

targets = sample sheet (read IDATs)

RGSet with pData (Rgset_original & pd_original)

Mset, RatioSet, RawBetas, gRatioSet

Quality checks for *samples*

- calculate **detection p values**
→ exclude samples with **mean detection p-value >.05**
= RGSet_qual & RawBetas_qual + detP & detP_qual
 - Run **minfi QC report** & take a look at it
 - **distribution artefacts** (this is a visual check done by user)
 - **sex mismatches**
- **RGSet_clean** & RawBetas_clean & **pd_clean** & annot_clean & detP_clean

Expertise needed for interpretation

Normalization

- **Stratified Quantile** normalization (**quantileN & quantileNBetas & quantileNMs**)
 - quantileN is a GenomicRatioSet
 - quantileNBetas are the beta values
 - quantileNMs are the M values
- Afterwards **BMIQ** (**BMIQ.quantileN**)
 - **Betas** after quantile + BMIQ normalization

Lot of different options. See R Script comments.



Filter out probes

- Detection **p-value >.01** (in one or more samples)
 - Probes on the **sex chromosomes**
 - Probes affected by **common SNPs**
 - **Cross-reactive** probes and polymorphic probes (Chen et al.)
 - **Cross-hybridizing** and polymorphic probes on EPIC (McCartney et al.)
- **BMIQ.quantileN_filtered**
- **quantileN_filtered & Betas_quantileN_filtered & Ms_quantileN_filtered**



Remove batch effects with Combat:

- NA **probes without variance** (M-values)
 - PCA on the M values → PCobj
 - determine which batches show strongest association with PCs
 - **Remove batches with combat** (→M_combat_1plate/etc..., last latest number)
- **Betas_combated & Betas_combated_ExprSet**



Remove batch effects with Combat:

- NA probes without variance (M-values)
- PCA on the M values → PCobj
- determine which batches show strongest association with PCs
- Remove batches with combat (→M_combat_1plate/etc..., last latest number)

→ **BMIQ.quantileN_combated&
BMIQ.quantileN_combated_ExprSet**

This is added for cases where all probes are needed!



- Apply **MixedUpMapper** to check mismatches between genotypes and methylation
- exclude mismatches from Betas_combated & Betas_combated_ExprSet & BMIQ.quantileN
- *if* exclusions: Betas_combated_mixc & Betas_combated_ExprSet_mixc & BMIQ.quantileN_combated_mixc



- Apply gaphunter to identify probes showing clustered distributions
- Betas_gapped



Additional checks for specific data set

- Cord blood: contamination maternal blood
- Contaminated IDs excluded in: **Betas_combated_CC**, Betas_combated_ExprSet_CC, BMIQ.quantileN_combated_CC, BMIQ.quantileN_combated_ExprSet_CC, pd_cleaned_CC



ID checks for final data set

- methylation_betas_XX in subfolder finalData
- And a sample-cleaned RGSet named RGSet_cleaned_final



estimate cell type proportions

- *reference-free: RefFreeEWAS*
- *reference-based: e.g. for cord blood Salas et al. or Bakulski et al.*