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.

[Claim 3]. The method of any one of claims 1or 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]. The method of any one of claims 1-3, 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]. The method of any one of claims 1-4, 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.

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

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

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

[Claim 9]. The method of any one of claims 1-8, 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.

[Claim 10]. The method of any one of claims 1-9, 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.

[Claim 11]. The method of any one of claims 1-10, 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]. The method of any one of claims 1-11, 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]. The method of any one of claims 1-12, 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]. The method of any one of claims 1-13,

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.

[Claim 15]. The method of claim 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.

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

[Claim 17]. The method of any one of claims 1-16, 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]. The method of any one of claims 1-17, comprising determining a treatment recommendation for the subject based on the copy number of the SMN1 gene determined.

[Claim 19]. 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.

[Claim 20]. The system of claim 19, 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.