WHAT IS CLAIMED IS:

1. A method for determining *GBA* status comprising:

under control of a hardware processor:

receiving a first plurality of sequence reads generated from a sample obtained from a subject;

aligning the first plurality of sequence reads to a reference genome sequence to obtain a second plurality of sequence reads aligned to *GBA* gene or *GBAP1* gene in the reference genome sequence;

determining a number of the sequence reads of the second plurality of sequence reads aligned to a unique region between *GBA* gene and *GBAP1* gene in the reference genome sequence;

determining a normalized number of the sequence reads aligned to the unique region between *GBA* gene and *GBAP1* gene in the reference genome sequence;

determining a total copy number of *GBA* gene and *GBAP1* gene using a Gaussian mixture model comprising a plurality of Gaussians each representing a different integer copy number, given the normalized number of the sequence reads aligned to the region between *GBA* gene and *GBAP1* gene;

phasing one or more haplotypes originating from *GBA* gene or *GBAP1* gene in a region of *GBA* gene, or a corresponding region of *GBAP1* gene, comprising a plurality of *GBA*/*GBAP1* differentiating bases using sequence reads of the second plurality of sequence reads aligned to the region, or the corresponding region, comprising the plurality of *GBA*/*GBAP1* differentiating bases;

determining a copy number of each of the one or more haplotypes using the total copy number of *GBA* gene and *GBAP1* gene and a number of sequence reads of the second plurality of sequence reads each comprising one or more of the plurality of *GBA*/*GBAP1* differentiating bases that support the haplotype; and

determining a *GBA* status of the subject using the one or more haplotypes originating from *GBA* gene or *GBAP1* gene in the region of *GBA* gene, or the corresponding region of *GBAP1* gene, and/or the copy number of each of the one or more haplotypes.

2. The method of claim 1, wherein the unique region between *GBA* gene and *GBAP1* gene in the reference genome sequence comprises a unique region about 10 kilobases in length.

3. The method of any one of claims 1-2, wherein the unique region between *GBA* gene and *GBAP1* gene in the reference genome sequence comprises chr1:155220429-155230539

of hg38 or a corresponding region of a reference human genome sequence.

4. The method of any one of claims 1-3, wherein determining the normalized number of the sequence reads aligned to the unique region between *GBA* gene and *GBAP1* gene in the reference genome sequence comprises: determining the normalized number of the sequence reads aligned to the unique region between *GBA* gene and *GBAP1* gene in the reference genome sequence using (1a) a depth of the sequence reads aligned to the unique region between the *GBA* gene and *GBAP1* gene, (1b) a length of the unique region, (2a) a depth of sequence reads of the first plurality of sequence reads aligned to each of a plurality of regions of the reference genome sequence other than a genetic locus comprising *GBA* gene and *GBAP1* gene, and (2b) a length of each of the plurality of regions of the reference genome other than the genetic locus comprising *GBA* gene and *GBAP1* gene.

5. The method of any one of claims 1-4, comprising determining a normalized, corrected number of the sequence reads aligned to the unique region between *GBA* gene and *GBAP1* gene in the reference genome sequence from the normalized number of the sequence reads aligned to the unique region between *GBA* gene and *GBAP1* gene in the reference genome sequence using (1) a GC content of the unique region between the *GBA* gene and *GBAP1* gene and optionally (2) a GC content of each of one or more regions of the reference genome sequence other than a genetic locus comprising *GBA* gene and *GBAP1* gene, wherein determining the total copy number of *GBA* gene and *GBAP1* gene comprises: determining the total copy number of *GBA* gene and *GBAP1* gene using the Gaussian mixture model, given the normalized, corrected number of the sequence reads aligned to the region between *GBA* gene and *GBAP1* gene.

6. The method of any one of claims 1-5, wherein determining the total copy number of *GBA* gene and *GBAP1* gene comprises: determining a copy number of the region between *GBA* gene and *GBAP1* gene using the Gaussian mixture model, given the normalized number of the sequence reads aligned to the region between *GBA* gene and *GBAP1* gene, and wherein the total copy number of *GBA* gene and *GBAP1* gene is the copy number of the region between *GBA* gene and *GBAP1* gene plus two.

7. The method of any one of claims 1-6, wherein determining the total copy number of *GBA* gene and *GBAP1* gene comprises: determining the total copy number of *GBA* gene and *GBAP1* gene using a Gaussian mixture model and a predetermined posterior probability threshold, given the normalized number of the sequence reads aligned to the region between *GBA* gene and *GBAP1* gene, optionally wherein the predetermined posterior probability threshold is 0.95.

8. A method for determining *CYP21A2* status comprising:

under control of a hardware processor:

receiving a first plurality of sequence reads generated from a sample obtained from a subject;

aligning the first plurality of sequence reads to a reference genome sequence to obtain a second plurality of sequence reads aligned to *CYP21A2* gene or *CYP21A1P* pseudogene in the reference genome sequence;

determining a number of the sequence reads of the second plurality of sequence reads aligned to *CYP21A2* gene or *CYP21A1P* pseudogene in the reference genome sequence;

determining a normalized number of the sequence reads aligned to *CYP21A2* gene or *CYP21A1P* pseudogene in the reference genome sequence;

determining a total copy number of *CYP21A2* gene and *CYP21A1P* pseudogene using a Gaussian mixture model comprising a plurality of Gaussians each representing a different integer copy number, given the normalized number of the sequence reads aligned to CYP21A2 gene or CYP21A1P pseudogene;

phasing one or more haplotypes originating from *CYP21A2* gene or *CYP21A1P* pseudogene in a region of *CYP21A2* gene, or a corresponding region of *CYP21A1P* pseudogene, comprising a plurality of *CYP21A2*/*CYP21A1P* differentiating bases using sequence reads of the second plurality of sequence reads aligned to the region, or the corresponding region, comprising the plurality of *CYP21A2*/*CYP21A1P* differentiating bases;

determining a copy number of each of the one or more haplotypes using the total copy number of *CYP21A2* gene and *CYP21A1P* pseudogene and a number of sequence reads of the second plurality of sequence reads each comprising one or more of the plurality of *CYP21A2*/*CYP21A1P* differentiating bases that support the haplotype; and

determining a *CYP21A2* status of the subject using the one or more haplotypes originating from *CYP21A2* gene or *CYP21A1P* pseudogene in the region of *CYP21A2* gene, or the corresponding region of *CYP21A1P* pseudogene, and/or the copy number of each of the one or more haplotypes.

9. The method of claim 8, wherein determining the normalized number of the sequence reads aligned to *CYP21A2* gene or *CYP21A1P* pseudogene in the reference genome sequence comprises: determining the normalized number of the sequence reads aligned to *CYP21A2* gene or *CYP21A1P* pseudogene in the reference genome sequence using (1a) a depth of the sequence reads aligned to *CYP21A2* gene or *CYP21A1P* pseudogene, (1b) a length of the unique region, (2a) a depth of sequence reads of the first plurality of sequence reads aligned to

each of a plurality of regions of the reference genome sequence other than a genetic locus comprising *CYP21A2* gene and *CYP21A1P* pseudogene, and (2b) a length of each of the plurality of regions of the reference genome other than the genetic locus comprising *CYP21A2* gene and *CYP21A1P* pseudogene.

10. The method of any one of claims 8-9, comprising determining a normalized, corrected number of the sequence reads aligned to *CYP21A2* gene or *CYP21A1P* pseudogene in the reference genome sequence from the normalized number of the sequence reads aligned to *CYP21A2* gene or *CYP21A1P* pseudogene in the reference genome sequence using (1) a GC content of *CYP21A2* gene or *CYP21A1P* pseudogene and optionally (2) a GC content of each of one or more regions of the reference genome sequence other than a genetic locus comprising *CYP21A2* gene and *CYP21A1P* pseudogene, wherein determining the total copy number of *CYP21A2* gene and *CYP21A1P* pseudogene comprises: determining the total copy number of *CYP21A2* gene and *CYP21A1P* pseudogene using the Gaussian mixture model, given the normalized, corrected number of the sequence reads aligned to *CYP21A2* gene or *CYP21A1P* pseudogene.

11. The method of any one of claims 8-10, wherein determining the total copy number of *CYP21A2* gene and *CYP21A1P* pseudogene comprises: determining the total copy number of *CYP21A2* gene and *CYP21A1P* pseudogene using a Gaussian mixture model and a predetermined posterior probability threshold, given the normalized number of the sequence reads aligned to CYP21A2 gene or CYP21A1P pseudogene, optionally wherein the predetermined posterior probability threshold is 0.95.

12. The method of any one of claims 8-11, wherein the Gaussian mixture model comprises a one-dimensional Gaussian mixture model.

13. The method of any one of claims 8-12, wherein the plurality of Gaussians of the Gaussian mixture model represents integer copy numbers 0 to 10.

14. The method of any one of claims 8-12, wherein the plurality of Gaussians of the Gaussian mixture model comprises 5 Gaussians.

15. A system for determining a gene recombinant variant comprising:

non-transitory memory configured to store executable instructions and a first plurality of sequence reads generated from a sample obtained from a subject; and

a hardware processor in communication with the non-transitory memory, the hardware processor programmed by the executable instructions to perform:

aligning the first plurality of sequence reads to a reference sequence to obtain a second plurality of sequence reads aligned to a gene or a gene paralog, or a region therebetween, in the reference sequence;

determining a total copy number of the gene and the gene paralog using a Gaussian mixture model comprising a plurality of Gaussians each representing a different integer copy number, given a number of the sequence reads aligned to the gene or the gene paralog, or a region therebetween;

phasing one or more haplotypes originating from the gene, comprising a recombinant variant of the gene, or the gene paralog, or a region of the gene or a corresponding region of the gene paralog, comprising a plurality of gene/gene paralog differentiating bases using sequence reads of the second plurality of sequence reads aligned to the region, or the corresponding region, comprising the plurality of gene/gene paralog differentiating bases; and

determining a copy number of each of the one or more haplotypes using the total copy number of the gene and the gene paralog and a number of sequence reads of the second plurality of sequence reads each comprising one or more of the plurality of gene/gene paralog differentiating bases that support the haplotype.

16. The system of claim 15, wherein the gene recombinant variant comprises a reciprocal recombinant variant.

17. The system of claim 15, wherein the gene recombinant variant comprises a non-reciprocal recombinant variant.

18. The system of any one of claims 15-17, wherein the reference sequence comprises a reference genome sequence.

19. The system of any one of claims 15-16, wherein the hardware processor is programmed by the executable instructions to perform: determining a number of the sequence reads aligned to the gene or the gene paralog, or a region therebetween.

20. The system of any one of claims 15-19, wherein the number of the sequence reads aligned to the gene or the gene paralog, or a region therebetween, comprises a normalized and/or GC corrected number of the sequence reads aligned to the gene or the gene paralog, or a region therebetween.