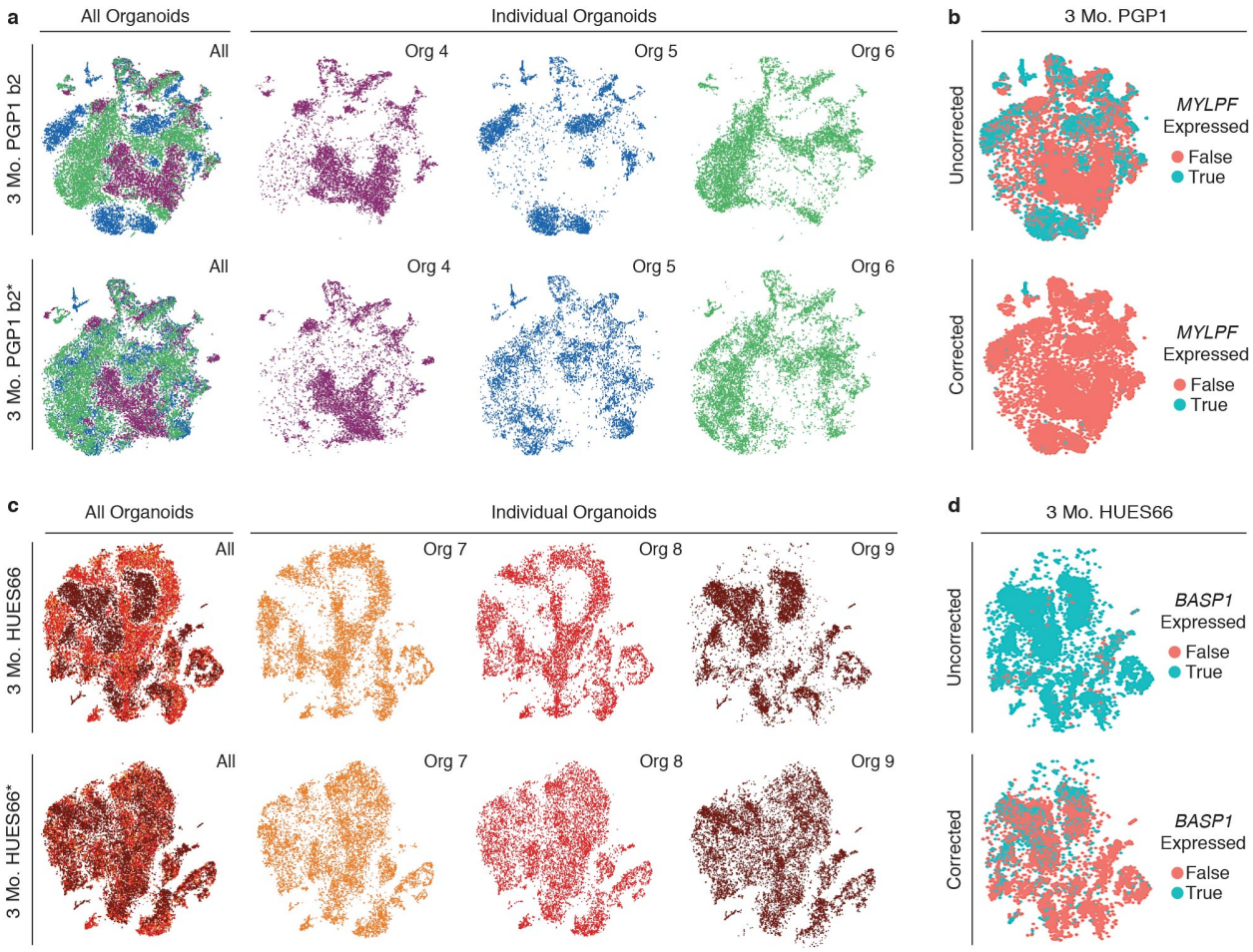

Supplementary Information Figure 1. Correcting for ambient RNA contamination improves co-clustering of organoids in the 3 month HUES66 and PGP1 batch 2 datasets. a) Co-clustering of the three organoids in the 3 month PGP1 Batch 2 dataset before (top) and after (bottom) removal of 15 mesodermal genes identified as contributing to ambient RNA contamination from the list of variable genes used for clustering. **b)** Expression calls for the *MYLPF* gene in cells from Orgs 4-6 before (top) and after (bottom) ambient RNA correction. **c)** Co-clustering of the three organoids in the 3 month HUES66 dataset before (top) and after (bottom) removal of 17 genes identified as contributing to ambient RNA contamination from the list of variable genes used for clustering. **d)** Expression calls for the *BASP1* gene in cells from Org 9 before (top) and after (bottom) ambient RNA correction.

We have provided batch-corrected data from all of the timepoint-matched organoids together in the main figures (Fig. 1 and 2). Before batch correction, we noted that in two cases there were single organoids that, while making all of the same cell types, displayed a distinct

cluster pattern (Supplementary Information Fig. 1A and C, top rows). We determined that the distinct cluster pattern did not in fact reflect a biological difference, but was due to ambient RNA contamination, a common source of noise in single-cell RNA-seq experiments. We evaluated this effect using the SoupX R package v0.3.0, which uses "empty" droplets in the single-cell RNA-seq dataset (nUMI < 10) to identify which mRNAs are present in the ambient media (Young & Behjati, "SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data", bioRxiv 303727). Default parameters were used. Several genes were found to significantly contribute to ambient RNA in the lanes containing the affected organoids (Org 5 and Org 9). To correct for this, many of these genes were removed from the list of highly variable genes used for clustering in that batch. For PGP1, these genes were mainly related to mesodermal functions, and were MYLPF, ACTC1, TNNC2, TNNI1, TNNC1, MEF2C, TPM2, ACTA1, MYL4, DES, MYL1, TNNI2, MYH3, TNNT3, and ENO3. For HUES66, these genes were H1FX, BASP1, GFBP2, RNF187, UBE2S, TCEAL5, FJX1, SRM, SMS, IER5, ID4, DUSP5, SFRP1, RPRM, CITED1, YBX3, and KCNG1.

After correcting for ambient RNA, the organoids from these batches overlap closely (Supplementary Information Fig. 1A and C, bottom row). Note that this does not mask all clustering differences, such as the increased number of corticofugal projection neurons found in Org 4 (bottom of purple t-SNE in Supplementary Information Fig. 1A). The expression of representative genes that were found to contribute to organoid-specific gene expression differences is shown before and after ambient RNA correction in Supplementary Information Fig. 1B and D (one out of 15 genes for PGP1 b2, one out of 17 genes for HUES66).

While this demonstrates where the source of variation originated, we decided to be conservative in the way we analyzed our organoids and retained the original data for all organoids in the manuscript (rather than use the SoupX corrected data) to avoid any possibility of distortion from the ambient mRNA correction algorithm. The batch corrected data presented in the main figures of this manuscript have not been corrected for ambient RNA, but readers who wish to use the raw data from this manuscript should be aware of this source of technical variation.