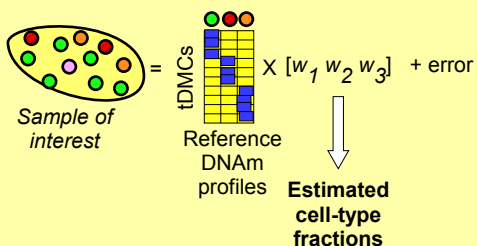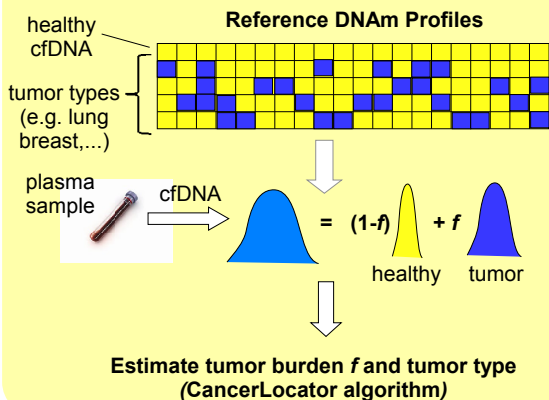# DNA methylation analysis of cell-type heterogeneity

a — *Estimating cell-type fractions*
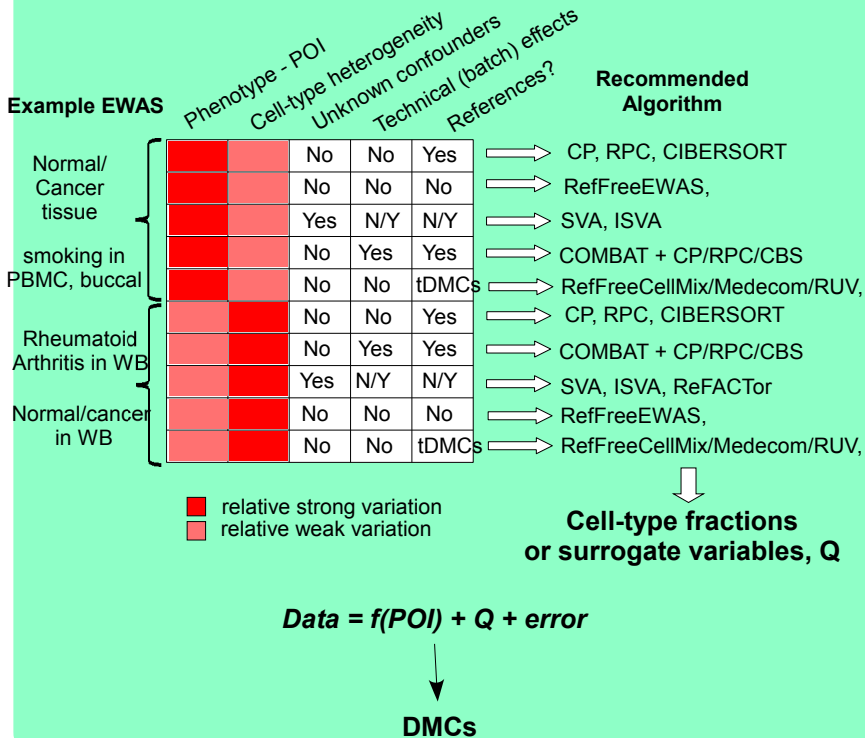
1. Complex tissue (e.g. whole blood / breast)

Sample of interest = tDMCs (Reference DNAm profiles) $\times$ $[w_1\ w_2\ w_3]$ + error $\rightarrow$ **Estimated cell-type fractions**

2. Infering tumor burden and tissue of origin from cell-free DNA in plasma

Reference DNAm Profiles

healthy cfDNA

tumor types (e.g. lung breast,...)

plasma sample $\xrightarrow{\text{cfDNA}}$ = $(1\text{-}f)$ + f

healthy tumor

Estimate tumor burden f and tumor type (CancerLocator algorithm)

b — *Identifying differentially methylated CpGs*

Example EWAS	Phenotype - POI	Cell-type heterogeneity	Unknown confounders	Technical (batch) effects	References?	Recommended Algorithm
Normal/ Cancer tissue	strong	weak	No	No	Yes	CP, RPC, CIBERSORT
	strong	weak	No	No	No	RefFreeEWAS,
	strong	weak	Yes	N/Y	N/Y	SVA, ISVA
smoking in PBMC, buccal	strong	weak	No	Yes	Yes	COMBAT + CP/RPC/CBS
	strong	weak	No	No	tDMCs	RefFreeCellMix/Medecom/RUV,
Rheumatoid Arthritis in WB	weak	strong	No	No	Yes	CP, RPC, CIBERSORT
	weak	strong	No	Yes	Yes	COMBAT + CP/RPC/CBS
	weak	strong	Yes	N/Y	N/Y	SVA, ISVA, ReFACTor
Normal/cancer in WB	weak	strong	No	No	No	RefFreeEWAS,
	weak	strong	No	No	tDMCs	RefFreeCellMix/Medecom/RUV,

■ relative strong variation
■ relative weak variation

$\rightarrow$ **Cell-type fractions or surrogate variables, Q**

Data = f(POI) + Q + error

$\downarrow$

DMCs

c — *Estimating clonal epigenetic heterogeneity*

● methylated CpG
● unmethylated CpG

Low clonal diversity $\rightarrow$ **WGBS/RRBS**

All reads concordant (CR)

$\downarrow$

PDR=0

healthy/good outcome

High clonal diversity $\rightarrow$ **WGBS/RRBS**

Discordant reads (DR)

$\downarrow$

high PDR

PDR=CR/(CR+DR)

cancer/ poor outcome