CLAIMS:

1. A computer-implemented method under control of a processor executing instructions, comprising:

receiving genomic sequence data of a biological sample, wherein the genomic sequence data comprises a plurality of sequence reads, each sequence read being associated with a unique molecular identifier of a plurality of unique molecular identifiers;

identifying errors in the genomic sequence data for the plurality of unique identifiers, based on sequence disagreement within a first subset of the plurality of sequence reads associated with a particular unique molecular identifier, sequence disagreement between the first subset and a second subset of the plurality of sequence reads having a complementary unique molecular identifier to the particular unique molecular identifier, or both, to generate an error rate of the genomic sequence data;

identifying a plurality of potential sequence variants in the genomic sequence data relative to a reference sequence;

classifying false positive sequence variants of the plurality of potential sequence variants based on the generated error rate of the genomic sequence data; and

eliminating the classified false positive sequence variants from the plurality of potential sequence variants to yield a plurality of sequence variants.

2. The method of claim 1, wherein the error rate comprises individual error rates based on a type of nucleotide change.

3. The method of claim 2, wherein the classifying comprises applying a function of the error rate for an individual type of nucleotide change and a read coverage of an individual potential sequence variant to determine that the individual potential sequence variant is a false positive sequence variant.

4. The method of claim 1, wherein identifying errors in the genomic sequence data based on the sequence disagreement within the first subset of the plurality of sequence reads comprises identifying differences between a first read direction or a second read direction associated with an individual unique molecular identifier, wherein the identified differences contribute to the error rate.

5. The method of claim 4, comprising collapsing the sequence reads of the first subset to yield a collapsed first subset sequence read indicative of a majority consensus sequence of the sequence reads of the first subset.

6. The method of claim 5, wherein identifying errors in the genomic sequence data based on sequence disagreement between the first subset and the second subset of the plurality of sequence reads complementary to the first subset comprises identifying second differences between the collapsed first subset sequence read and the sequence reads of the second subset, wherein the identified second differences contribute to the error rate.

7. The method of claim 6, comprising collapsing the sequence reads of the second subset and the collapsed first subset sequence read to yield a duplex collapsed sequence read indicative of a majority consensus sequence of the sequence reads of the first subset and the second subset.

8. The method of claim 7, wherein the duplex collapsed sequence read is compared to the reference sequence to identify a potential sequence variant of the plurality of potential sequence variants.

9. The method of claim 1, wherein the classifying comprises applying a function based at least in part on the error rate to each potential sequence variant to

generate a score and determining a false positive or true positive classification of each potential sequence variant based on the score.

10. The method of claim 9, wherein the error rate is a global error rate for the sequence data, and wherein the function is based on the error rate and a read coverage at a site of each individual potential sequence variant.

11. The method of claim 9, wherein the error rate is selected from a plurality of error rates at each potential sequence variant site based on a type of potential sequence variant, and wherein the function is based on the selected error rate and a read coverage at each individual potential sequence variant site.

12. The method of claim 9, wherein the error rate is weighted based on a type of the biological sample, and wherein the function is based on the weighted error rate and a read coverage at each individual potential sequence variant site.

13. The method of claim 1, comprising providing an indication of the plurality of sequence variants on a display.

14. The method of claim 1, wherein the error rate is independent of positive or negative strand identification.

15. A computer-implemented method under control of a processor executing instructions, comprising:

receiving genomic sequence data of a first biological sample, wherein the genomic sequence data comprises a plurality of sequence reads, each sequence read being associated with a unique molecular identifier of a plurality of unique molecular identifiers;

identifying first sequence differences within a first subset of the plurality of sequence reads associated with a first unique molecular identifier;

collapsing the first subset to yield a collapsed first subset sequence read, wherein the collapsing comprises eliminating sequence differences present in a minority of the sequencing reads of the first subset;

identifying second sequence differences within a second subset of the plurality of sequence reads associated with a second unique molecular identifier, the second unique molecular identifier being complementary at least in part to the first unique molecular identifier;

collapsing the second subset to yield a collapsed second subset sequence read, wherein the collapsing comprises eliminating sequence differences present in a minority of the sequencing reads of the second subset; and

determining that a sequence variant relative to a baseline in the collapsed first subset, the collapsed second subset, or a duplex of the collapsed first subset and the collapsed second subset is valid based on a function of an error rate of the genomic sequence data, wherein the error rate is determined based in part on the identified first sequence differences and the identified second sequence differences.

16. The method of claim 15, comprising determining that an additional sequence variant in a third subset associated with a third unique molecular identifier is valid based on the function of the error rate.

17. The method of claim 15, comprising determining that an additional sequence variant in a third subset associated with a third unique molecular identifier is a false positive based on the function of the error rate.

18. The method of claim 17, comprising eliminating the additional sequence variant from an indication of sequence variants in the genomic sequence data.

19. A sequencing device configured to identify sequence variants in genomic sequence data of a biological sample, comprising:

a memory device comprising executable application instructions stored therein; and

a processor configured to execute the application instructions stored in the memory device, wherein the application instructions comprise instructions that cause the processor to:

receive genomic sequence data of a biological sample, wherein the genomic sequence data comprises a plurality of sequence reads, each sequence read being associated with a unique molecular identifier of a plurality of unique molecular identifiers;

identify a plurality of errors in the genomic sequence data based on sequence disagreement between sequence reads associated with each unique molecular identifier of the plurality of unique molecular identifiers to generate an error rate of the genomic sequence data;

identify a plurality of potential sequence variants in the genomic sequence data relative to a reference sequence; and

determine a validity of the plurality of potential sequence variants based at least in part on the generated error rate.

20. The sequencing device of claim 19, wherein the validity is based on a function of the error rate and a sequence coverage of an individual potential sequence variant.

21. The sequencing device of claim 19, comprising a user interface configured to receive user input, wherein the user input comprises a sample type of the biological sample.

22. The sequencing device of claim 21, wherein the error rate is weighted based on the sample type.