WHAT IS CLAIMED IS:

[Claim 1]. A method for determining a copy number of survival of motor neuron 1 (SMN1) gene comprising:

under control of a hardware processor:

receiving sequence data comprising a plurality of sequence reads obtained from a sample of a subject aligned to survival of motor neuron 1 (SMN1) gene or survival of motor neuron 2 (SMN2) gene;

determining (i) a first number of sequence reads of the plurality of sequence reads aligned to a first SMN1 or SMN2 region comprising at least one of exon 1 to exon 6 of the SMN1 gene or the SMN2 gene, respectively, and (ii) a second number of sequence reads of the plurality of sequence reads aligned to a second SMN1 or SMN2 region comprising at least one of exon 7 and exon 8 of the SMN1 gene or the SMN2 gene, respectively;

determining (i) a first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and (ii) a second normalized number of the sequence reads aligned to the second SMN1 or SMN2 region using (i) a length of the first SMN1 or SMN2 region and (ii) a length of the second SMN1 or SMN2 region, respectively;

determining (i) a copy number of total survival of motor neuron (SMN) genes, each being an intact SMN1 gene, an intact SMN2 gene, a truncated SMN1 gene, or a truncated SMN2 gene, and (ii) a copy number of any intact SMN genes, each being the intact SMN1 gene or the intact SMN2 gene, using a Gaussian mixture model comprising a plurality of Gaussians each representing a different integer copy number, given (i) the first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and (ii) the second normalized number of the sequence reads aligned to the second SMN1 or SMN2 region, respectively;

for one of a plurality of SMN1 gene-specific bases associated with the intact SMN1 gene, determining a most likely combination, of a plurality of possible combinations each comprising a possible copy number of the SMN1 gene and a possible copy number of the SMN2 gene summed to the copy number of any intact SMN genes determined, given (a) a number of sequence reads of the plurality of sequence reads with bases that support the SMN1 gene-specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support a SMN2 gene-specific base of the SMN2 gene corresponding to the SMN1 gene-specific base; and

determining a copy number of the SMN1 gene using the most likely combination

of the possible copy number of the SMN1 gene and the possible copy number of the SMN2 gene determined for the SMN1 gene-specific base.

[Claim 2]. The method of claim 1, wherein the sequencing data comprises whole genome sequencing (WGS) data or short-read WGS data.

~~3. The method of any one of claims 1-2, wherein the subject is a fetal subject, a neonatal subject, a pediatric subject, an adolescent subject, or an adult subject.~~

~~4. The method of any one of claims 1-3, wherein the sample comprises cells or cell-free DNA.~~

~~5. The method of any one of claims 1-4, wherein the sample comprises fetal cells or cell-free fetal DNA.~~

[Claim 3~~6~~]. The method of any one of claims 1~~-5~~or 2, wherein a sequence read of the plurality of sequence reads is aligned to the first SMN1 or SMN2 region or the second SMN1 or SMN2 region with an alignment quality score of about zero.

[Claim 4~~7~~]. The method of any one of claims 1-~~6~~3, wherein the first SMN1 or SMN2 region comprises the exon 1 to the exon 6 of the SMN1 gene or the SMN2 gene, respectively, and is about 22.2 kb in length, wherein the second SMN1 or SMN2 region comprises the exon 7 and the exon 8 of the SMN1 gene or the SMN2 gene, respectively, and is about 6 kb in length.

[Claim 5~~8~~]. The method of any one of claims 1-~~7~~4, wherein determining (i) the first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and (ii) the second normalized number of the sequence reads aligned to the second region comprises: determining (i) the first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and (ii) the second normalized number of the sequence reads aligned to the second SMN1 or SMN2 region using (i) the length of the first SMN1 or SMN2 region and (ii) the length of the second SMN1 or SMN2 region, respectively, and (iii) a depth of sequence reads of a region of a genome of the subject other than genetic loci comprising the SMN1 gene and the SMN2 gene in the sequence data.

~~9. The method of claim 8, wherein determining (i) the first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and (ii) the second normalized number of the sequence reads aligned to the second SMN1 or SMN2 region comprises:~~

~~determining (i) a first SMN1 or SMN2 region-length normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and (ii) a second SMN1 or SMN2 region-length normalized number of the sequence reads aligned to the second SMN1 or SMN2 region using (i) the length of the first SMN1 or SMN2 region and (ii) the length of the second SMN1 or SMN2 region, respectively; and~~

~~determining (i) a first normalized depth of the sequence reads aligned to the first~~

~~region SMN1 or SMN2 and (ii) a second normalized depth of the sequence reads aligned to the second SMN1 or SMN2 region from (i) the first SMN1 or SMN2 region length normalized number and (ii) the second SMN1 or SMN2 region length normalized number, respectively, using the depth of the sequence reads of the region of the genome of the subject other than genetic loci comprising the SMN1 gene and the SMN2 gene, the first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and the second normalized number of the sequence reads aligned to the second SMN1 or SMN2 region being the first normalized depth and the second normalized depth, respectively.~~

~~10. The method of any one of claims 1-9, wherein determining (i) the first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and (ii) the second normalized number of the sequence reads aligned to the second region comprises: determining (i) the first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and (ii) the second normalized number of the sequence reads aligned to the second SMN1 or SMN2 region using (i) a GC content of the first SMN1 or SMN2 region and (ii) a GC content of the second SMN1 or SMN2 region, respectively, and (iii) a depth of sequence reads of a region of a genome of the subject other than genetic loci comprising the SMN1 gene and the SMN2 gene in the sequence data, and (iv) a GC content of the region of the genome.~~

~~11. The method of any one of claims 8-10, wherein the depth of the region comprises an average depth or a median depth of the sequence reads of the region of the genome of the subject other than the genetic loci comprising the SMN1 gene and the SMN2 gene in the sequencing data.~~

~~12. The method of claim 11, wherein the region comprises about 3000 pre-selected regions of about 2 kb in length each across the genome of the subject.~~

~~13. The method of any one of claims 1-12, wherein (i) the first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and/or (ii) the second normalized number of the sequence reads aligned to the second SMN1 or SMN2 region is about 30 to about 40.~~

[Claim 6~~14~~]. The method of any one of claims 1-~~13~~5, wherein the Gaussian mixture model comprises a one-dimensional Gaussian mixture model.

[Claim 7~~15~~]. The method of any one of claims 1-~~14~~6, wherein the plurality of Gaussians of the Gaussian mixture model represents integer copy numbers 0 to 10.

[Claim 8~~16~~]. The method of any one of claims 1-~~15~~7, wherein a mean of each of the plurality of Gaussians is the integer copy number represented by the Gaussian.

[Claim 9~~17~~]. The method of any one of claims 1-~~16~~8, wherein determining (i) the copy number

of the total SMN genes and (ii) the copy number of any intact SMN genes comprises determining (i) the copy number of the total SMN genes and (ii) the copy number of any intact SMN genes using the Gaussian mixture model, and a first predetermined posterior probability threshold, given (i) the first normalized number of the sequence reads aligned to the first SMN1 or SMN2 region and (ii) the second normalized number of the sequence reads aligned to the second SMN1 or SMN2 region, respectively.

~~18. The method of claim 17, wherein the first predetermined posterior probability threshold is 0.95.~~

[Claim 10~~9~~]. The method of any one of claims 1-~~18~~9, comprising determining a copy number of truncated SMN genes using (i) the copy number of the total SMN genes determined and (ii) the copy number of the intact SMN genes determined.

~~20. The method of claim 19, wherein the copy number of the truncated SMN genes is a difference of (i) the copy number of the total SMN genes determined and (ii) the copy number of the intact SMN genes determined.~~

~~21. The method of any one of claims 1-20, wherein the SMN1 gene-specific base is a splicing enhancer.~~

~~22. The method of any one of claims 1-21, wherein the SMN1 gene-specific base is a base at c.840 of the SMN1 gene.~~

[Claim 11~~23~~]. The method of any one of claims 1-~~22~~10, wherein the most likely combination of the possible copy number of the SMN1 gene and the possible copy number of the SMN2 gene is associated with a highest posterior probability, relative to other combinations of the plurality of combinations given (a) the number of sequence reads of the plurality of sequence reads with bases that support the SMN1 gene-specific base and (b) the number of sequence reads of the plurality of sequence reads with bases that support the corresponding SMN2 gene-specific base.

[Claim 12~~4~~]. The method of any one of claims 1-~~23~~11, wherein determining the most likely combination of the possible copy number of the SMN1 gene and the possible combination of the SMN2 gene comprises: determining the most likely combination, of the plurality of possible combinations each comprising a possible copy number of the SMN1 gene and a possible copy number of the SMN2 gene summed to the copy number of any intact SMN genes determined, given a ratio of (a) a number of sequence reads of the plurality of sequence reads with bases that support the SMN1 gene-specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support the SMN2 gene-specific base of the SMN2 gene corresponding to the SMN1 gene-specific base.

[Claim 13~~25~~]. The method of any one of claims 1-~~24~~12, wherein determining the most likely combination of the possible copy number of the SMN1 gene and the possible combination of

the SMN2 gene comprises:

determining (a) a number of sequence reads of the plurality of sequence reads with bases that support the SMN1 gene- specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support the SMN2 gene-specific base of the SMN2 gene corresponding to the SMN1 gene-specific base;

determining the ratio of (a) a number of sequence reads of the plurality of sequence reads with bases that support the SMN1 gene-specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support the SMN2 gene- specific base of the SMN2 gene corresponding to the SMN1 gene-specific base; and determining the most likely combination, of the plurality of possible combinations each comprising a possible copy number of the SMN1 gene and a possible copy number of the SMN2 gene summed to the copy number of any intact SMN genes determined based on the ratio of (a) a number of sequence reads of the plurality of sequence reads with bases that support the SMN1 gene-specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support the SMN2 gene- specific base of the SMN2 gene corresponding to the SMN1 gene-specific base.

[Claim 14~~26~~]. The method of any one of claims 1-~~25~~13,

wherein determining the most likely combination of the possible copy number of the SMN1 gene and the possible combination of the SMN2 gene comprises: for each of the plurality of SMN1 gene-specific bases, determining a most likely combination, of a plurality of possible combinations each comprising a possible copy number of the SMN1 gene and a possible copy number of the SMN2 gene summed to the copy number of any intact SMN genes determined, associated with a highest posterior probability given (a) a number of sequence reads of the plurality of sequence reads with bases that support the SMN1 gene-specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support a SMN2 gene-specific base of the SMN2 gene corresponding to the SMN1 gene-specific base, and

wherein determining the copy number of the SMN1 gene comprises: determining the copy number of the SMN1 gene based on the possible copy number of the SMN1 gene of the most likely combination of the possible copy number of the SMN1 gene and the possible copy number of the SMN2 gene determined for each of the plurality of SMN1 gene-specific bases.

~~27. The method of claim 26, wherein the SMN1 gene-specific base has a concordance with each of the plurality of SMN1 gene-specific bases other than the SMN1 gene-specific base above a predetermined concordance threshold.~~

~~28. The method of claim 27, wherein the concordance threshold is 97%.~~

~~29. The method of any one of claims 26-28, wherein the plurality of SMN1 gene-specific bases comprises 8 SMN1 gene-specific bases.~~

~~30. The method of any one of claims 26-29, wherein each of the plurality of SMN1 gene-specific bases is on intron 6, the exon 7, intron 7, or the exon 8 of the SMN1 gene.~~

~~31. The method of any one of claims 26-30, wherein the plurality of SMN1 gene-specific bases if the subject is of a first race, the plurality of SMN1 gene-specific bases if the subject is of a second race, and the plurality of SMN1 gene-specific bases if the subject is of an unknown race are different.~~

~~32. The method of any one of claims 26-31, wherein a race of the subject is unknown, and wherein the plurality of SMN1 gene-specific bases is not race specific.~~

~~33. The method of any one of claims 26-31, wherein a race of the subject is known, and wherein the plurality of SMN1 gene-specific bases is specific to the race of the subject.~~

[Claim 15~~34~~]. The method of ~~any one of~~ claim~~s~~ ~~26-33~~14, comprising:

receiving race information of the subject; and

selecting the plurality of SMN1 gene-specific bases from pluralities of SMN1 gene-specific bases based on the race information received.

~~35. The method of any one of claims 1-34, wherein determining the copy number of the SMN1 gene comprises: determining the copy number of the SMN1 gene and a copy number of the SMN2 gene using the most likely combination of the possible copy number of the SMN1 gene and the possible copy number of the SMN2 gene determined for each of the plurality of SMN1 gene-specific bases.~~

~~36. The method of any one of claims 1-35, wherein determining the copy number of the SMN1 gene comprises: determining the copy number of the SMN1 gene using the most likely combination of the possible copy number of the SMN1 gene and the possible copy number of the SMN2 gene determined for the SMN1 gene-specific base and a second predetermined posterior probability threshold for the combination of the possible copy number of SMN1 gene and the possible copy number of the SMN2 gene.~~

~~37. The method of claim 36, wherein the second predetermined posterior probability threshold is 0.6 or 0.8.~~

~~38. The method of any one of claims 26-37, wherein a majority of the possible copy numbers of the SMN1 gene determined agree, and wherein the copy number of the SMN1 gene determined is the agreed possible copy number of the SMN1 gene.~~

~~39. The method of claim 38, comprising: determining a possible combination comprising a possible copy number of the~~

~~SMN1 gene and a possible copy number of the SMN2 gene summed to the copy number of any intact SMN genes determined given (a) a number of sequence reads of the plurality of sequence reads with bases that support any of the plurality of SMN1 gene- specific bases and (b) a number of sequence reads of the plurality of sequence reads with bases that support any of the plurality of corresponding SMN2 gene-specific bases; and~~

~~determining the possible copy number of the possible combination is the agreed possible copy number of the SMN1 gene.~~

~~40. The method of any one of claims 1-39, wherein determining the copy number of the SMN1 gene comprises determining the copy number of the SMN1 gene to be zero, one, or more than one.~~

[Claim ~~4~~16]. The method of any one of claims 1-~~40~~15, comprising determining a spinal muscular atrophy (SMA) status of the subject based on the copy number of the SMN1 gene.

~~42. The method of claim 41, wherein the SMA status of the subject comprises SMA, SMA carrier/not SMA, and not SMA carrier.~~

[Claim 17~~43~~]. The method of any one of claims 1-~~42~~16, comprising determining subject is a silent SMA carrier using a number of sequence reads of the plurality of sequence reads aligned to g.27134 of the SMN1 gene and the bases of the sequence reads aligned to the g.27134 of the SMN1 gene.

[Claim 18~~44~~]. The method of any one of claims 1-~~43~~17, comprising determining a treatment recommendation for the subject based on the copy number of the SMN1 gene determined.

~~45. The method of claim 44, wherein the treatment recommendation comprises administering Nusinersen and/or Zolgensma to the subject.~~

~~46. A method for genotyping cytochrome P450 family 2 subfamily D member 6 (CYP2D6) gene comprising:~~

~~under control of a hardware processor:~~

~~receiving sequence data comprising a plurality of sequence reads obtained from a sample of a subject aligned to cytochrome P450 family 2 subfamily D member 6 (CYP2D6) gene or cytochrome P450 Family 2 Subfamily D Member 7 (CYP2D7) gene;~~

~~determining (i) a first number of sequence reads of the plurality of sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene;~~

~~determining (i) a first normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene using (i) a length of the CYP2D6 gene or the CYP2D7 gene, respectively;~~

~~determining (i) a total copy number of the CYP2D6 gene and the CYP2D7 gene using a Gaussian mixture model comprising a plurality of Gaussians each representing a~~

different integer copy number, given (i) the first normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene;

for one of a plurality of CYP2D6 gene-specific bases, determining a most likely combination, of a plurality of possible combinations each comprising a possible copy number of the CYP2D6 gene and a possible copy number of the CYP2D7 gene summed to the total copy number of the CYP2D6 gene and the CYP2D7 gene determined, given (a) a number of sequence reads of the plurality of sequence reads with bases that support the CYP2D6 gene-specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support a CYP2D7 gene-specific base of corresponding to the CYP2D6 gene-specific base; and

determining an allele of the CYP2D6 gene the subject has using the most likely combination of the possible copy number of the CYP2D6 gene and the possible copy number of the CYP2D7 gene determined for the CYP2D6 gene-specific base.

47. The method of claim 46, wherein the sequencing data comprises whole genome sequencing (WGS) data or short-read WGS data.

48. The method of any one of claims 46-47, wherein the subject is a fetal subject, a neonatal subject, a pediatric subject, an adolescent subject, or an adult subject.

49. The method of any one of claims 46-48, wherein the sample comprises cells or cell-free DNA.

50. The method of any one of claims 46-49, wherein the sample comprises cells or cell-free DNA.

51. The method of any one of claims 46-50, wherein a sequence read of the plurality of sequence reads is aligned to the CYP2D6 gene or the CYP2D7 gene with an alignment quality score of about zero.

52. The method of any one of claims 46-51, wherein determining (i) the first number of sequence reads of the plurality of sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene comprises: determining (i) the first number of sequence reads of the plurality of sequence reads aligned to at least one exon or intron of the CYP2D6 gene or at least one of exon or intron of the CYP2D7 gene.

53. The method of any one of claims 46-52, wherein determining (i) the first normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene comprises: determining (i) the first normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene using (i) the length of the CYP2D6 gene or the CYP2D7 gene, respectively, and (iii) a depth of sequence reads of a region of a genome of the subject other than genetic loci comprising the CYP2D6 gene and the CYP2D7 gene in the sequence data.

54. The method of claim 53, wherein determining (i) the first normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene and (ii) the second normalized number of the sequence reads aligned to the second region comprises:

determining (i) a first CYP2D6 gene or the CYP2D7 gene length normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene using (i) the length of the CYP2D6 gene or the CYP2D7 gene, respectively; and

determining (i) a first normalized depth of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene from (i) the CYP2D6 gene or the CYP2D7 gene length normalized number using the depth of the sequence reads of the region of the genome of the subject other than genetic loci comprising the CYP2D6 gene and the CYP2D7, the first normalized depth of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene being the first normalized number of the sequence reads aligned to the CYP2D6 gene or CYP2D7 gene, respectively.

55. The method of any one of claims 46-54, wherein determining (i) the first normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene comprises: determining (i) the first normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene using (i) a GC content of the CYP2D6 gene or the CYP2D7 gene and (iii) a depth of sequence reads of a region of a genome of the subject other than genetic loci comprising the CYP2D6 gene and the CYP2D7 gene in the sequence data, and (iv) a GC content of the region of the genome.

56. The method of any one of claims 53-55, wherein the depth of the region comprises an average depth or a median depth of the sequence reads of the region of the genome of the subject other than the genetic loci comprising the CYP2D6 gene and the CYP2D7 gene in the sequencing data.

57. The method of claim 56, wherein the region comprises about 3000 pre-selected regions of about 2 kb in length each across the genome of the subject.

58. The method of any one of claims 46-57, wherein (i) the first normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene and/or (ii) the second normalized number of the sequence reads aligned to the second region is about 30 to about 40.

59. The method of any one of claims 46-58, wherein the Gaussian mixture model comprises a one-dimensional Gaussian mixture model.

60. The method of any one of claims 46-59, wherein the plurality of Gaussians of the Gaussian mixture model represents integer copy numbers 0 to 10.

61. The method of any one of claims 46-60, wherein a mean of each of the plurality of Gaussians is the integer copy number represented by the Gaussian.

62. The method of any one of claims 46-61, wherein determining (i) the total copy number of the CYP2D6 gene and the CYP2D7 gene comprises determining (i) the total copy number of the CYP2D6 gene and the CYP2D7 gene using the Gaussian mixture model and a first predetermined posterior probability threshold, given (i) the first normalized number of the sequence reads aligned to the CYP2D6 gene or the CYP2D7 gene.

63. The method of claim 62, wherein the first predetermined posterior probability threshold is 0.95.

64. The method of any one of claims 46-63, wherein the most likely combination of the possible copy number of the CYP2D6 gene and the possible copy number of the CYP2D7 gene is associated with a highest posterior probability, relative to other combinations of the plurality of combinations given (a) the number of sequence reads of the plurality of sequence reads with bases that support the CYP2D6 gene-specific base and (b) the number of sequence reads of the plurality of sequence reads with bases that support the corresponding CYP2D7 gene-specific base.

65. The method of any one of claims 46-64, wherein determining the most likely combination of the possible copy number of the CYP2D6 gene and the possible copy number of the CYP2D7 gene comprises: determining the most likely combination, of the plurality of possible combinations each comprising a possible copy number of the CYP2D6 gene and a possible copy number of the CYP2D7 gene summed to the total copy number of the CYP2D6 gene and the CYP2D7 gene determined, given a ratio of (a) the number of sequence reads of the plurality of sequence reads with bases that support the CYP2D6 gene-specific base and (b) the number of sequence reads of the plurality of sequence reads with bases that support a CYP2D7 gene-specific base of corresponding to the CYP2D6 gene-specific base.

66. The method of any one of claims 46-65, wherein determining the most likely combination of the possible copy number of the CYP2D6 gene and the possible copy number of the CYP2D7 gene comprises:

determining (a) the number of sequence reads of the plurality of sequence reads with bases that support the CYP2D6 gene-specific base and (b) the number of sequence reads of the plurality of sequence reads with bases that support a CYP2D7 gene-specific base of corresponding to the CYP2D6 gene-specific base;

determining a ratio of (a) the number of sequence reads of the plurality of sequence reads with bases that support the CYP2D6 gene-specific base and (b) the number of sequence reads of the plurality of sequence reads with bases that support a CYP2D7 gene-specific base of corresponding to the CYP2D6 gene-specific base; and determining the most likely combination, of the plurality of possible

combinations each comprising a possible copy number of the CYP2D6 gene and a possible copy number of the CYP2D7 gene summed to the total copy number of the CYP2D6 gene and the CYP2D7 gene determined, given the ratio (a) a number of sequence reads of the plurality of sequence reads with bases that support the CYP2D6 gene-specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support a CYP2D7 gene-specific base of corresponding to the CYP2D6 gene-specific base.

67. The method of any one of claims 46-66, wherein determining the allele of the CYP2D6 gene the subject has comprises: determining one or more structural variants of the CYP2D6 gene the subject has using the most likely combination of the possible copy number of the CYP2D6 gene and the possible copy number of the CYP2D7 gene determined for the CYP2D6 gene-specific base.

68. The method of claim 67,

wherein determining the most likely combination of the possible copy number of the CYP2D6 gene and the possible copy number of the CYP2D7 gene comprises: for each of the plurality of CYP2D6 gene-specific bases, determining a most likely combination, of a plurality of possible combinations each comprising a possible copy number of the CYP2D6 gene and a possible copy number of the CYP2D7 gene summed to the total copy number the CYP2D6 gene and the CYP2D7 gene determined, associated with a highest posterior probability given (a) a number of sequence reads of the plurality of sequence reads with bases that support the CYP2D6 gene-specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support a CYP2D7 gene-specific base of the CYP2D7 gene corresponding to the CYP2D6 gene- specific base, and

wherein determining the one or more structural variants of the CYP2D6 gene the subject has comprises determining the one or more structural variants of the CYP2D6 gene the subject has using the most likely combination of the possible copy number of the CYP2D6 gene and the possible copy number of the CYP2D7 gene determined for each of the plurality of CYP2D6 gene-specific bases.

69. The method of claim 68, wherein determining the one or more structural variants of the CYP2D6 gene the subject has comprises: determining the one or more structural variants of the CYP2D6 gene the subject has based on the copy numbers of the CYP2D6 gene of the most likely combinations determined for two or more of the plurality of CYP2D6 gene-specific bases that are different and the positions of the two or more CYP2D6 gene-specific bases.

70. The method of any one of claims 68-69, wherein the CYP2D6 gene-specific base

has a concordance with each of the plurality of CYP2D6 gene-specific bases other than the CYP2D6 gene-specific base above a predetermined concordance threshold.

71. The method of claim 70, wherein the concordance threshold is 97%.

72. The method of any one of claims 68-71, wherein the plurality of CYP2D6 gene-specific bases comprises 118 CYP2D6 gene-specific bases.

73. The method of any one of claims 68-72, wherein the plurality of CYP2D6 gene-specific bases if the subject is of a first race, the plurality of CYP2D6 gene-specific bases if the subject is of a second race, and the plurality of CYP2D6 gene-specific bases if the subject is of an unknown race are different.

74. The method of any one of claims 68-73, wherein a race of the subject is unknown, and wherein the plurality of CYP2D6 gene-specific bases is not race specific.

75. The method of any one of claims 68-73, wherein a race of the subject is known, and wherein the plurality of CYP2D6 gene-specific bases is specific to the race of the subject.

76. The method of any one of claims 68-75, comprising:

receiving race information of the subject; and

selecting the plurality of CYP2D6 gene-specific bases from pluralities of CYP2D6 gene-specific bases based on the race information received.

77. The method of any one of claims 67-76, comprising:

determining (ii) a second number of sequence reads of the plurality of sequence reads aligned to a spacer region between the CYP2D7 gene and a repetitive element REP7 downstream of the CYP2D7 gene;

determining (ii) a second normalized number of the sequence reads aligned to the spacer region using (ii) a length of the spacer region; and

determining (ii) a copy number of the spacer region using the Gaussian mixture model given (ii) the second normalized number of the sequence reads aligned to the spacer region,

wherein determining the structural variant of the CYP2D6 gene the subject has comprises: determining the allele of the CYP2D6 gene the subject has using the most likely combination of the possible copy number of the CYP2D6 gene and the possible copy number of the CYP2D7 gene determined for the CYP2D6 gene-specific base and the copy number of the spacer region.

78. The method of claim 77, wherein the one or more structural variants comprise a CYP2D6/CYP2D7 fusion allele with the spacer region and the repetitive element REP7 downstream of the CYP2D6/CYP2D7 fusion allele.

79. The method of any one of claims 46-78, comprising: determining one or more

small variants of the CYP2D6 gene the subject has using the sequence data received.

80. The method of claim 79, wherein determining the one or more small variants of the CYP2D6 gene the subject has comprises: for a small variant position of the CYP2D6 gene associated with a small variant allele of the CYP2D6 gene, determining a most likely combination of a possible copy number of the small variant allele of the CYP2D6 gene at the small variant position and a possible copy number of a reference allele of the CYP2D6 gene summed to a copy number of the CYP2D6 gene at the small variant position, given (a) a number of sequence reads with bases supporting the small variant allele of the CYP2D6 gene at the small variant position and (b) a number of sequence reads with bases supporting the reference allele of the CYP2D6 gene at the small variant position, wherein the possible copy number of the small variant allele of the CYP2D6 gene of the most likely combination at the small variant position indicates the one or more small variants of the CYP2D6 gene.

81. The method of claim 79, wherein determining the one or more small variants of the CYP2D6 gene the subject has comprises: for each of a plurality of small variant positions of the CYP2D6 gene, the small variant position being associated with a small variant allele of the CYP2D6 gene, determining a most likely combination of a possible copy number of the small variant allele of the CYP2D6 gene at the small variant position and a possible copy number of a reference allele of the CYP2D6 gene at the small variant position summed to a copy number of the CYP2D6 gene at the small variant position, given (a) a number of sequence reads with bases supporting the small variant allele of the CYP2D6 gene at the small variant position and (b) a number of sequence reads with bases supporting the reference allele of the CYP2D6 gene at the small variant position, wherein the possible copy numbers of the small variant alleles of the CYP2D6 gene of the most likely combinations at the plurality of small variant positions indicate the one or more small variants of the CYP2D6 gene.

82. The method of any one of claims 46-78, comprising:

for a small variant position of the CYP2D6 gene associated with a small variant allele of the CYP2D6 gene, determining a most likely combination of a possible copy number of the small variant allele of the CYP2D6 gene at the small variant position and a possible copy number of a reference allele of the CYP2D6 gene at the small variant position summed to a copy number of the CYP2D6 gene at the small variant position, given (a) a number of sequence reads aligned to the CYP2D6 gene that overlap the small variant position and with bases supporting the small variant allele of the CYP2D6 gene at the small variant position and (b) a number of sequence reads aligned to the CYP2D6 gene that overlap the small variant position and with bases supporting the reference allele of the CYP2D6 gene at the small variant position; and

determining one or more small variants the CYP2D6 gene using the possible copy number of the small variant allele of the CYP2D6 gene of the most likely combination determined.

83. The method of any one of claims 46-78, comprising:

for each of a plurality of small variant positions of the CYP2D6 gene, the small variant position being associated with a small variant allele of the CYP2D6 gene, determining a most likely combination of a possible copy number of the small variant allele of the CYP2D6 gene at the small variant position and a possible copy number of a reference allele of the CYP2D6 gene at the small variant position summed to a copy number of the CYP2D6 gene at the small variant position, given (a) a number of sequence reads aligned to the CYP2D6 gene that overlap the small variant position and with bases supporting the small variant allele of the CYP2D6 gene at the small variant position and (b) a number of sequence reads aligned to the CYP2D6 gene that overlap the small variant position and with bases supporting the reference allele of the CYP2D6 gene at the small variant position; and

determining one or more small variants the CYP2D6 gene using the possible copy numbers of the small variant alleles of the CYP2D6 gene of the most likely combinations at the plurality of small variant positions determined.

84. The method of any one of claims 68-83, wherein the small variant position is in a CYP2D6/CYP2D7 homology region, wherein determining the most likely combination comprises determining the most likely combination of the possible copy number of the small variant allele of the CYP2D6 gene at the small variant position and the possible copy number of the reference allele of the CYP2D6 gene at the small variant position summed to the copy number of the CYP2D6 gene at the small variant position given (a) a number of sequence reads aligned to the CYP2D6 gene or CYP2D7 gene with bases supporting the small variant allele of the CYP2D6 gene at the small variant position and/or (b) a number of sequence reads aligned to the CYP2D6 gene or CYP2D7 gene with bases supporting the reference allele of the CYP2D6 at the small variant position.

85. The method of any one of claims 68-83, wherein the small variant position is not in a CYP2D6/CYP2D7 homology region, wherein determining the most likely combination comprises determining the most likely combination of the possible copy number of the small variant allele of the CYP2D6 gene at the small variant position and the possible copy number of the reference allele of the CYP2D6 gene at the small variant position summed to the copy number of the CYP2D6 gene at the small variant position given (a) a number of sequence reads aligned to the CYP2D6 gene and not to the CYP2D7 gene with bases supporting the small

variant allele of the CYP2D6 gene at the small variant position and/or (b) a number of sequence reads aligned to the CYP2D6 gene and not CYP2D7 gene with bases supporting the reference allele of the CYP2D6 at the small variant position.

86. The method of any one of claims 80-85, comprising determining the copy number of the CYP2D6 gene at the small variant position

87. The method of any one of claims 80-86, wherein the copy number of the CYP2D6 gene at the small variant position comprises a copy number of the CYP2D6 gene.

88. The method of any one of claims 80-87, wherein the copy number of the CYP2D6 gene at the small variant position comprises a copy number of the CYP2D6 gene of possible copy numbers of the CYP2D6 gene of the most likely combinations determined.

89. The method of any one of claims 80-88, wherein the copy number of the CYP2D6 gene at the small variant position comprises a copy number of the CYP2D6 gene of possible copy numbers of the CYP2D6 gene of the most likely combinations determined and closest to the small variant position.

90. The method of any one of claims 80-89, wherein the copy number of the CYP2D6 gene at the small variant position comprises a copy number of the CYP2D6 gene at a 5' position or 3' position of the small variant position.

91. The method of any one of claims 80-90, comprising: (a) determining the number of sequence reads with bases supporting the small variant allele of the CYP2D6 gene; and (b) determining the number of sequence reads with bases supporting the reference allele of the CYP2D6 gene.

92. The method of any one of claims 46-91, wherein determining the allele of the CYP2D6 gene the subject has comprises: determining alleles of the CYP2D6 gene the subject has.

93. The method of any one of claims 46-92, wherein determining the allele of the CYP2D6 gene the subject has comprises: determining a star allele and/or a haplotype of the CYP2D6 gene the subject has using the one or more structural variants of the CYP2D6 gene determined and/or the one or more small variants of the CYP2D6 gene determined, optionally wherein the star allele is associated with a known function.

94. The method of any one of claims 46-93, comprising: determining a level of CYP2D6 enzymatic activity in the subject using the allele of the CYP2D6 gene determined.

95. The method of claim 94, wherein the enzymatic activity is poor, intermediate, normal, or ultrarapid.

96. The method of any one of claims 46-95, comprising determining a dosage recommendation of a treatment and/or a treatment recommendation for the subject based on the

~~allele of the CYP2D6 gene the subject has.~~

[Claim 19~~7~~]. A system for paralog genotyping comprising:

non-transitory memory configured to store executable instructions and sequence data comprising a plurality of sequence reads obtained from a sample of a subject aligned to a first paralog or a second paralog; and

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

determining a copy number of paralogs of a first type using a Gaussian mixture model comprising a plurality of Gaussians each representing a different integer copy number given (i) a first number of sequence reads aligned to a first region;

for one of a plurality of first paralog-specific bases, determining a most likely combination, of a plurality of possible combinations each comprising a possible copy number of a first paralog of the first type and a possible copy number of a second paralog of the first type summed to the copy number of the paralogs of the first type determined, given (a) a number of sequence reads of the plurality of sequence reads with bases that support the first paralog-specific base and (b) a number of sequence reads of the plurality of sequence reads with bases that support a second paralog-specific base of the second paralog corresponding to the first paralog-specific base; and

determining a copy number or an allele of first paralog using the most likely combination of the possible copy number of the first paralog and the possible copy number of the second paralog determined for the first paralog-specific base.

~~98. The system of claim 97, wherein the hardware processor is programmed by the executable instructions to perform: determining (i) a first number of sequence reads of a plurality of sequence reads in sequence data obtained from a sample of a subject aligned to the first region.~~

~~99. The system of any one of claims 97-98, wherein the hardware processor is programmed by the executable instructions to perform: determining (i) a first normalized number of the sequence reads aligned to the first region using (i) a length of the first region, wherein determining the copy number of the first type of paralogs comprises: determining the copy number of the first type of paralogs using the Gaussian mixture model given (i) the first normalized number of the sequence reads aligned to the first region.~~

~~100. The system of any one of claims 97-99, wherein the hardware processor is programmed by the executable instructions to perform: receiving the sequence data comprising the plurality of sequence reads aligned to the first region.~~

~~101. The system of any one of claims 97-100, wherein the hardware processor is~~

~~programmed by the executable instructions to perform: determining a copy number of one or more paralogs of a second type using the Gaussian mixture given (ii) a second number of sequence reads aligned to a second region.~~

~~102. The system of claim 101, wherein determining the copy number or the allele of the first paralog comprises: determining the copy number or the allele of the first paralog using the most likely combination of the possible copy number of the first paralog and the possible copy number of the second paralog determined for the first paralog-specific base and the copy number of the one or more paralogs of the second type.~~

~~103. The system of claim 101, wherein the hardware processor is programmed by the executable instructions to perform: determining a copy number of paralogs of a third type from the copy number of the paralogs of the first type and the copy number of the paralogs of the second type, and wherein determining the copy number or the allele of the first paralog comprises: determining the copy number or the allele of the first paralog using the most likely combination of the possible copy number of the first paralog and the possible copy number of the second paralog determined for the first paralog-specific base.~~

[Claim ~~1~~2~~0~~4]. The system of ~~any one of~~ claim~~s~~ 19~~7-103~~, wherein the first paralog is survival of motor neuron 1 (SMN1) gene, wherein the second paralog is survival of motor neuron 2 (SMN2) gene, wherein the first region comprises at least one exon 1 to exon 6 of the SMN1 gene and at least one exon 1 to exon 6 of the SMN2 gene, wherein the second region comprises at least one of exon 7 and exon 8 of the SMN1 gene and at least one of exon 7 and exon 8 of the SMN2 gene, wherein the paralogs of the first type comprise an intact SMN1 gene and an intact SMN2 gene, wherein the one or more paralogs of the second type comprise the intact SMN1 gene, the intact SMN2 gene, a truncated SMN1 gene, or a truncated SMN2 gene, and wherein the copy number of the first paralog comprises a copy number of the SMN1 gene.

~~105. The system of any one of claims 97-103, wherein the first paralog is Cytochrome P450 Family 2 Subfamily D Member 6 (CYP2D6) gene, wherein the second paralog is Cytochrome P450 Family 2 Subfamily D Member 7 (CYP2D7) gene, wherein the first region comprises the CYP2D6 gene and the CYP2D7 gene, wherein the second region comprises a spacer region between the CYP2D7 gene and a repetitive element REP7 downstream of the CYP2D7 gene, wherein the paralogs of the first type comprise the CYP2D6 gene and the CYP2D7 gene, wherein the one or more paralogs of the second type comprise a CYP2D6/CYP2D7 fusion allele with the spacer region and the repetitive element REP7 downstream of the CYP2D6/CYP2D7 fusion allele, and wherein the allele of the first paralog comprises an allele of the CYP2D6 gene is a small variant or a structural variant of the CYP2D6 gene.~~

~~106. The system of any one of claims 97-105, wherein the first paralog and the second paralog have a sequence identity of at least 90%.~~