EMTAB3929: Data Characteristics

Margaret R. Starostik

(1) Histogram showing the number of reads per cell that mapped to human reference genome GRCh38.84.

```
# (1) Generate histogram of the number of reads that mapped
# to reference genome GRCh38.84.
mapping_plot01 <- ggplot(metasheet, aes(x = Sample, y = `Mapped Reads`,
  fill = EStage)) + geom_bar(stat = "identity")
mapping_plot02 <- mapping_plot01 + labs(title = "", x = "Cells",
  y = "Mapped Reads") + theme_bw() + theme(axis.text.x = element_blank(),
  axis.ticks.x = element_blank(), plot.title = element_text(hjust = 0.5))
ggsave(paste0(project_folder, "Results/01_DataSummary/EMTAB3929_MappedReads.pdf"),
  plot = mapping_plot02, device = "pdf", width = 6, height = 4,
  units = "in")
print(mapping_plot02)
```

(2) Dot plot depicting sample size for each cell lineage used in this project.

```
# (2) Generate dot plot summarizing the samples in this study
dotplot_data <- as.data.frame(metasheet %>% group_by(`Revised lineage (this study)`,
  EStage) %>% summarise(EmbryoCount = length(unique(Embryo)),
  CellCount = length(Cell)))
colnames(dotplot_data)[1] <- "Lineage"
dotplot_data$Lineage <- factor(dotplot_data$Lineage, levels = c("Undefined",
  "ICM", "Trophectoderm", "Intermediate", "Epiblast", "Primitive Endoderm"))

dotplot01 <- ggplot(dotplot_data, aes(x = EStage, y = Lineage)) +
  geom_point(aes(size = EmbryoCount, color = CellCount))
dotplot02 <- dotplot01 + labs(title = "", x = "\nDevelopmental Stage",
  y = "Cell Lineage\n") + scale_color_continuous(name = "Cells") +
  scale_size_continuous(name = "Embryos")
ggsave(paste0(project_folder, "Results/01_DataSummary/EMTAB3929_SampleSummary.pdf"),
  plot = dotplot02, device = "pdf", width = 8, height = 4.5,
  units = "in")
print(dotplot02)
```
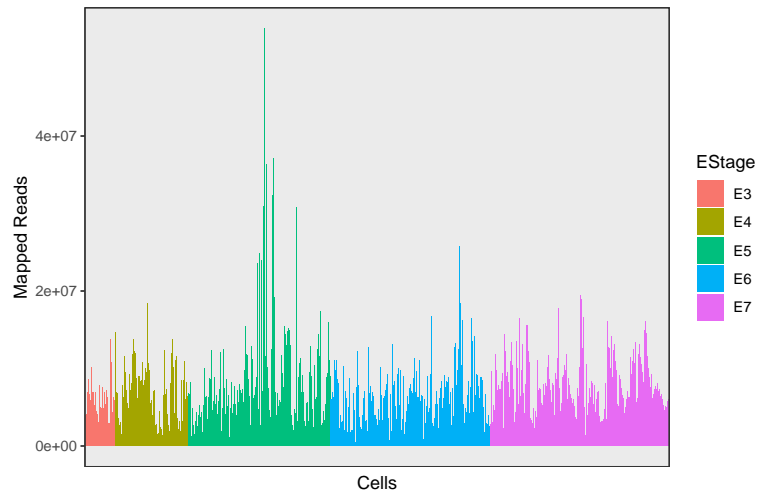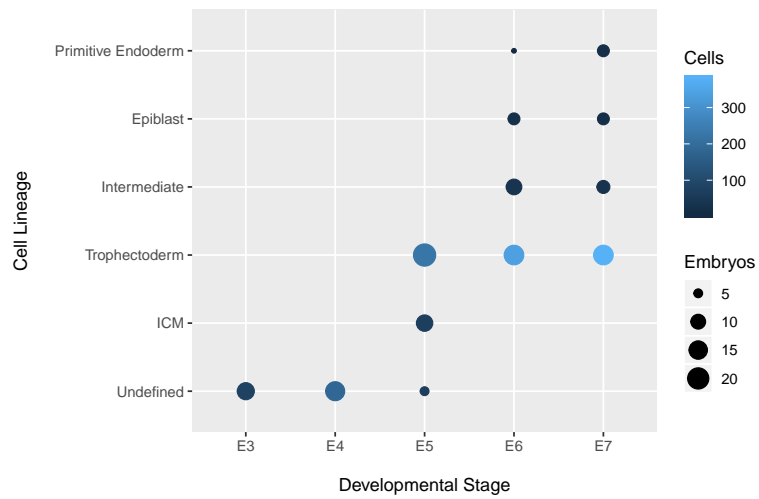
Figure 1: Reads Mapped to GRCh38.84

Figure 2: Sample Summary

```
knitr::kable(dotplot_data[order(dotplot_data$EStage), ], caption = "Sample Summary",
    row.names = FALSE)
```

Table 1: Sample Summary

Lineage	EStage	EmbryoCount	CellCount
Undefined	E3	13	78
Undefined	E4	16	185
ICM	E5	12	66
Trophectoderm	E5	22	227
Undefined	E5	5	67
Epiblast	E6	7	25
Intermediate	E6	11	36
Primitive Endoderm	E6	4	8
Trophectoderm	E6	17	336
Epiblast	E7	7	20
Intermediate	E7	8	32
Primitive Endoderm	E7	7	22
Trophectoderm	E7	17	379

```
write.table(dotplot_data[order(dotplot_data$EStage), ], file = paste0(project_folder,
    "Results/01_DataSummary/EMTAB3929_SampleSummary.txt"), quote = FALSE,
    sep = "\t", row.names = FALSE)
```