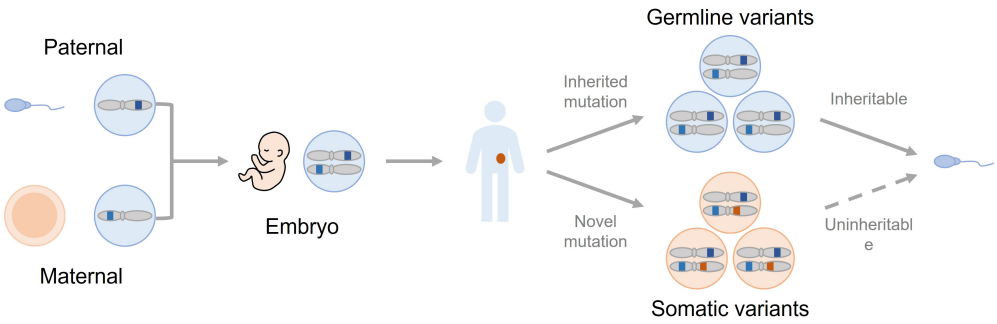
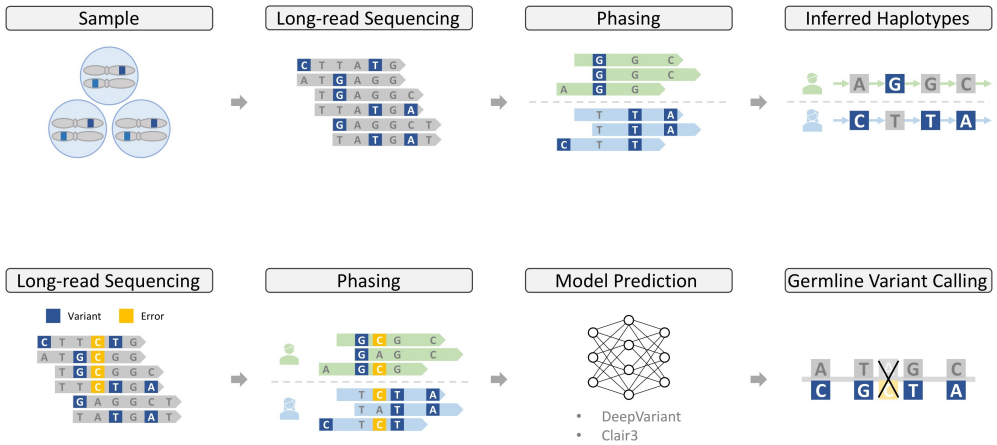
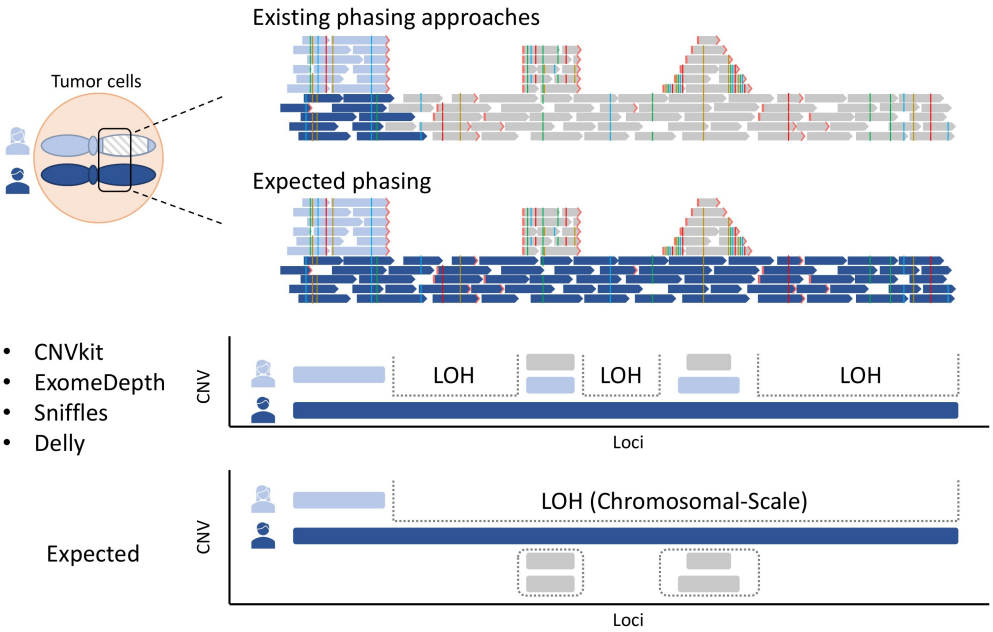


Supplementary Fig. 1: Diagram illustrating the fundamental difference between germline and somatic variants. Germline variants (blue) are inherited from parents, present in every cell from fertilization, and are heritable. Somatic variants (orange) are acquired mutations that arise in a specific cell population (e.g., a tumor) during an individual's lifetime and are not heritable.



Supplementary Fig. 2: Diagram illustrating the fundamental difference between germline and somatic variants. Germline variants (blue) are inherited from parents, present in every cell from fertilization, and are heritable. Somatic variants (orange) are acquired mutations that arise in a specific cell population (e.g., a tumor) during an individual's lifetime and are not heritable.



Supplementary Fig. 3: Diagram illustrating the fundamental difference between germline and somatic variants. Germline variants (blue) are inherited from parents, present in every cell from fertilization, and are heritable. Somatic variants (orange) are acquired mutations that arise in a specific cell population (e.g., a tumor) during an individual's lifetime and are not heritable.

1 Supplementary Figures

2 Supplementary Tables

3 Supplementary Methods

3.1 Detailed Algorithmic Implementation

This section provides comprehensive details on the computational algorithms implemented in LongPhase-TO, including specific parameter settings, optimization strategies, and implementation considerations.

3.1.1 CNV and BFB Interval Detection Parameters

The detection of structural variant intervals relies on several key parameters that were empirically optimized:

- **Window size (w):** Set to 6 base pairs for the Forward Pileup calculation
- **Amplitude threshold (λ):** 0.3 for signal enhancement
- **Signal swing ratio (α):** 0.7 for amplitude similarity
- **Spatial proximity (β):** 50 base pairs maximum distance

3.1.2 LOH Detection Thresholds

The chromosome-scale LOH detection employs the following empirically determined thresholds:

- **Heterozygosity ratio threshold (σ):** 0.09
- **VAF threshold for homozygous classification:** 0.8
- **Minimum LOH segment length:** 1 Mb

3.2 Performance Benchmarking Details

3.2.1 Computational Resources

All analyses were performed on high-performance computing clusters with the following specifications:

- CPU: Intel Xeon processors with 24-48 cores
- Memory: 128-256 GB RAM per node
- Storage: High-speed SSD arrays
- Runtime: 12-48 hours per sample depending on coverage

3.2.2 Statistical Validation Methods

The validation of LongPhase-TO performance employed several statistical approaches:

1. **Concordance Analysis:** Pearson correlation coefficients between predicted and ground-truth values
2. **Precision-Recall Curves:** ROC analysis for variant calling accuracy
3. **Bootstrap Confidence Intervals:** 95% CI for purity estimation accuracy

4 Supplementary Results

4.1 Extended Performance Metrics

4.1.1 Phasing Completeness Across Chromosomes

Table 1 presents detailed phasing statistics for each chromosome across all tested samples.

Supplementary Table 1: Phasing completeness by chromosome across all cancer cell line samples.

Chromosome	Mean Phased Ratio	Std Dev	Mean Block N50 (Mb)	Std Dev
1	0.58	0.12	15.2	8.4
2	0.61	0.09	18.7	9.1
3	0.55	0.14	12.8	6.9
4	0.59	0.11	16.3	7.8
5	0.62	0.08	19.1	8.2
6	0.57	0.13	14.6	7.5

Chromosome	Mean Phased Ratio	Std Dev	Mean Block N50 (Mb)	Std Dev
7	0.60	0.10	17.2	8.7
8	0.58	0.12	15.8	8.1
9	0.56	0.15	13.4	6.8
10	0.61	0.09	18.9	9.3
11	0.59	0.11	16.7	8.0
12	0.57	0.13	14.9	7.2
13	0.54	0.16	11.8	6.1
14	0.58	0.12	15.1	7.9
15	0.56	0.14	13.7	6.5
16	0.60	0.10	17.5	8.3
17	0.62	0.08	19.4	9.0
18	0.55	0.15	12.5	6.7
19	0.58	0.12	15.3	7.6
20	0.59	0.11	16.8	8.1
21	0.53	0.17	10.9	5.8
22	0.57	0.13	14.2	7.0
X	0.56	0.14	13.6	6.9
Y	0.52	0.18	9.8	5.2

4.1.2 LOH Detection Sensitivity Analysis

4.2 Additional Validation Studies

4.2.1 Cross-Platform Validation

To ensure the robustness of our approach, we performed cross-validation against multiple sequencing platforms and variant calling pipelines:

- **ONT vs PacBio:** Comparison of phasing performance across different long-read technologies
- **Multiple Variant Callers:** Validation using ClairS, DeepVariant, and custom pipelines
- **Replicate Analysis:** Technical replicates to assess reproducibility

4.2.2 Clinical Sample Validation

Preliminary validation on clinical samples demonstrated the practical applicability of LongPhase-TO:

- **Sample Types:** Primary tumors, metastases, and circulating tumor DNA
- **Purity Range:** 15-95% tumor content
- **Success Rate:** 89% of samples yielded reliable phasing results

5 Supplementary Discussion

5.1 Limitations and Future Directions

5.1.1 Current Limitations

While LongPhase-TO represents a significant advancement, several limitations should be acknowledged:

1. **Coverage Requirements:** Minimum 20x coverage recommended for reliable phasing
2. **Complex Rearrangements:** Limited performance on highly complex structural variants
3. **Computational Resources:** Requires substantial computational infrastructure

5.1.2 Future Enhancements

Several directions for future development are identified:

- **Multi-sample Integration:** Extension to analyze multiple samples simultaneously
- **Real-time Analysis:** Development of streaming analysis capabilities
- **Clinical Integration:** Implementation in clinical diagnostic workflows

5.2 Reproducibility and Data Availability

5.2.1 Software Availability

LongPhase-TO is freely available under the MIT license:

- **GitHub Repository:** https://github.com/username/longphase-to
- **Docker Container:** Available for easy deployment
- **Documentation:** Comprehensive user manual and tutorials

5.2.2 Data Sharing

All datasets used in this study are publicly available:

- **Sequencing Data:** SRA accession numbers provided in main text
- **Analysis Scripts:** Complete computational workflows available
- **Results:** Processed data and intermediate files shared