scRNA-seq pipeline

Design → Raw Data → Preprocessing: quantification, QC, filtering, normalization → Dimension Reduction

Cell Types ← Marker Gene DE ← Clustering ← Alignment

Cell Types ← Cell Type Prediction ← Alignment

Cell Types → Diff. Abundance Analysis

Cell Types → Aggregation → Diff. State Analysis

Cell Types → Diff. State Analysis

Diff. State Analysis → Geneset analyses

Diff. State Analysis → Aggregation of related cell types
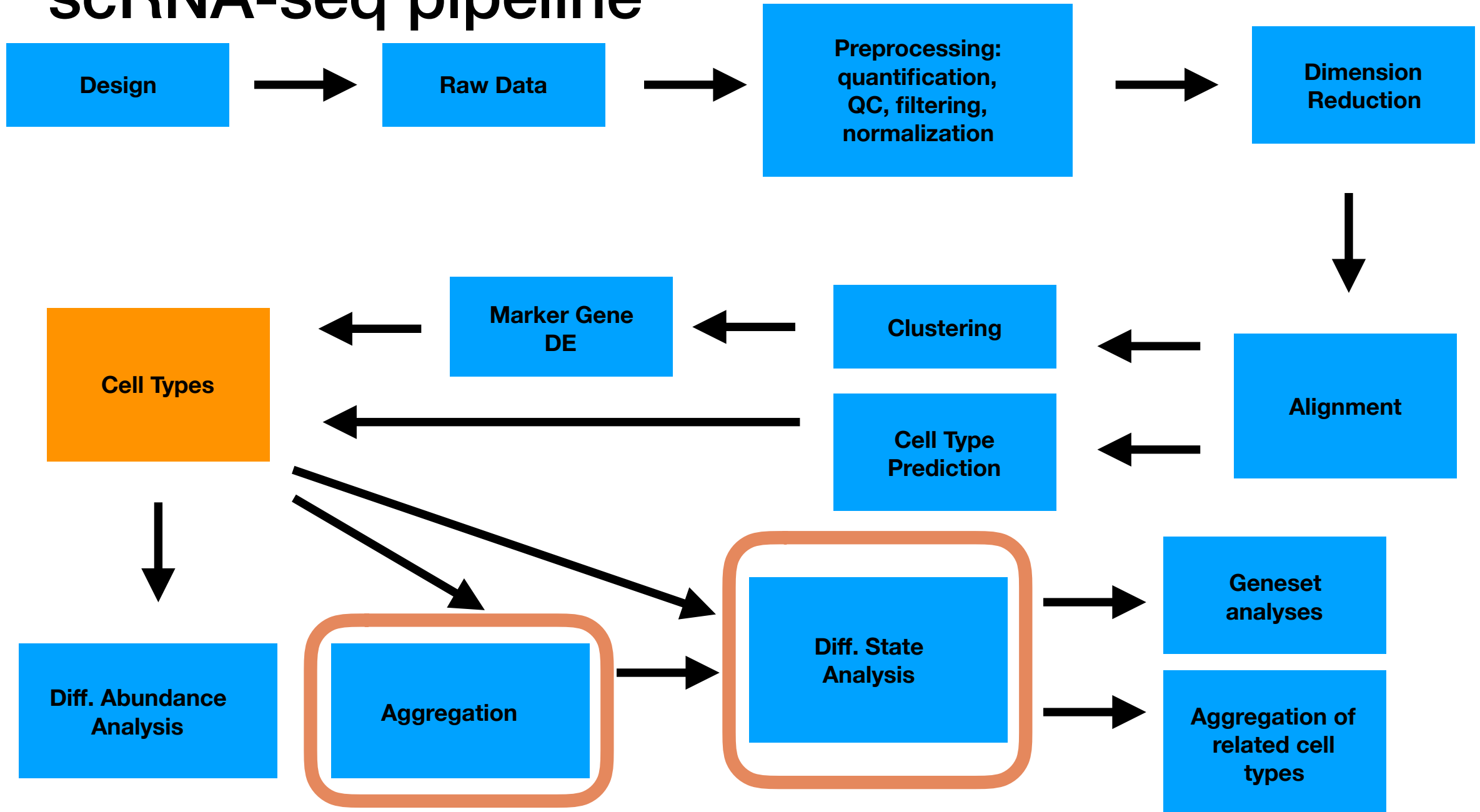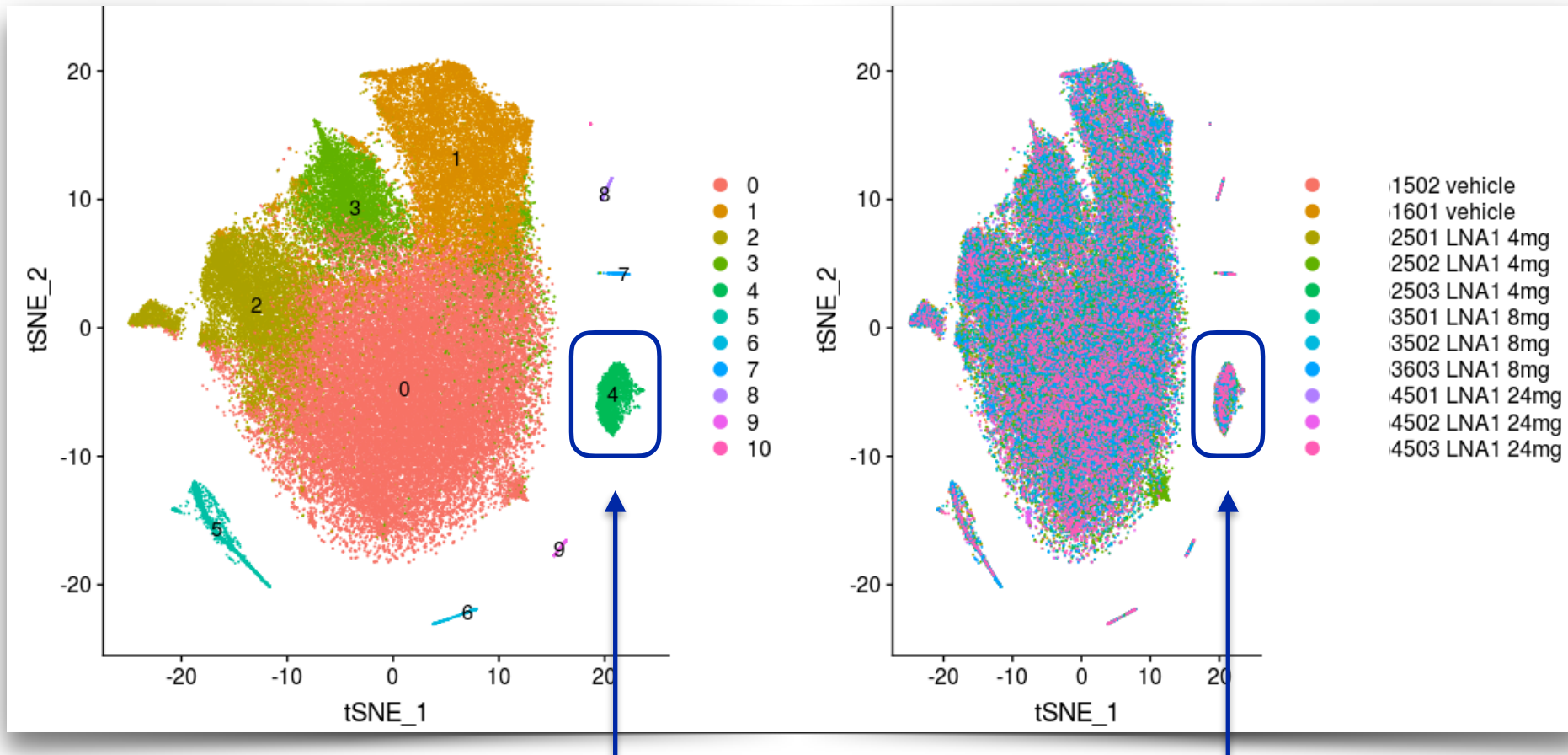
Two types of differential expression: marker gene DE, <u>differential state analysis</u>

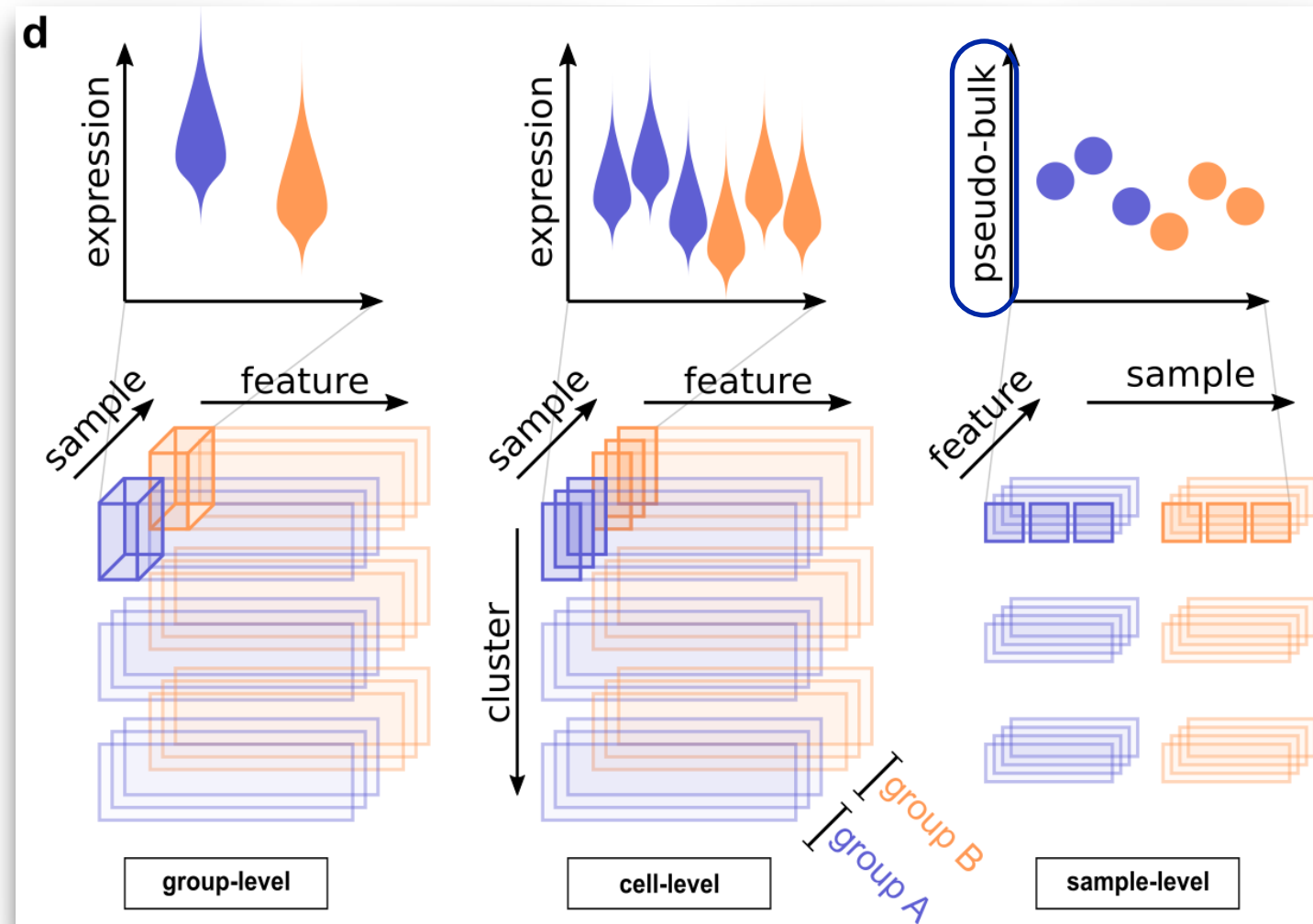repeat for each population ..
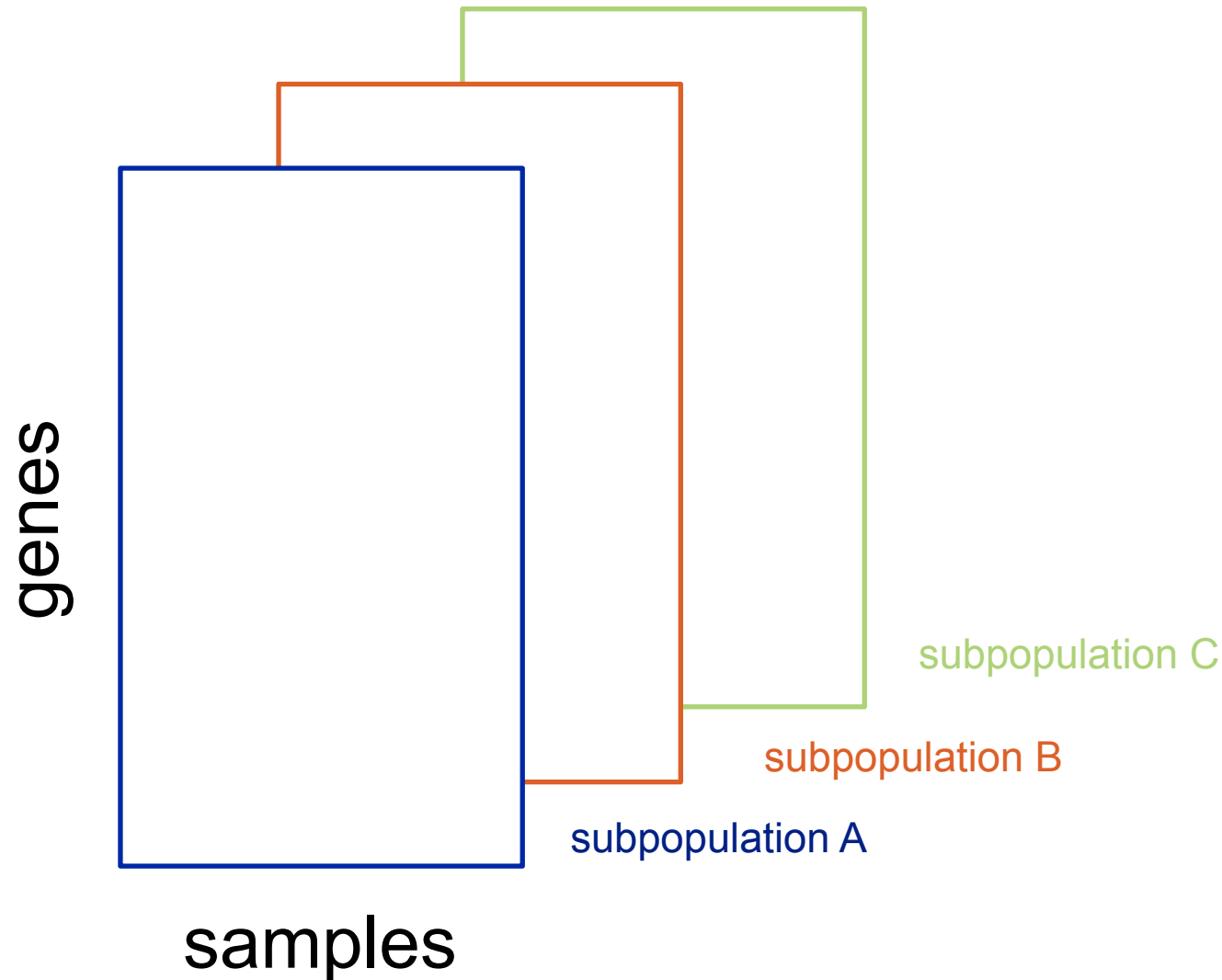
Focus: Marker gene DE

Focus: cross-sample DE

After "Cell Type Prediction" / "Clustering", various ways to view the inference problem

Multi-sample
Multi-condition
Multi-population

Results suggesting reducing the problem to a set of bulk RNA-seq DE analyses is fast an powerful

genes

samples

subpopulation A

subpopulation B

subpopulation C

- At least 1 modelling question remains: should you model each population on its own, or all populations at once?

- Currently done as the former; in the latter, one could model a main effect (common response) in addition to interactions (subpopulation specific deviations)