

Accelerating cancer genomics with cloud-scale Bioconductor

Vincent Carey (BWH), Aedin Culhane (DFCI), Rafa Irizarry (DFCI)

- Aim 1: Infrastructure for scalable multiomics
 - Remote data, remote analysis, autonomous system provisioning
- Aim 2: Onco-nuggets for cloud-based app development for cancer genomics
- Aim 3: Renovate *MLInterfaces* for scalable statistical learning with cancer applications
- Tools 1: HDF5/server, EC2, BigTable/BigQuery, AWS Lambda, GDC, GA4GH: break R bottlenecks inside and outside session concept
- Tools 2: TCGA, e.g., as a “MultiAssayExperiment”; operators, gadgets, plotly
- Tools 3: mlR, sparklyr, H2O
- Challenge: meet goals without losing integrity assurances of familiar local/in-memory programming patterns