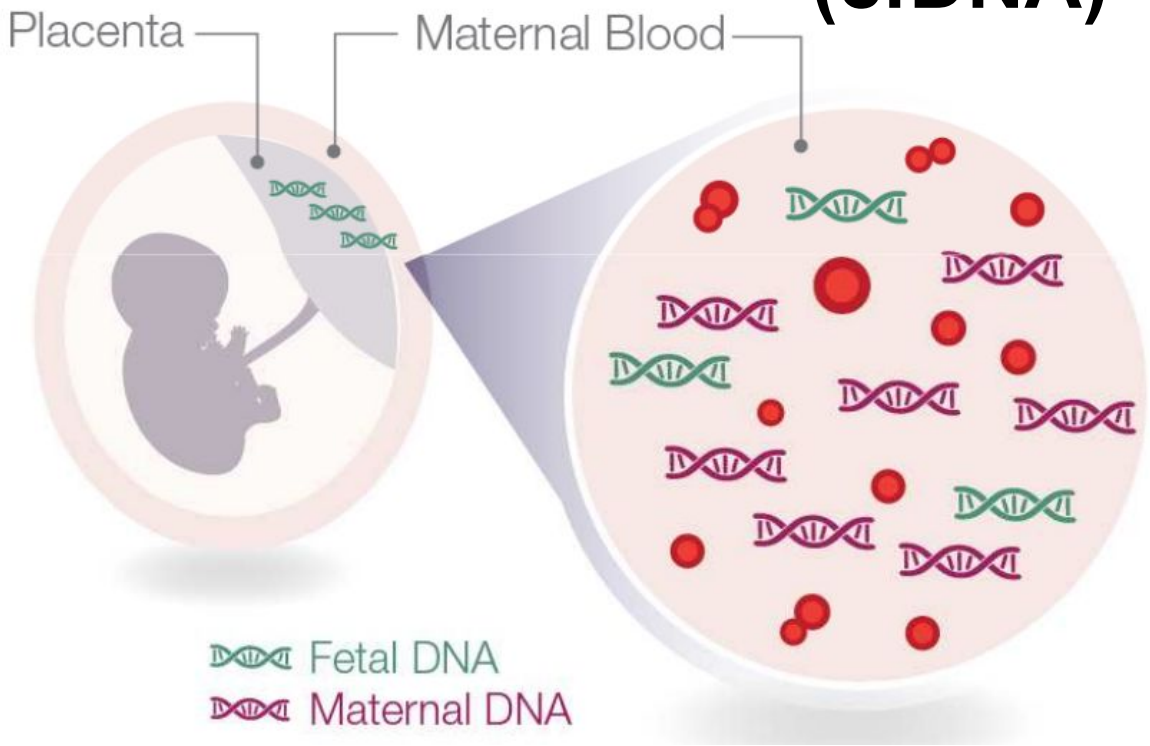
Development of Non-Invasive Prenatal Testing (NIPT) analysis pipeline.

Polina Kozyulina

Project leader:
Andrey Glotov

Ott Research Institute of Obstetrics, Gynecology & Reproductology

Cell free DNA (cfDNA)

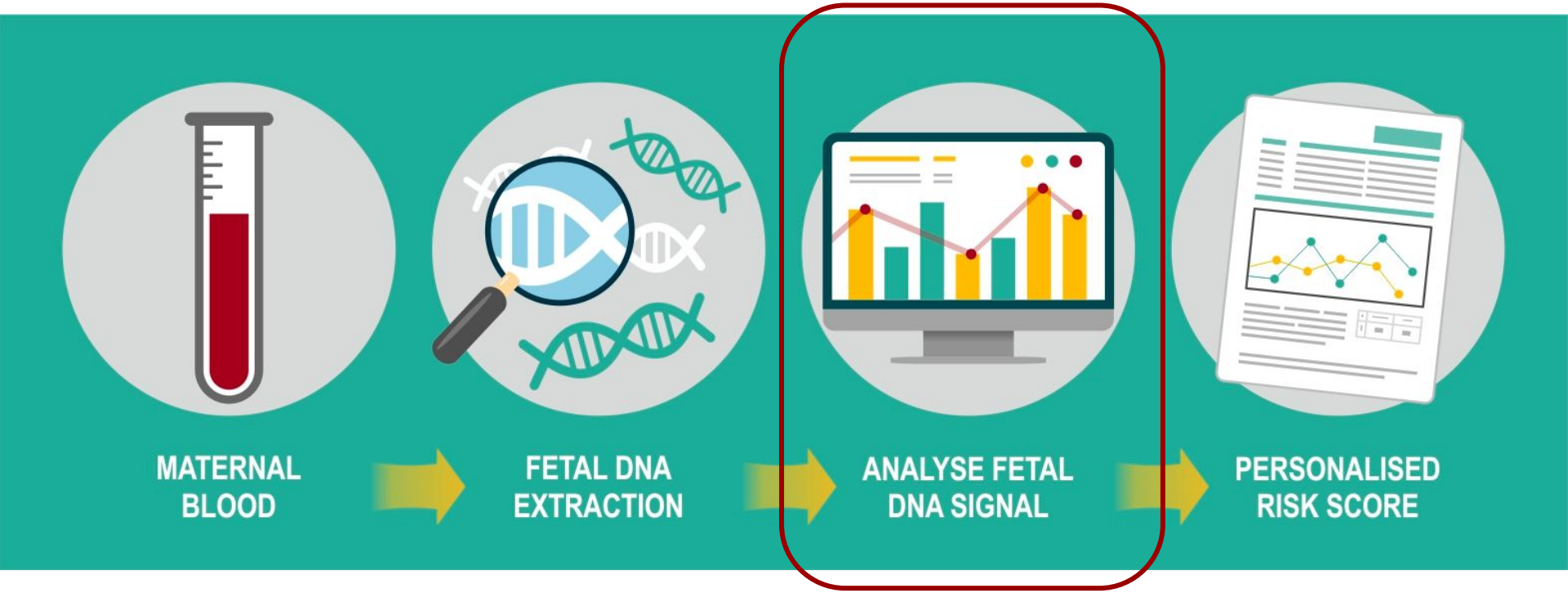


cfDNA comes from apoptotic cells derived from:

- Maternal Circulation
  - Adipocytes
  - White Blood Cells
- Fetal
  - Placental cells (trophoblasts) in the maternal circulation

Fetal fraction = % of fetal cfDNA in a sample

What is NIPT?



Non-Invasive Prenatal Testing

Screening method of aneuploidy detection

NIPT approaches

WGS

- ➔ Read lengths >= 36–52 bp
- ➔ Total coverage ~ 2–10 million reads per sample
- ➔ Free analysis tools
- ➔ Expensive

Targeted

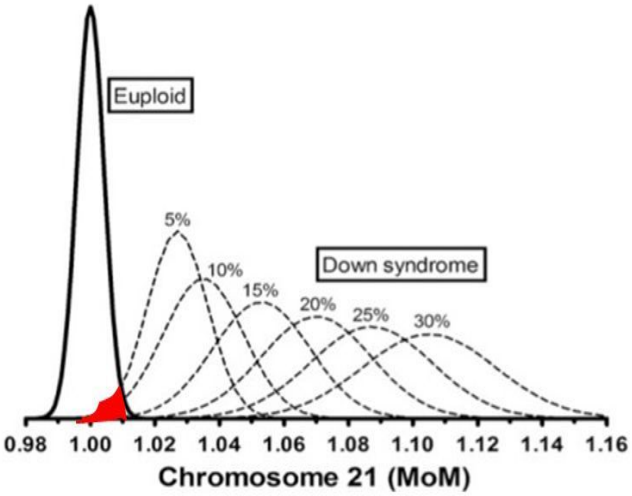
- ➔ Panels of non-polymorphic loci of 13, 18, 21 and X(Y) chromosomes
- ➔ High coverage
- ➔ PCR duplicates
- ➔ Cheaper

SNP-based

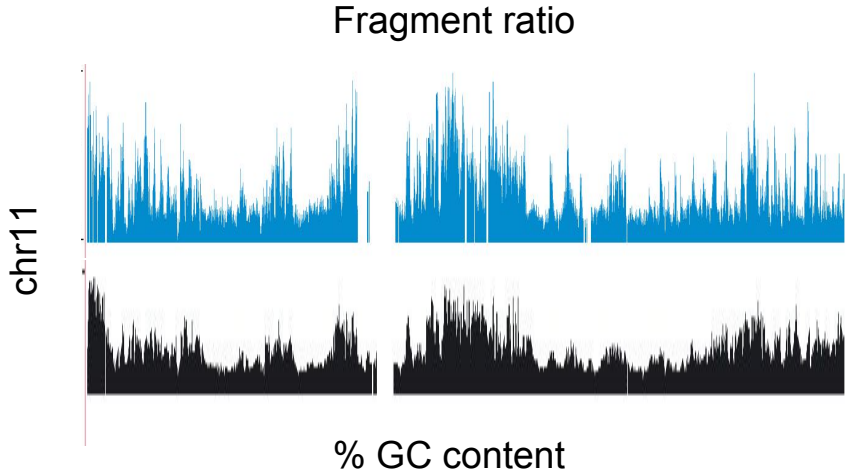
- ➔ High coverage
- ➔ Sequencing of 2 samples: cfDNA and maternal DNA
- ➔ PCR duplicates
- ➔ Cheaper

Problems

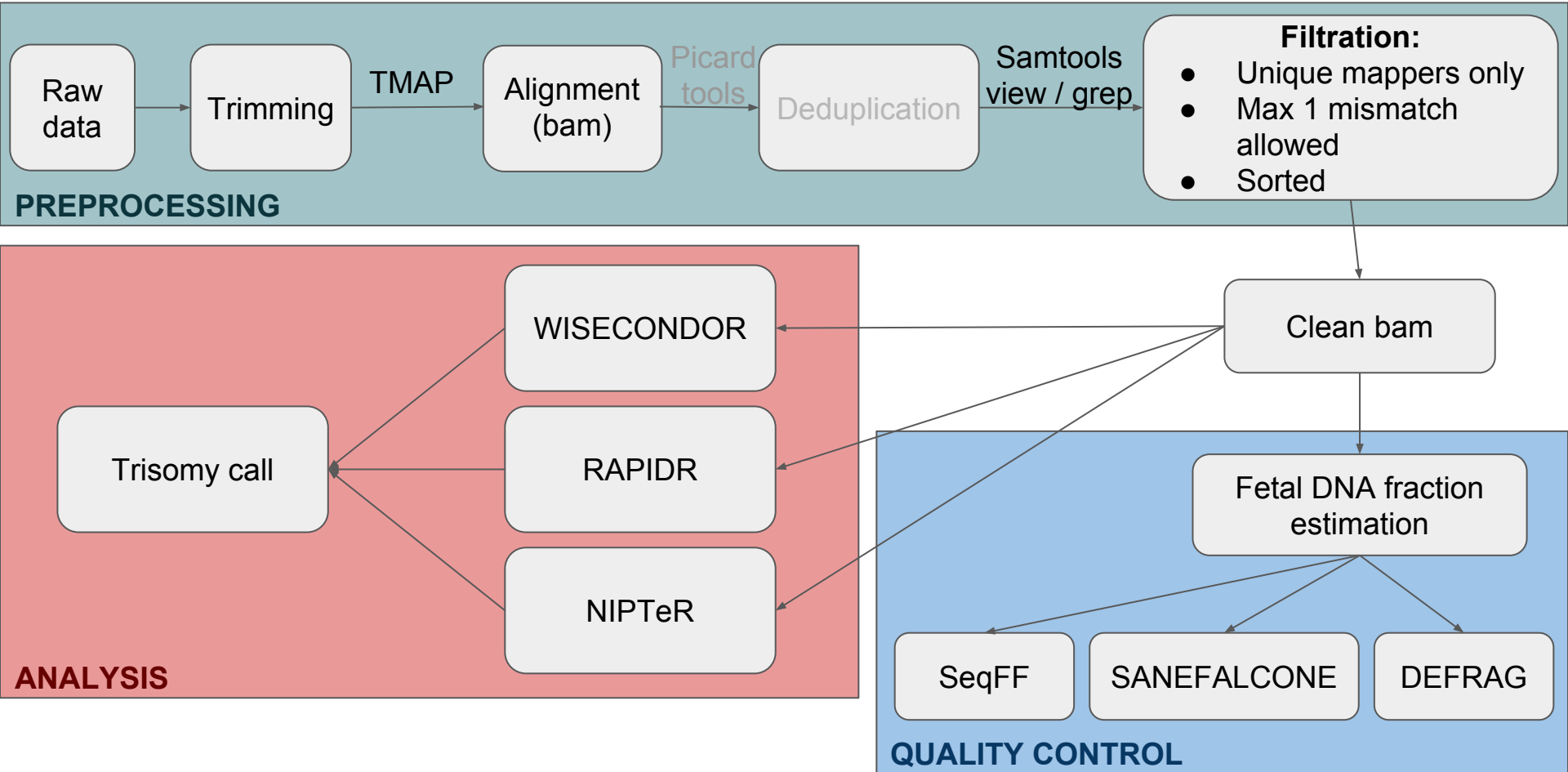
Fetal fraction < 4%



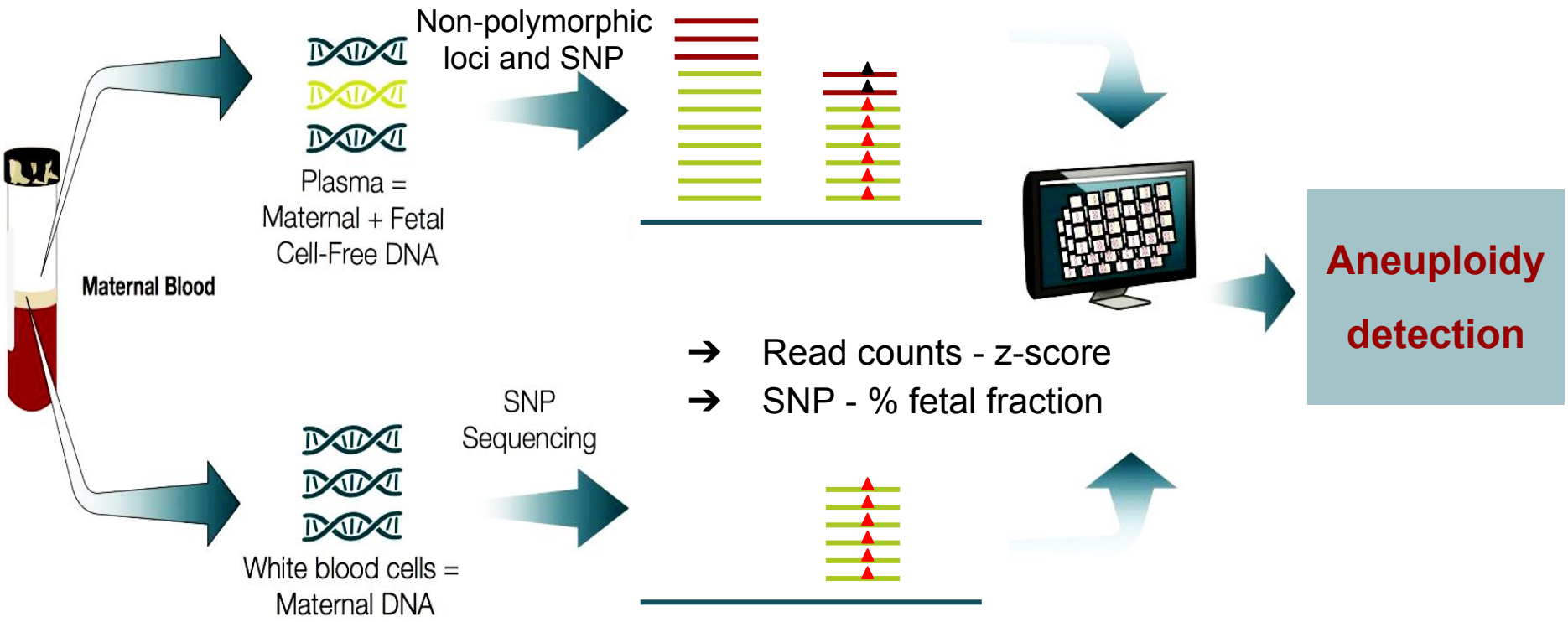
GC bias



WGS analysis pipeline



Target panels



Pipeline and panel design

Panel design

Non-polymorphic loci

- ➔ No SNP or CNV (MAF > 0.01)
- ➔ Average GC content [45; 55]
- ➔ No repeats
- ➔ No homopolymers
- ➔ Uniform locus-specific oligo melting temperatures

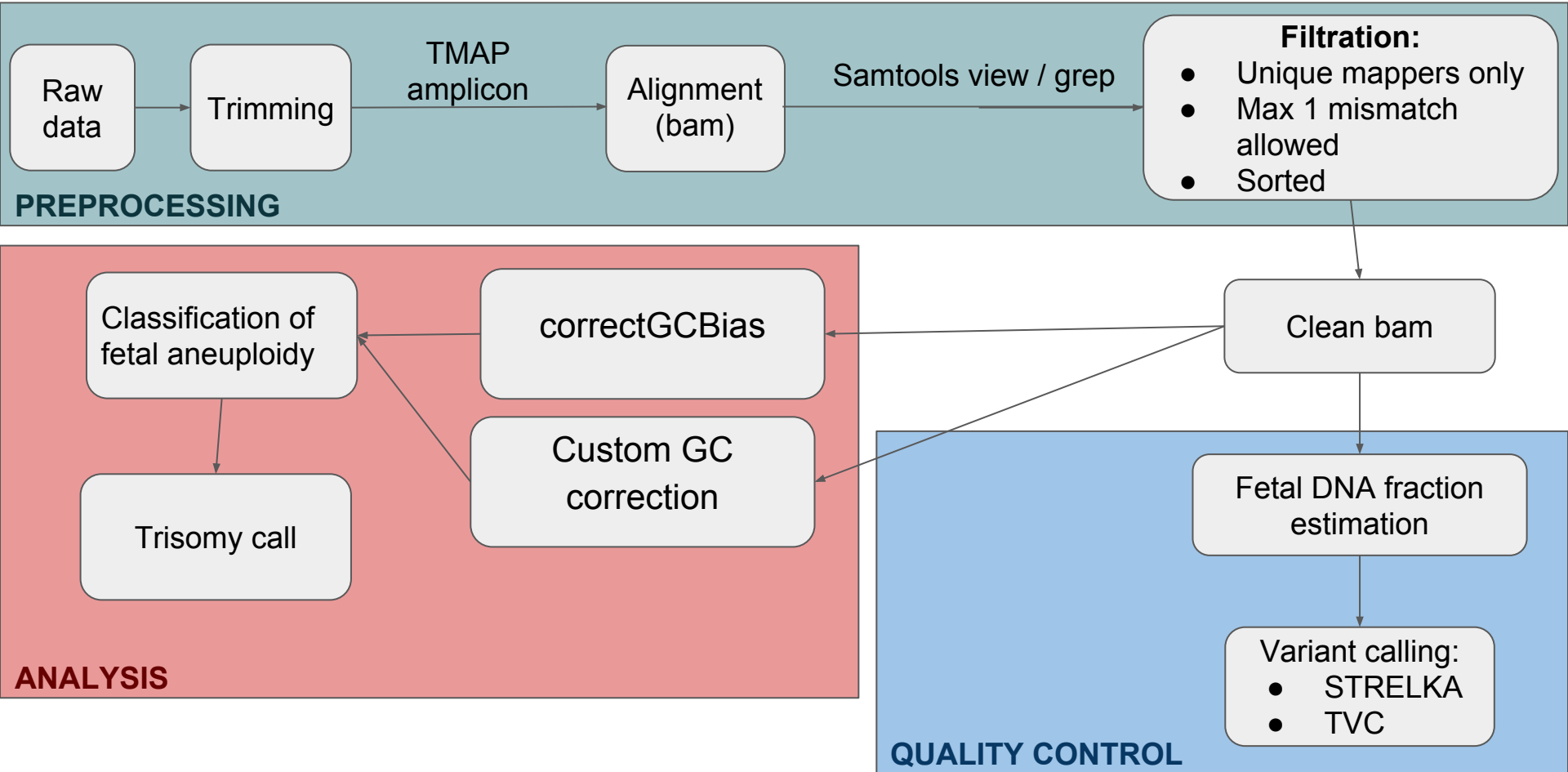
~ 340-380 loci per chromosome (13, 18, 21, X, Y)

SNP

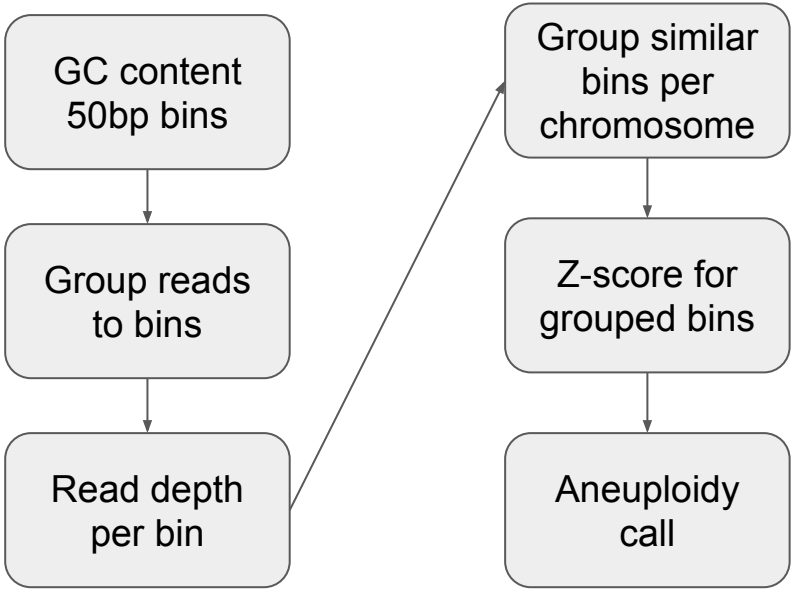
- ➔ SNP (MAF >= 0.20 and present in slavic population)
- ➔ Average GC content [45; 55]
- ➔ No repeats and homopolymers
- ➔ Uniform locus-specific oligo melting temperatures

~ 2000 loci (chromosomes 1-12)

Target analysis pipeline



Custom GC correction and Classification of fetal aneuploidy



$$binRepr_i = \frac{binTotalRD_i}{\sum_{j=4} binTotalRD_j}$$

TotalRD = total read depth

j = bins from same category for the rest of chromosomes.

$$Z = x - \mu\sigma$$

x = sample bin representation

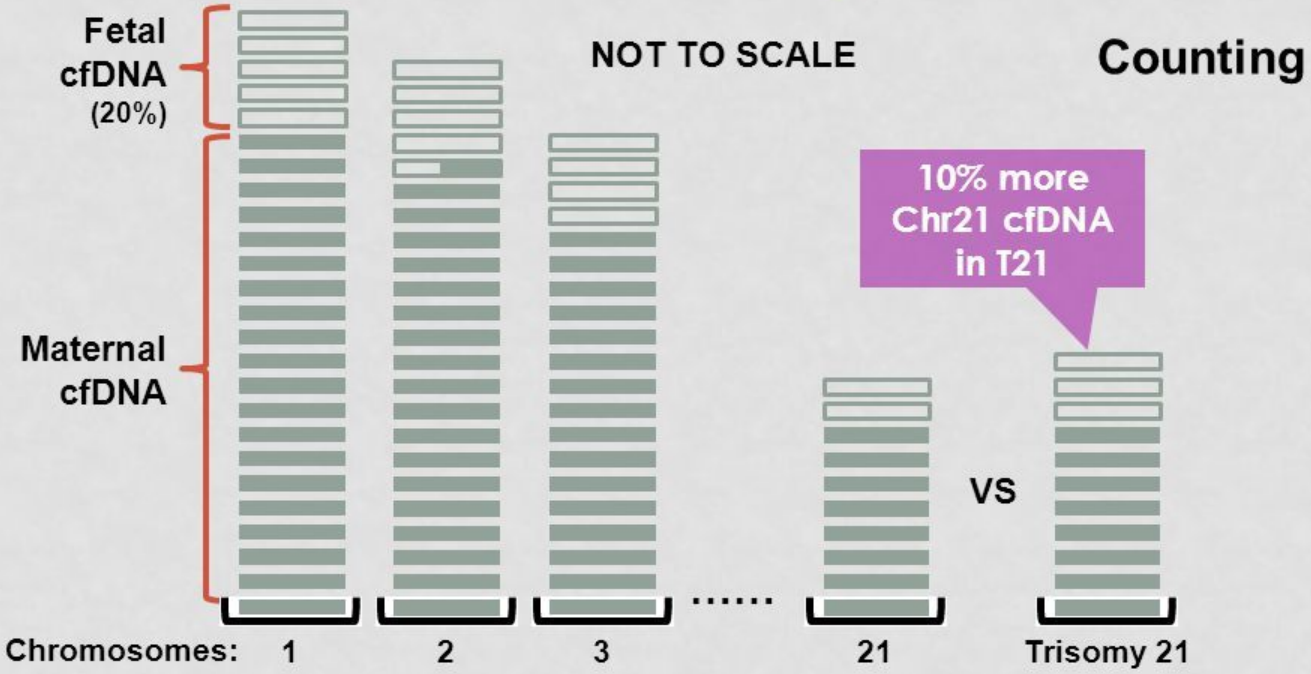
μ = bin representation run median

σ = bin representation median empirical deviation (calculated from set of all samples per run).

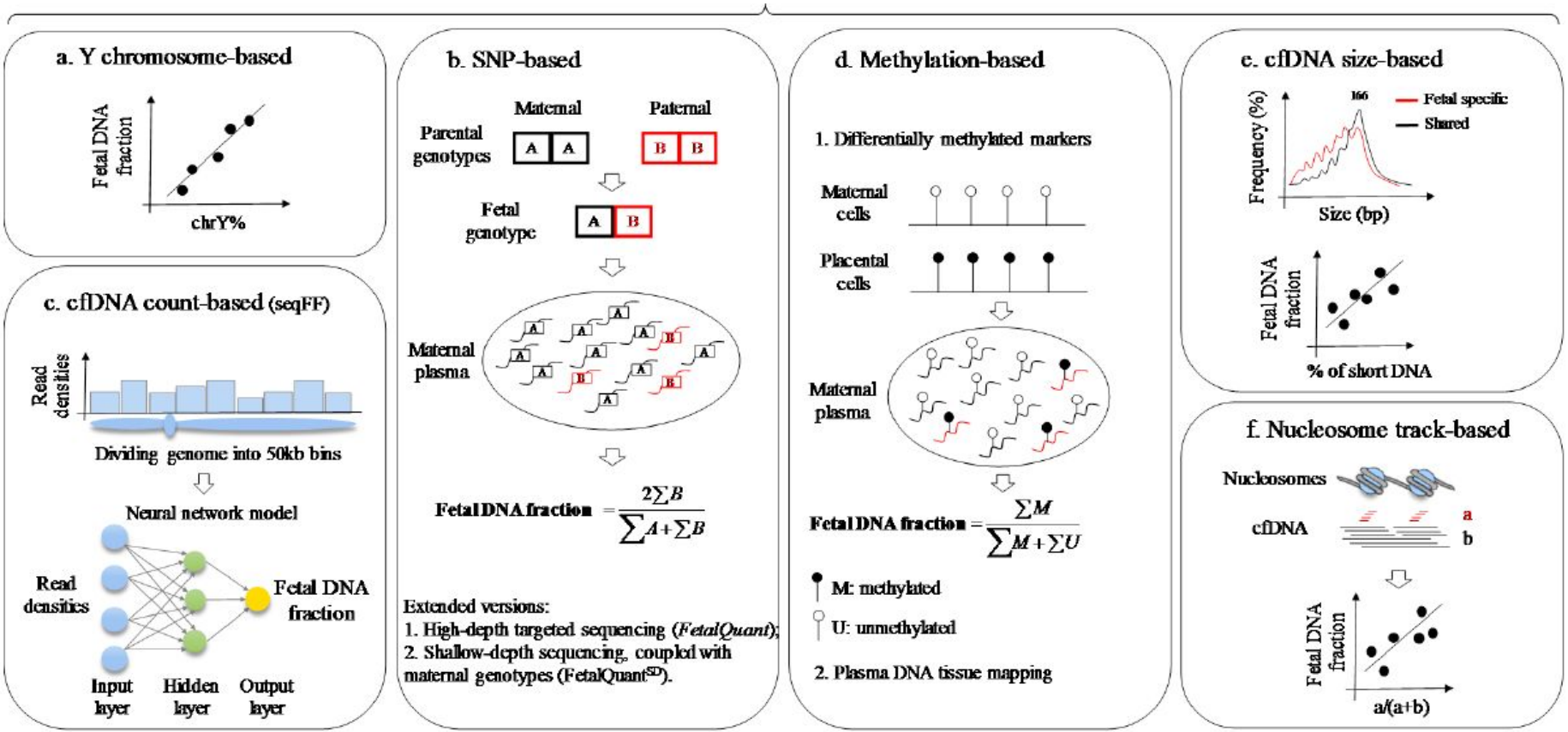
Thank you!

DETECTION OF FETAL ANEUPLOIDY

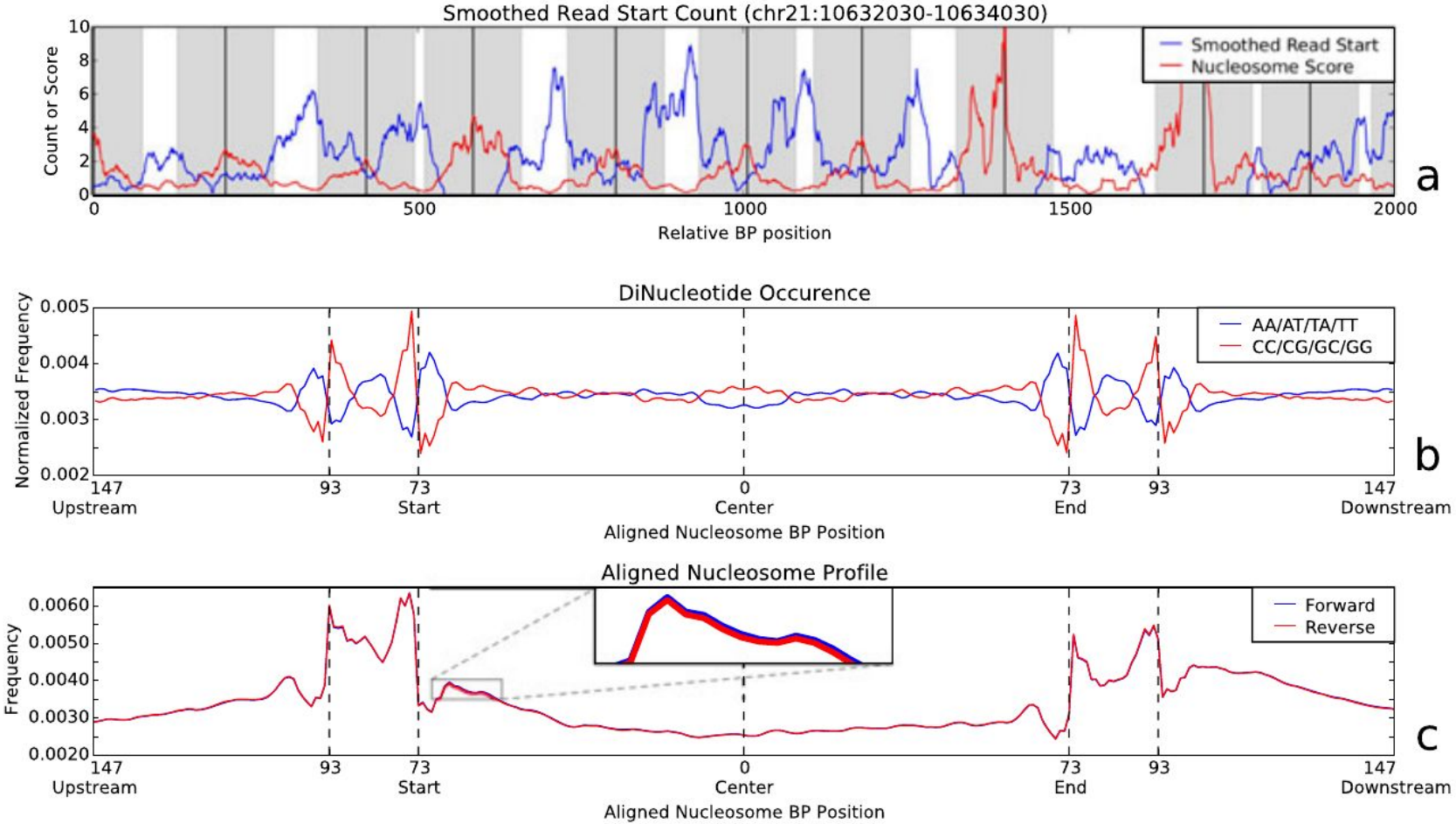
MPS Enables Precise Molecular Counting



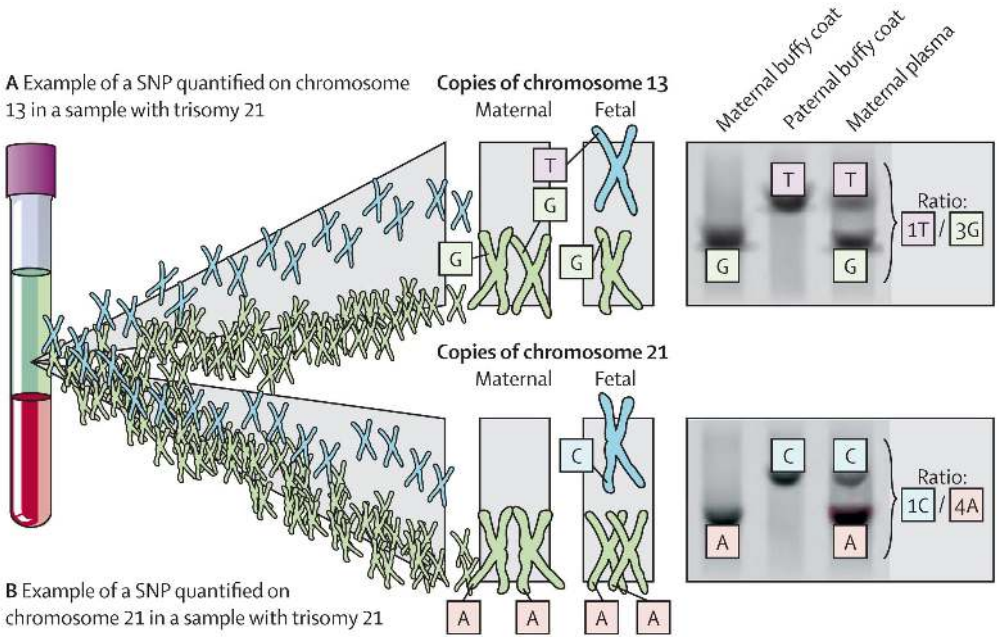
Fetus DNA fraction estimation



SANEFALCONE



NEEDS TRAINING - about 300 for training in total, split over 12 different batches of about 25 samples each



Why false- positive or negative results?

- confined placental mosaicism
- vanishing twin syndrome
- fetal or maternal mosaicism
- tumors
- maternal duplication
- technical issues

Z-score

- standard Z-score
- normalized chromosome value (NCV)
- median-absolute-deviation-based Z-score

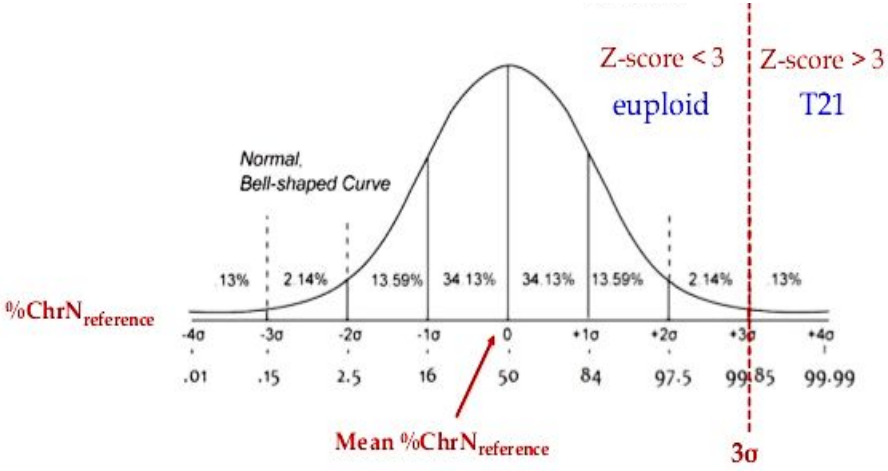
$$Z_k = \frac{y_k - \mathrm{E}[y_k|\text{normal control}]}{\sqrt{\mathrm{var}[y_k|\text{normal control}]}}$$

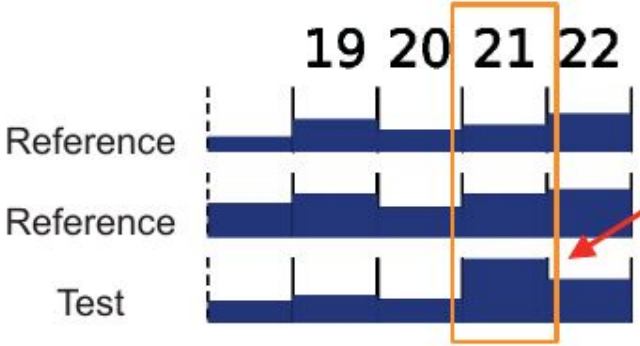
Denote yk as the ratio of chromosome read number to the total read number. For a sample with unknown fetal karyotype , yk was normalized by the mean and standard deviation acquired from normal controls

$$\xi_k \equiv \frac{s_k - \mathrm{E}[s_k|\text{normal control}]}{\sqrt{\mathrm{var}[s_k|\text{normal control}]}}$$

s_k = yk/yR , where yR is read number of reference chromosome to total read number

NVC z score





Chr 21/Chr1 Sample

Chr21/Chr1 Reference Mean

$$\mathrm{NCV}_{ij} = \frac{x_{ij} - \hat{\mu}_j}{\hat{\sigma}_j},$$

Chr21/Chr1 Reference SD

Z-score approach (advanced)

1) Reads --> to 50kbp “bins”.
2) Select bins contain sequences unique to that chromosomal region.
3) Z-score → for each bin
4) Z-scores averaged across the entire chromosome
5) Calculate chromosomal representation

Chromosome representations:

$$chrRep_i = \frac{chrTotalRC_i}{\sum_{j=1\ldots22} chrTotalRC_j}$$

TotalRC - total read counts

Chromosome Z scores:

$$Z = x - \mu\sigma$$

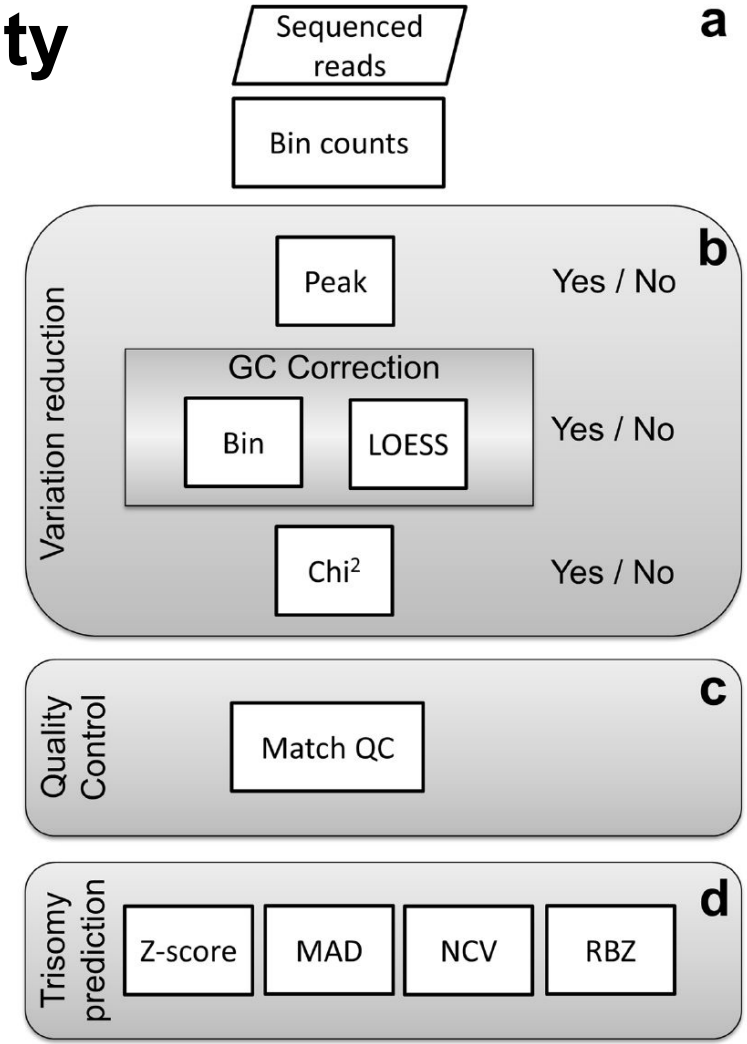
x: sample chromosome representation

μ: chromosome representation plate median

σ: chromosome representation MADs, median absolute deviation, as calculated using a training set of all samples.

Improve z-score sensitivity (NIPTeR)

- chi-squared-based variation reduction (χ 2 VR)
- regression-based Z-score (RBZ)
- Match QC score



Z-score based tools

Tool	Lang	date	input	comment
WISECONDOR	Python	2014	bam	Included in IonNIPT package
RAPIDR	R	2014	bam	
NIPTeR	R	2017	bam	Improved sensitivity