Simultaneous analysis of protein and transcripts

Combined single cell sorting and deep sequencing reveals ILC plasticity

Summary

Single-cell RNA-Sequencing (scRNA-Seq) is a new and incredibly powerful technique for elucidating the heterogeneity and function of biological systems, with unprecedented resolution. However, current software tools used for discovery of transcriptional profiles are assay-specific, require coding expertise, or facilitate only a limited depth of analysis. Herein, we detail a meta-analysis combining surface protein and transcript data, to discover expression profiles between cell types.

Innate Lymphoid Cells (ILCs) are part of the early immune response, and have a role in homeostasis and inflammation. Recent studied have called into question the definition of ILC subpopulations and their heterogeneity.[2] We show here that ILC populations are much more plastic than previously appreciated.

Experimental Methods Summary

Raw data was generously provided by collaborators, Björklund et al, of the Karolinska Institute, who were interested in expanding the current understanding of ILC's role in gut health.[3]

Human ILCs from adenoid tissues (helper groups 1-3 and NK) were index sorted using conventional FACS, and sequenced using the SMART-Seq2 protocol. Generating >64,000 gene transcript parameters.

Analysis

Flow cytometry analysis was performed in FlowJo® followed by combination and meta-analysis in SeqGeq™. Sequencing data was normalized using the ComBat protocol, in order to reduce batch effects from different runs.[4] Using the cell ID columns in common with both FACS and Sequencing data-sets, data files were merged to create a single, CSV data file with flow cytometry and single cell gene expression data. That data file was then imported for analysis into SeqGeq.

Flow Cytometry Analysis

Gating described by Mjoesberg et al was performed to obtain ILC subsets, using FACS parameters, as seen in Figure 1.[5]

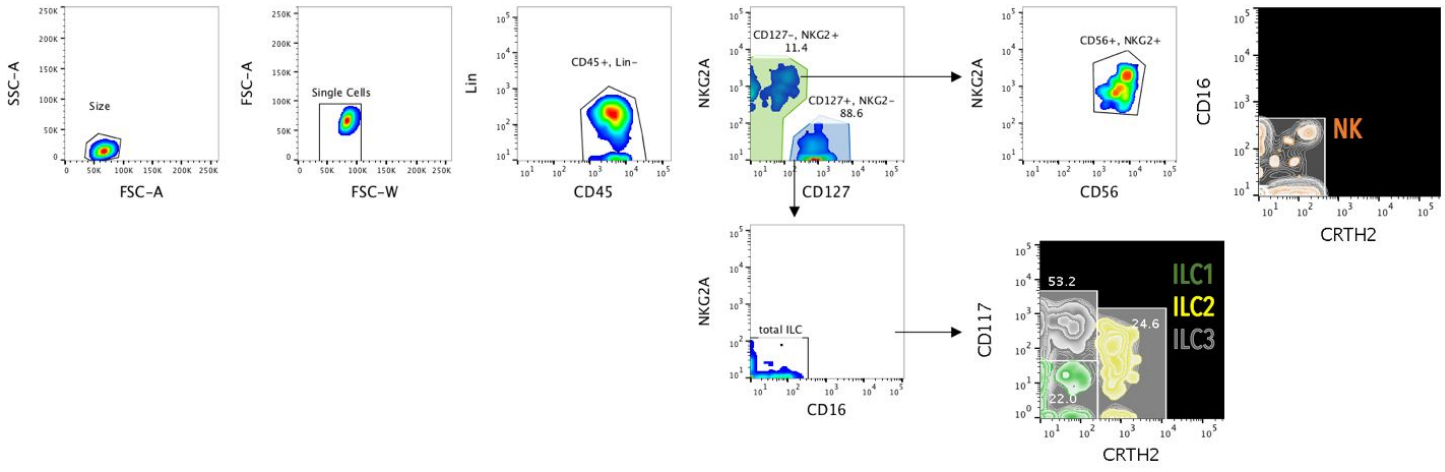

Figure 1 - Gating ILCs by phenotype.

Tech-Note: If you have a CSV file containing new parameter column and a "cell ID" column, which matches the cell ID column in your scRNA-Seq data file, you can simply drag and drop that CSV file onto your corresponding data in SeqGeq to *automatically* join or "stitch" these data matrices together.

Transcript Analysis

During scRNA-Seq protocol, much of the true transcription signal is obscured in the process of amplification. In order to find gene parameters expressed significantly above this noise, a statistical technique developed by Brennecke et al, was employed in R.[6] This yielded 847 genes said to be differentially expressed above statistical noise.

Tech-Note: Given the number of parameters generated in scRNA-Seq experiments, dimensionality reduction is a key tool for characterizing and discovering cell populations.[8] SeqGeq has comprehensive suite of dimensionality reduction tools, the results of which can be explored and gene sets extracted.

Dimensionality reduction, using principal component analysis (PCA) was performed on those 847 parameters. Principle components were then used to generate clusters in t-SNE[7] parameters, and overlaid with phenotypes defined by flow cytometry (Figure 2). More than 10% of events were found to lie outside clusters predicted by FACS analysis, implying functional plasticity among these subsets, as their phenotype did not strictly match their transcriptional profile.

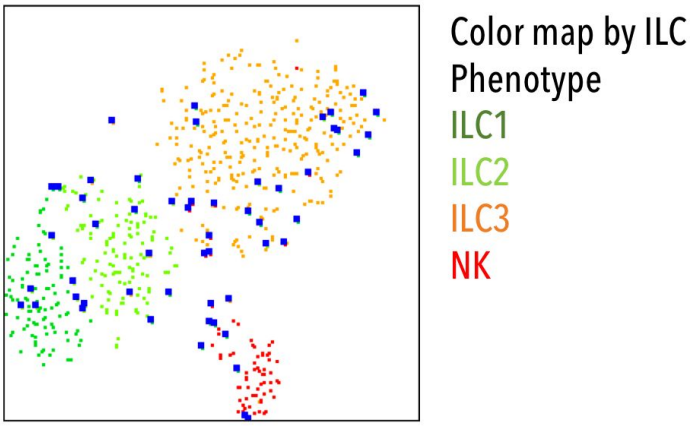

Figure 2 - t-SNE clustering mapped by surface protein phenotype. Cells whose transcriptional profile is discordant with cluster predicted by phenotype are highlighted in blue.

Differential Expression

A pivot analysis was used to assess differential expression of genes from each phenotype as defined in FACS, with respect to the rest of the data-set in Figure 3a.

Tech-Note: The pivot function in SeqGeq allows researchers to quickly investigate normalized expression of all genes in a data matrix with respect to two populations, viewed on the x and y axes. These genes can be gated, creating new gene sets, which can be investigated further.

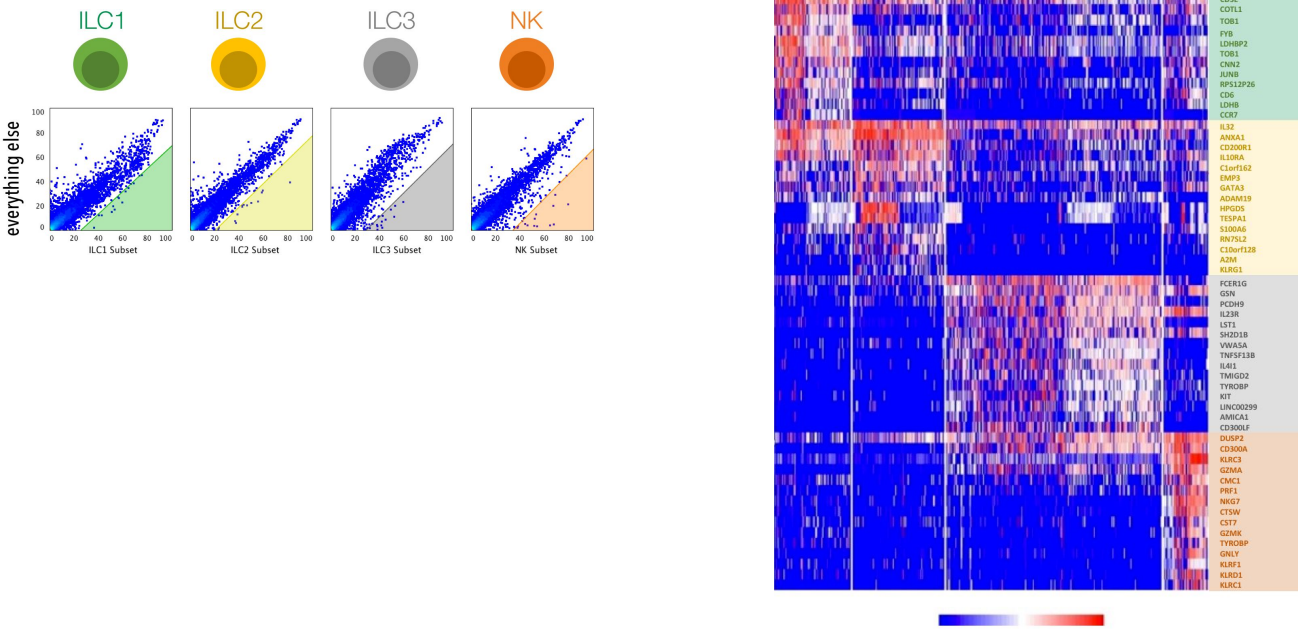

Figure 3 (a) Gating of the top 15 most differentially expressed genes for each ILC phenotype. (b) ILC subsets were heatmapped versus these gene-sets in order to confirm their significance in figure 3-b.

Characterization

As a final piece in the analysis process, these genes of interest were investigated individually in the available literature, with a particular interest in their association with our phenotypes of interest, and known or undescribed immune functions. Illustrated in Figure 4.

reported
unreported
unreported immune genes

ILC1	ILC2	ILC3	NK
CCR7	GATA3	KIT	CD300A
COTL1	IL10RA	FCER1G	KLRC3
RP11-466H18.1	ANXA1	IL23R	GZMA
RP11-613F7.1	HPGDS	LST1	PRF1
LDHBP2	CD200R1	SH2D1B	NKG7
TOB1	A2M	TMIGD2	CTSW
CNN2	EMP3	TYROBP	CST7
RPS12P26	C1orf162	AMICA1	GZMK
LDHB	EMP3	CD300LF	TYROBP
RP11-864N7.2	ADAM19	GSN	GNLY
TC2N	TESPA1	PCDH9	KLRF1
FAIM3	S100A6	VWA5A	KLRD1
CD3E	C10orf128	TNFSF13B	KLRC1
JUNB	KLRG1	LINC00299	DUSP2
CD6	IL32	IL4I1	CMC1

Figure 4. ILC Gene Sets. Showing genes differentially expressed in each indicated cell population, categorized by reported in literature, unreported, and unreported immune genes.

Discussion

Genes differentially expressed in each of the ILC phenotypes open the door to future research topics. Notably, we uncovered genes corresponding to surface receptors not yet characterized in ILC subsets, and those not yet associated with immune functions.

Some unexpected immune genes found in group 1 ILCs are characteristic of helper T-cells, and further investigation of these cells is warranted, given a current controversy regarding whether ILC1 are in fact T-cells.[2] In a broader sense, this illustrates the power of single cell RNA sequencing.

SeqGeq is a powerful tool in the hands of any researcher, to quickly and intuitively explore single cell gene expression data.

Resources

Sign up for a free 30 day trial of SeqGeq: www.flowjo.com/solutions/seqgeq/free-trial

Access SeqGeq software here: www.flowjo.com/solutions/seqgeq

This entire dataset is publicly available: tinyurl.com/taylorilcs

SeqGeq documentation: docs.flowjo.com/seqgeq

If you have questions about this software please write to SeqGeq technical support: seqgeq@flowjo.com

References

1. Neill, Flynn. *Origins and evolution of innate lymphoid cells: Wardens of barrier immunity*. Parasite Immunol. Apr 19, 2017
2. Simoni, Fehlings, Kloverpris et al. *Human Innate Lymphoid Cell Subsets Possess Tissue-Type Based Heterogeneity in Phenotype and Frequency*. Immunity. Jan 17, 2017
3. Björklund, Forkel, Picelli, et al. *The heterogeneity of human CD127+ innate lymphoid cells revealed by single-cell RNA sequencing*. Nat. Immunol. Mar 17, 2016
4. Johnson, Rabinovic, Li. *Adjusting batch effects in microarray expression data using Empirical Bayes methods*. Biostatistics. 2007.
5. Mjoesberg, Trifari, Crellin, et al. *Human IL-25- and IL-33-responsive type 2 innate lymphoid cells are de ned by expression of CRTH2 and CD161*. Nat. Immunol. Sept 11, 2011
6. Brennecke, Anders, Kim, et al. *Accounting for technical noise in single-cell RNA-seq experiments*. Nat. Methods. Sept 22, 2013.
7. Van der Maaten. *Accelerating t-SNE using Tree-Based Algorithms*. Journal of Machine Learning Research. Oct 15, 2014.
8. Newell, Cheng. *Mass cytometry: blessed with the curse of dimensionality*. Nat Immunol. Jul 19, 2016
9. Taylor, Velazquez-Palafox, Halpert et al. *Splice of Life: Deep Single Cell RNA-Sequencing of Human Innate Lymphoid Cells Reveals Functional Plasticity*. CYTO Oral Presentation. June 12, 2017

FlowJo, LLC • USA: 800-366-6045 • International: 1-541-201-0022 • support@flowjo.com • www.flowjo.com