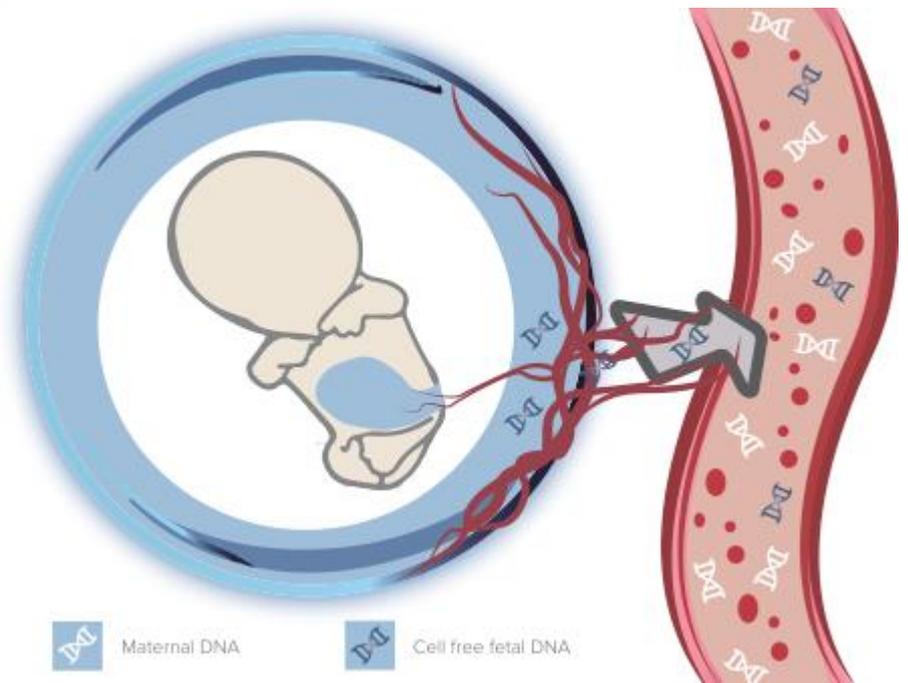
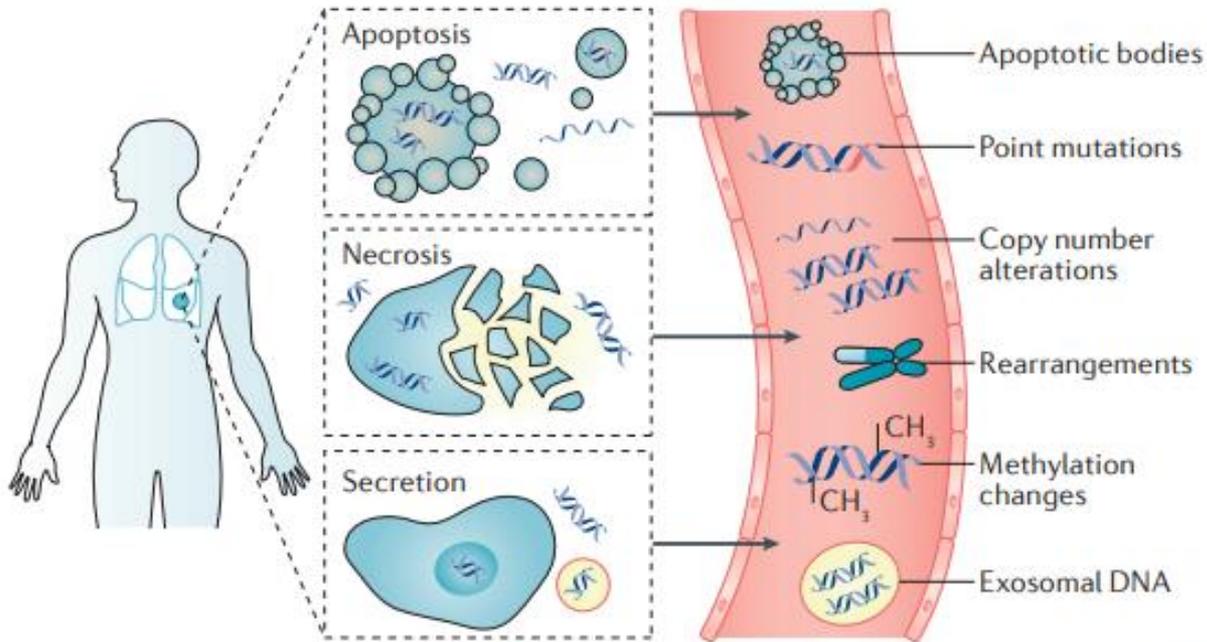


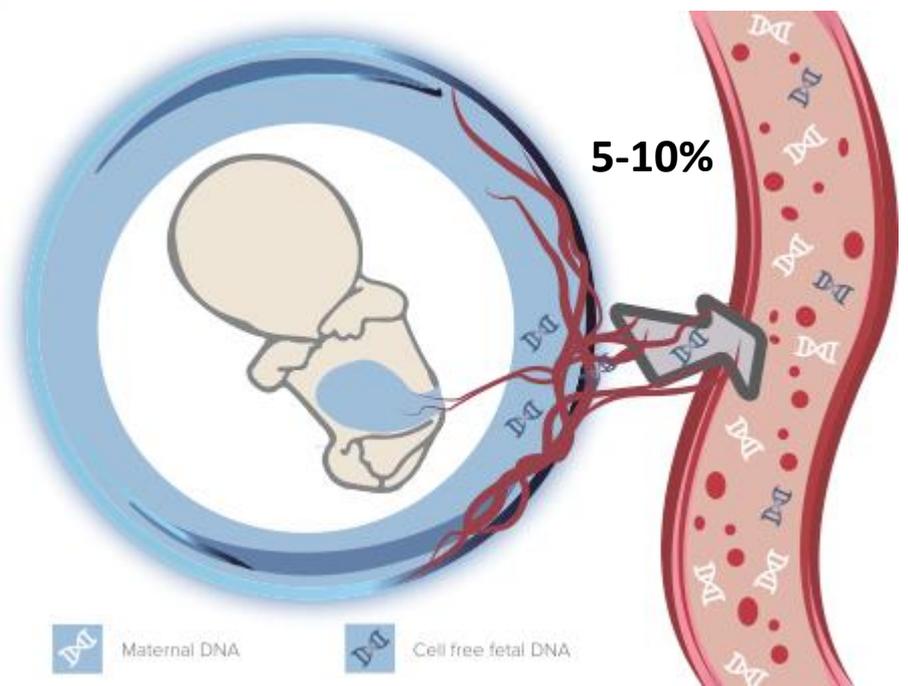
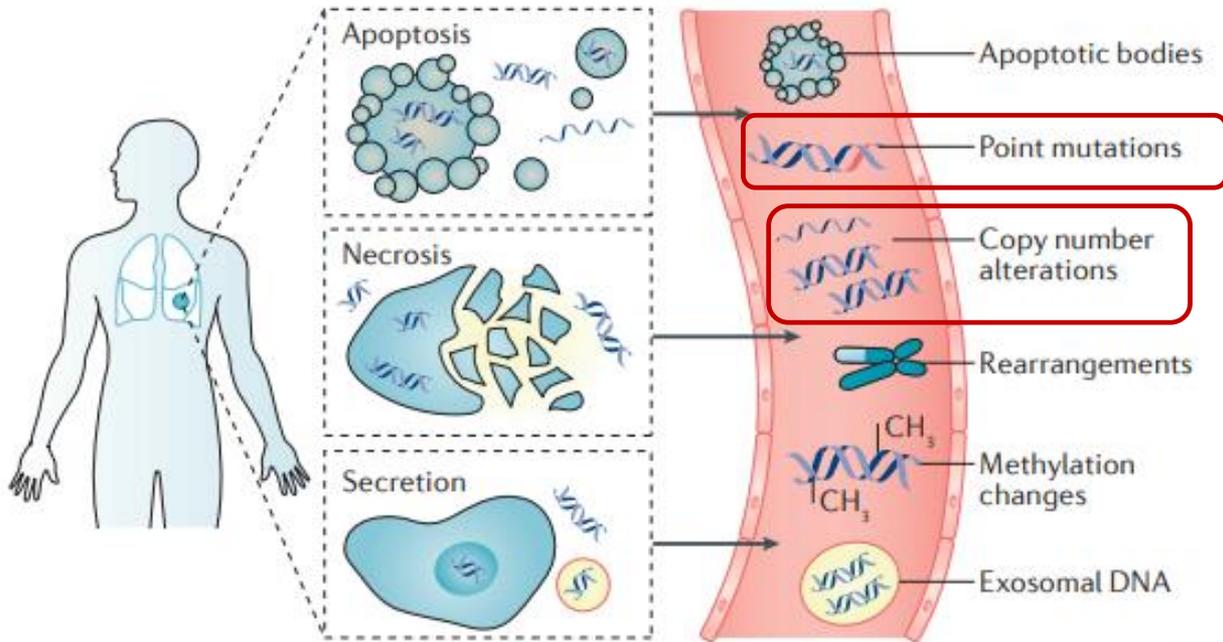
Анализ данных секвенирования внеклеточной ДНК

Шубина Екатерина м.н.с. лаборатория молекулярно-генетических методов «Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии им. В.И.Кулакова»

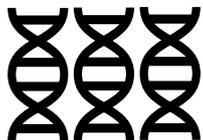
внеклеточная ДНК



внеклеточная ДНК



Неинвазивный пренатальный ДНК-скрининг (НИПС)



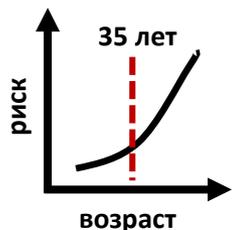
Анеуплоидия – число хромосом не кратное гаплоидному



Синдром Дауна – лишняя 21 хромосома, **1:700-1000** родов

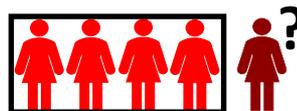
Синдром Эдвардса – лишняя 18 хромосома, **1:6000-8000** родов

Синдром Патау – лишняя 13 хромосома, **1:10000-16000** родов



Риск наличия анеуплоидий у плода увеличивается с возрастом матери

Комбинированный скрининг 1-го триместра
косвенные маркеры (УЗИ + биохимия)



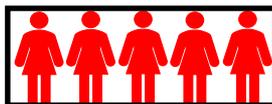
Выявляет не более **80%** анеуплоидий

Высокий риск



Более 80% ложноположительных результатов

Неинвазивный пренатальный ДНК-скрининг
прямой анализ ДНК плода в крови матери



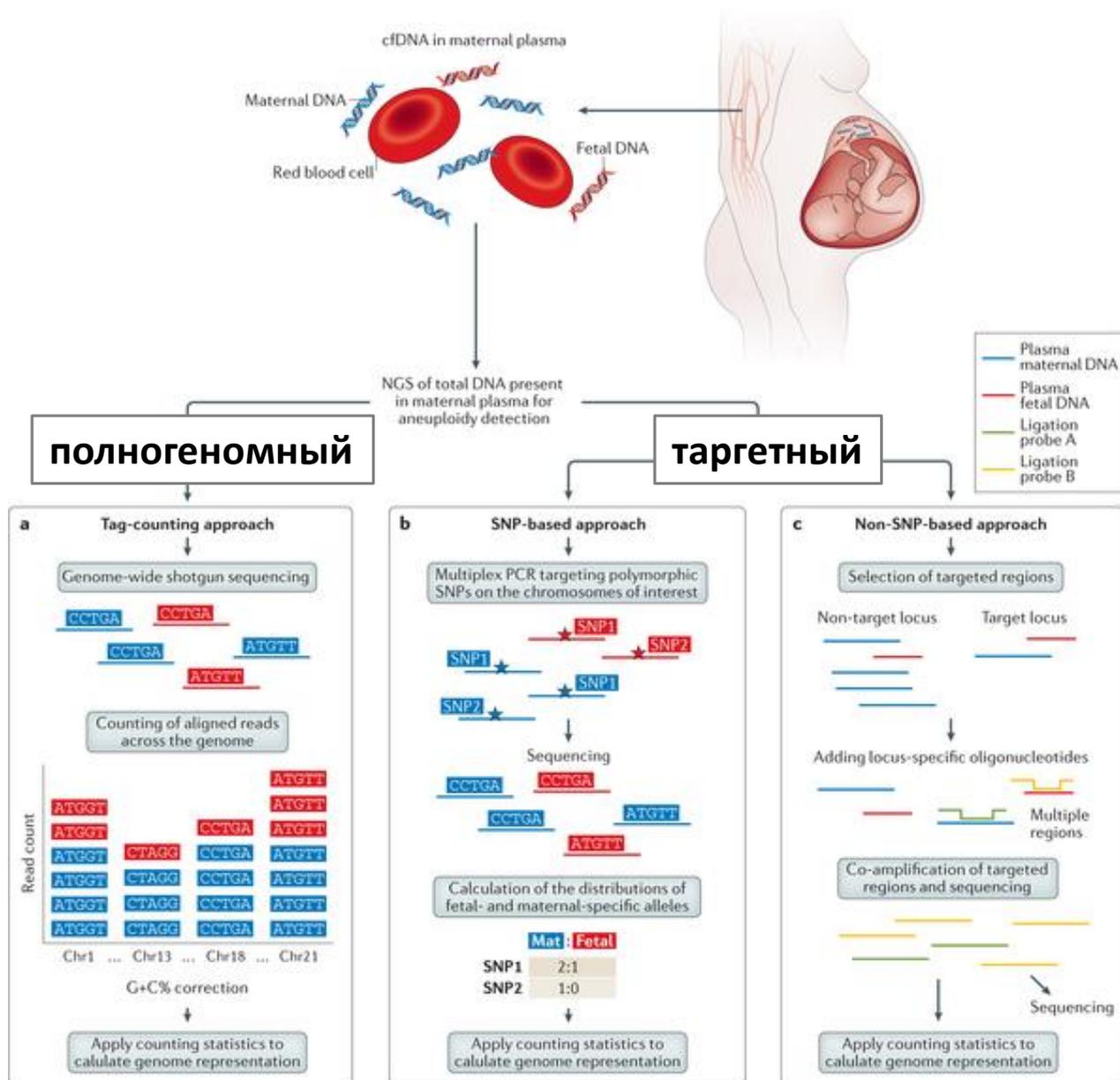
Выявляет **до 99%** анеуплоидий

Высокий риск



Менее 10% ложноположительных результатов

Подходы к проведению НИПС



Этапы анализа данных НИПС

- Выравнивание на референсный геном
- Фильтрация ПЦР-дубликатов
- Подсчет количества ридов для фрагментов фиксированной длины
- Фильтрация неуникальных фрагментов
- GC-коррекция
- Определение риска наличия анеуплоидии

Выравнивание на референсный геном

Bwa, bowtie2, TMAP....

Фильтрация ПЦР-дубликатов

samtools rmdup, Picard ...

Подсчет количества ридов для фрагментов фиксированной длины

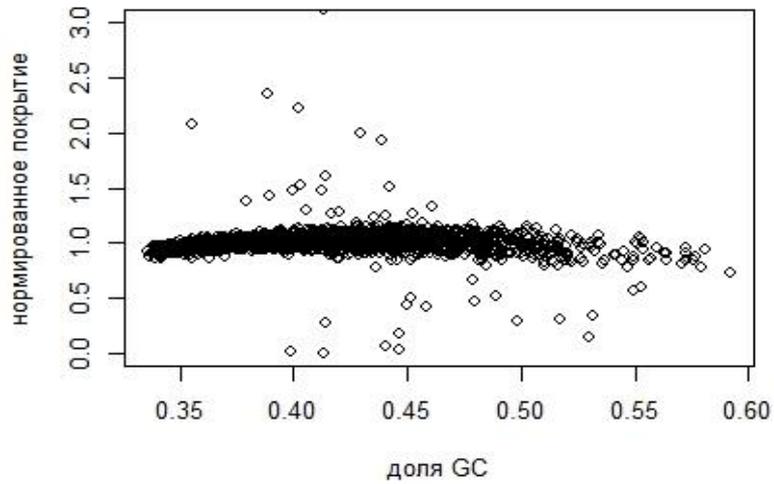
MQ>x, NM<y

Фильтрация неуникальных фрагментов

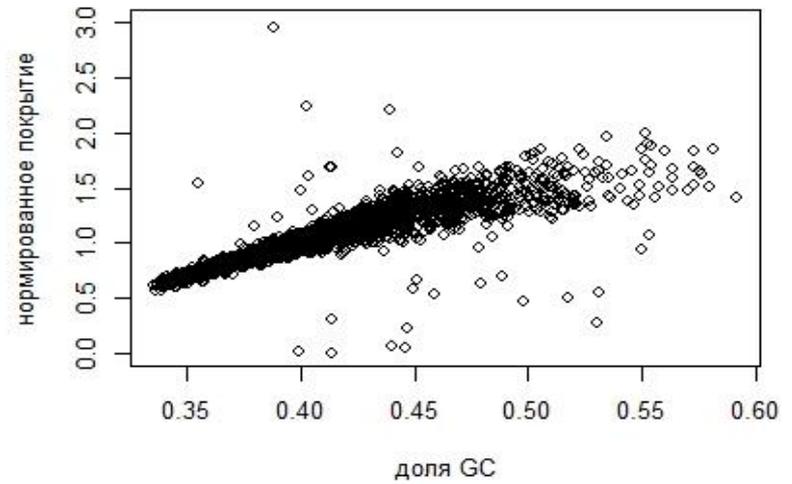
repeats/blacklists, «маппируемость» участка генома

GC-коррекция

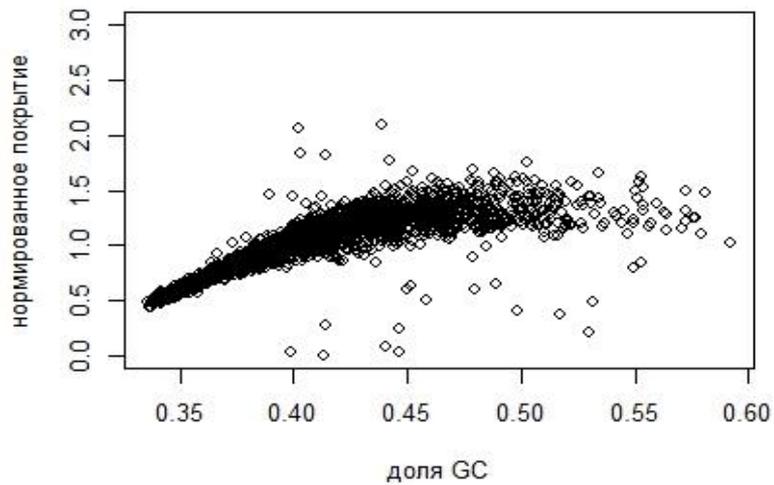
17_3_66



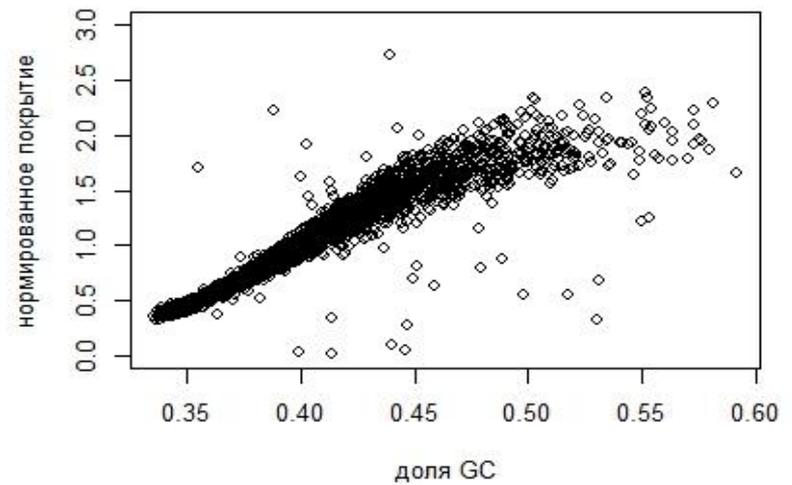
18_4_17

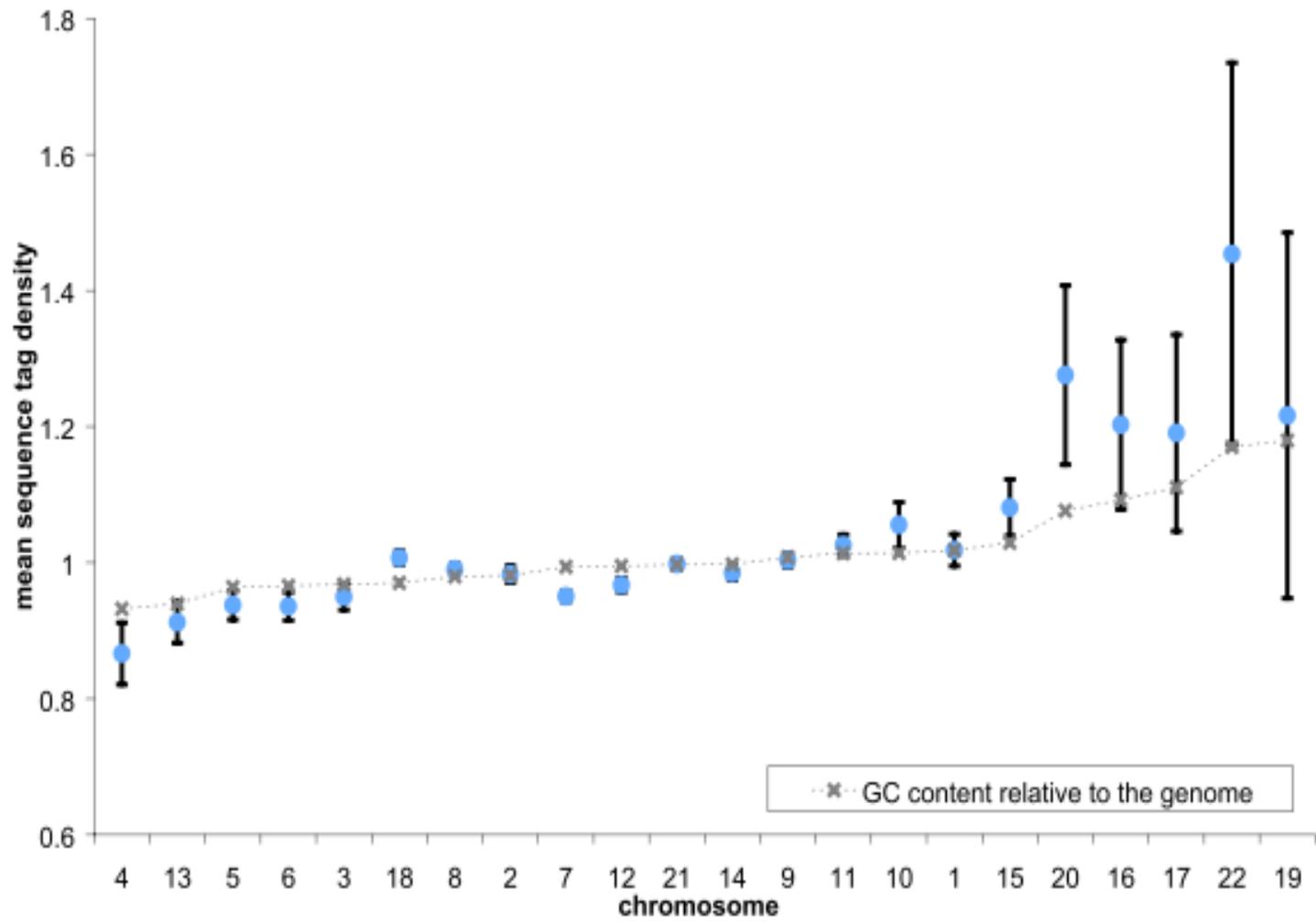


16_98_100

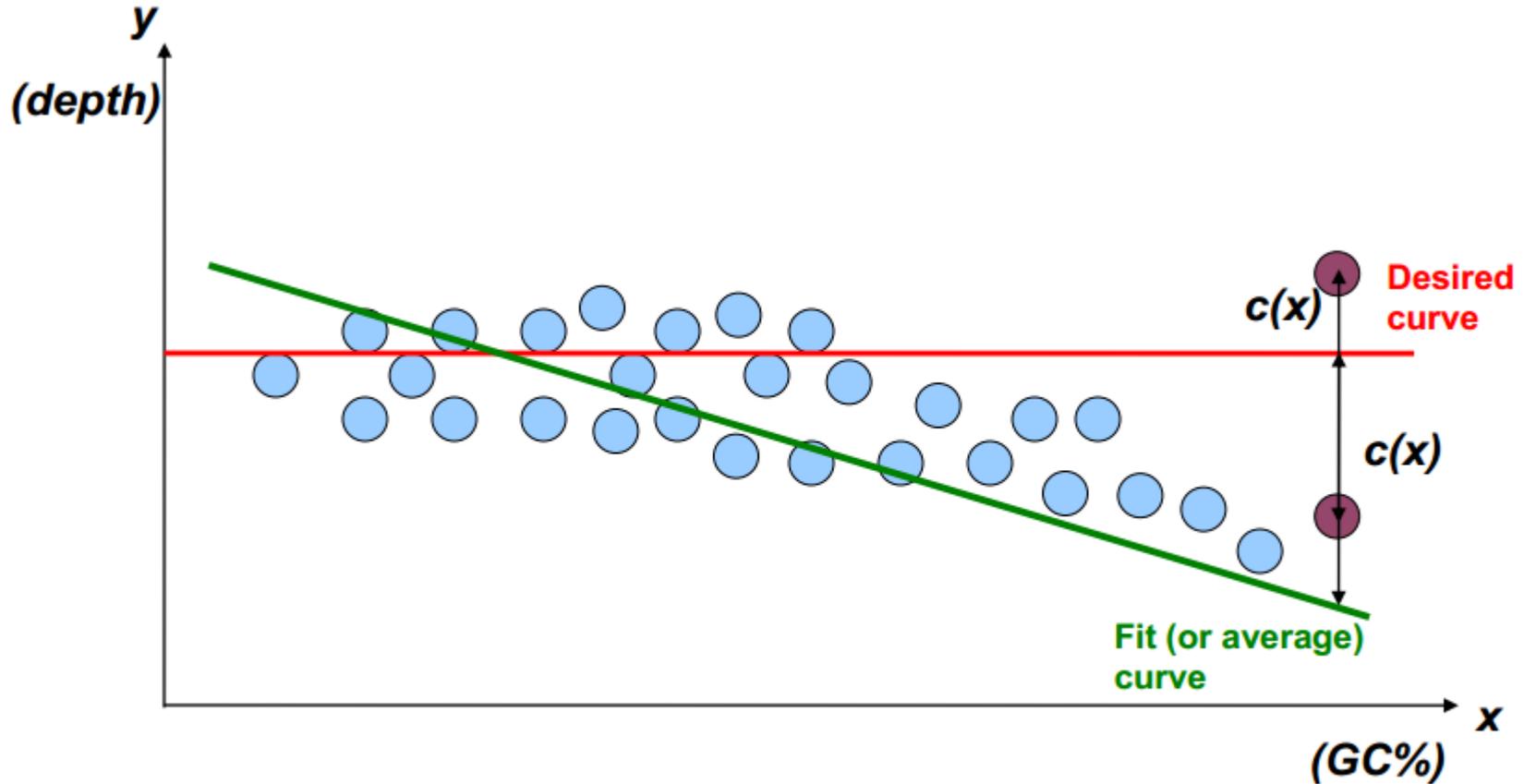


17_98_100



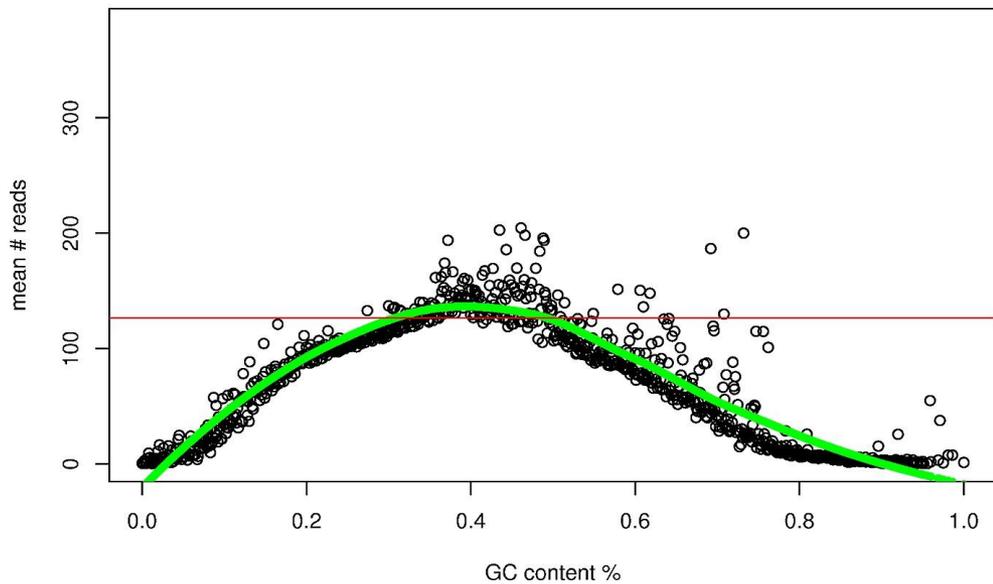


GC коррекция (LOESS)

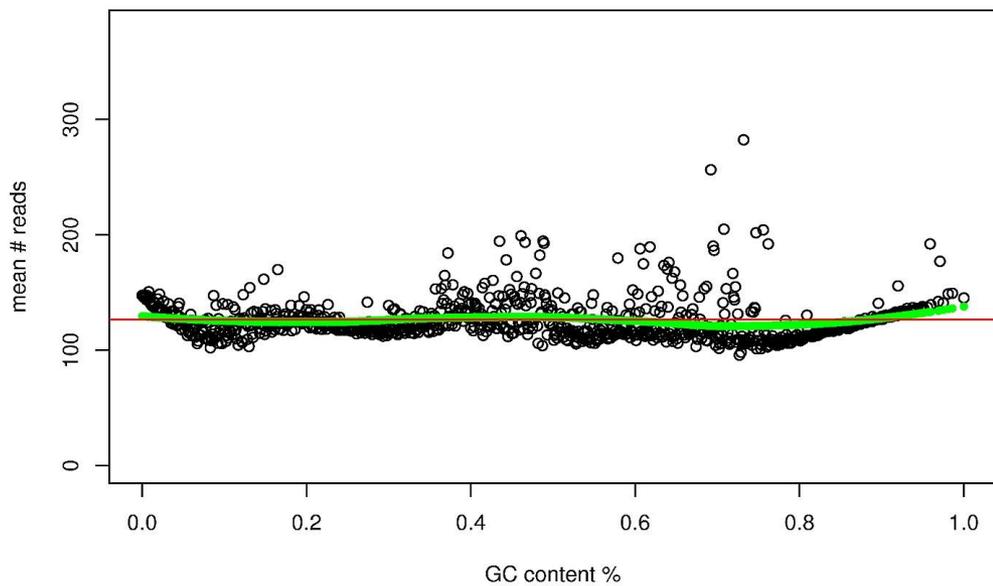


$$y' = y - c(x)$$
$$c(x) = f(x) - e(x)$$

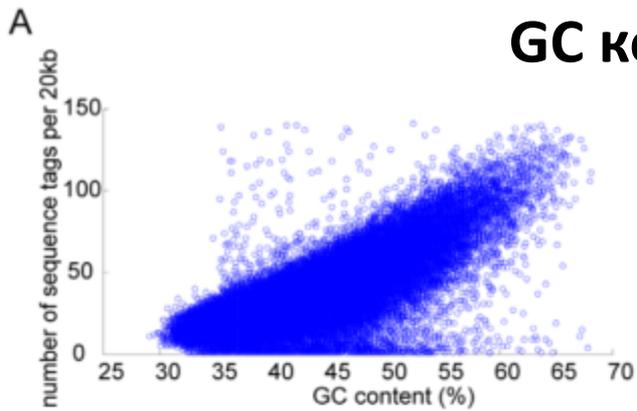
GC content bias – raw Data



GC content bias – post LOESS correction



GC коррекция (весовые коэффициенты)

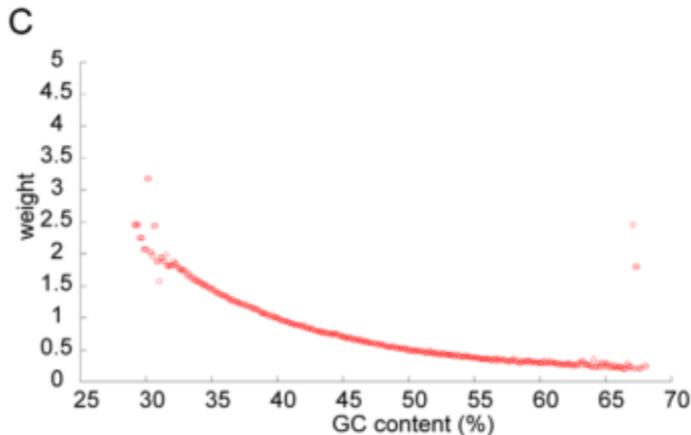
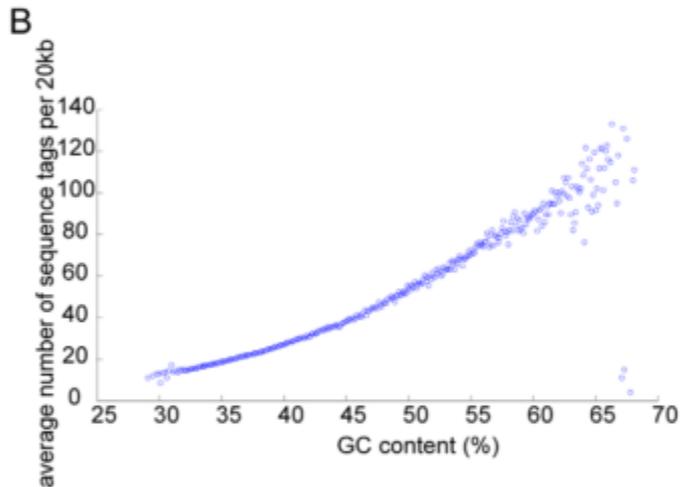


$$w = \frac{\bar{M}}{M_i}$$

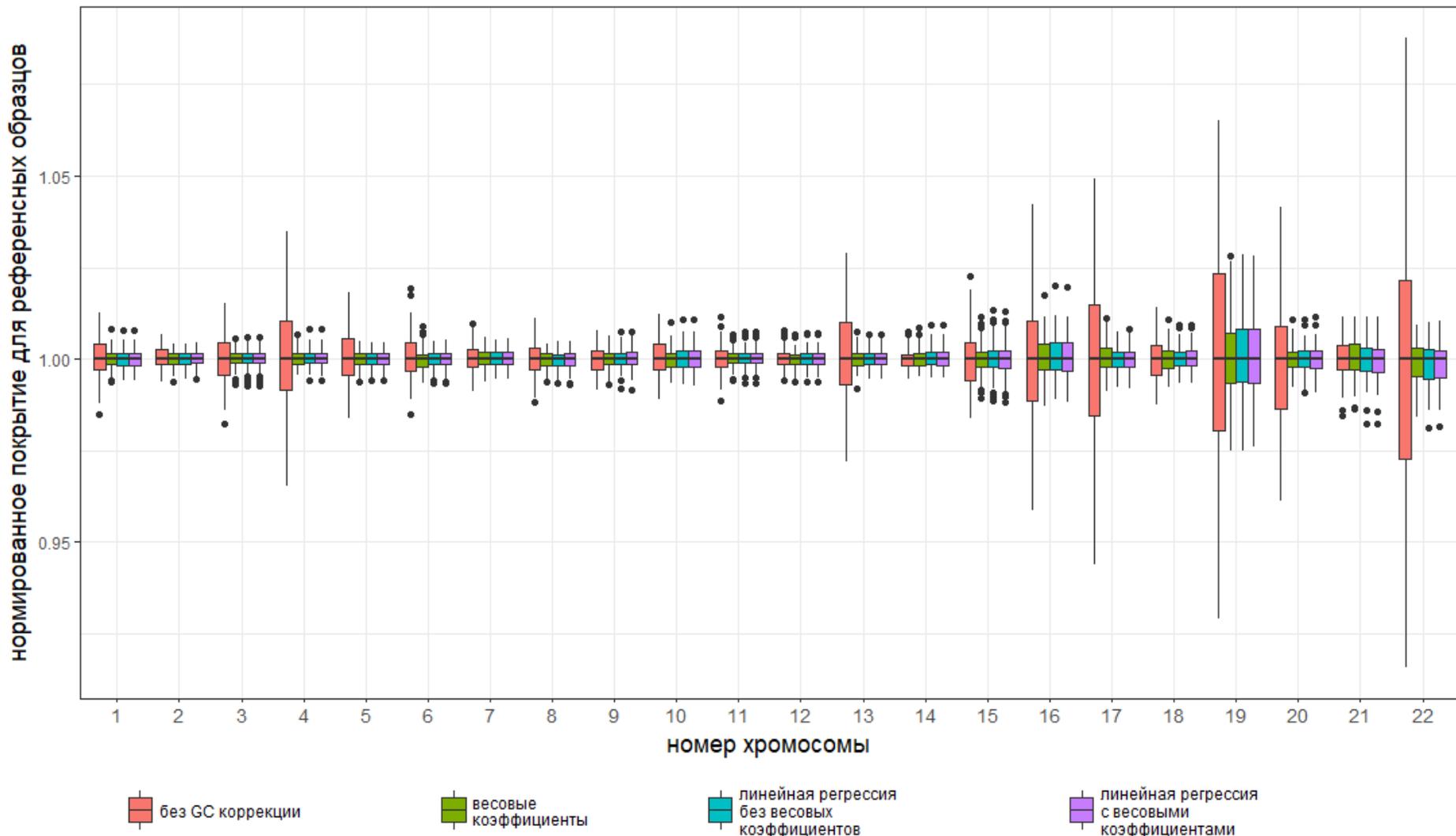
\bar{M} – среднее покрытие фрагмента фиксированной длины

M_i – среднее покрытие фрагментов с GC составом i

$$R_{corr} = R * w$$



Сравнение нормированного покрытия для всех аутосом референсных образцов с использованием различных способов GC-коррекции.



Определение риска наличия анеуплоидии

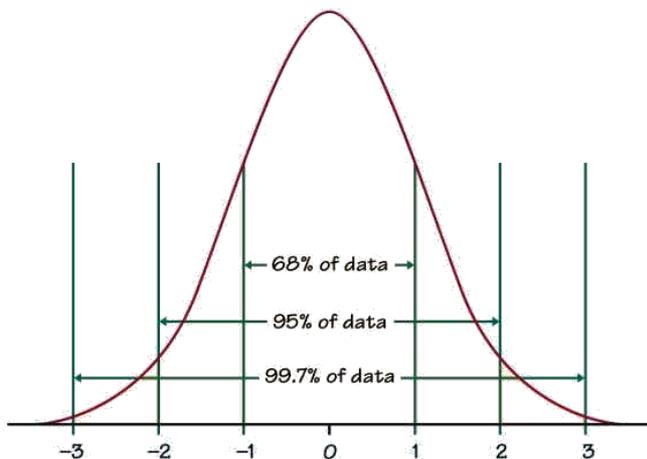
$$Z_{chrN} = \frac{\%chrN - \mu_{\%vhrNref}}{\sigma_{\%chrNref}}$$

$\%chrN$ – доля фрагментов, приходящаяся на исследуемую хромосому;

$\mu_{\%vhrNref}$ – среднее значение доли фрагментов, приходящихся на исследуемую хромосому для набора референсных образцов;

$\sigma_{\%chrNref}$ – стандартное отклонение доли фрагментов, приходящихся на исследуемую хромосому для набора референсных образцов.

$|Z| > 3$ – высокий риск анеуплоидии



Проблемы:

CNV у мамы – риск ложноположительных и ложноотрицательных результатов

Возможны «редкие анеуплоидии»

Normalized chromosome value

$$NCV_j = \frac{x_j - \mu_j}{\sigma_i}$$

x_j – рассчитанное отношение количества фрагментов для хромосомы j к количеству фрагментов для «референсной хромосомы» в исследуемом образце;

μ_j среднее значение отношений для хромосомы j в референсной выборке образцов,

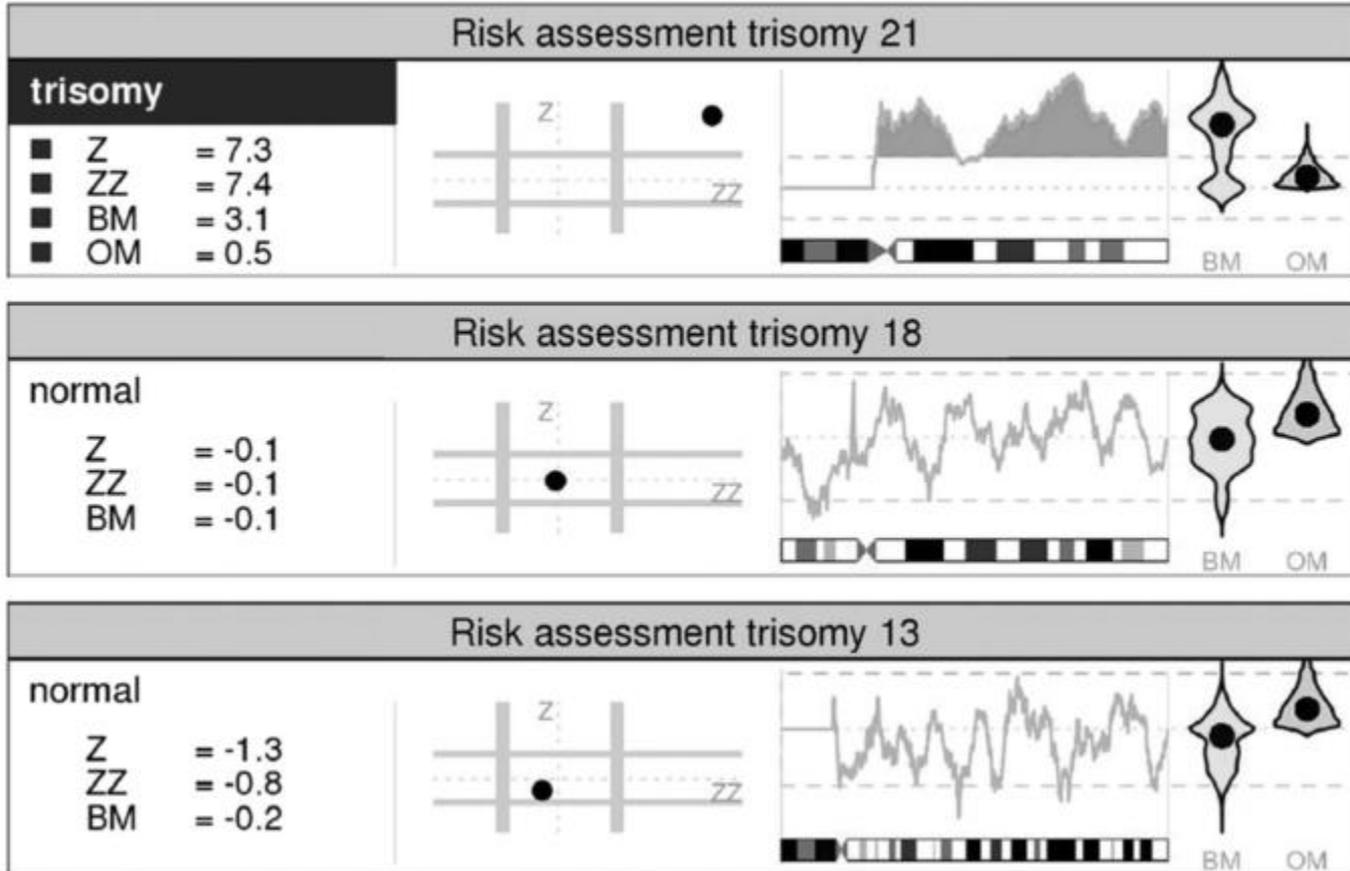
σ_i – стандартное отклонение отношений для хромосомы j в референсной выборке образцов.

Проблемы:

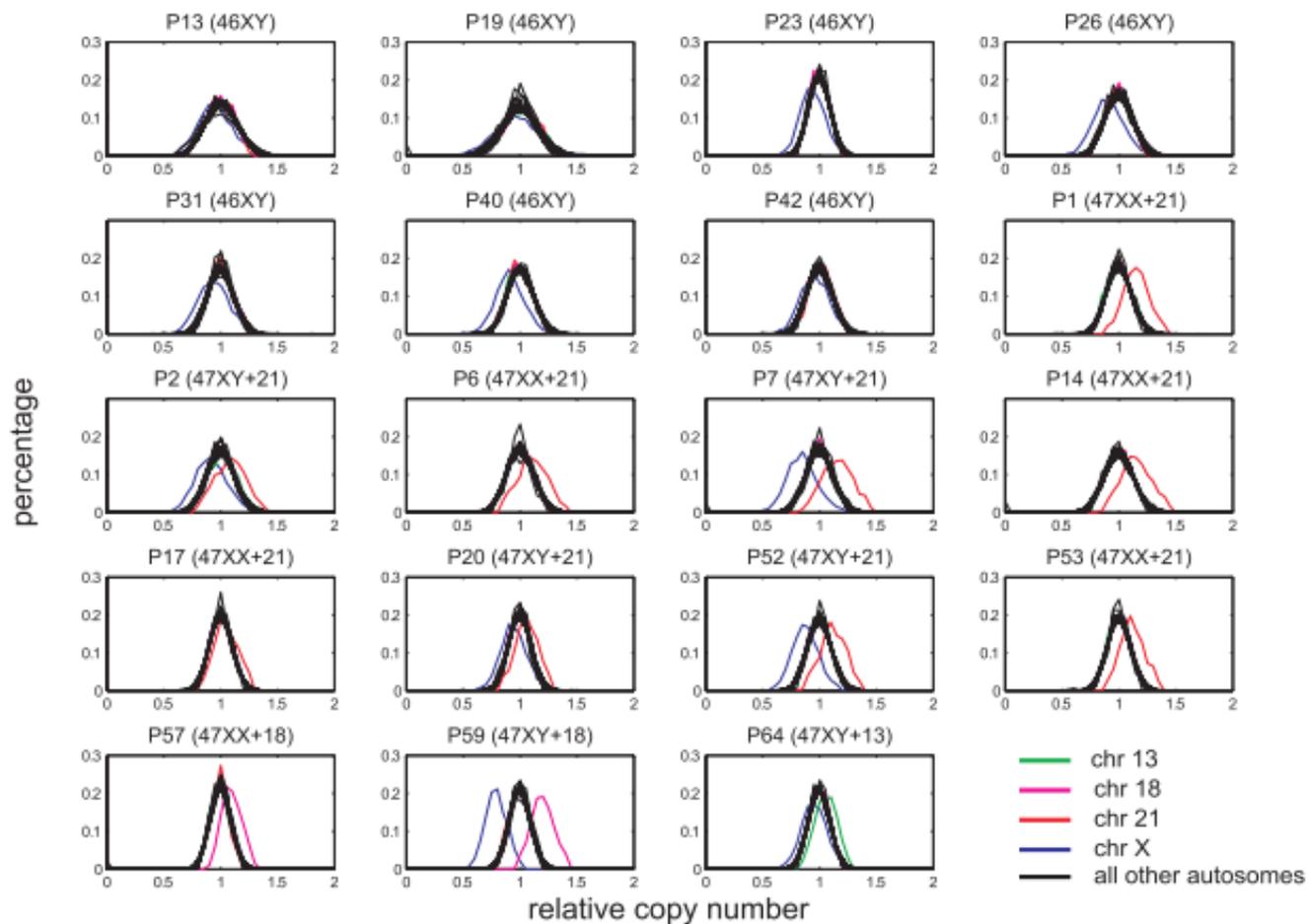
CNV у мамы – риск ложноположительных и ложноотрицательных результатов

Возможны «редкие анеуплоидии»

ZZ-score

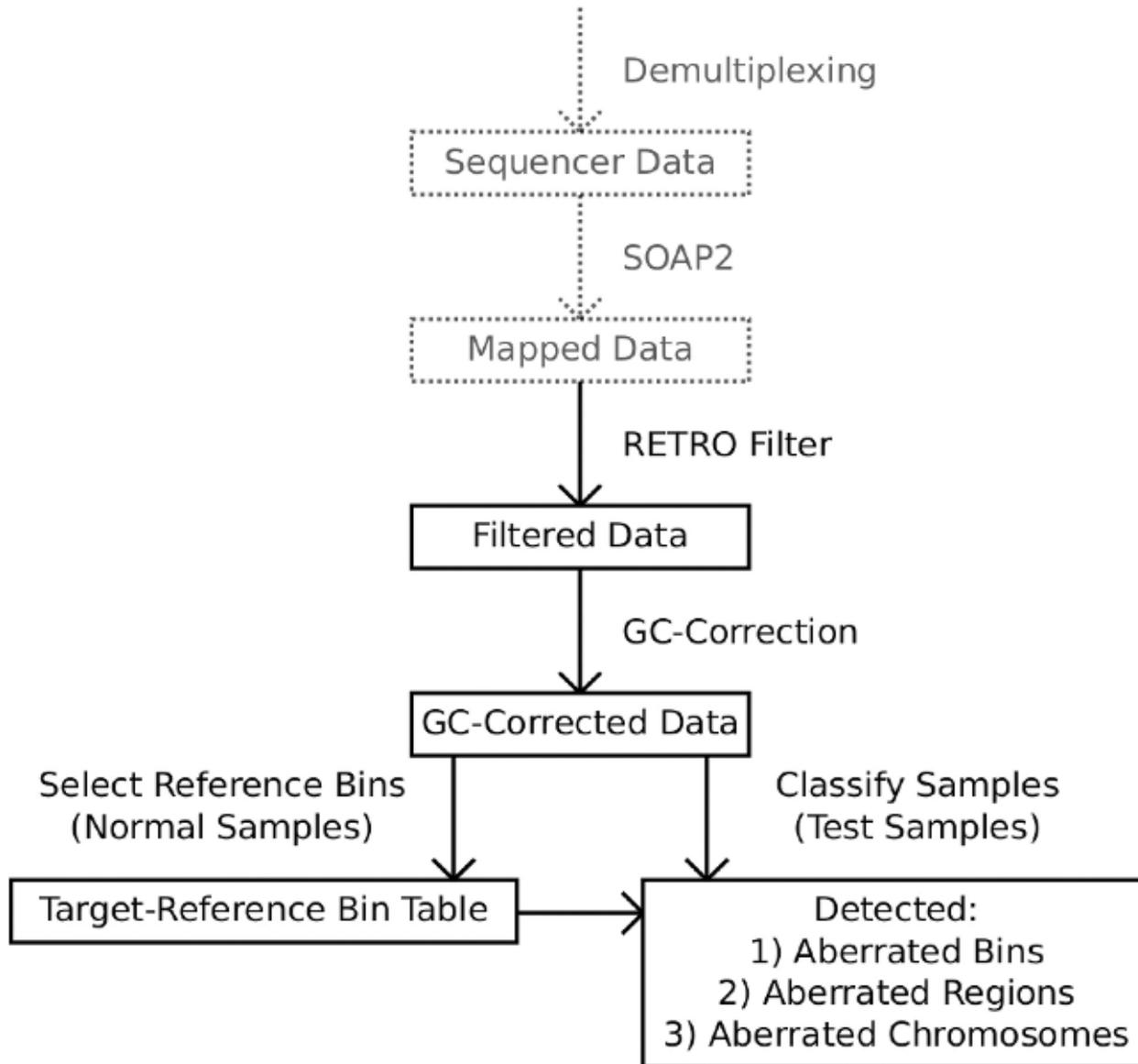


Тест стьюдента

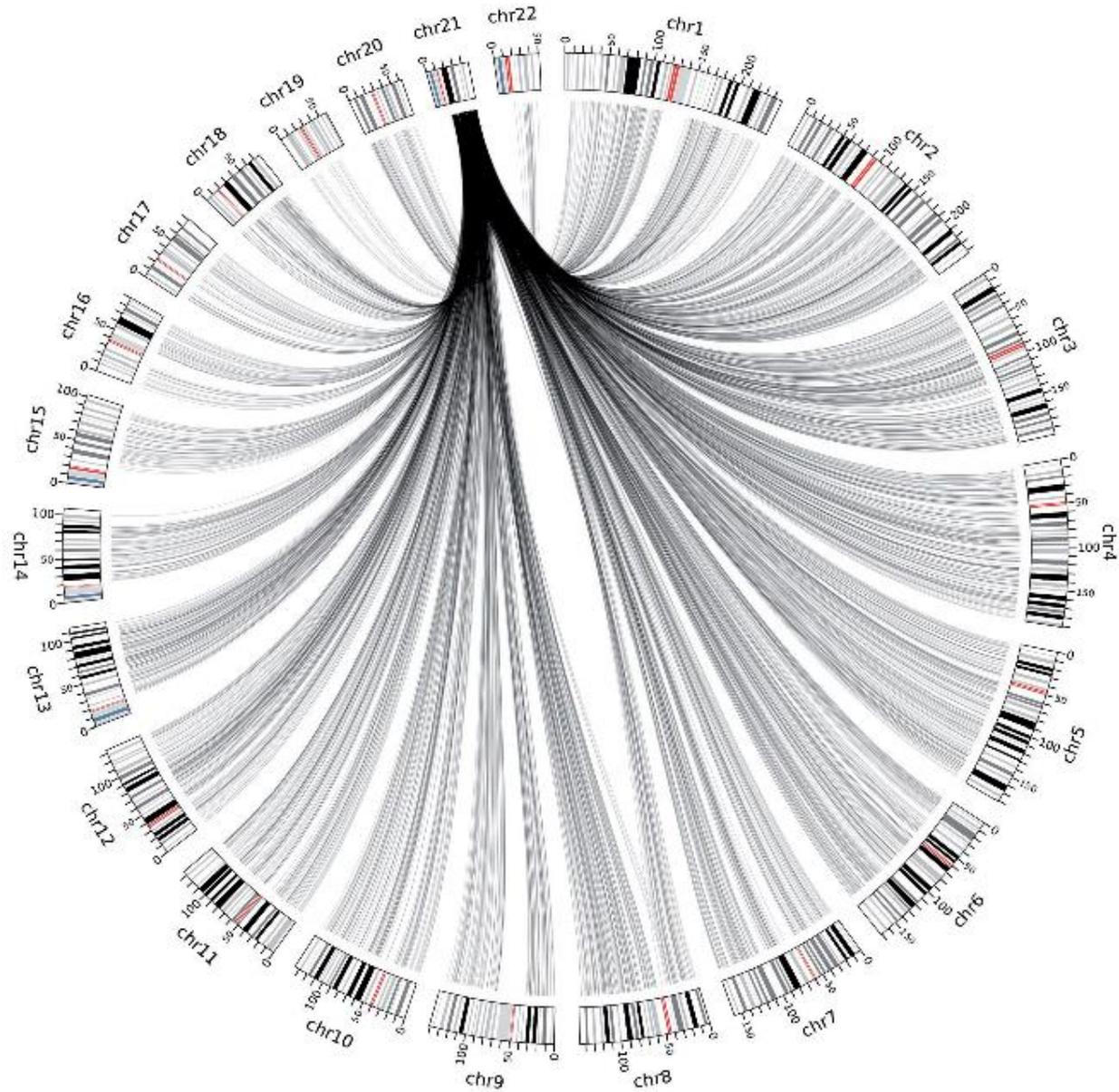


$$T_{xi} = \frac{\mu_x - \mu_i}{2 \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_i^2}{n_i}}}$$

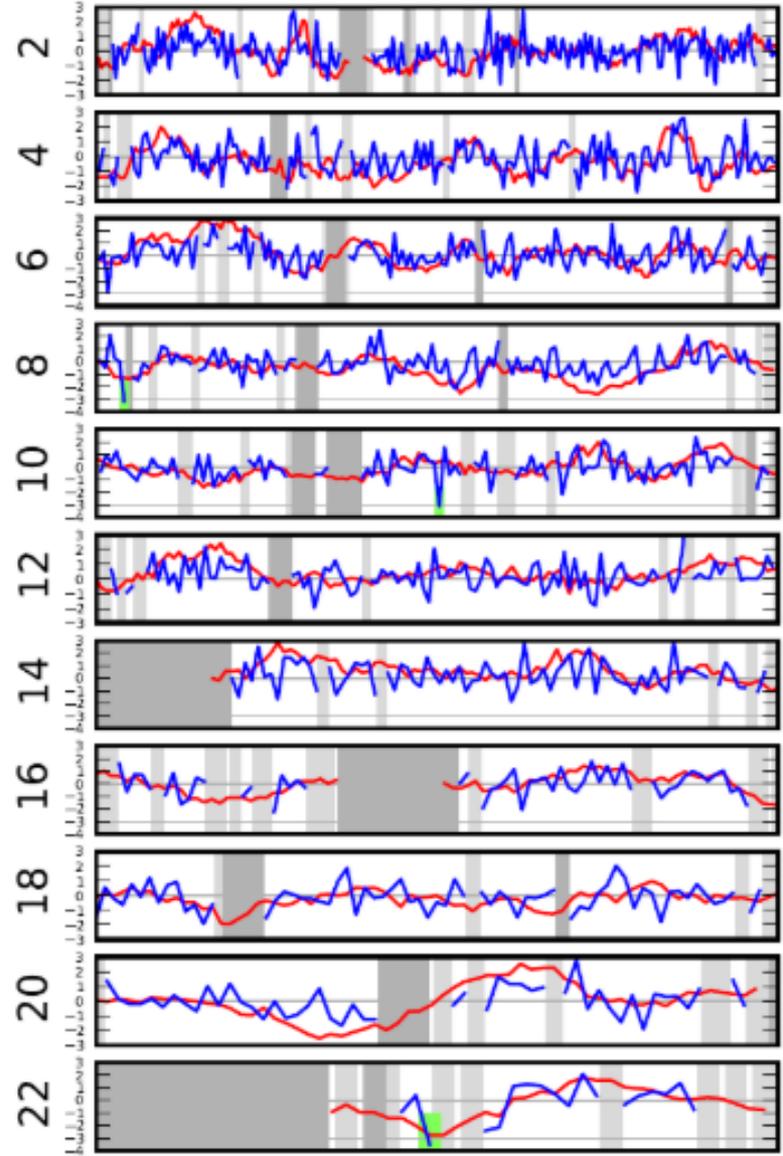
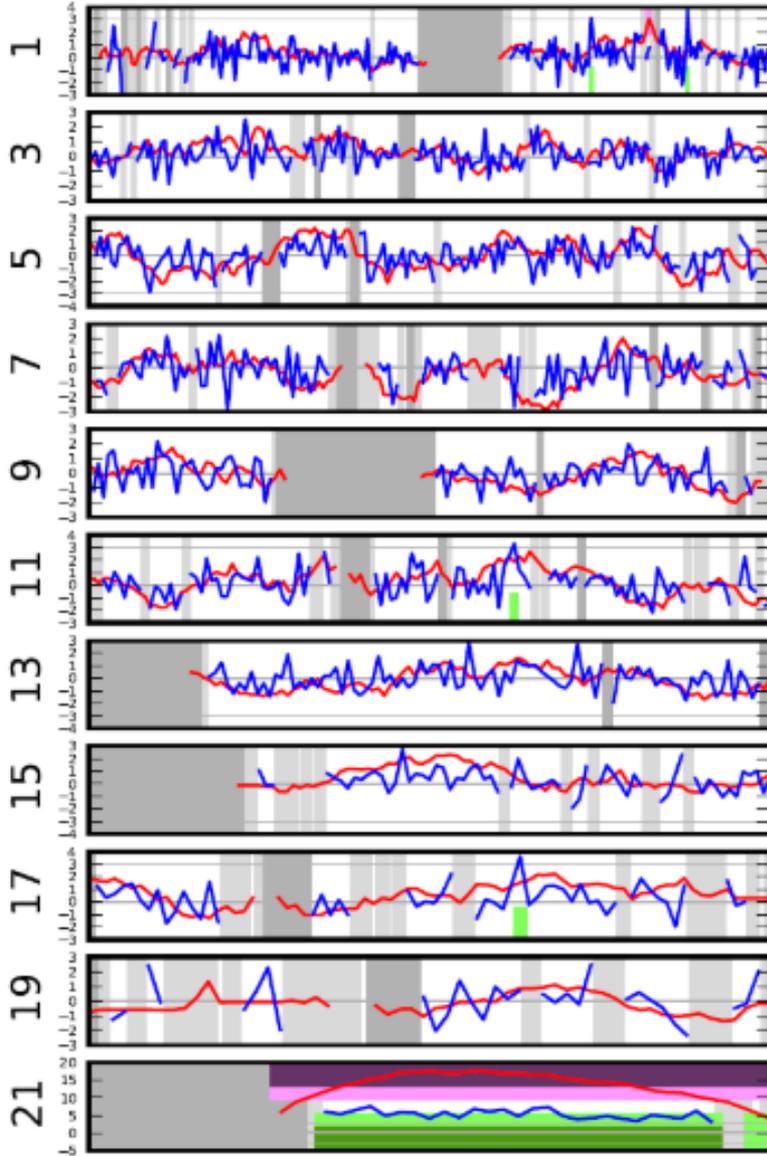
Wisecondor



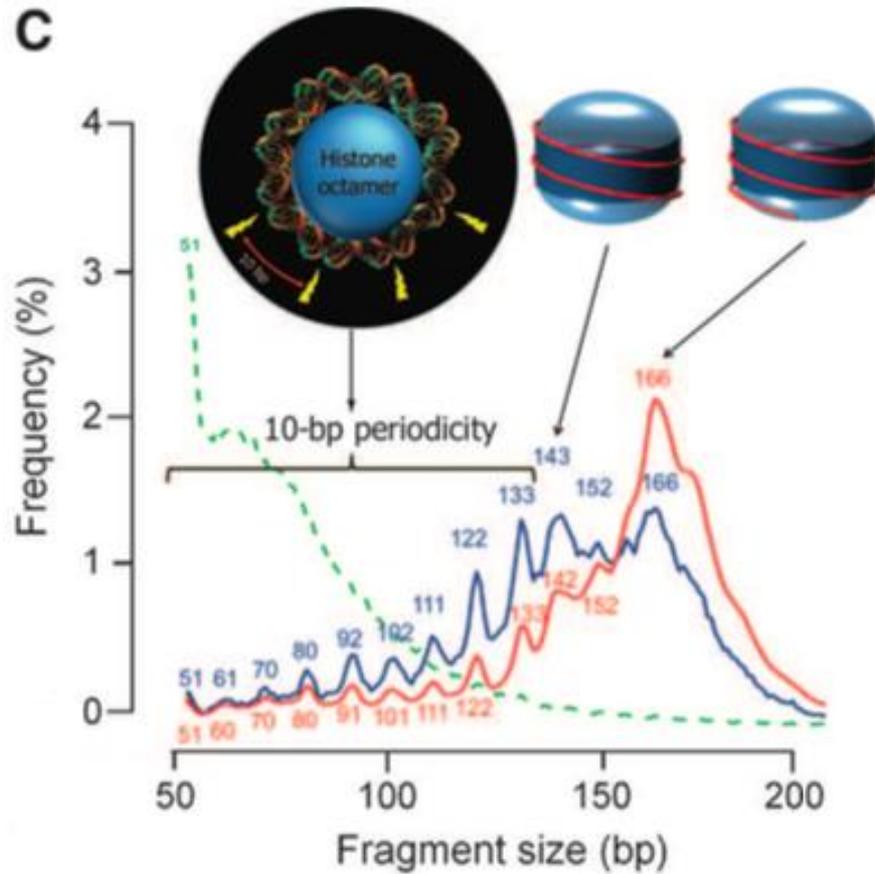
Wisecondor



Wisecondor



Использование информации о длине чтений



Lo et. al
Prenat Diag. 2010

Size-based Z-score

$$\Delta F_{chrN} = P(\leq 150)_{chrN} - P(\leq 150)_{chrRef}$$

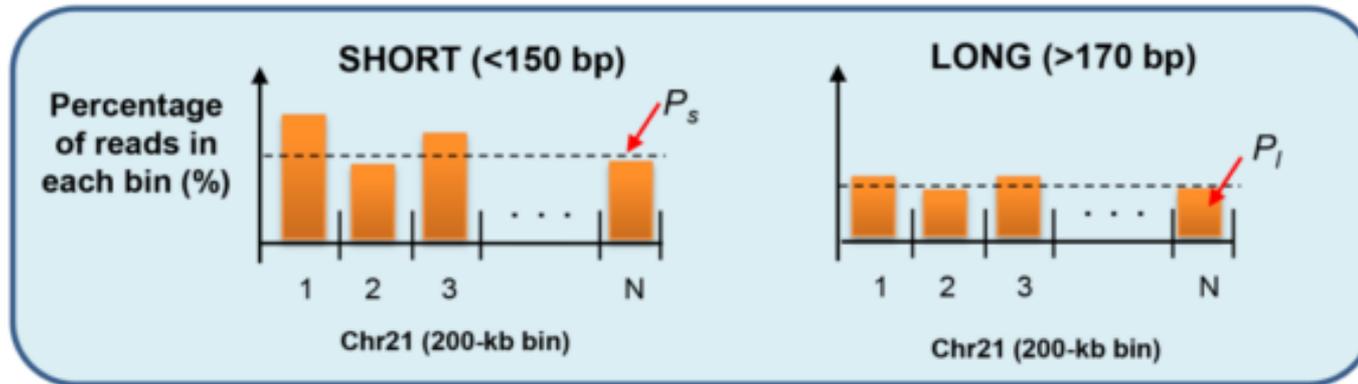
$P(\leq 150)_{chrN}$ Доля ридов короче 150 п.н. на исследуемой хромосоме

$P(\leq 150)_{chrRef}$ Доля ридов короче 150 п.н. на референсных хромосомах (все аутосомы кроме 13, 18, 21)

$$z\text{-score}_{chrN} = \frac{\Delta F_{chrN_sample} - mean \Delta F_{chrN_ref}}{SD \Delta F_{chrN_ref}}$$

COFFEE

Assigning DNA fragments into bins along chromosome 21 according to sizes

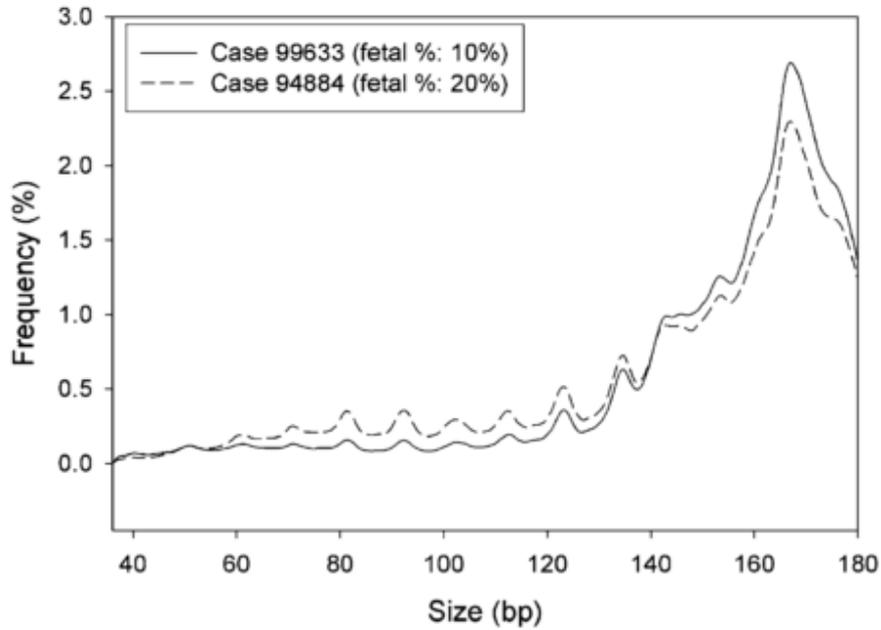


**Paired Mann-Whitney rank-sum test
across bins between short and long DNA
categories (i.e. P_s v.s. P_l)**

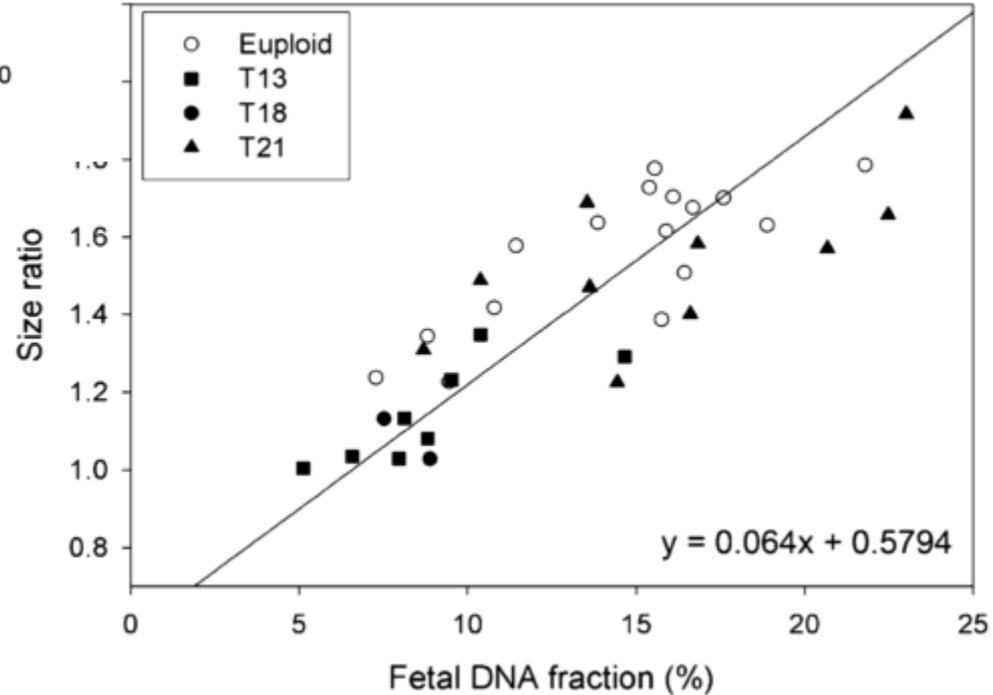
Trisomy: p-value < 0.001

Euploidy: p-value \geq 0.001

Size-based определение доли плодовой ДНК



$$\text{Size ratio} = \frac{P(100 - 150)}{P(163 - 169)}$$

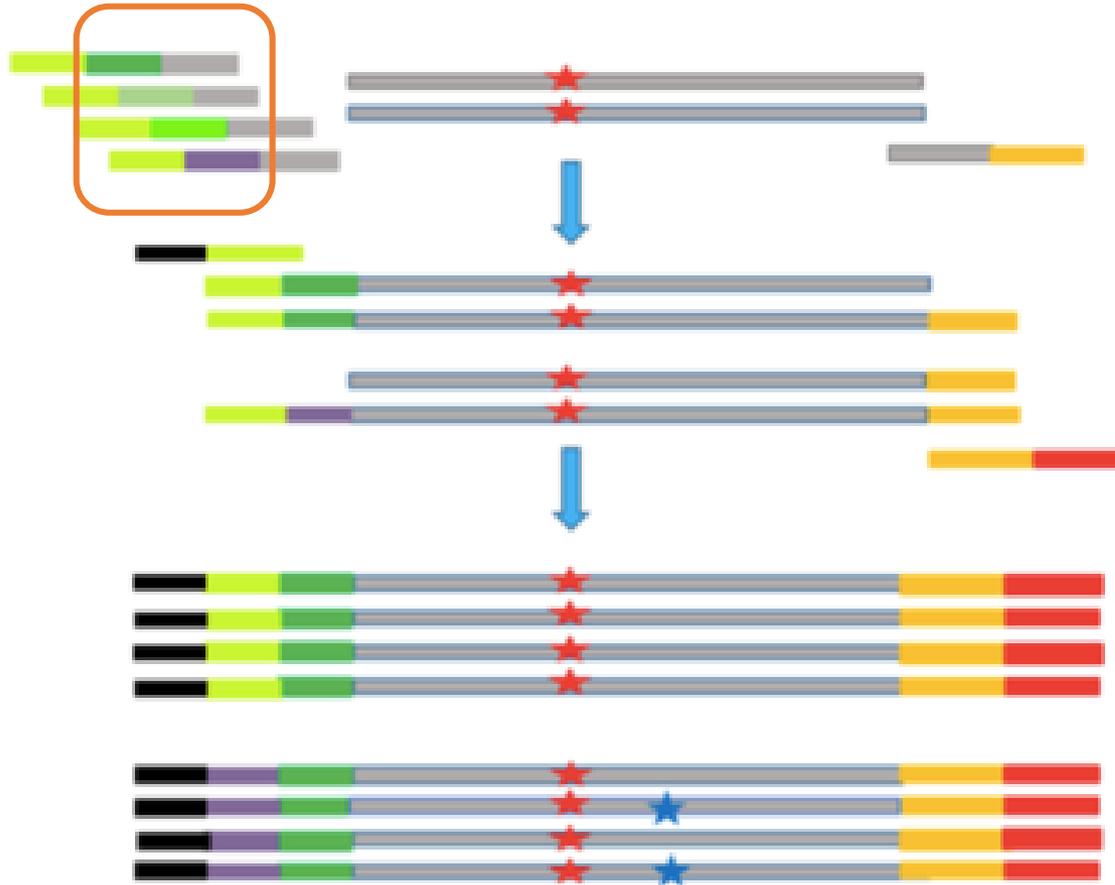


Определение наличия точечных мутаций

Проблемы:

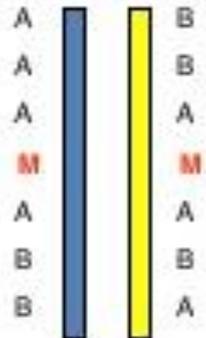
- Ошибки секвенирования
- Точность подсчета доли мутантных аллелей
- Allele dropout

Ошибки секвенирования/точность подсчета – молекулярные индексы

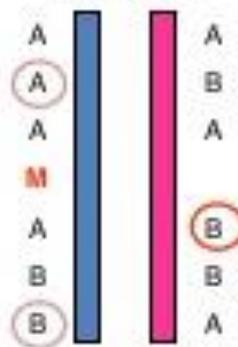


Allele dropout - гаплотипирование

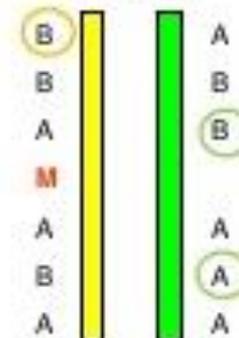
Affected child



Mother



Father



Спасибо за внимание!

