

Running a core without a budget

Michael Poidinger – PI Bioinformatics
SIgN, Singapore



SigN was launched in January **2008** by Singapore's A*STAR with the aim of expanding and strengthening immunology in Singapore.

Mission & Vision

- To be a leading international research centre in **Human Immunology**
- Expand the knowledge base in immunology research, ensuring the translation of research discoveries towards medical applications.

Bioinformatics team and resources

- ▶ 5 (soon 9) post-doctoral bioinformatic scientists
 - 2 with pharma industry experience
 - Experience with the cross analysis of heterogeneous data sets
 - Experience in biology and specifically immunoinformatics
- ▶ Commercial grade hardware
 - Multi-node 64bit clusters
 - Many Terabytes of storage
 - ASTAR Internal Cloud

Funding/work model

- ▶ Each RI has a (generous), ongoing core fund (~70%) with access to various competitive grant schemes and industry collaboration (~30%)
- ▶ Staff work in close contact with the 25 PIs and
 - 180 researchers in the institute.
- ▶ Every analysis is special
 - Biologists ask us what can we do with the data
- ▶ Projects often extended due to findings

BIOinformatics

Clinical studies

- ▶ SIgN is involved with a number of clinical studies
 - Allergies
 - Sepsis
 - Kidney rejection
 - Urinary tract infection
 - HIV
 - Elderly cohort
 - Nasopharyngeal carcinoma
 - H1N1 vaccination

NGS studies

- ▶ RNAseq on sorted immune cells from healthy, latent and active TB infection
 - mRNA and miRNA
- ▶ Variant detection in yeast mutants
- ▶ RNAseq on immune cells from bats
- ▶ ChIPseq and methylseq on underweight babies
- ▶ Microbiome of ICU patients and prediction of candidiasis
- ▶ Tcell and Bcell repertoire in response to Dengue

Heterologous Data Integration

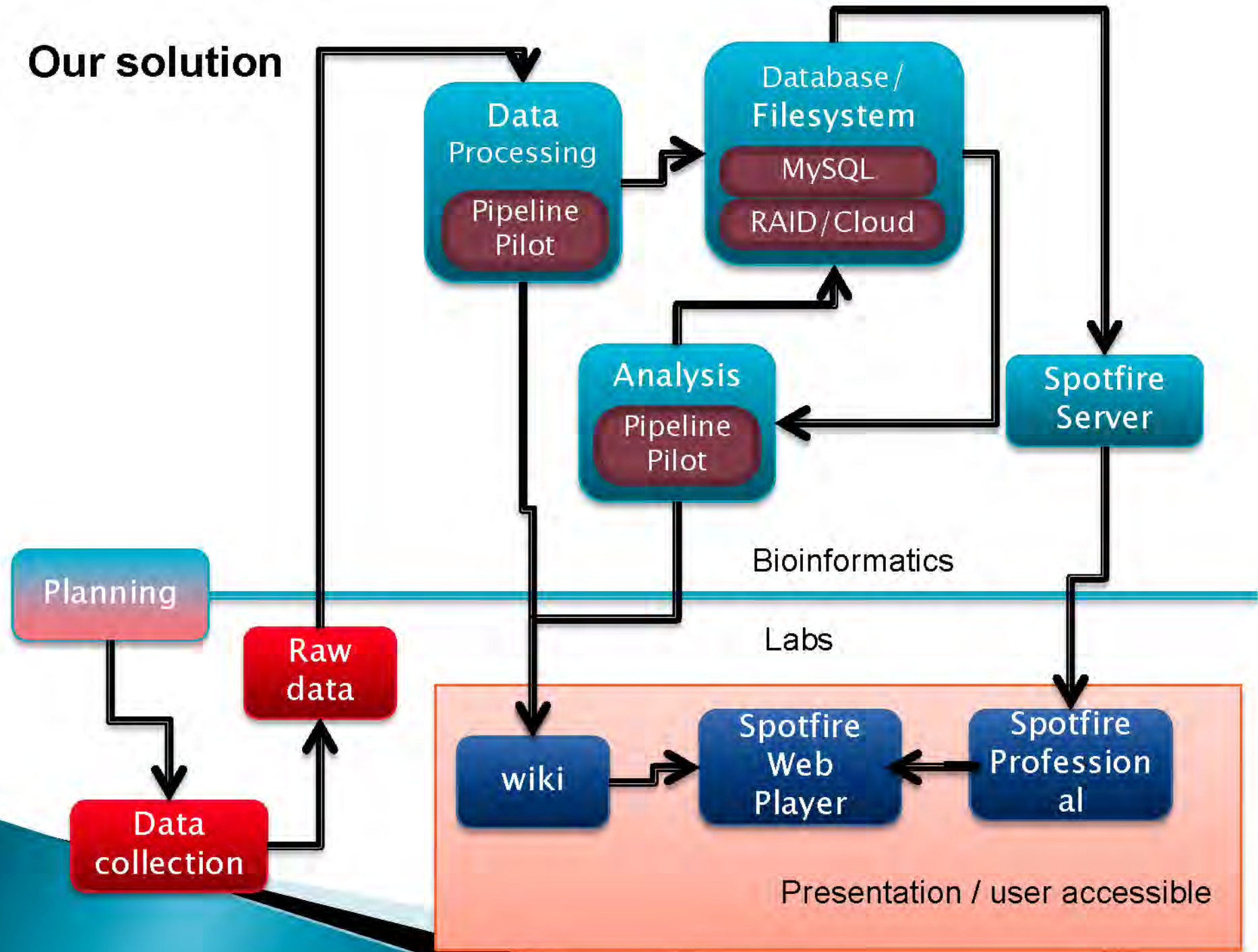
- ▶ Many projects comprise data from a range of source
 - Clinical survey
 - Clinical tests (e.g. skin prick for allergy)
 - Cell based assays
 - Usual suspects of HT data (microarray, NGS)
 - Luminex (measures analytes in blood)
 - Flow cytometry
 - Mass cytometry

We don't....

- ▶ Crank the handle on generic protocols
- ▶ Work on our own research
 - Unless it derives specifically from an existing project
- ▶ Build, install or require systems like Galaxy, that push analysis back to the biologist

... not that there's anything wrong with that

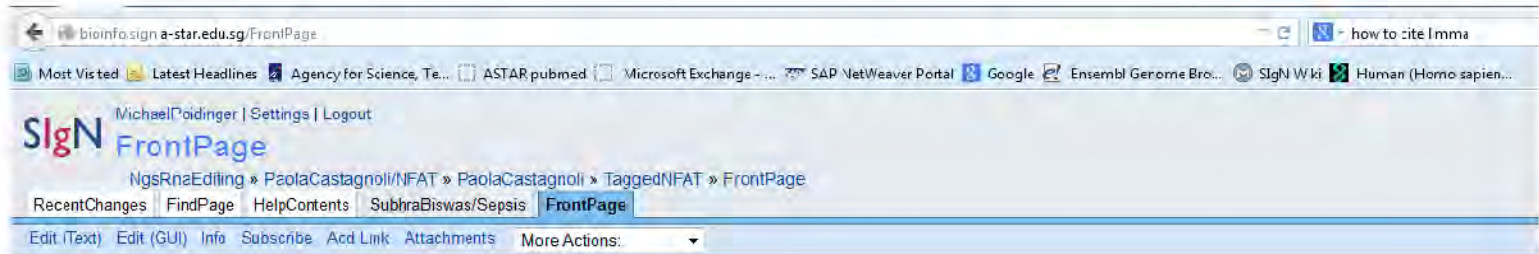
Our solution



Software/tools

- ▶ Commercial licenses for a range of products
 - Thompson Reuters/ GeneGO
 - Ingenuity Pathway Analysis
 - Oncomine
 - Transfac
- ▶ Pipeline Pilot from ~~Accelrys~~ Dassault
- ▶ Spotfire from ~~Tibco~~ Perkin Elmer

Reporting via Moin Moin Wiki



SlgN Bioinfo wiki

Welcome to the SlgN Bioinfo wiki. All experiments and analyses involving the Core and/or Bioinformatics teams will be stored here. Experiments are stored by experiment name and experimental data is restricted according to the directions of the PI.

To gain access to this wiki, use your system username/password (the same one you use to access email). Anyone that had a specific wiki account created before we changed over require access to any of the experimental pages, or would like to utilise this wiki for the management of your own data, please send a request to Michael.

Please note this is an intranet site, access from outside requires VPN.

Information

- [System administration information](#)
- [Information - general information and knowledge sharing](#)
- [Infrastructure](#)
- [Functional Genomics Laboratory](#)
- [Functional Genomics Laboratory Technical Information](#)

Analysis results

- [AlessandraMortellaro](#)
- [AnnaMarieFairhurst](#)
- [ChengWang](#)
- [EeCheeRen](#)
- [EvanNewell](#)
- [FlorentGinhoux](#)
- [FlowCore](#)

Wiki: Project reporting

JanuaryCohort/NeutrophilRNAQuality - SigN Bioinfo

JanuaryCohort/NeutrophilRNAQ...

bioinfo.sign.a-star.edu.sg/JanuaryCohort/NeutrophilRNAQuality

Google

Updated analysis

The Spotfire visualization has been updated to include the latest January cohort data. The updated visualization is available at:

http://spotfire.sign.a-star.edu.sg/SpotfireWeb/ViewAnalysis.aspx?file=/cohort_studies/january%20cohort/jan_cohort_neutrophil_rna_quality_microarray_updated



20110923 Initial analysis

The RNA quality was assessed by the Microarray Core group and was made available as an [Excel spreadsheet](#). This was processed and stored into the database prior to being combined with the rest of the cohort data to determine if the samples suitable for microarray had any bias in them.

The results are located in the following Spotfire visualization:

http://spotfire.sign.a-star.edu.sg/SpotfireWeb/ViewAnalysis.aspx?file=/cohort_studies/january%20cohort/jan_cohort_neutrophil_rna_quality_microarray

The findings are as follows:

1. Kruskal-Wallis tests were conducted to determine if any of the continuous parameters was significantly different between samples suitable for microarray and not.
 1. The RNA RIN value and yield showed significant results and that is to be expected since they were used to classify whether the samples were suitable for microarray.
 2. 
2. Chi-square tests were conducted for discrete parameters.
 1. The first 3 parameters are to be expected since they are part of the parameters that determined microarray suitability.
 2. 3 SNPs were also found to be significant but again they are to be expected without any multiple testing correction.
3. Spearman correlation tests were also conducted against the RNA yield.
 1. Although many of the tests were significant, the R^2 is very low for most of the significant results. R^2

Links to visualizations

Link to raw data

Analysis findings

Pipeline Pilot

The screenshot displays the Pipeline Pilot software interface. On the left is a tree view of components under the 'NGS' folder, including 'Easter cluster' and various analysis tools like 'bowtie easter component', 'Cuff union component', 'fastQC on easter', 'gsnap easter component', and 'tophat and cufflinks easter component'. The main workspace shows a workflow diagram with nodes such as 'Repository Query', 'Property Value T...', 'Reference ID and Name ...', 'keep('transcriptID',refer...', 'Cache Writer', 'Don't Pass Data', 'merge outputs', 'merge gene stats onto...', 'Join Data from Cache', 'NodeSetType (DataRoot)', 'originalTranscriptID = tr...', 'Reference ID and Name ...', 'experimentName := \$(ne...', and 'Add Features to Repository'. The interface includes a status bar with 'Elapsed Time:' and a task bar with several open protocols.

The help window is open, displaying the following text:

Returns the selected output type for each input region
 Queries a repository for the specified output type contained within a region or regions. The *Repository Name* is specified by component parameters, and the region is specified by properties on the input data record. The output type is one of

- subsequence of the specified reference sequence
- all features from the specified feature set(s)
- subsequence of the specified reference sequence plus all features from the specified feature set(s)
- mapped reads from the specified experiment(s)
- alignment of all mapped reads from the specified experiment(s)
- consensus sequence for the alignment of all mapped reads from the specified experiment(s)

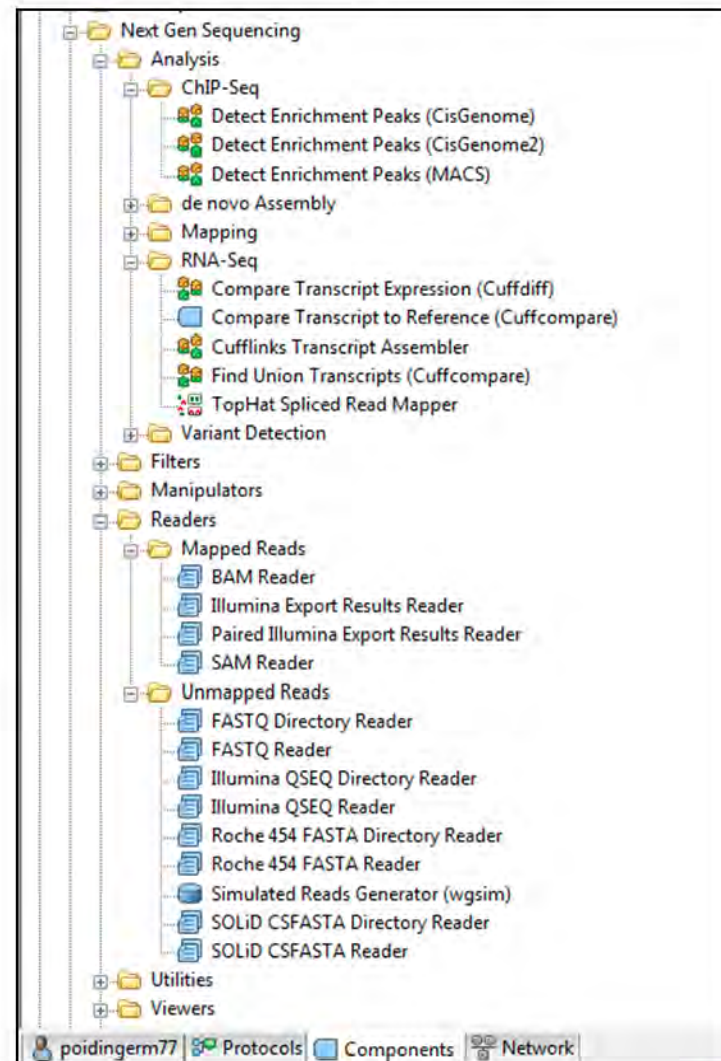
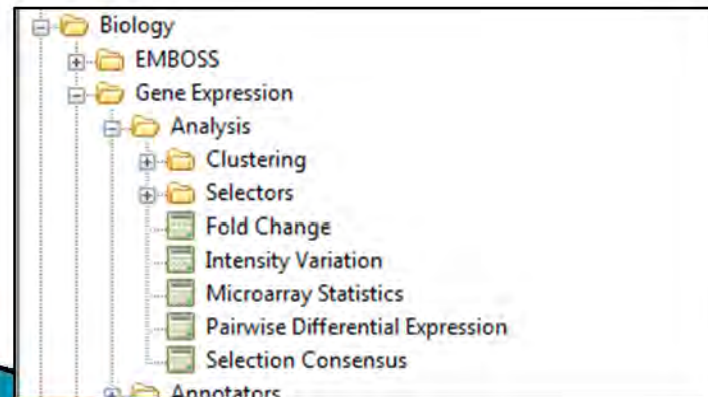
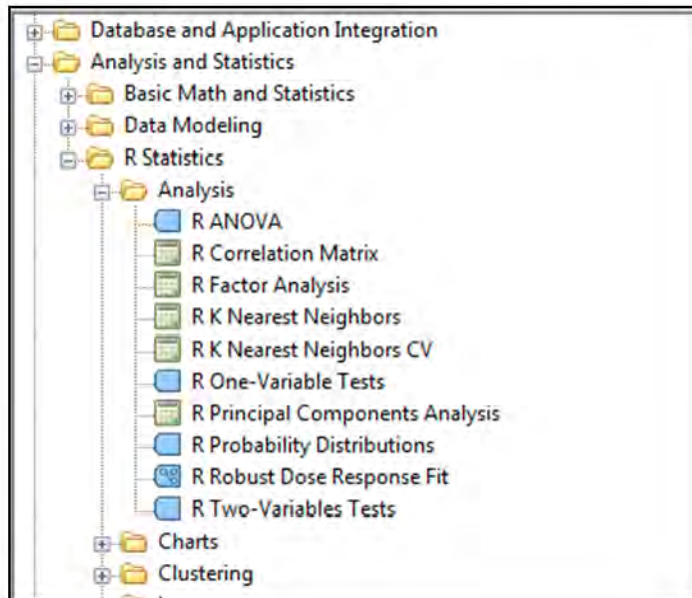
A region is defined by the three properties *referenceName*, *start*, and *end*. Data records with region parameters can be generated by *Region Generator* or *Region Splitter*.

The help window also includes a 'Parameters' table:

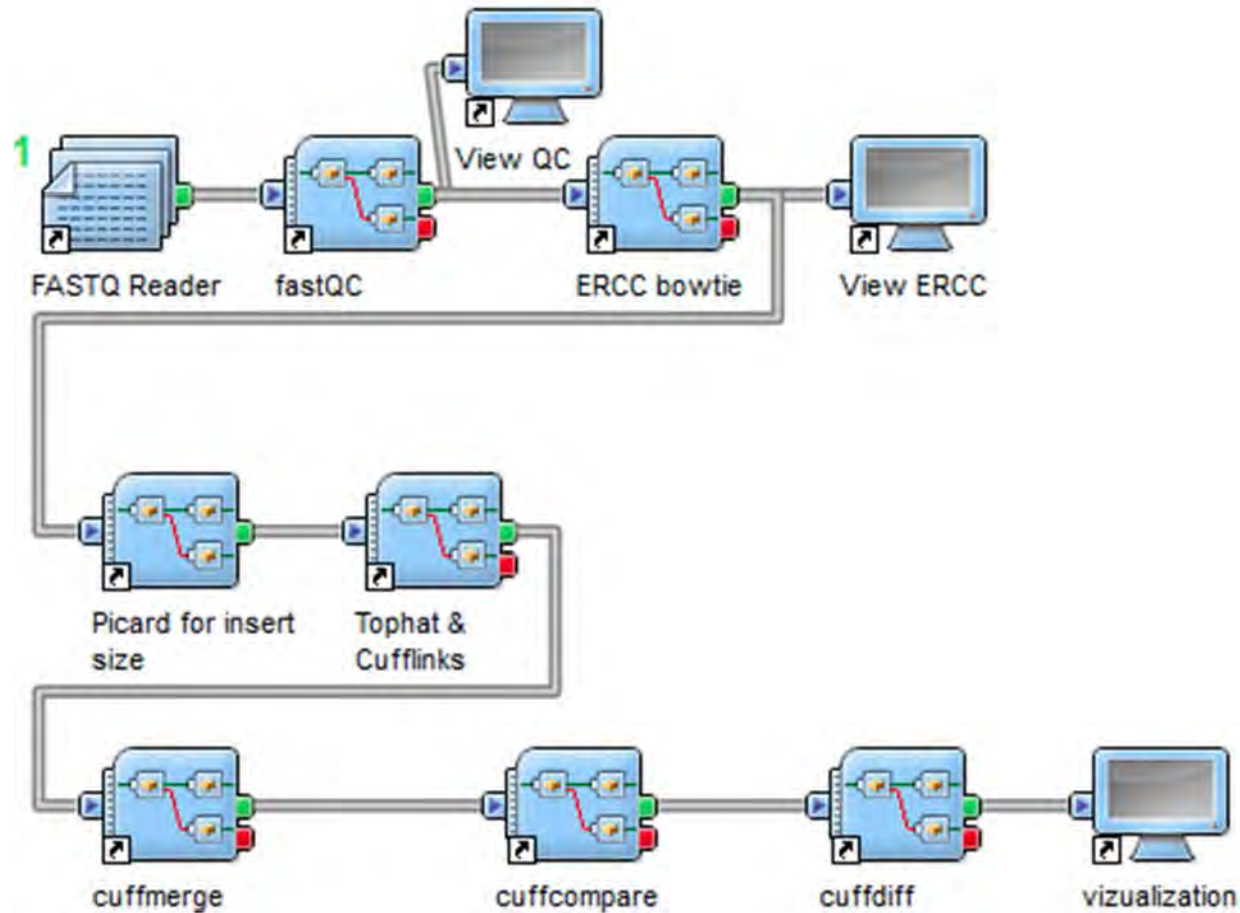
Parameters	
Repository Name	\$(Repository Name)
Scope	\$(Scope)
User Specified Directory	\$(User Specified Directory)
Region	
Reference	
Start	
End	
Output Type	
Feature Set	\$(Features for cuffdiff)
Parse GFF Attributes	True
Include Features	Overlapping

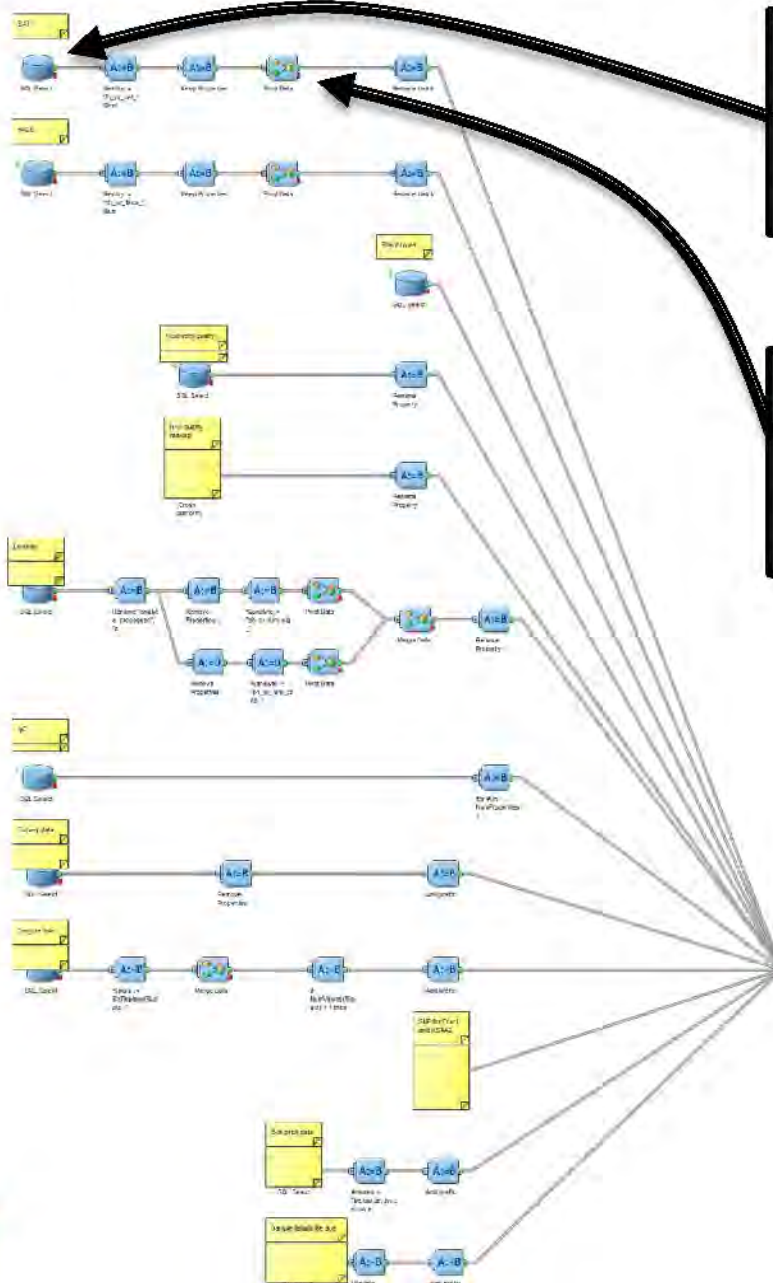
At the bottom of the help window, there are tabs for 'Parameters' and 'Implementation', and a 'Ready' status indicator.

Pipeline Pilot components



RNAseq Workflow





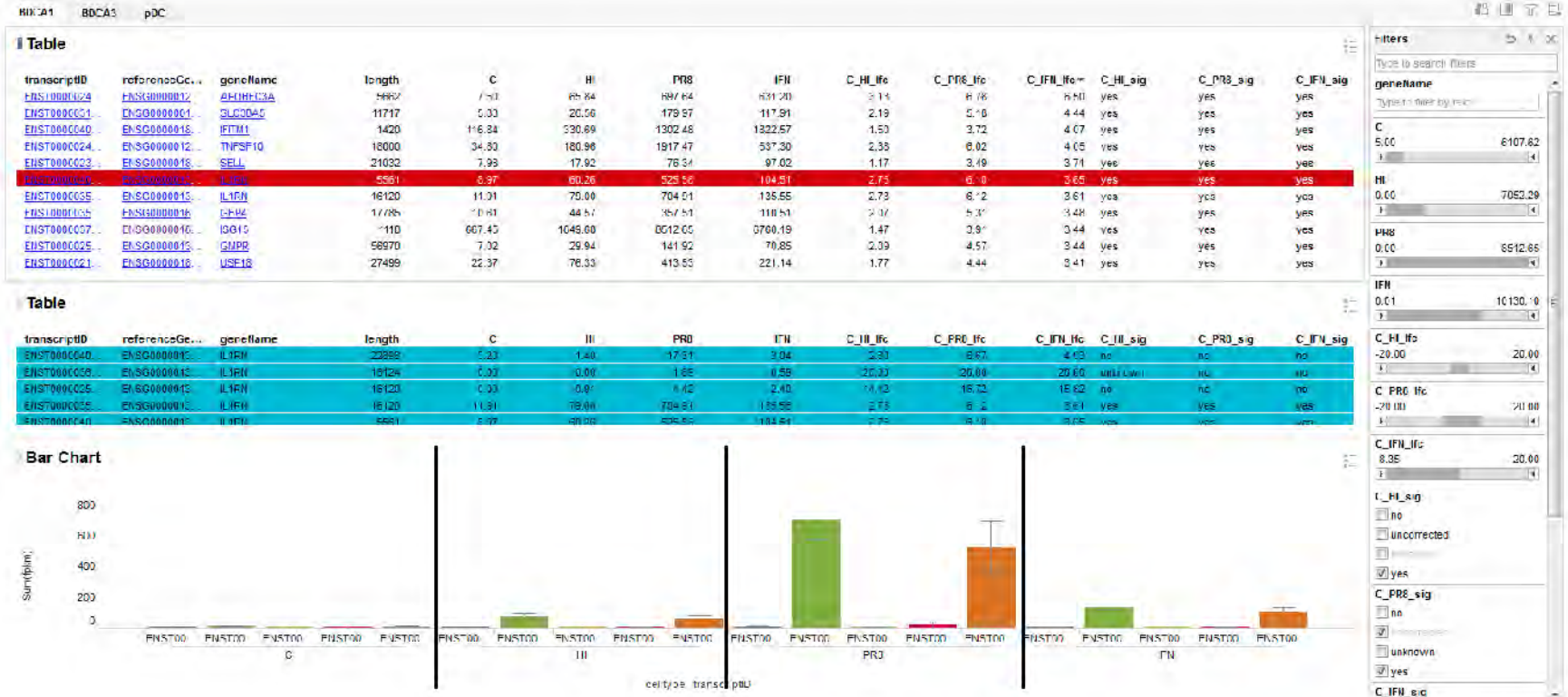
Parameters	
Data Source	cohort_august
SQL Statement	SELECT cohort_august.adrian_bat_data_normalized.cell,cohort_august.adrian_bat_...
Maximum	
Select Behavior	Multiple Data
Additional Options	
Separate Connection	False

Origins of data

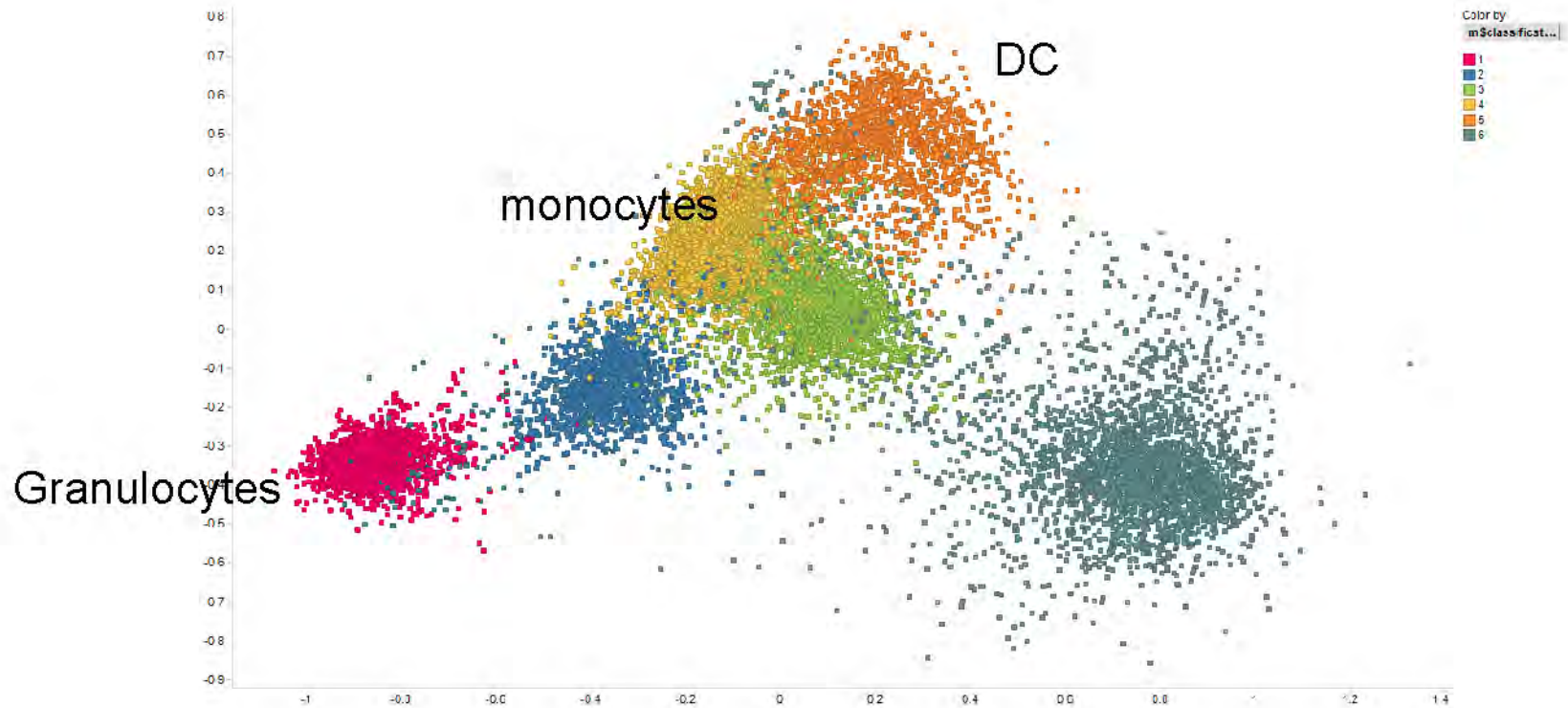
Parameters	
PivotName	entity
PivotValue	reading_normalized
MergeUsing	sample_batch
Data Are Sorted	False

Parameter values are captured

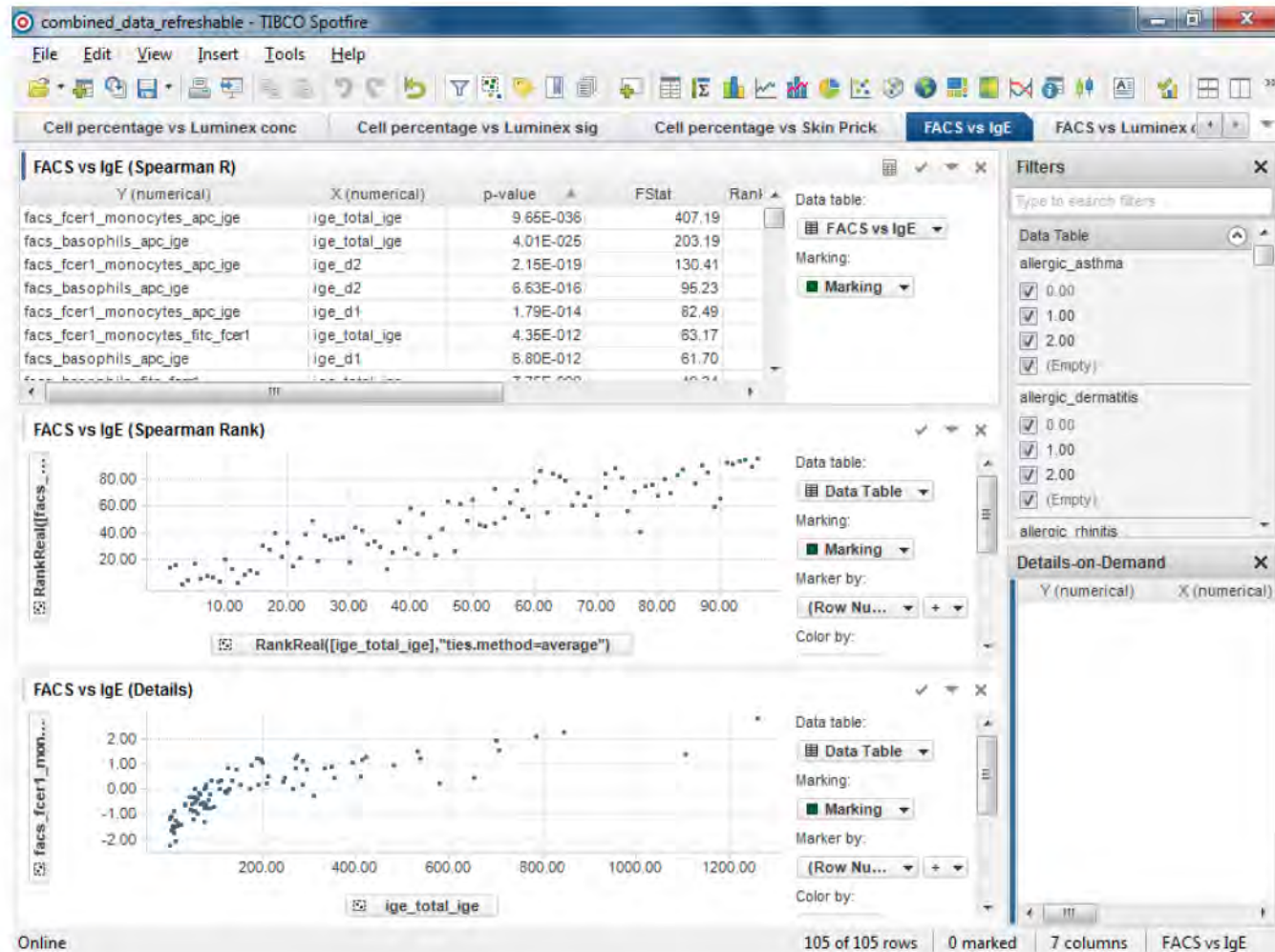
Spotfire – RNAseq



Spotfire – mass cytometry



Spotfire Pairwise correlations



Spotfire – Signature profiles



Spotfire – Frequency distribution

