

CLAIMS

We Claim:

1. A method for facilitating prenatal diagnosis of a genetic disorder from a maternal sample associated with a pregnant woman, the method comprising:
 - adding, to the maternal sample, a set of quality control template (QCT) molecules associated with the genetic disorder, the set of QCT molecules comprising:
 - target-associated regions with sequence similarity to a target sequence region of endogenous target molecules, and
 - variation regions with sequence dissimilarity to a sequence region of the endogenous target molecules;
 - generating a co-amplified mixture based on co-amplifying the set of QCT molecules and nucleic acid molecules comprising the target sequence region;
 - sequencing the co-amplified mixture;
 - computationally determining a unique number of the set of QCT molecules, based on a number of the variation regions that are distinct and detected from QCT molecule sequence reads from the sequencing, wherein the QCT molecule sequence reads correspond to the set of QCT molecules;
 - calculating the average QCT sequencing depth based on dividing a number of the QCT molecule sequence reads by the unique number of QCT molecules;
 - determining an absolute count of the endogenous target molecules based on dividing a total read count for the endogenous target molecules by the average QCT sequencing depth;
 - determining an absolute count of endogenous reference molecules based on dividing a total read count for the endogenous reference molecules by the average QCT sequencing depth; and
 - facilitating the prenatal diagnosis of the genetic disorder based on a comparison between the absolute count of endogenous target sequences and the absolute count of endogenous reference sequences.
2. The method of Claim 1,
 - wherein the genetic disorder comprises a single gene disorder,
 - wherein determining the absolute count of the endogenous target molecules comprises determining the absolute count of the endogenous target molecules comprising a mutation

associated with the single gene disorder, based on dividing the total read count for the endogenous target molecules by the average QCT sequencing depth,

- wherein determining the absolute count of the endogenous reference molecules comprises determining the absolute count of the endogenous reference molecules lacking the mutation, based on dividing the total read count for the endogenous reference molecules by the average QCT sequencing depth, and
- wherein facilitating the prenatal diagnosis of the genetic disorder comprises facilitating the prenatal diagnosis of the single gene disorder based on the comparison.

3. The method of Claim 1,

- wherein the genetic disorder comprises a chromosomal abnormality,
- wherein determining the absolute count of the endogenous target molecules comprises determining the absolute count of the endogenous target molecules associated with a first chromosome, based on dividing the total read count for the endogenous target molecules by the average QCT sequencing depth,
- wherein determining the absolute count of the endogenous reference molecules comprises determining the absolute count of the endogenous reference molecules associated with a second chromosome, based on dividing the total read count for the endogenous reference molecules by the average QCT sequencing depth, and
- wherein facilitating the prenatal diagnosis of the genetic disorder comprises facilitating the prenatal diagnosis of the chromosomal abnormality based on the comparison.

4. The method of Claim 1, wherein the genetic disorder comprises a chromosomal microdeletion,

- wherein determining the absolute count of the endogenous target molecules comprises determining the absolute count of the endogenous target molecules associated with a microdeletion region, based on dividing the total read count for the endogenous target molecules by the average QCT sequencing depth,
- wherein determining the absolute count of the endogenous reference molecules comprises determining the absolute count of the endogenous reference molecules associated with a second chromosomal region not expected to have a microdeletion, based on dividing the total read count for the endogenous reference molecules by the average QCT sequencing depth, and

- wherein facilitating the prenatal diagnosis of the genetic disorder comprises facilitating the prenatal diagnosis of the chromosomal microdeletion based on the comparison.
5. The method of Claim 1, wherein the genetic disorder comprises a copy number variation,
 - wherein determining the absolute count of the endogenous target molecules comprises determining the absolute count of the endogenous target molecules associated with a region that may have copy number variation, based on dividing the total read count for the endogenous target molecules by the average QCT sequencing depth,
 - wherein determining the absolute count of the endogenous reference molecules comprises determining the absolute count of the endogenous reference molecules associated with a region not expected to have a copy number variation, based on dividing the total read count for the endogenous reference molecules by the average QCT sequencing depth, and
 - wherein facilitating the prenatal diagnosis of the genetic disorder comprises facilitating the prenatal diagnosis of the copy number variation based on the comparison.
 6. The method of Claim 1, wherein the average QCT sequencing depth used in determining the absolute count of the endogenous target molecules and the absolute count of endogenous reference molecules is determined separately from their corresponding QCTs.
 7. The method of Claim 1, wherein facilitating the prenatal diagnosis comprises facilitating the prenatal diagnosis of the genetic disorder based on a fetal fraction measurement, the absolute count of endogenous target sequences, and the absolute count of endogenous reference sequences.
 8. A method for identifying contamination associated with at least one of sequencing library preparation and high throughput sequencing, the method comprising:
 - generating a set of quality control template (QCT) molecules, each QCT molecule comprising:
 - a target-associated region with sequence similarity to a target sequence region of a biological target, and
 - a variation region with sequence dissimilarity to a sequence region of the biological target; and
 - computationally determining a set of QCT sequence read clusters based on the variation regions of the set of QCT molecules,

- wherein the set of QCT sequence read clusters comprises QCT molecule sequence reads derived from the high throughput sequencing corresponding to a set of QCT mixtures generated based on the set of QCT molecules and a set of samples comprising the biological target, and
 - wherein the sequencing library preparation comprises co-amplification, of the set of QCT molecules and nucleic acid molecules comprising the biological target, based on the sequence similarity of the target-associated region and the target sequence region of the biological target; and
 - based on the set of QCT sequence read clusters, determining a contamination parameter describing the contamination associated with the at least one of the sequencing library preparation and the high throughput sequencing.
9. The method of Claim 8, wherein computationally determining the set of QCT sequence read clusters comprises:
- clustering a first QCT molecule sequence read and a second QCT molecule sequence read into a QCT sequence read cluster, of the set of QCT sequence read clusters, based on a variation region sequence similarity satisfying a first condition, and
 - for each QCT sequence read cluster of the set of QCT sequence read clusters, determining an assignment of the QCT sequence read cluster to a sample identifier of a set of sample identifiers identifying the set of samples,
 - wherein determining the contamination parameter comprises determining the contamination parameter based on the set of QCT sequence read clusters and the assignments of the QCT sequence read clusters to the sample identifiers of the set of sample identifiers.
10. The method of Claim 9, wherein determining the contamination parameter comprises:
- identifying a first and a second QCT sequence read cluster corresponding to a shared variation region sequence, wherein the assignments of the first and the second QCT sequence read clusters are to distinct sample identifiers of the set of sample identifiers;
 - generating a read depth comparison between a first read depth associated with the first QCT sequence read cluster and a second read depth associated with the second QCT sequence read cluster; and
 - based on the read depth comparison, determining the contamination parameter associated with a sample identified by a distinct sample identifier of the distinct sample identifiers.

11. The method of Claim 9, wherein clustering the first and the second QCT sequence reads comprises clustering the first and the second QCT sequence reads into the QCT sequence read cluster based on the variation region sequence similarity of fewer than three point substitutions, and based on a read depth associated with the QCT sequence read cluster satisfying a second condition.
12. The method of Claim 8, wherein determining the contamination parameter comprises:
 - determining a first molecular fingerprint associated with first amplification in a first instance of the sequencing library preparation, based on the set of QCT sequence read clusters;
 - determining a second molecular fingerprint associated with second amplification in a second instance of the sequencing library preparation, based on an additional set of QCT sequence read clusters; and
 - based on a comparison between the first and the second molecular fingerprints, determining a carry-over contamination parameter describing carry-over contamination from the first instance to the second instance.
13. The method of Claim 8, wherein the contamination parameter describes a degree of index misassignment associated with the high throughput sequencing.
14. The method of Claim 13, wherein the contamination parameter is adapted for use in determination of a diagnostic outcome for assays associated with at least one of noninvasive prenatal testing and liquid biopsies.
15. The method of Claim 8, further comprising generating a single QCT library comprising the set of QCT molecules, wherein the single QCT library is adapted for deployment, at a single stage of the at least one of the sequencing library preparation and the high throughput sequencing, of less than 0.00001 nanograms of amplifiable QCT molecules for each sample of the set of samples.
16. The method of Claim 8,

- wherein each variation region of the set of QCT molecules comprises an embedded molecular identifier (EMI) region comprising a set of variable "N" bases, wherein each "N" base is selected from any one of an "A" base, a "G" base, a "T" base, and a "C" base,
- wherein each QCT molecule of the set of QCT molecules further comprises an additional EMI region comprising an additional set of variable "N" bases, wherein the additional EMI region is separated from the EMI region by a sequence region of the QCT molecule, wherein the set of variable "N" bases and the additional set of variable "N" bases each comprise greater than three "N" bases, and
- wherein determining the contamination parameter comprises determining the contamination parameter based on the set of QCT sequence read clusters derived based on the EMI regions and the additional EMI regions of the set of QCT molecules.

17. A method for characterization associated with at least one of sequencing library preparation and sequencing, the method comprising:

- generating a set of quality control template (QCT) molecules, each QCT molecule comprising:
 - a target-associated region matching a target sequence region of a target molecule; and
 - a variation region not matching a corresponding region of the target molecule, the variation region comprising an embedded molecular identifier having a randomly selected nucleotide sequence
- generating a mixture comprising the set of QCT molecules and nucleic acid molecules of a sample;
- co-amplifying the mixture;
- sequencing the co-amplified mixture to generate sequencing data;
- computationally determining a set of QCT sequence read clusters, using the sequence data, based on a number of distinct variation regions of the set of QCT molecules identified in the sequencing data;
- computationally determining a count of QCT molecules that were in the mixture prior to the co-amplification based on the set of QCT sequence read clusters; and
- based on the count of QCT molecules, determining the count of QCT molecules associated with the at least one of the sequencing library preparation and the sequencing.

18. The method of Claim 17,

- wherein the each QCT molecule comprises a first QCT identifier region shared amongst the set of QCT molecules and adapted to identifying the QCT molecule,
- wherein the method further comprises generating a set of additional QCT molecules, each additional QCT molecule comprising a second QCT identifier region shared amongst the set of additional QCT molecules and adapted to identifying the additional QCT molecule; and
- wherein computationally determining the set of QCT sequence read clusters comprises determining the set of QCT sequence read clusters based on the first and the second QCT identifier regions.

19. The method of Claim 18,

- wherein the set of QCT molecules is adapted for deployment at a first stage of the at least one of the sequencing library preparation and the sequencing,
- wherein the set of additional QCT molecules is adapted for deployment at a second stage of the at least one of the sequencing library preparation and the sequencing,
- wherein computationally determining the set of QCT sequence read clusters comprises:
 - determining a first subset of the set of QCT sequence read clusters, wherein the first subset corresponds to the first QCT identifier region and is associated with the first stage, and
 - determining a second subset of the set of QCT sequence read clusters, wherein the second subset corresponds to the second QCT identifier region and is associated with the second stage; and
- wherein determining the sequencing-related parameter comprises determining a sample tracking parameter associated with loss-of-sample, based on the first and the second subsets of the set of QCT sequence read clusters.

20. The method of Claim 18, wherein determining the sequencing-related parameter comprises:

- determining a first absolute count and a second absolute count corresponding to the set of QCT molecules and the set of additional QCT molecules, respectively, based on the set of QCT sequence read clusters, and
- determining at least one of a pipette error parameter and a quantification error parameter based on the first and the second absolute counts.

21. The method of Claim 18, wherein determining the sequencing-related parameter comprises:

- identifying QCT sequence reads that are not assigned to a QCT sequence read cluster of the set of QCT sequence read clusters; and
- determining at least one of a sequencing error rate and a polymerase error rate from a number of the QCT sequence reads that are not assigned and a total number of QCT sequence reads.

22. The method of Claim 18,

- wherein the variation region of the each QCT molecule comprises a first embedded molecular identifier (EMI) region separated from a second EMI region by at least the first QCT identifier region,
- wherein the each additional QCT molecule comprises a first additional EMI region separated from a second additional EMI region by at least the second QCT identifier region,
- wherein the first, the second, the first additional, and the second additional EMI regions comprise a set of variable "N" bases, and wherein each "N" base is selected from any one of an "A" base, a "G" base, a "T" base, and a "C" base, and
- wherein computationally determining the set of QCT sequence read clusters comprises determining the set of QCT sequence read clusters based on the first and the second QCT identifier regions, and on the first, the second, the first additional, and the second additional EMI regions.

23. The method of Claim 22,

- wherein, for each QCT molecule, the corresponding QCT molecule sequence is characterized by full sequence similarity to a first sequence template of the biological target except for the first QCT identifier region, the first EMI region, and the second EMI region;
- wherein, for each additional QCT molecule, the corresponding additional QCT molecule sequence is characterized by full sequence similarity to a second sequence template except for the second QCT identifier region, the first additional EMI region, and the second additional EMI region.

24. The method of Claim 17,

- wherein determining the set of QCT sequence read clusters comprising determining a filtered subset of QCT sequence read clusters based on read depths corresponding to the filtered subset of QCT sequence read clusters,
- wherein determining the target molecule count comprises:
 - determining a QCT molecule count based on the filtered subset of QCT sequence read clusters;
 - determining a correction factor ratio based on the QCT molecule count and the QCT molecule sequence reads; and
 - determining the target molecule count based on the correction factor ratio and target molecule sequence reads derived from the sequencing, the target molecule sequence reads associated with the biological target.

25. The method of Claim 24, further comprising adaptively determining a read depth threshold based on read depth distribution features for the QCT molecule sequence reads, wherein determining the filtered subset of QCT sequence read clusters comprises determining the filtered subset based on the satisfaction of the adaptively determined read depth threshold by the read depths.

26. The method of Claim 24, wherein each read depth of the read depths corresponds to greater than twenty reads for the corresponding QCT sequence read cluster of the filtered subset of QCT sequence read clusters.

27. The method of Claim 17, wherein determining an absolute abundance of the target molecule count comprises determining the target molecule count for facilitating diagnosis associated with at least one of noninvasive prenatal testing and liquid biopsies.

28. The method of Claim 17, wherein determining the sequencing-related parameter comprises determining a contamination parameter comprising at least one of: a cross-contamination parameter describing cross-contamination across sample compartments associated with the at least one of the sequencing library preparation and the sequencing, a carry-over contamination parameter describing carry-over contamination across a plurality of instances of the at least one of the sequencing library preparation and the sequencing, and an index-hopping contamination parameter describing index-hopping contamination associated with index-hopping primers.

29. The method of Claim 28,

- wherein the set of QCT molecules is adapted for the sequencing,
- wherein generating the set of QCT molecules comprises:
 - amplifying a first subset of QCT molecules of the set of QCT molecules; and
 - amplifying a second subset of QCT molecules of the set of QCT molecules,
- wherein the QCT molecule sequencing reads are derived from the sequencing corresponding to:
 - the QCT mixture generated based on the first subset of QCT molecules and the sample comprising the biological target, and
 - an additional QCT mixture generated based on the second subset of QCT molecules and an additional sample comprising the biological target, wherein the sample and the additional sample respectively correspond to a first sample compartment and a second sample compartment of the sample compartments.