Supplementary Results

Variable expression of secretory lineage transcription factors in intestinal epithelial stem cells

Having validated VarID by recovering variable expression of lineage-biased gene networks in hematopoietic progenitors, I wanted to test my method by applying it to the murine intestinal epithelium[1], where lineage-biased states within the stem cell compartment have not yet been characterized. Absorptive enterocytes and secretory lineages comprising antimicrobial Paneth cells, mucus producing goblet cells, lumen-sensing tuft cells, and hormone producing enteroendocrine cells differentiate from an Lgr5+ intestinal stem cell[2]. The intestinal epithelium undergoes constant turnover, where stem cells, residing at the bottom of the intestinal crypt, migrate upward along the crypt-villus axis. In this process they become transiently amplifying and eventually commit to one of the different lineages. Intestinal stem cells are believed to be a homogenous population without apparent lineage-biased sub-types[1,3]. I ran VarID on intestinal epithelial single-cell RNA-seq data from a recent study[1] and recovered a number of cell types, comprising *Lgr5*+ intestinal stem cells, *Krt19*+ cells, *Dclk1*+ tuft cells, *Muc2*+/*Agr2*+/*Zg16*+ goblet cells, *Chgb*+/*Sst*+/*Gcg*+ enteroendocrine cells, and *Sox4*+/*Dll1*+ secretory progenitors (Supplementary Fig. 6a,b). I did not identify mature Paneth cells although secretory progenitors in cluster 8 show limited upregulation of Paneth cell markers, e.g. *Defa24* and *Lyz1*.

Louvain clustering on the full knn network and application of Seurat gave very similar results and both approaches failed to resolve a number of populations, e.g., the different enteroendocrine subtypes or the *Zg16*+ and *Zg16*- goblet cell subtypes identified by VarID (Supplementary Fig. 7).

The transition probabilities between clusters recover known differentiation trajectories (Supplementary Fig. 6c), such as the emergence of enterocytes (cluster 17and 5) via transiently amplifying cells (cluster 4) from intestinal stem cells (cluster 10), which independently give rise to secretory progenitors (cluster 8) of goblet, Paneth, *Chgb*+ enterochromaffin and *Gcg*+ enteroendocrine cells. Tuft cells (cluster 15) and *Sst*+ enteroendocrine cells (cluster 7) are predicted to differentiate independently of the *Sox4*+ progenitor from intestinal stem cells.

I next performed deconvolution of the baseline variability by regressing a second order polynomial to the variance-mean relation across all genes in logarithmic space (Supplementary Fig. 8a). Dividing the variance by the polynomial fit effectively eliminates the systematic variance-mean dependence (Supplementary Fig. 6d). For this dataset, the alternative approach using Pearson residuals from the negative binomial generalized linear model with log link function and the total transcript count as independent variable (Supplementary Fig. 8b-d) also leads to variance estimates without systematic mean dependence (Supplementary Fig. 6e). Using both methods I predicted genes with increased variability in the *Lgr5*+ stem cell population, i.e. cluster 10 versus the remaining populations (one-sided Wilcoxon rank sum-test $P<0.001$, Benjamini-Hochberg corrected, see Methods, $\log_2$-foldchange >1.25). Increased variance of Pearson residuals predicted many more genes with enhanced variability

than the first approach, but comprised 90% of the genes predicted by this method (Supplementary Fig. 6f). Hence, direct inference of corrected variances by the first approach appears to be more conservative. Genes predicted by the first method comprised sub-groups of distinct expression dynamics (Supplementary Fig. 6g). While some genes were predominantly expressed in cluster 10 (e.g. *Gkn3*, *Rgcc*, *Ascl2*), others were upregulated in mature populations of distinct lineages (e.g. *Rnf167* in enterocytes, *Maged1* in enterochromaffin cells, and *Car8* in goblet and enteroendocrine cells). Only a minor fraction of differentially variable genes would have been identified based on differential gene expression (Supplementary Fig. 6h). To explore whether lineage determining genes are stochastically expressed and exhibit increased variability in the stem cell population, I focused on the subset of transcription factors[4] and performed regulatory network inference using GENIE3 (Methods, Supplementary Fig. 6i and Supplementary Fig. 8e). Only 12 of these 31 transcription factors were identified as differentially expressed, demonstrating the power of VarID to detect genes with enhanced stochastic expression within the stem cell population.

The predicted gene regulatory network is centered around the transcription factor *Sox4* required for differentiation of tuft and enteroendocrine cells[5] and is up-regulated in these populations (Supplementary Fig. 8e). *Sox4* connects to *Hes1* which is most strongly expressed in intestinal stem cells and involved in the regulation of proliferation and differentiation[6]. On the other hand, *Sox4* links to *Tox3* which is markedly upregulated in enteroendocrine progenitors (Supplementary Fig. 6j) and was only recently validated by CRISPR modification to yield an enteroendocrine-specific phenotype[7]. Finally, I observed a link between *Sox4* and *Foxa3*, which become upregulated in goblet cells (Supplementary Fig. 8e). *Hopx* was also among the transcription factors with increased local variability in stem cells, but did not link to any of the other factors in the network. *Hopx* was upregulated in the enteroendocrine and tuft cell lineages (Supplementary Fig. 6k).

In summary, VarID recovers stochastic activity of transcription factors known to be required for the differentiation of distinct secretory lineages, suggesting that intestinal stem cells show lineage priming towards secretory fates.

References

1. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).
2. van der Flier, L. G. & Clevers, H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. *Annu. Rev. Physiol.* **71**, 241–60 (2009).
3. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–5 (2015).
4. Hu, H. *et al.* AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, D33–D38 (2019).

5. Gracz, A. D. *et al.* Sox4 Promotes Atoh1-Independent Intestinal Secretory Differentiation Toward Tuft and Enteroendocrine Fates. *Gastroenterology* **155**, 1508-1523.e10 (2018).
6. Ueo, T. *et al.* The role of Hes genes in intestinal development, homeostasis and tumor formation. *Development* **139**, 1071–1082 (2012).
7. Gehart, H. *et al.* Identification of Enteroendocrine Regulators by Real-Time Single-Cell Differentiation Mapping. *Cell* **176**, 1158-1173.e16 (2019).