CLAIMS

WHAT IS CLAIMED IS:

1. A method, comprising:

(a) providing a sample comprising a set of double-stranded polynucleotide molecules, each double-stranded polynucleotide molecule including first and second complementary strands;

(b) tagging said double-stranded polynucleotide molecules with a set of duplex tags, wherein each duplex tag differently tags said first and second complementary strands of a double-stranded polynucleotide molecule in said set;

(c) sequencing at least some of said tagged strands to produce a set of sequence reads;

(d) reducing and/or tracking redundancy in said set of sequence reads;

(e) sorting sequence reads into paired reads and unpaired reads, wherein (i) each paired read corresponds to sequence reads generated from a first tagged strand and a second differently tagged complementary strand derived from a double-stranded polynucleotide molecule in said set, and (ii) each unpaired read represents a first tagged strand having no second differently tagged complementary strand derived from a double-stranded polynucleotide molecule represented among said sequence reads in said set of sequence reads;

(f) determining quantitative measures of (i) said paired reads and (ii) said unpaired reads that map to each of one or more genetic loci; and

(g) estimating with a programmed computer processor a quantitative measure of total double-stranded polynucleotide molecules in said set that map to each of said one or more genetic loci based on said quantitative measure of paired reads and unpaired reads mapping to each locus.

2. The method of claim 1, further comprising (h) detecting copy number variation in said sample by determining a normalized total quantitative measure determined in step (g) at each of said one or more genetic loci and determining copy number variation based on the normalized measure.

3. The methods of claim 1 or claim 2, wherein the said sample comprises double-stranded polynucleotide molecules sourced from a tumour sample.

4. The method of claim 1 or claim 2, wherein said sample comprises double-stranded polynucleotide molecules sourced substantially from cell-free nucleic acids.

5. The method of any one of claims 1 to 4, wherein said duplex tags are not sequencing adaptors.

6. The method of any one of claims 1 to 5, wherein the duplex tags are double-stranded tags that are Y-shaped with a hybridized portion at one end of the tag and a non-hybridized portion at the opposite end of the tag.

7. The method of any one of claims 1 to 6, wherein the duplex tags contain molecular barcodes, optionally wherein the number of different molecular barcodes is no more than 100, 50, 40, 30, 20 or 10 molecular barcodes.

8. The method of any one of claims 1 to 7, wherein reducing redundancy in said set of sequence reads comprises collapsing sequence reads produced from amplified products of an original polynucleotide molecule in said sample back to said original polynucleotide molecule.

9. The method of claim 8, wherein the collapsing back to said original polynucleotide molecule is based on their tags and one or more endogenous sequences.

10. The method of claim 9, wherein the one or more endogenous sequences is a terminal sequence of the polynucleotides to be analysed, or the length of the sequence.

11. The method of any one of claims 8 to 10, further comprising determining a consensus sequence for said original polynucleotide molecule.

12. The method of claim 11, further comprising identifying polynucleotide molecules at one or more genetic loci comprising a sequence variant.

13. The method of claim 11, further comprising determining a quantitative measure of paired reads that map to a locus, wherein both strands of said pair comprise a sequence variant.

14. The method of claim 12 or claim 13, further comprising determining a quantitative measure of paired molecules in which only one member of said pair bears a sequence variant and/or determining a quantitative measure of unpaired molecules bearing a sequence variant.

15. The methods of any one of claims 12 to 14 wherein the sequence variant is selected from the group consisting of a single nucleotide variant, an indel, a transversion, a translocation, an inversion, a deletion, a chromosomal structure alteration, a gene fusion, a chromosome fusion, a gene truncation, a gene amplification, a gene duplication and a chromosomal lesion.