

Programme's name	Features	Language	link
BaseVar	Variant calling, call SNP from NIPT sequencing data (поиск SNP в последовательностях из исследования, определение частот аллелей)	python	https://github.com/ShujiaHuang/basevar – package itself https://pubmed.ncbi.nlm.nih.gov/30290141/ - article of interest
Mapper	Инструмент для выравнивания коротких последовательностей фиксированной длины для NIPT и CNV, когда задействовано низкое покрытие. Геном человека индексируется до начала выравнивания. Не справляется с мисматчами и пропусками <ul style="list-style-type: none"> Fixed-length short sequences Precise matching (mismatches and gaps are not supported) index-based sequence alignment 	C, C++	https://github.com/jia-zhuang/mapper Статью не нашла
hoobari	Поиск SNPs and indels, их наследование Байесовский алгоритм. 1. Определение fetal fraction, снижение глубины прочтений 2. Генотипирование родителей 3. Собственно входные данные: первичные вероятности на основе родительских последовательностей и фетальной фракции. Выходные данные: homozygous to the reference allele (0/0), heterozygous (0/1), and homozygous to the alternate allele (1/1)	python	https://github.com/nshomron/hoobari https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6396420/ - ссылка на статью непосредственно в репозитории, 2019 https://www.sciencedirect.com/science/article/pii/S2001037020305572 - 2021
NIPT-FFMLP	Модель на основе нейронной сети, обученная на признаках в числе прочтений в окне для определения фетальной фракции в НИПТ.	Python, Perl	https://github.com/happenywong/NIPT-FFMLP Статью не нашла, поэтому подробнее информации нет
Wisecondor (Within-Sample Copy Number aberration DetectOR)	Определение copy number variation (CNV) в неглубоком WGS (low-coverage). Сравнение числа прочтений на тестовом образце с участками на других хромосомах этого же организма, которые ведут себя таким же образом. 1. Разделить референсный геном на бины, определить глубину прочтения для каждого 2. Определение GC-содержание каждого бина, использовать LOWESS функцию (GC-count/read depth). Получить GC-corrected read frequency. 3. Для каждого бина провести сравнение нормализованных частот прочтения с таковыми из набора нормальных (диплоидных) образцов с использованием z-score. 4. Посчитать Евклидово расстояние для каждого из бинов (by summing the squared GC-normalized read frequency differences over a set of normal samples). 5. «Внутриобразцовое» ('within-sample') сравнение бинов с похожим поведением (т.к. разные районы генома различаются в характеристиках частоты прочтений от одного образца к другому, что осложняет межобразцовое сравнение). Кроме того, это дает возможность больше доверять результатам, так как все регионы находятся в одинаковых условиях эксперимента Bin size 1 Mb	Python	https://github.com/VUmcCGP/wisecondor https://academic.oup.com/nar/article/42/5/e31/1063962 - Article of interest
Wisecondor X	1. For every sample, every bin was tested against its own set of reference bins within the same sample using a z-score method based on GC-normalized read frequencies Практически то же самое, что и предыдущий метод, однако способен работать с половыми хромосомами, более продвинутый	Python, R	https://github.com/CenterForMedicalGeneticsGhent/WisecondorX https://academic.oup.com/nar/article/47/4/1605/5253050

NIPTUNE	<ol style="list-style-type: none">1. Выравнивание последовательностей2. Используя скрипт <i>gsc.py</i> из Wisecondor, провести коррекцию по GC-содержанию.3. Разделение каждого образца на бины длиной 1 Mb.4. Применение функции LOESS с использованием референса GRCh37 human genome для получения нормализованного GC-содержания (counts) на один бин.5. Использование Principal component analysis (PCA) для определения образцов с несбалансированным содержанием бинов как потенциальные ошибки качества библиотеки или секвенирования, ошибки выравнивания или патологию матери6. Авторский метод MagicY, оценивающий пропорцию прочтений семи У-специфичных регионов по отношению к числу прочтений У-хромосомы позволяет определить пол плода7. Использование gaussian mixture model с 2 gaussians, а также проекция порога MagicY на ось фетальной фракции и дальнейшее внедрение KNN позволило отнести образцы к определенному полу.8. Чтобы посчитать фетальную фракцию, были использованы количества прочтений У-хромосомы и Seqff на основе предобученных бинов в модуле <i>peso.py</i>.9. Модуль <i>sao.py</i> содержит WisecondorX (хромосомные aberrации), threshold(Z-score)=5	Python	https://github.com/UCA-MSI/NIPTUNE/blob/master/README.md (upon the request for access) https://academic.oup.com/bib/article/23/1/bbab380/6370845
---------	--	--------	--

Programme's name	Features	Language	link
seqFF	<p>Multivariate regression algorithm with a fixed bin size of 50 kb of autosomal sequence counts. The model was designed directly on bin-wise copy number features.</p> <p>ElasticNet and Weighted rank selection criterion – weighting of bins that are predictive of fetal DNA fraction.</p> <p>Подсчет количества чтений, выровненных в определенных аутосомных регионах и применение схемы взвешивания, полученной на основе многомерной модели. Универсальная применимость к плодам XX, XY.</p>	Python	https://github.com/HyunbinCho/seqff https://obgyn.onlinelibrary.wiley.com/doi/10.1002/pd.4615 - Article of interest
GenomomFF	<p>Measuring fetal fractions using multiple regression model (Elasticnet).</p> <p>Подход незначительно отличается от того, что используется в seqFF, за исключением того, что используется больший размер бина, а корреляция была выше (fetal fraction x read size). Используются так же FFY, SNP</p>	Python, R	https://github.com/TheragenGenomecare/GenomomFF https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9110588 - Article of interest
PREFACE	<p>Principal component analysis, neural network (NN) weighs the computed principle components (PCs) to model fetal-induced variance.</p> <p>Modeling of copy number profiles and FFY – fetal fraction Y reads.</p> <p>Обучение модели происходит на ограниченном количестве ретроспективных данных поверхностного полногеномного секвенирования (shallow-depth WGS) , что исключает межлабораторную погрешность</p>	R	https://github.com/CenterForMedicalGeneticsGhent/PREFACE https://www.ncbi.nlm.nih.gov/pmc/articles/31219182/ - Article of interest
Fetalfraction-SNPimpute	<ol style="list-style-type: none"> 1. SNP imputation (because of the reduction of number of reads and lowering the coverage) 2. Performing MAF (Minor allele frequency) filtering to reduce errors. It helps to remove SNV with an extremely low MAF <p>Более высокое покрытие и большее количество SNP приводит к более точному сравнению фрагментов ДНК матери и плода. Однако это дорого и долго. Поэтому, снизив покрытие, можно уменьшить стоимость этапа секвенирования, но точность измерения фракции плода снижается по мере уменьшения охвата. Чтобы это компенсировать, провели процедуру замещения пустующих данных на основе известного распределения (импутация).</p>	Python (run on ubuntu)	https://github.com/KMJ403/fetalfraction-SNPimpute https://pubmed.ncbi.nlm.nih.gov/29126132/ - Article of interest
NIPTeR	<ol style="list-style-type: none"> 1. Создать NIPTSample с использованием функции bin_bam_sample 2. Снижение GC-смещения LOESS-функцией и bin weight based функцией 3. Метод хи-квадрат определяет перераспределенные бины в контрольной группе и корректирует их как в контроле, так и в интересующем образце <p>С использованием долей чтений, картированных на определенные хромосомы, можно определить наличие или отсутствие трисомий для <i>NIPTSample</i> на основе <i>NIPTControlGroup</i> с использованием 3 разных алгоритмов: Z-score, NCV (Normalized Chromosome Value), RBZ (Regression based Z-score) – прогнозирование хромосомной фракции для образца и контрольной группы.</p> <p>Match QC score – статистика для контрольных групп (насколько хорошо образец вписывается в контрольную группу на основе доли чтений, сопоставленных с различными хромосомами)</p>	R	https://github.com/molgenis/NIPTeR https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-018-2557-8 https://cran.r-project.org/web/packages/NIPTeR/vignettes/NIPTeR.html