CLAIMS

1. A computer-implemented method comprising:

generating an initial reference panel comprising marker-variant indicators for marker variants at genomic coordinates corresponding to genomic samples of different haplotypes;

adding, to the initial reference panel, at least one target-variant position comprising a data field for target-variant indicators indicating a presence or absence of a target variant at a genomic coordinate within the genomic samples;

phasing, based on the marker variants, alleles of the genomic samples to determine a presence or absence of the target variant at the genomic coordinate in corresponding alleles present on maternal haplotypes and paternal haplotypes; and

generating a target-variant-reference panel comprising (i) the marker-variant indicators of the initial reference panel and (ii) the target-variant indicators within the at least one target-variant position for the phased alleles of the genomic samples.

2. The computer-implemented method of claim 1, wherein the at least one target-variant position comprises a target-variant position for target-variant indicators of a biallelic target variant or a multi-allelic target variant.

3. The computer-implemented method of claim 1 or 2, wherein phasing the alleles of the genomic samples comprises phasing heterozygous alleles of a subset of the genomic samples.

4. The computer-implemented method of any one of claims 1-3, wherein generating the initial reference panel comprises generating a phased reference panel comprising the marker-variant indicators for marker variants phased according to the maternal haplotypes and the paternal haplotypes of the genomic samples.

5. The computer-implemented method of any one of claims 1-4, wherein adding the at least one target-variant position comprises adding, to the initial reference panel, the at least one target-variant position comprising the data field not included in the initial reference panel.

6. The computer-implemented method of any one of claims 1-5, wherein the target variant comprises a repeat expansion.

7. The computer-implemented method of any one of claims 1-6, wherein the target variant comprises a deletion, an insertion, a duplication, an inversion, a translocation, or a copy number variation (CNV) transmitted within a population.

8. The computer-implemented method of any one of claims 1-7, wherein the target variant satisfies one or more of a threshold carrier frequency, a threshold linkage disequilibrium (LD) with respect to particular marker variants, or a threshold mutation rate.

9. A system comprising:

at least one processor; and

a non-transitory computer readable medium comprising instructions that, when executed by the at least one processor, cause the system to:

identify nucleotide reads corresponding to a target genomic sample;

access a target-variant-reference panel comprising (i) marker-variant indicators for marker variants from an initial reference panel and (ii) target-variant indicators within a data field of at least one target-variant position added to the initial reference panel for phased alleles of genomic samples of different haplotypes, the target-variant indicators indicating a presence or absence of a target variant at a genomic coordinate within the genomic samples; and

impute a genotype call for the target variant within the target genomic sample based on a comparison of the target-variant-reference panel and the nucleotide reads corresponding to the target genomic sample.

10. The system of claim 9, wherein the target-variant indicators indicate a presence or absence of the target variant in the at least one target-variant position for the phased alleles of the genomic samples.

11. The system of claim 9 or 10, further comprising instructions that, when executed by the at least one processor, cause the system to:

determine phased alleles of the target genomic sample based on the comparison of the target-variant-reference panel and the nucleotide reads corresponding to the target genomic sample; and

impute the genotype call by imputing a phased genotype call for the target variant within the target genomic sample based on the phased alleles of the target genomic sample.

12. The system of any one of claims 9-11, further comprising instructions that, when executed by the at least one processor, cause the system to impute the genotype call for the target variant by generating a prediction of whether the target genomic sample comprises the target variant.

13. The system of claim 12, further comprising instructions that, when executed by the at least one processor, cause the system to generate the prediction by predicting whether the target genomic sample comprises a pathogenic variant at an allele present on a maternal haplotype or a paternal haplotype.

14. The system of any one of claims 9-13, further comprising instructions that, when executed by the at least one processor, cause the system to impute the genotype call by:

identifying, within the nucleotide reads corresponding to the target genomic sample, one or more single-nucleotide polymorphisms (SNPs) as one or more marker variants within the target-variant-reference panel for the target variant; and

determining the genotype call further based on the one or more SNPs within the nucleotide reads.

15. The system of any one of claims 9-14, further comprising instructions that, when executed by the at least one processor, cause the system to impute the genotype call for the target variant by imputing the genotype call for a repeat expansion.

16. A non-transitory computer-readable medium storing instructions that, when executed by at least one processor, cause a computing device to:

identify nucleotide reads corresponding to a target genomic sample;

access a target-variant-reference panel comprising (i) marker-variant indicators for marker variants from an initial reference panel and (ii) target-variant indicators within a data field of at least one target-variant position added to the initial reference panel for phased alleles of genomic samples, the target-variant indicators indicating a presence or absence of a target variant at a genomic coordinate within the genomic samples; and

impute a genotype call for the target variant within the target genomic sample based on a comparison of the target-variant-reference panel and the nucleotide reads corresponding to the target genomic sample.

17. The non-transitory computer-readable medium of claim 16, wherein the target-variant indicators indicate a presence or absence of the target variant in the at least one target-variant position for the phased alleles of genomic samples.

18. The non-transitory computer-readable medium of claim 16 or 17, further comprising instructions that, when executed by the at least one processor, cause the computing device to:

determine phased alleles for the target genomic sample based on the comparison of the target-variant-reference panel and the nucleotide reads corresponding to the target genomic sample; and

impute the genotype call by imputing a phased genotype call for the target variant within the target genomic sample based on the phased alleles of the target genomic sample.

19. The non-transitory computer-readable medium of any one of claims 16-18, further comprising instructions that, when executed by the at least one processor, causes the computing device to impute the genotype call for the target variant by imputing the genotype call for a Replication Factor C Subunit 1 (RFC1) gene, a Cytochrome P450 Family 2 Subfamily D Member 6 (CYP2D6) gene, Cytochrome P450 Family 2 Subfamily B Member 6 (CYP2B6) gene, Cytochrome P450 Family 21 Subfamily A Member 2 (CYP21A2) gene, Survival Motor Neuron 1

(SMN1) gene, Survival Motor Neuron 2 (SMN2) gene, Glucosylceramides Beta (GBA) gene, Blood Group Rh(CE) (RHCE) gene, Lipoprotein(A) (LPA) gene, a Fragile X Mental Retardation 1 (FMR1) gene, a Hexosaminidase Subunit Alpha (HEXA) gene, Hemoglobin Subunit Alpha 1 (HBA1) gene, Hemoglobin Subunit Alpha 2 (HBA2) gene, or a Hemoglobin Subunit Beta (HBB) gene.

20. The non-transitory computer-readable medium of any one of claims 16-19, wherein the target-variant-reference panel comprises marker-variant indicators for marker variants within a target genomic region for the target variant and does not comprise additional marker-variant indicators for additional marker variants outside of the target genomic region.