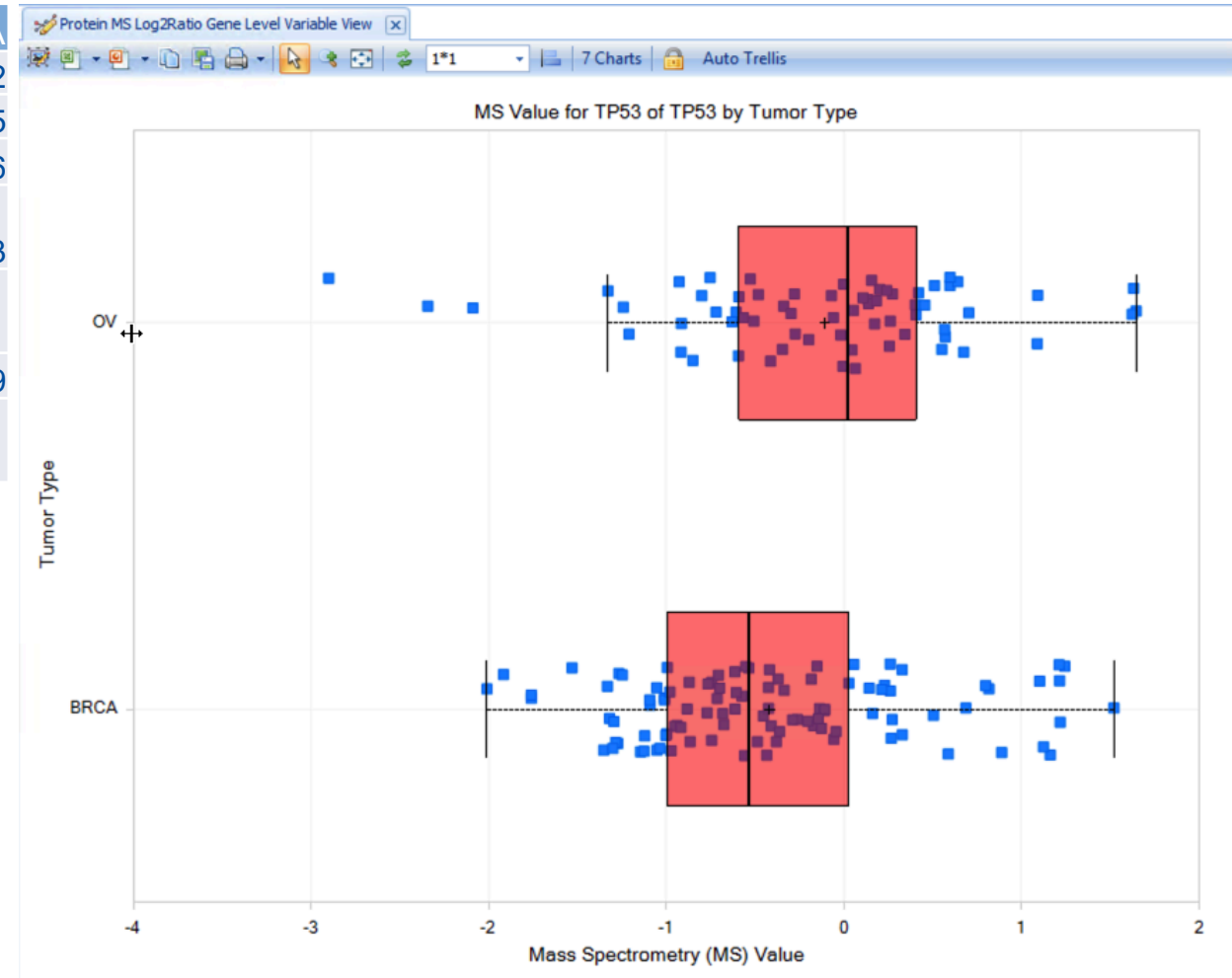
Cloud analysis support with Microsoft Azure



Multi-Omic internal Lands - Add protein MS data to internal Lands

Supports Protein and Modification views

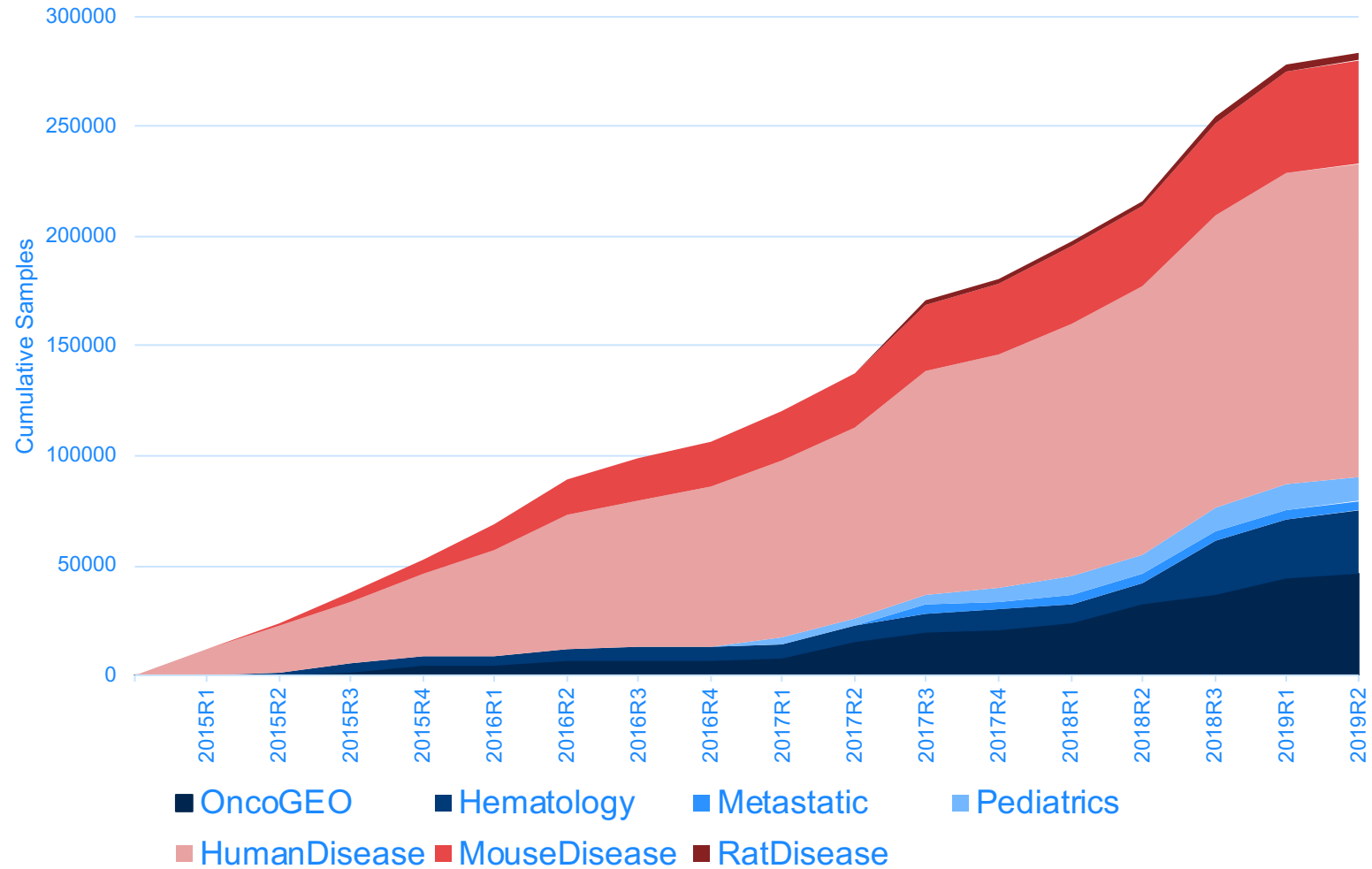
GeneID	ID	09-1664-01A	13-1484-01A	13-1488-01A	13-1489-01A
A1BG	A1BG	0.3354	-0.8444	-0.075	0.0142
A2M	A2M	-0.0316	-0.4225	0.1958	-0.4135
AAAS	AAAS	-0.0311	-0.4959	0.1128	0.0346
AAAS	NP_001166937.1_pS462		0.6563		-0.1613
AAAS	NP_001166937.1_pS508				
AAK1	AAK1	-0.339	0.0169	-0.0568	0.0069
AAK1	NP_055726.3_pS637		0.4128		



Agenda

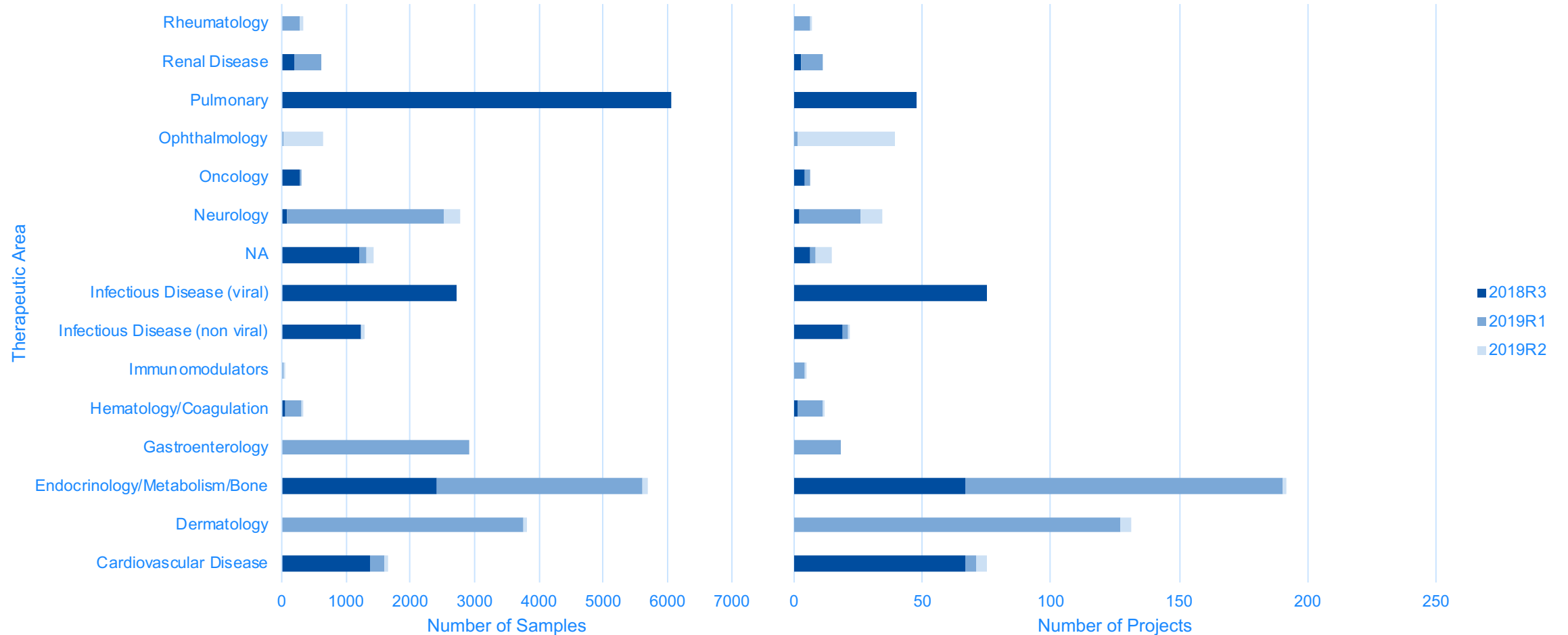
- Interface
- Framework
- **Content**
- On the horizon

Project-based OmicLand content growth



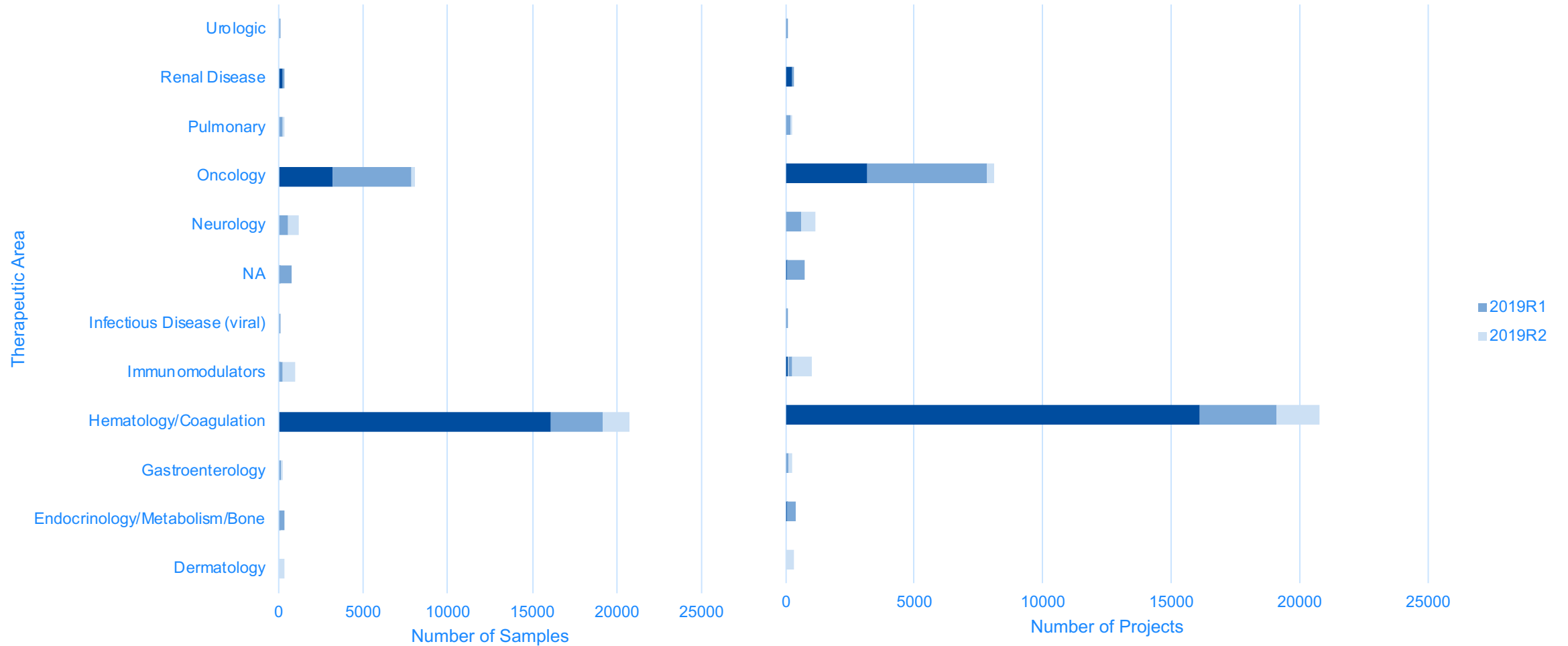
DiseaseLand content growth, 2018R3-2019R2

30508 samples, 690 new projects



OncoLand content growth, 2018R3-2019R2

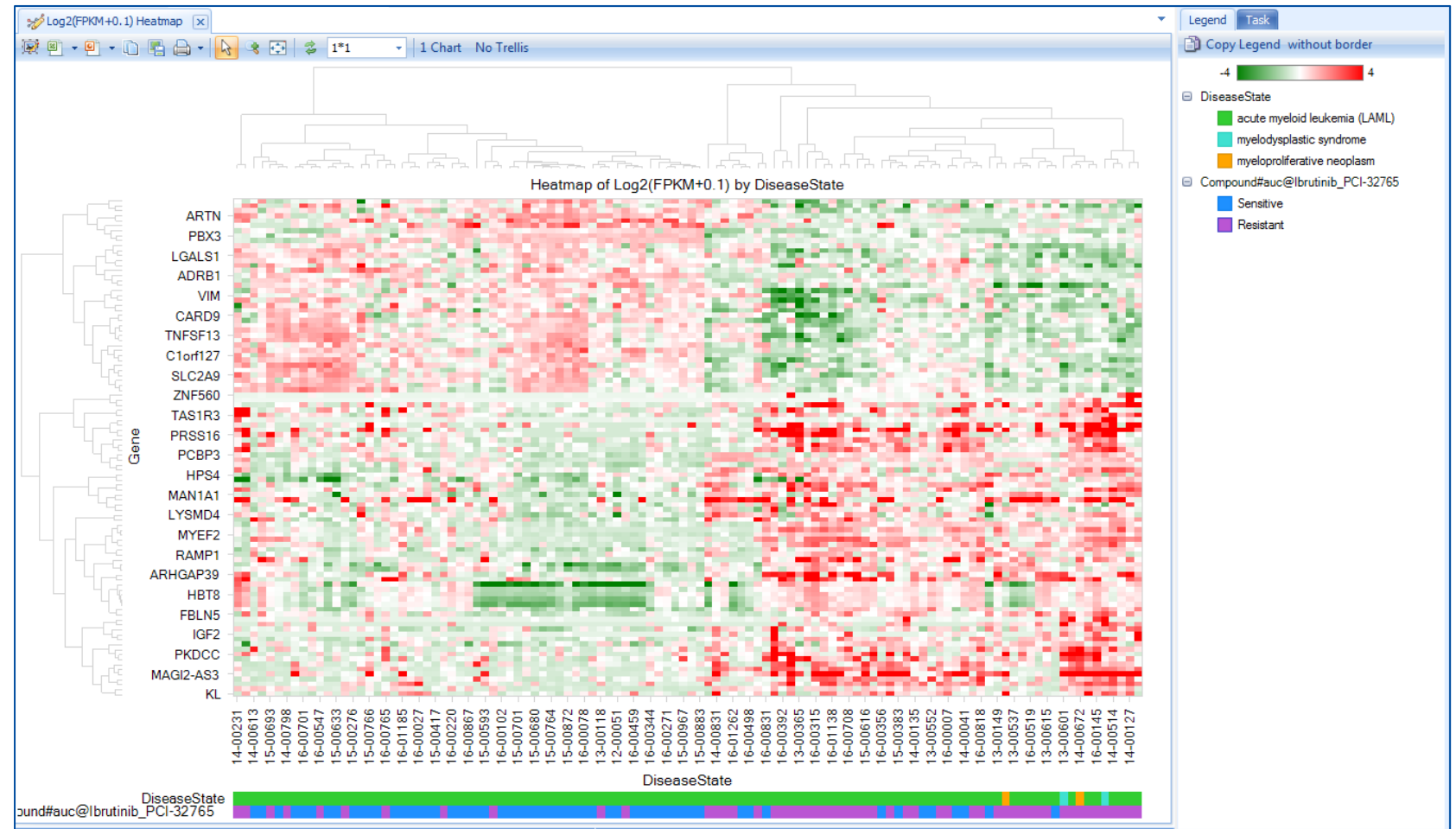
33160 samples, 407 new projects



Just released: BeatAML Land

Explore RNA-seq, DNA-seq and ex vivo drug treatment measurements from 562 acute myeloid leukemia patients

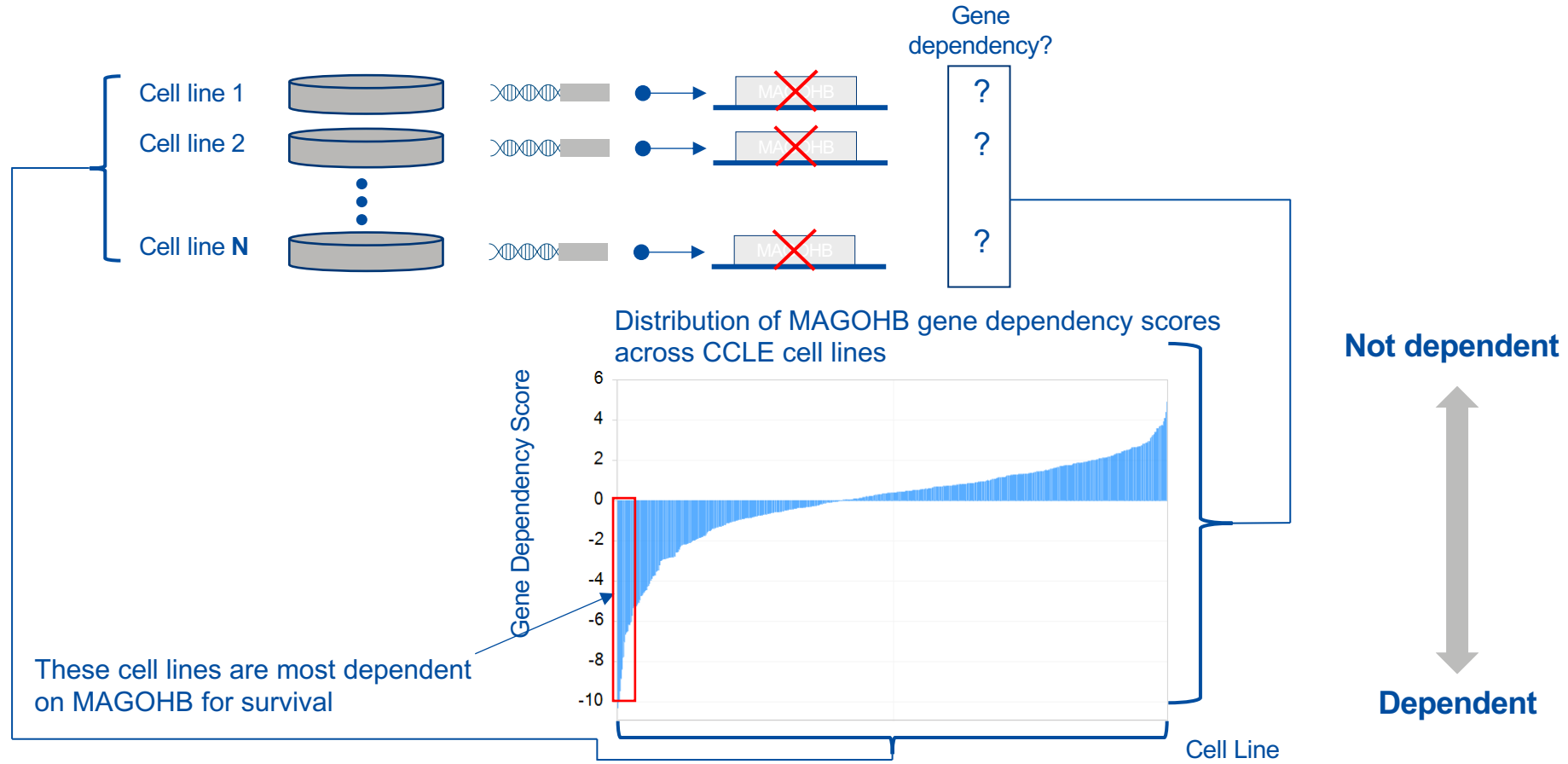
Learn more this afternoon



Upcoming release: 2019 R3 focus

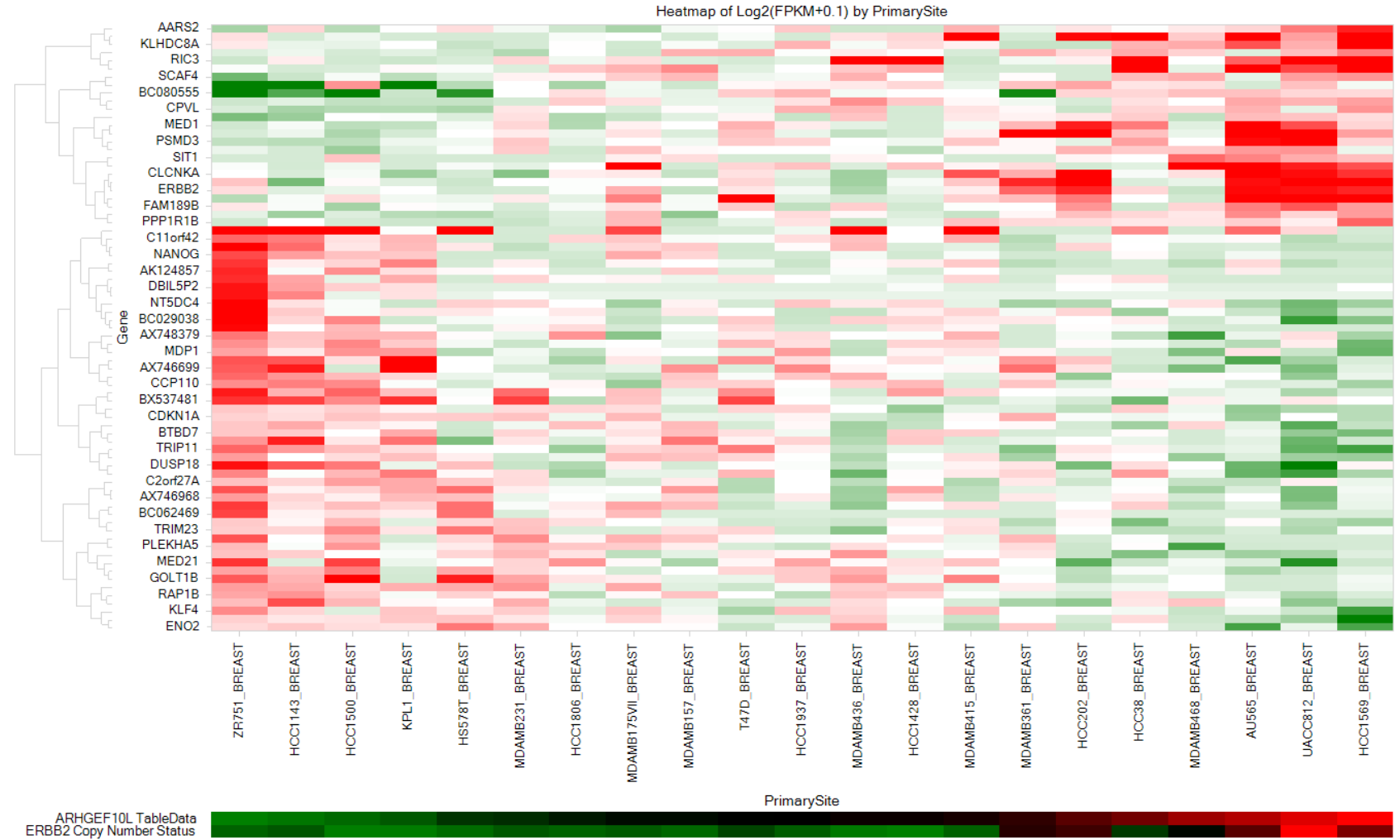
- OncoLand
 - New Land: OncoMouse
 - Focus on cancers of reproductive system, GI system, respiratory system, urinary system, skin, CNS
- DiseaseLand
 - CNS disease (ALS, multiple sclerosis, Alzheimer's disease, Huntington's disease)
 - HIV

Upcoming: CCLE Land with CRISPR/RNAi dependency data



CCLE Land with RNAi/CRISPR dependency data

- CCLE 'omics data
 - RNA expression
 - Protein levels
 - CNV, mutation data
- Gene Dependency Scores
 - RNAi/CRISPR
 - Project DRIVE
 - Project Achilles
- Measurement data support
 - Drug sensitivity
 - GDSC, CTRP
 - CCLE metabolite data
 - Internal measurements



Agenda

- Interface
- Framework
- Content
- On the horizon

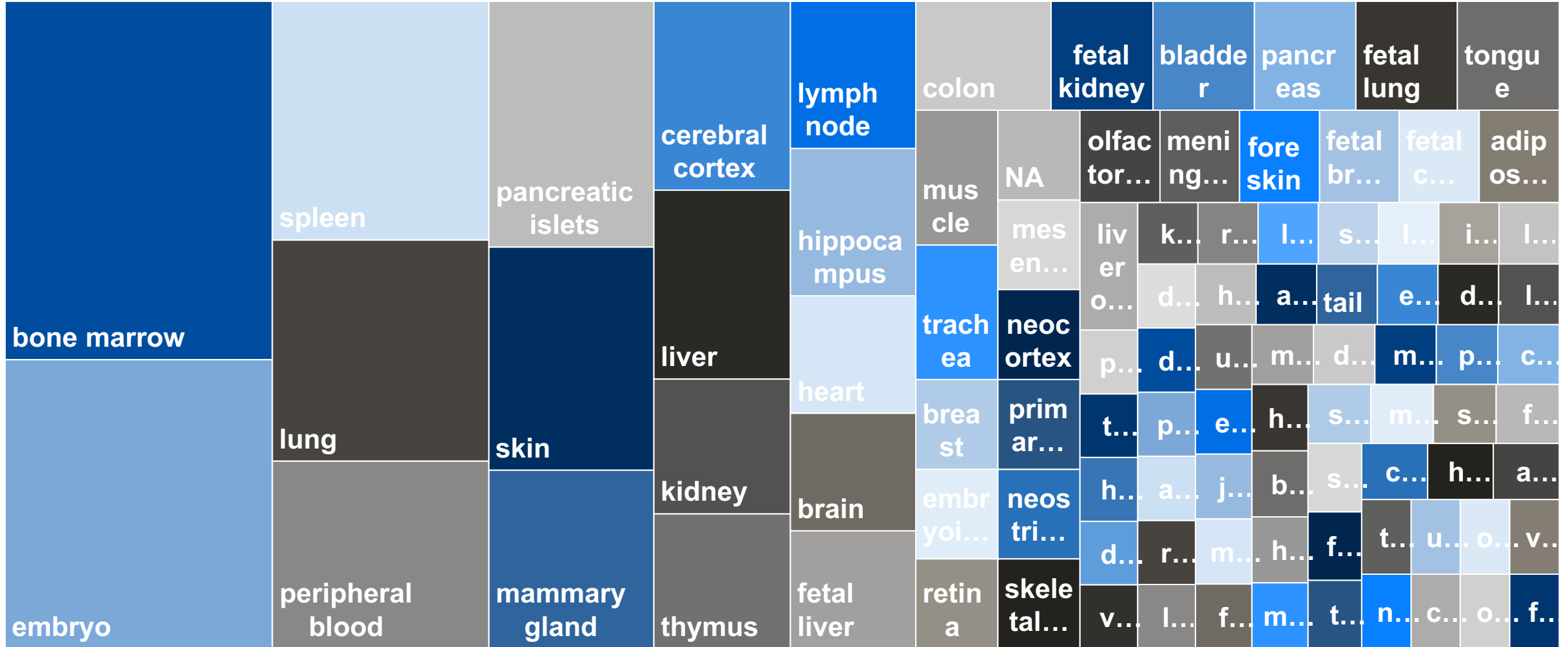
On the horizon

- Update Lands for latest genome/gene models
- New Land Content
 - New curated datasets for Lands
 - Update GTEx to v8
- Single-cell Land Version 2
- External tool support on HPC and Cloud
- Improved Land APIs

On the horizon: Single-Cell Land V2 framework

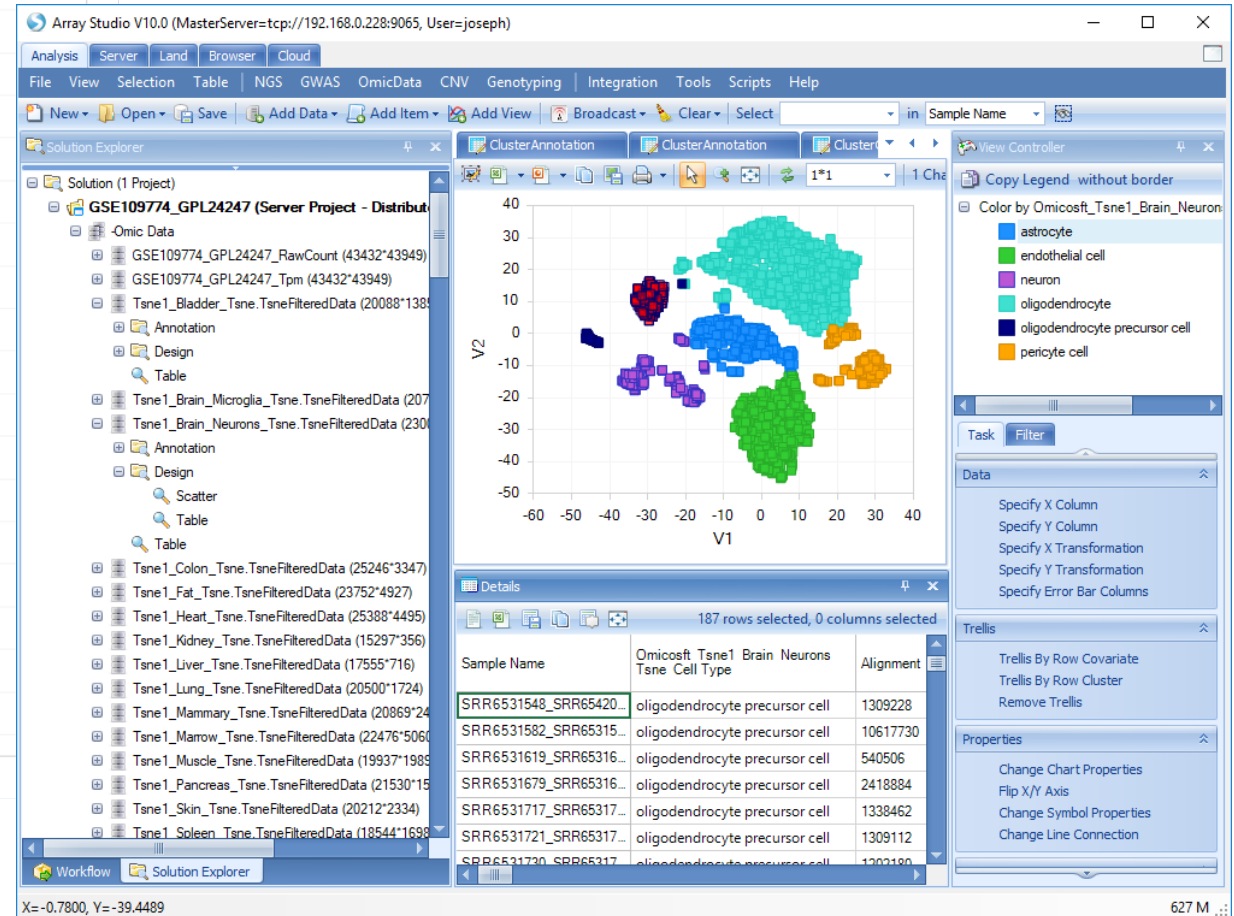
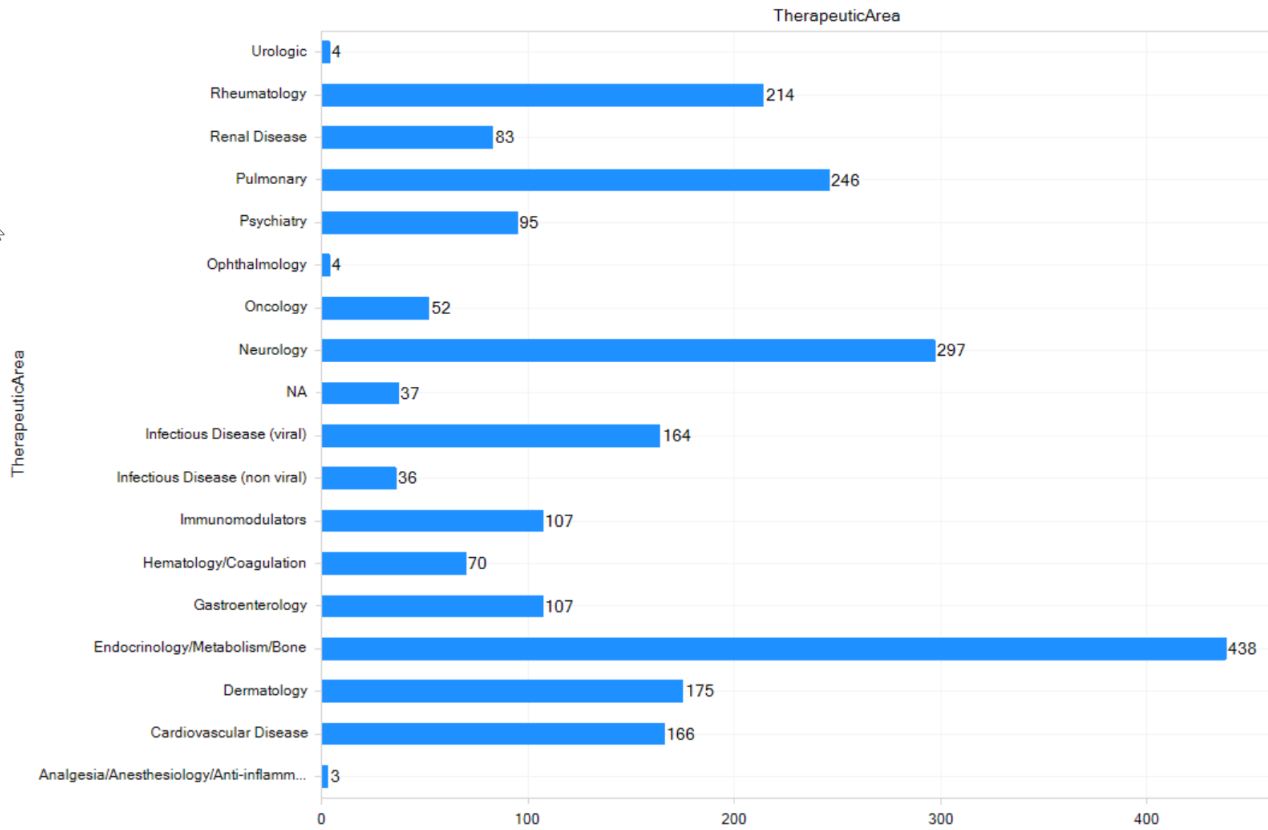
- Browse curated single-cell RNA projects by project, sample and cluster metadata
- Explore tSNE plots for each project with cluster annotations, overlay expression of marker genes
- Search by genes to find cell clusters enriched or depleted for gene of interest

143 curated single-cell projects, 97 tissues represented



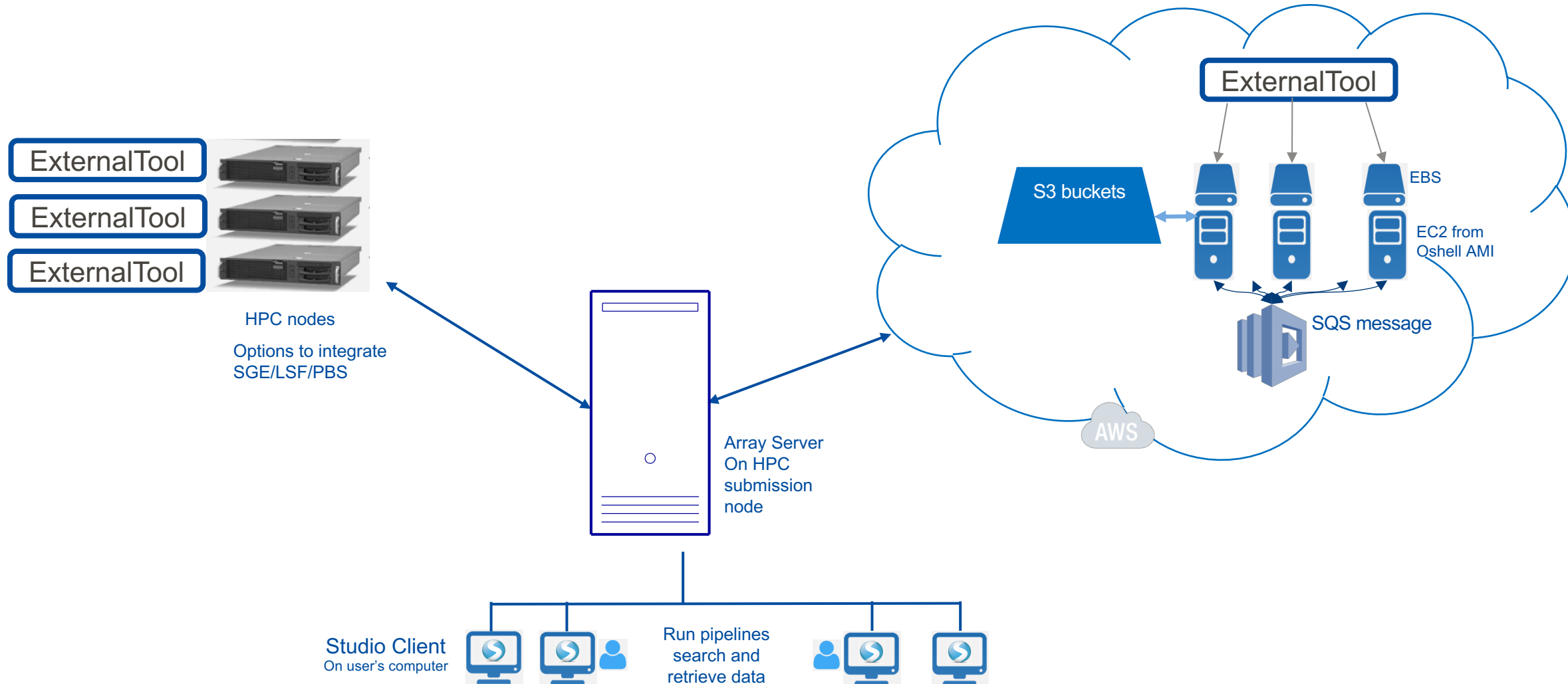
On the horizon: Retrieve full Array Studio projects with curated single-cell data

Download and explore carefully curated scRNA-seq projects with normalized cell-level expression data, using Array Studio's scRNA analysis support, including tSNE and Seurat analysis

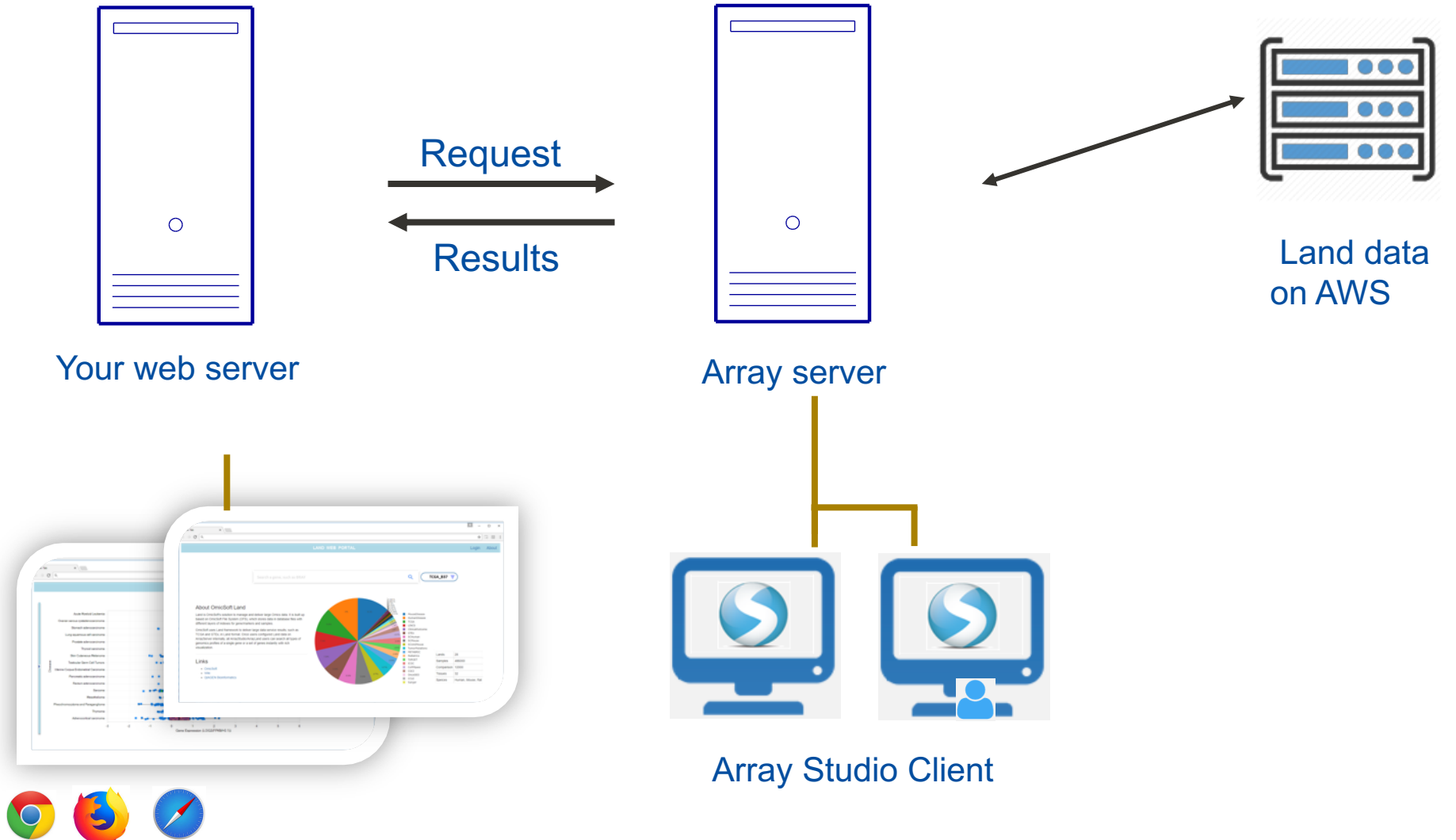


On the horizon: Cloud/HPC integration with External Tools

ExternalTool framework to launch Container Images with external executable/R package/Python packages



On the horizon: Improve Land APIs for web developer-friendly access to Land data



Q&A

