

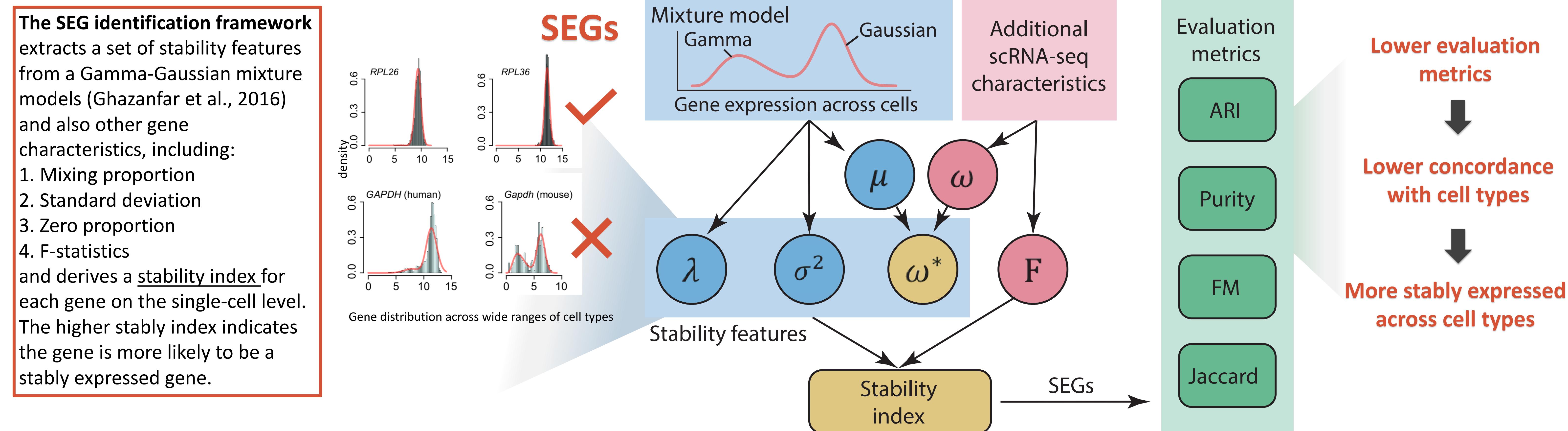


Evaluating stably expressed genes in single cells

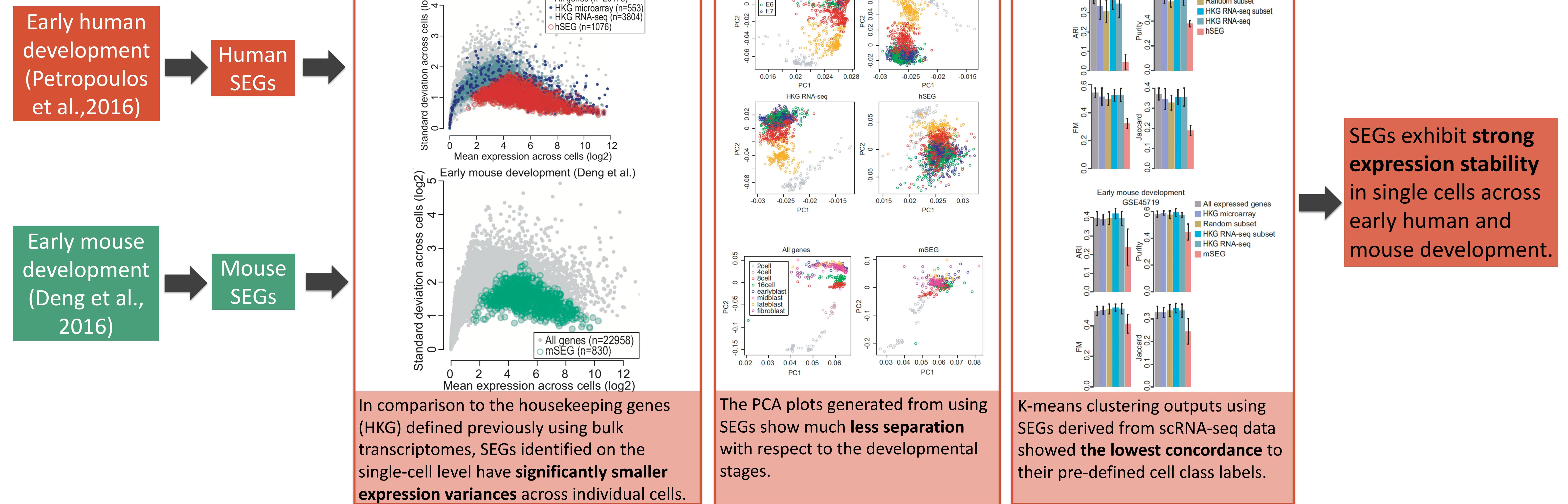
Yingxin Lin¹, Shila Ghazanfar^{1,2}, Dario Strbenac¹, Andy Yi-Yang Wang¹, Ellis Patrick^{1,3}, Dave M. Lin⁴, Terence P. Speed⁵, Jean Yee Hwa Yang^{1,2*}, Pengyi Yang^{1,2*}

¹School of Mathematics and Statistics, ²Charles Perkins Centre, ³Westmead Institute for Medical Research, University of Sydney, ⁴Department of Biomedical Sciences, Cornell University, ⁵Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, *Corresponding authors

Stably expressed genes (SEGs) identification and evaluation framework



Characterisation of stably expressed genes



Benchmark results using independent scRNA-seq datasets

Benchmark datasets					Peripheral blood mononuclear cells (Villani et al. 2017)				hPSCs and endoderm progenitors (Chu et al. 2016)			
ID	Publication	Description	Organism	# cell	All genes	h-bHK microarray	h-bHK RNA-seq	h-scHK	All genes	h-bHK microarray	h-bHK RNA-seq	h-scHK
GSE94820	(Villani et al. 2017)	Peripheral blood mononuclear cells	Human	1140	55±8	42±3	38±4	29±6	69±5	58±5	55±6	41±3
GSE75748	(Chu et al. 2016)	Pluripotent stem cells and endoderm progenitors	Human	1018	69±7	62±2	59±1	52±5	80±4	74±3	71±5	59±3
GSE72056	(Tirosh et al. 2016)	Multicellular metastatic melanoma	Human	4645	67±5	56±1	52±3	45±4	75±4	66±4	63±5	51±2
GSE67835	(Darmanis et al. 2015)	Adult and fetal brain	Human	466	49±6	39±1	35±2	29±4	60±5	48±4	46±6	34±2
GSE60361	(Zeisel et al. 2015)	Mouse cortex and hippocampus	Mouse	3005	31±5	18±2	18±1	15±1	53±7	50±3	39±4	36±3
GSE52583	(Treutlein et al. 2014)	Developmental lung epithelial cells	Mouse	198	80±5	73±1	74±1	71±1	82±3	76±4	74±3	68±2
E-MTAB-4079	(Scialdone et al. 2016)	Mesoderm diversification	Mouse	1205	51±3	39±2	40±1	37±1	62±6	59±2	50±3	47±3
GSE84133	(Baron et al. 2016)	Pancreas inter- and intra-cells	Mouse	822	32±2	22±2	24±1	21±1	44±6	41±2	33±3	30±3

SEGs maintain expression stability in drastically different tissues and biological systems.

Conclusion

SEGs identified in this study have immediate utility both for understanding variation and stability of single-cell transcriptomes and for practical applications such as scRNA-seq data normalization, such as, scMerge (Lin et al., 2018). The proposed computational framework can be applied to identify genes with stable expressions in other scRNA-seq datasets. By modelling single-cell transcriptomes, our approach uncovers and highlights the previously unappreciated gene stability at the single-cell level.

Further information

The SEGs Shiny app is available at <http://shiny.maths.usyd.edu.au/SEGs>; The codes for SEGs identification are available at <https://github.com/PengyiYang/Single-cell-housekeeping-genes>. Reference: Ghazanfar et al. BMC Systems Biology 2016, 10(Suppl 5):127 doi: 10.1186/s12918-016-0370-4 Lin et al. BioRxiv 2017, doi: <https://doi.org/10.1101/229815> Lin et al. BioRxiv 2018, doi: <https://doi.org/10.1101/393280> Contact: yingxin.lin@sydney.edu.au / pengyi.yang@sydney.edu.au / jean.yang@sydney.edu.au