



CASCADE 2027 NON-CONFIDENTIAL SUMMARY »

Long-Read DNA Sequencing for Precision Relationship and Ancestry Detection

Overview

Determining whether two people are biologically related — and how closely — is central to genetic disease diagnosis, genealogy, and family health communication, but today's leading methods rely on statistical matching of common DNA variants using population reference panels. The reference panels approach systematically underperforms for people from populations that are underrepresented in existing genomic databases, and it cannot reliably detect the short DNA segments that link distant genetic relationships because common-variant methods lose resolution below roughly 10 centimorgans. This means that current technology cannot consistently and reliably detect relatives that are more distant than 4th or 5th cousins. Our UW innovation uses long-read DNA sequencing to reconstruct an individual's two chromosome copies (haplotypes) directly from the physical sequencing reads themselves, without needing a population reference panel. By additionally identifying rare genetic variants — which are far less likely to be shared by chance than common variants — the system can classify genomic segments as inherited from a common ancestor with meaningfully higher resolution and accuracy than conventional approaches, allowing accurate connections between genetic relatives as far as 10th cousins.

- **Removes population bias:** because the method does not depend on reference panels built primarily from well-studied populations, it works equally well for individuals from underrepresented ancestries.
- **Higher resolution matching:** rare-variant analysis can provide substantially greater resolution for detecting shared genomic segments compared to common-variant approaches.
- **Broad application base:** the same underlying method supports relationship inference allowing people to identify thousands of additional biological relatives compared with current technology – which enables more family health communication and genealogy/ancestry services.

Description

The system ingests long-read sequencing data from two or more individuals and performs read-based haplotype phasing — assigning genetic variants to a specific chromosome copy based on their physical co-occurrence on individual sequencing reads, rather than by statistical inference from a reference database. This allows phasing of both common and rare variants in a single

COMOTION REFERENCE »

BDP 9343

INNOVATION MANAGER »

Jennifer S. McCullar

SECTOR »

Health Tech

LEAD INVESTIGATORS »

Brian Shirts, MD, PhD:
Department of Laboratory
Medicine and Pathology,
University of Washington and
Department of Pathology,
Microbiology, and Immunology,
Vanderbilt University Medical
Center

Nikhita Damaraju, MS:
Department of Laboratory
Medicine and Pathology,
University of Washington

Danny Erwin Miller, MD, PhD:
Department of Laboratory
Medicine and Pathology,
University of Washington

UNIVERSITY OF WASHINGTON COMOTION »

1100 NE Campus Parkway, # 200
Seattle, WA 98195
Tel: (206) 543-3970
uwcomotion@uw.edu
comotion.uw.edu



pass, including in complex or repetitive genomic regions that are typically inaccessible to conventional short-read sequencing.

Phased haplotypes and rare variants, that are unlikely to be shared by chance, allow identification of window as identical-by-descent (IBD), that are as small as 200kb (about 50 times smaller than current resolution of DNA analysis from companies like Ancestry and MyHeritage).

- **Modality:** research tool / bioinformatics software platform with downstream diagnostic and consumer-genealogy applications.
- **Validation:** benchmarked in a retrospective study using a cohort of parent-child sequencing trios with known genetic relationships as ground truth, comparing the approach against conventional reference-panel-based phasing. The framework demonstrated strong accuracy in distinguishing related from unrelated genomic segments, and separately produced longer detected shared-DNA segments and greater genome coverage than conventional reference-panel phasing across the majority of chromosomes tested.

Technology Readiness Level (TRL)

TRL 4 — System or subsystem components tested in a laboratory environment. The core read-based phasing and rare-variant IBD classification framework has been implemented in software and validated retrospectively against a multi-trio long-read sequencing dataset using known parent-child relationships as ground truth, demonstrating strong classification accuracy genome-wide. Testing to date has occurred in a research/laboratory setting; the system has not yet been evaluated in a clinical or consumer-facing operational environment.

Value Proposition

- Improves relationship-detection accuracy for individuals from populations underrepresented in standard genomic reference panels, addressing a documented equity gap in existing DNA-matching services.
- Delivers meaningfully higher-resolution detection of shared DNA segments than conventional common-variant approaches, enabling identification of more distant or smaller shared genomic regions.
- Requires no population reference panel or parental samples, simplifying data pipelines.
- Demonstrated strong classification accuracy for parent-child relationships in retrospective validation, using an objective, reproducible statistical framework.
- Applicable across multiple commercial pathways, including consumer genealogy/ancestry matching, diagnostic test accuracy improvement, and genetic counseling tools for rare-disease families.

Intellectual Property Position

Patent application pending.

Market Opportunity

- **Consumer genealogy and DNA-matching platforms:** (e.g., ancestry and relative-finder services) represent a direct commercial application. More than 50 million people have done a DNA test to find relatives. This is a paradigm shifting technology – allowing improved match accuracy and thousands of additional connections for users of these tools.

- **Clinical and diagnostic laboratories:** performing rare-disease testing, prenatal screening, and newborn screening could use the underlying rare-variant/family-information framework to improve confidence in low-coverage sequencing results at lower cost.
- **Cases involving uncertain biological relationships:** including adoption search, uncertain paternity, and unknown-parentage cases, represent an additional direct-to-consumer and forensic-adjacent application for the underlying rare-variant relatedness methodology.
- **Competitive landscape:** competing IBD-detection and genetic-genealogy platforms (e.g., large consumer DNA-testing companies) rely on common-variant, reference-panel-based matching, limiting their power to detect genetic relatives accurately beyond 3rd or 4th cousins. Because this technology is inherently much more powerful, it has potential to disrupt the market and replace older technologies.

References

1. Shirts, B.H., et al., "Family-specific variants and the limits of human genetic diversity," *Genetics in Medicine*, 2018. PMID: 30019097
2. Ranola, J.M.O., et al., "Modified capture-recapture estimates of the number of families with Lynch syndrome in Central Ohio," *Fam Cancer*. 2019 Jan;18(1):67-73. PMID: 30019097.
3. Shirts, B.H. "ConnectMyVariant: An Innovative Use of Technology and Social Networks to Realize the Benefits of Cascade Screening." *Public Health Genomics*. 2023;26(1):177-182. PMID: 37751715.
4. Damaraju N., et al. "Determinants of haplotype phasing accuracy in long-read human genome sequencing," *bioRxiv [Preprint]*. 2026 May 10:2026.05.04.722832 PMID: 42388792.

Additional peer-reviewed publication(s) describing the validation study referenced in this profile are anticipated but not yet publicly available; citations will be added upon publication.