

FlowJo Webinars: Machine Learning in FlowJo v10

Christian R. Aguilera-Sandoval, PhD
Field Applications Scientist, Informatics
Crbech.flowjo@bd.com

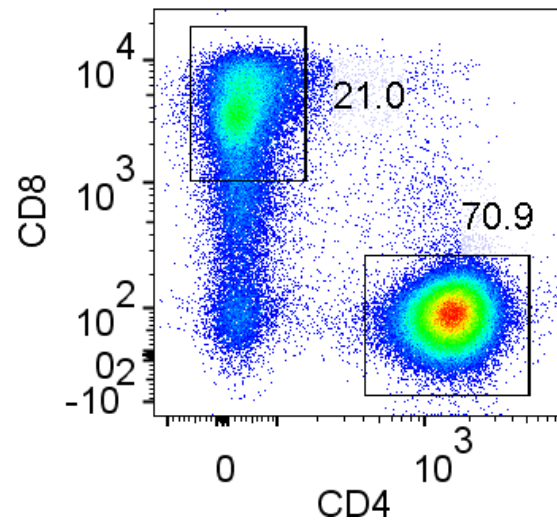
Jack Panopoulos, PhD
Field Applications Scientist, Informatics
Jack.Panopoulos@bd.com

Outline

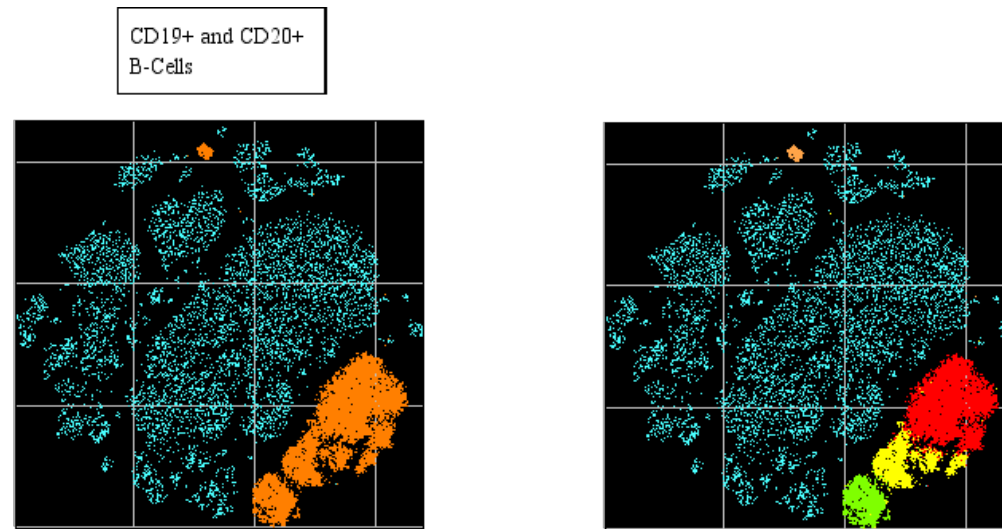
- Gating practices in flow cytometry
 - Manual Hierarchical gating
 - High-dimensionality reduction techniques and clustering algorithms
- Intro to Machine Learning
 - Supervised
 - Unsupervised
- Description of Dimreduc techniques and clustering algorithms
- Comparison of dimreduc techniques
- Comparison of clustering algorithms
- Which tool should I use and when?
- Live Demo of dimreduc techniques and clustering algorithms
- Coming Soon: Preview of unreleased plug-in
 - Cluster quality check

Current Gating Practices in Flow Cytometry

- Manual hierarchical gating



- High-dimensionality reduction and clustering



	Subset Name	Count
	Phenograph_96C5_29	3519
	Phenograph_96C5_16	9695
	Phenograph_96C5_6	485
	Phenograph_96C5_2	2986
	Ungated	113864

Manual Hierarchical Gating

- Manual hierarchical gating
 - Traditional gating using parent and child gates
 - Visual inspection of bivariate plots to identify cell populations
 - Problematic for high-dimensional settings
 - Subjectivity
 - Operator bias
 - Difficulty of identification for rare populations
 - Reproducibility
 - Populations could be overseen/ignored due to high parameter use

High-dimensionality Reduction and Clustering

- Pros
 - Efficient and time saving for high-dimensionality panels
 - Uncover multidimensional structures not seen in bivariate projections
 - Uses machine learning as the tool to identify cell phenotypes reducing bias, user error yet increasing reproducibility
- Cons:
 - Technical expertise
 - Advanced computer hardware
- Machine learning methods
 - Supervised
 - Unsupervised

Machine Learning - Supervised



Machine Learning - Unsupervised



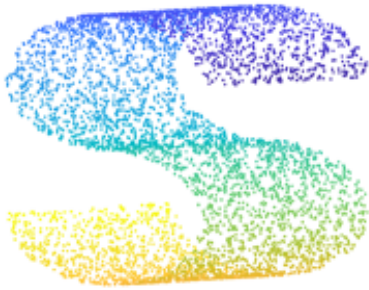
FlowJo's Machine Learning's Tools To Be Discussed Today

Method	Purpose	
tSNE ¹	DimRex	Method to visualize general structure of nonlinear data and assess a sample's heterogeneity through high resolution of local neighborhood structure
UMAP ²	DimRex	Method to quickly visualize the local and global structure of nonlinear data as a trade-off to resolution of local neighborhood structure and fidelity of global structure
TriMAP ³	DimRex	Method to visualize with high fidelity the global structure of nonlinear data
FlowSOM ⁴	Clustering	Self-organizing maps, followed by hierarchical consensus meta-clustering to merge clusters
Xshift ⁵	Clustering	Weighted k-nearest neighbor density estimation, detection of density centroids, cells linked by centroid via density-ascending paths
Phenograph ⁶	Clustering	Detection of k-nearest of neighbors of each cell, followed by partitioning of the map into clusters based on phenotype

¹= Maaten et al, 2008 ²= McInnes et al 2018 ³= Amid et al, 2019 ⁴= Van Gassen, 2015 ⁵=Samusik et al, 2016 ⁶= Levine et al, 2015

A Comparison of DimRex Algorithms

Original (AUC, GS)



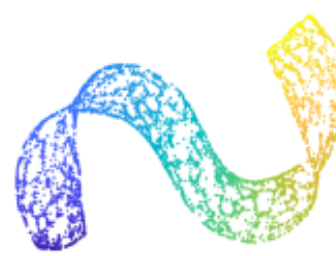
t-SNE (**0.18**, 0.18)



UMAP (0.16, 0.13)



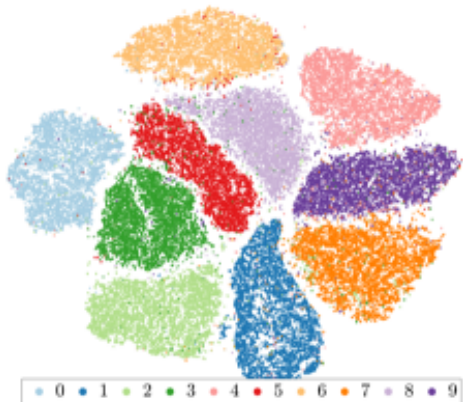
TriMap (0.15, **0.80**)



PCA (0.03, 1.00)



MNIST - t-SNE (0.97, 0.90)



MNIST - UMAP (0.95, 0.91)



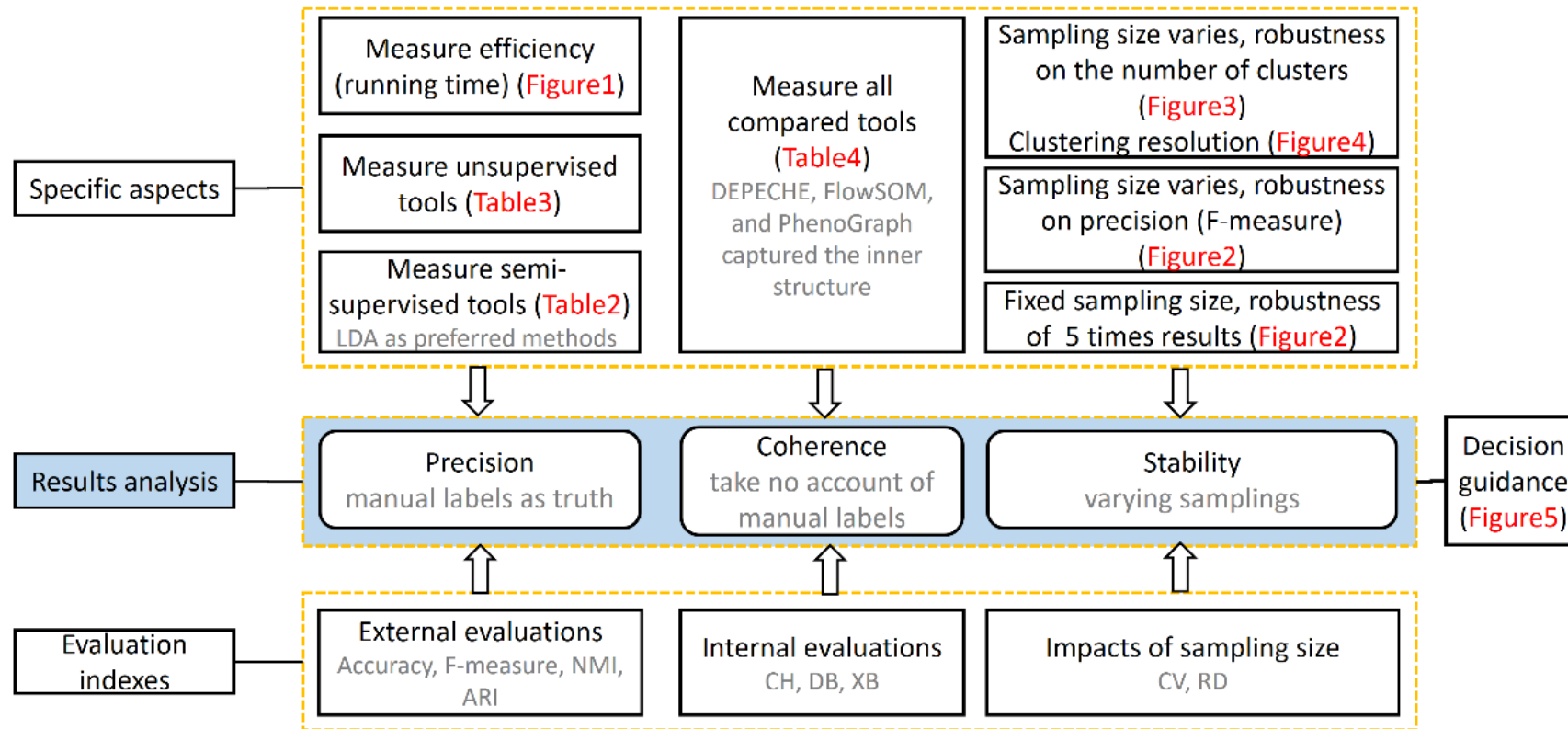
MNIST - TriMap (0.94, 0.92)



MNIST - PCA (0.38, 1.00)



Study Flowchart



LDA Is Superior to ACDC In Precision*

Table 2 Summary of external evaluations for semi-supervised methods

Datasets	Methods	External evaluations			
		Accuracy	<i>F</i> -measure	NMI	ARI
Cell Cycle	ACDC	0.8342 \pm 0.0071	0.8466 \pm 0.0093	0.4325 \pm 0.0212	0.5579 \pm 0.0129
	LDA	0.9095 \pm 0.0006	0.9110 \pm 0.0005	0.6189 \pm 0.0032	0.7225 \pm 0.0021
Colon	ACDC	0.7439 \pm 0.0026	0.7874 \pm 0.0076	0.5705 \pm 0.0088	0.5952 \pm 0.0041
	LDA	0.8576 \pm 0.0011	0.8587 \pm 0.0012	0.7410 \pm 0.0012	0.7626 \pm 0.0017
Levine13dim	ACDC	0.9010 \pm 0.0029	0.9275 \pm 0.0026	0.8635 \pm 0.0041	0.9011 \pm 0.0052
	LDA	0.9582 \pm 0.0005	0.9586 \pm 0.0005	0.9275 \pm 0.0008	0.9539 \pm 0.0007
Levine32dim	ACDC	0.9943 \pm 0.0006	0.9939 \pm 0.0007	0.9380 \pm 0.0052	0.9791 \pm 0.0020
	LDA	0.9809 \pm 0.0003	0.9807 \pm 0.0004	0.9595 \pm 0.0006	0.9830 \pm 0.0002
Muscle	ACDC	0.8787 \pm 0.0101	0.8784 \pm 0.0089	0.6750 \pm 0.0168	0.7593 \pm 0.0190
	LDA	0.9240 \pm 0.0011	0.9238 \pm 0.0011	0.7606 \pm 0.0031	0.8295 \pm 0.0031
Samusik01	ACDC	0.9682 \pm 0.0027	0.9731 \pm 0.0019	0.9347 \pm 0.0047	0.9616 \pm 0.0021
	LDA	0.9757 \pm 0.0002	0.9759 \pm 0.0002	0.9482 \pm 0.0004	0.9735 \pm 0.0005

Data shown as mean \pm standard deviation

Unsupervised Methods Outperform Semisupervised Methods In Coherence

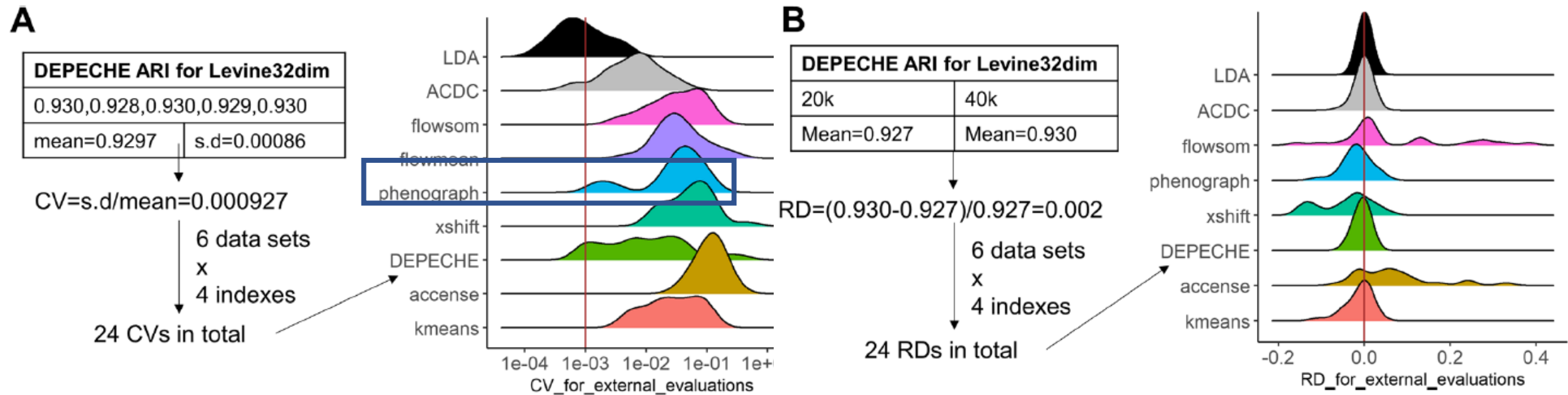
Table 4 Summary of internal evaluations for each compared methods

Datasets	Methods	Internal evaluations		
		CH	DB	XB
Levine13dim	Accense	3.4230 ± 0.0932	1.8832 ± 0.1408	1.2321 ± 0.0278
	PhenoGraph	4.0739 ± 0.0176	1.4645 ± 0.0346	1.3972 ± 0.1349
	Xshift	3.5106 ± 0.0289	2.4284 ± 0.0443	1.7868 ± 0.0476
	kmeans	3.8508 ± 0.0150	2.1550 ± 0.0546	1.6213 ± 0.1471
	flowMeans	4.0475 ± 0.0194	1.5030 ± 0.0849	1.4234 ± 0.1182
	FlowSOM	3.8486 ± 0.0071	1.7564 ± 0.0615	1.5043 ± 0.1531
	DEPECHE	4.2783 ± 0.0174	1.1677 ± 0.0342	1.3562 ± 0.0392
	ACDC	3.9638 ± 0.0110	1.4916 ± 0.0370	1.3109 ± 0.0948
	LDA	3.8288 ± 0.0106	2.0046 ± 0.0493	1.3828 ± 0.1167
Levine32dim	Accense	3.4621 ± 0.0901	2.3414 ± 0.0925	0.7891 ± 0.0950
	PhenoGraph	3.7401 ± 0.0081	1.8293 ± 0.0810	1.0009 ± 0.0479
	Xshift	3.6669 ± 0.0102	2.2576 ± 0.1324	0.8295 ± 0.1382
	kmeans	3.8761 ± 0.0166	2.0587 ± 0.0386	0.9972 ± 0.0441
	flowMeans	3.8546 ± 0.0393	1.6975 ± 0.2199	0.7985 ± 0.0709
	FlowSOM	3.8244 ± 0.0285	1.5974 ± 0.0863	0.8366 ± 0.0792
	DEPECHE	4.1480 ± 0.0009	1.4727 ± 0.0023	0.7575 ± 0.0351
	ACDC	3.6169 ± 0.0046	1.3974 ± 0.0049	0.7693 ± 0.1310
	LDA	3.8297 ± 0.0007	1.7011 ± 0.0099	0.7155 ± 0.0139

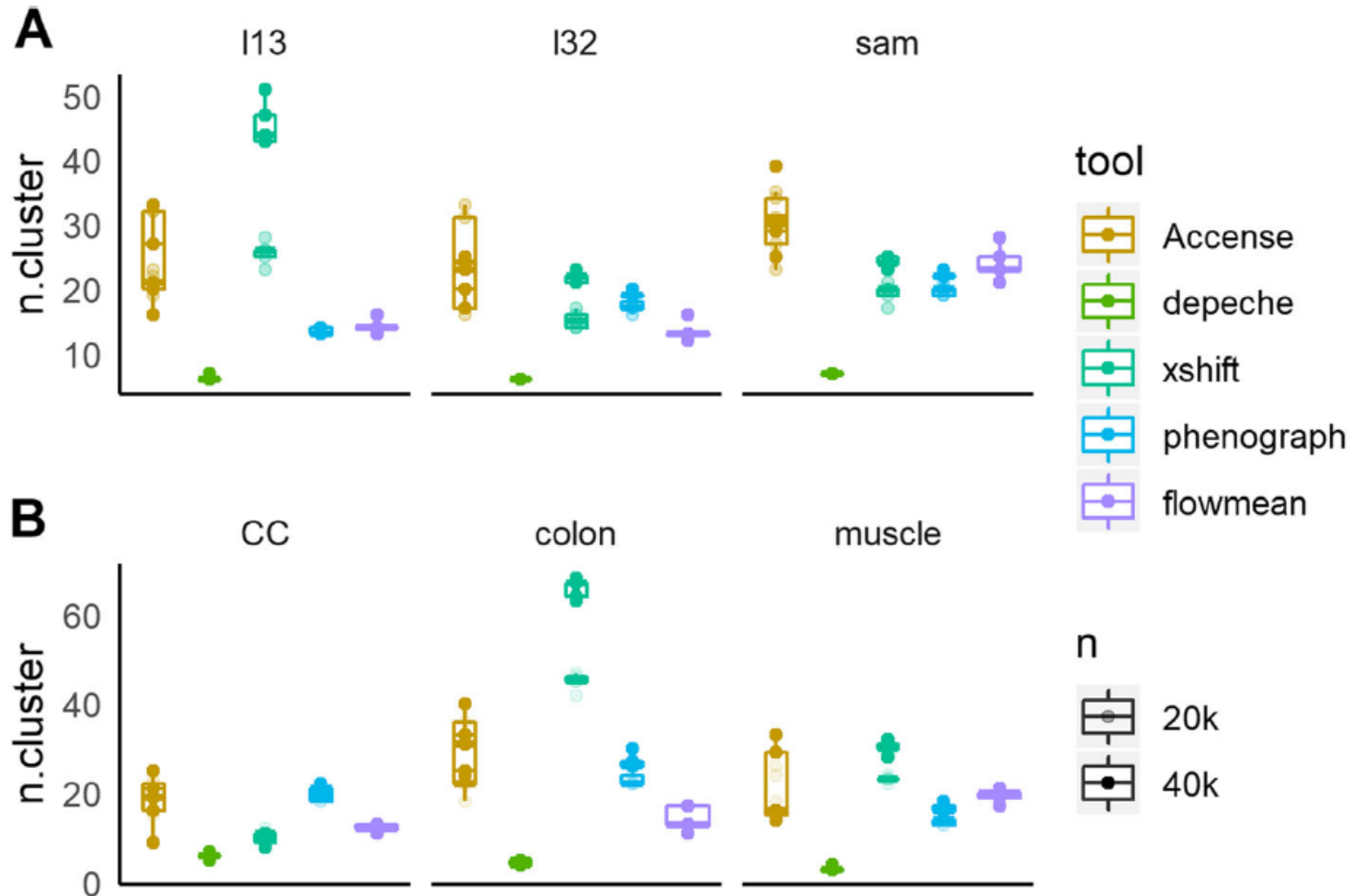
Best Unsupervised Clustering Methods for Inner Structure

- FlowSOM
- Phenograph
- DEPECHE

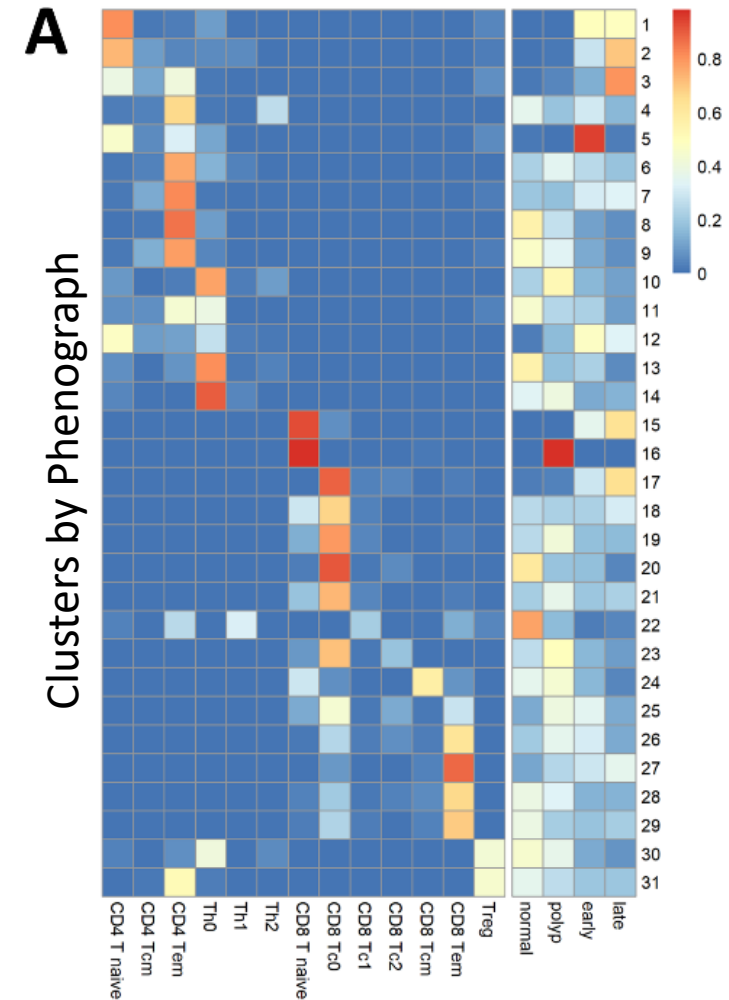
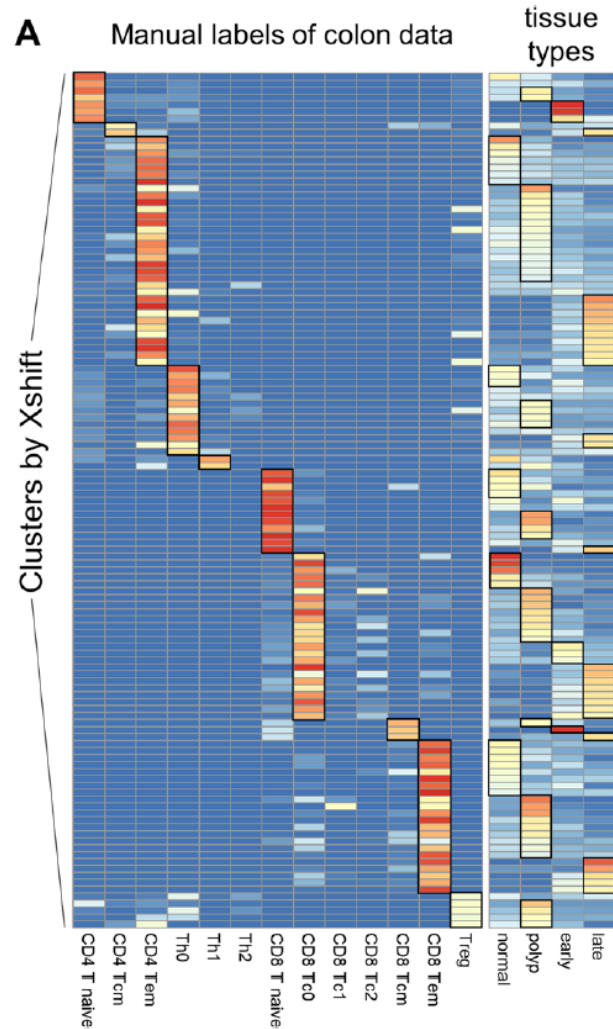
Phenograph and LDA Are The Most Robust Methods With Repetitive Tests or Varying Sample Sizes

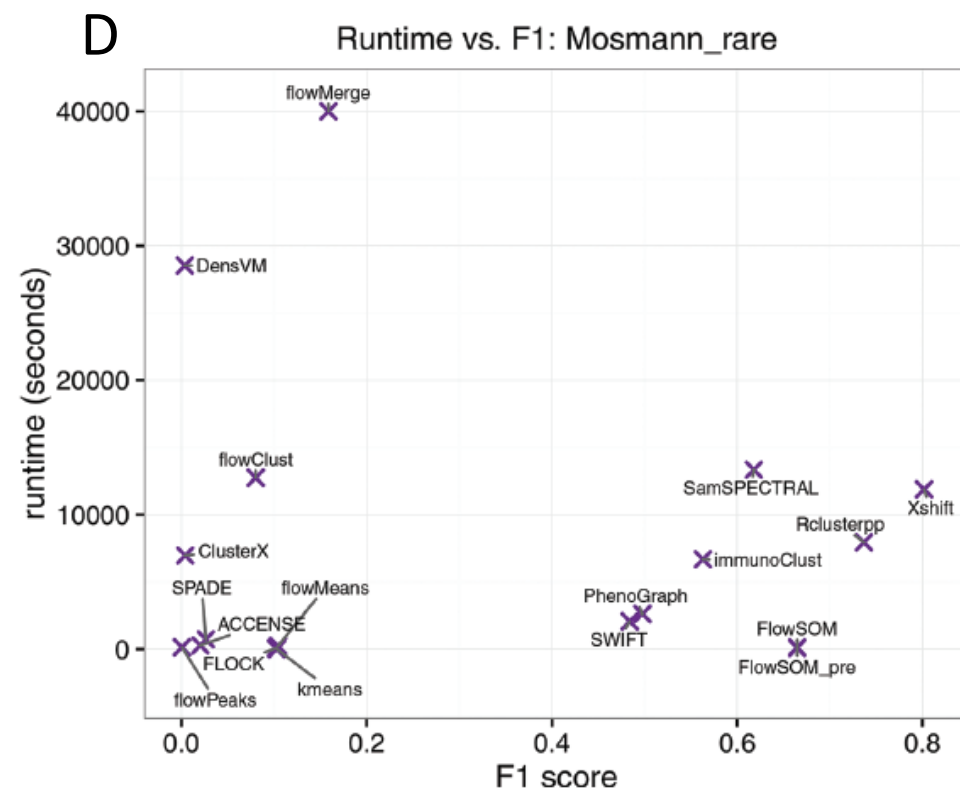
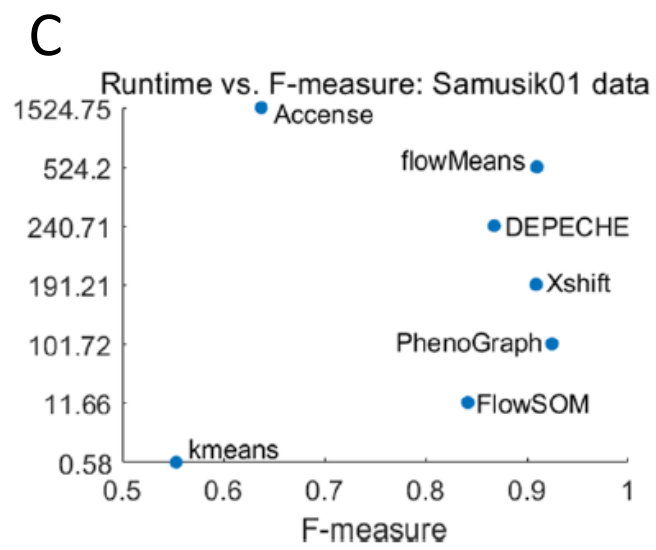
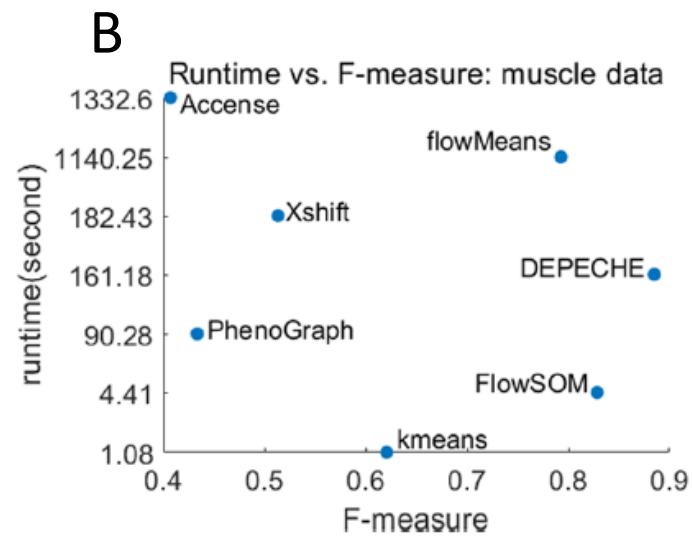
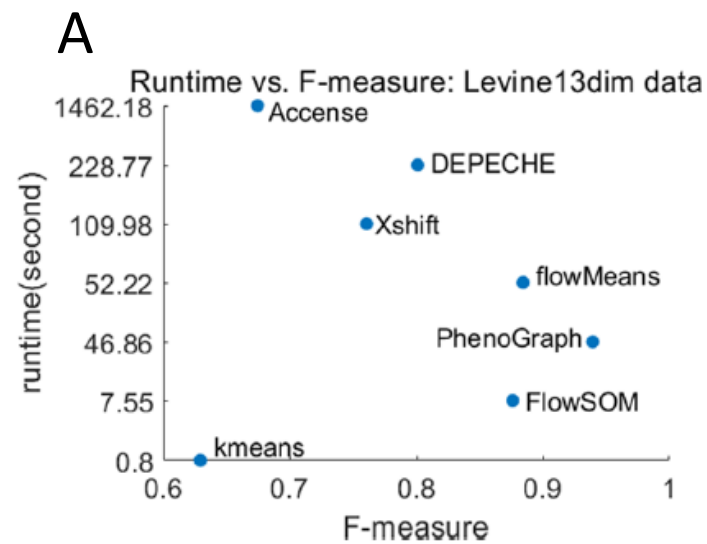


Phenograph Detects More Clusters Consistently In Varying Sample Sizes

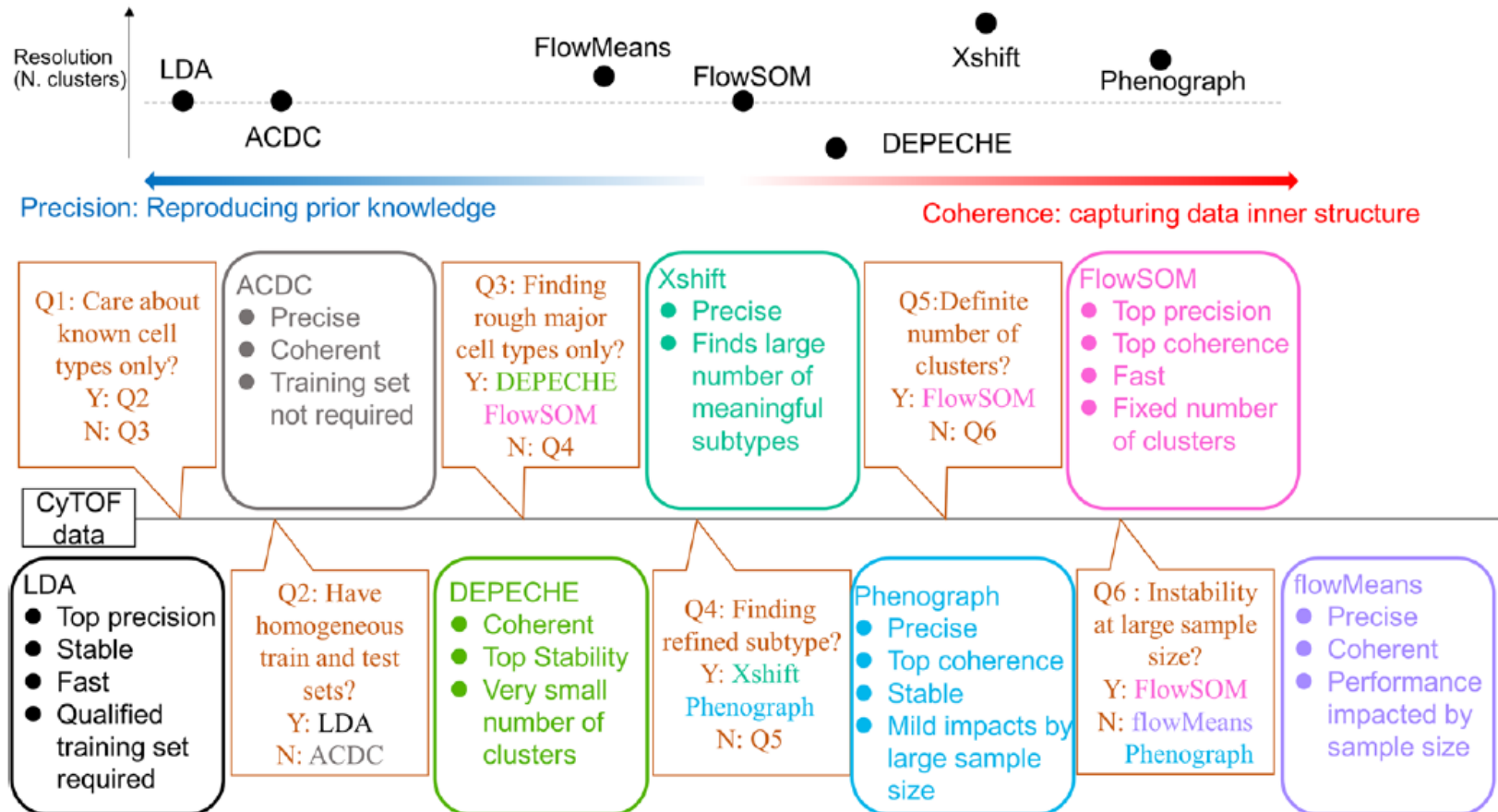


Xshift And Phenograph Best Identify Subsets Of Major Cell Types





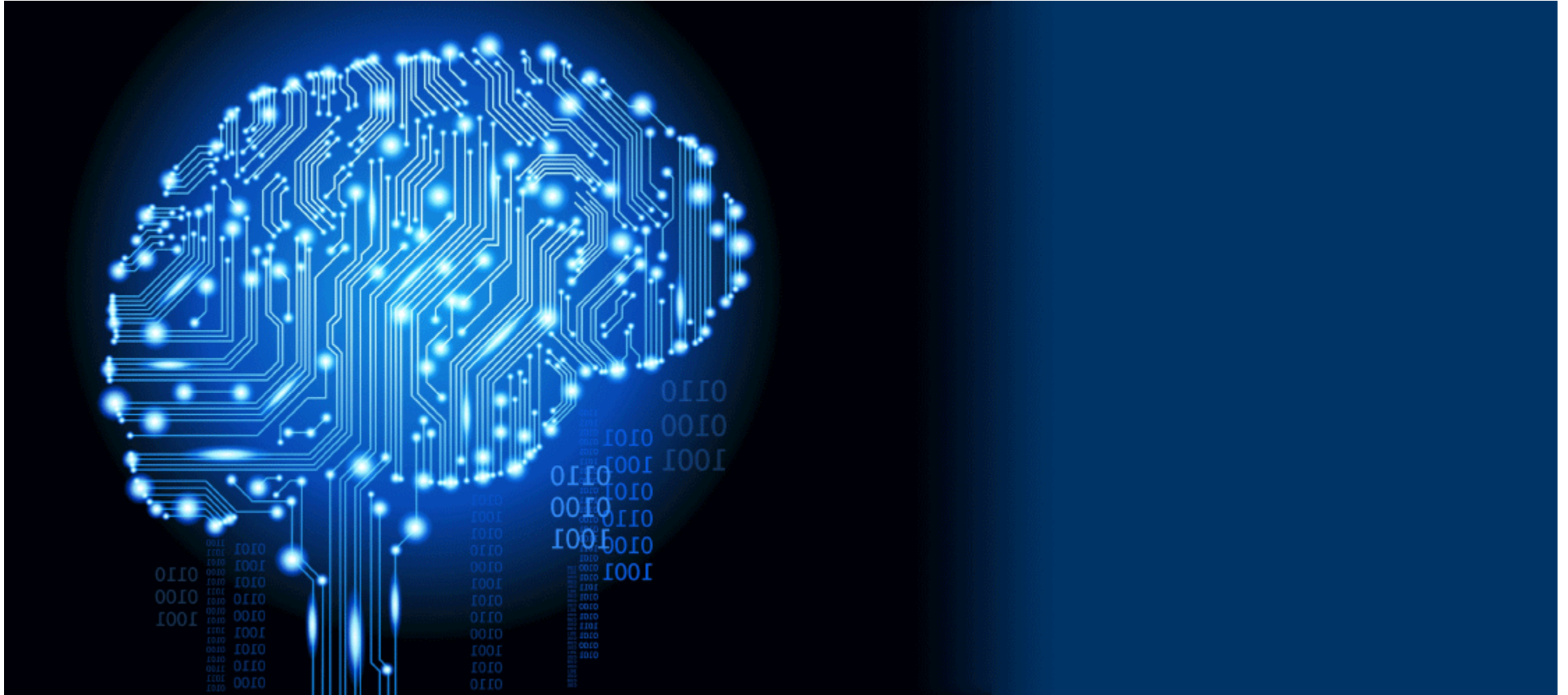
The Right Tool For The Right Job



Conclusions

- Top unsupervised tools:
 - FlowSOM and Phenograph
 - Precision
 - Coherence
 - Stability
- Top tools to detect subsets:
 - Phenograph and Xshift (using Elbow Plot Determination)
- FlowSOM best tool when dealing with very large datasets

Next: Live Demo



Thank You!!!

Christian R. Aguilera-Sandoval, PhD

Crbech.flowjo@bd.com

Serena Di Cecilia, PhD

Serena.Di.Cecilia@bd.com

Jack Panopoulos, PhD

Jack.Panopoulos@bd.com

Tech Support

techsupport@flowjo.com