Human Copy Number and Haplotype-Resolved Structural Variation Associate with Chronic Wound Bacterial Species

Khalid Omeir B.S.[1], Jacob Ancira M.S.[1,2], Rebecca Gabrilska B.S.[3], Craig Tipton PhD[1,2], Clint Miller[4], Ashley Noe B.S.[1], Kumudu Subasinghe M.S.[5], Megan Rowe M.S.[5], Nicole Phillips PhD[5], Joseph Wolcott MD[4], Caleb D. Phillips PhD[1,6]

[1] Department of Biological Sciences, Texas Tech University, Lubbock, USA, [2] RTL Genomics, MicroGenDX, Lubbock, Texas, USA, [3] Department of Surgery, Texas Tech University Health Sciences Center, Lubbock, Texas, USA, [4] Southwest Regional Wound Care Center, Lubbock, Texas, USA, [5] Microbiology, Immunology & Genetics, University of North Texas Health Science Center, Fort Worth, Texas, USA, [6] Natural Science Research Laboratory, Texas Tech University, Lubbock, Texas, USA

NIH Grant ID: 1R15GM141973-01

Introduction

- Chronic wound microbiomes influence wound chronicity and healing outcomes.
- Patient-to-patient differences in wound microbial composition suggest host factors may contribute to microbial variation.
- Prior work has linked host genetic variation, including nucleotide polymorphisms and heritable gene expression, to chronic wound bacteria.
- Copy number variants, or CNVs, are deletions or duplications of DNA segments that can alter gene dosage and regulation, making them strong candidates for host-microbiome effects.
- Unlike SNPs, CNVs can affect larger genomic regions, and haplotype-resolved CNVs may capture additional allelic complexity beyond total copy number alone.
- Here, we tested whether total CNV dosage and CNV haplotypes associate with chronic wound bacterial species.

Methodology

- **Genotyping:** Buccal DNA was genotyped at ~600k markers on the Illumina Global Screening Array.
- **Microbiome profiling:** Species-level relative abundance data were obtained from clinical 16S reports generated by MicroGenDX.
- **CNV calling and haplotype assignment:** Fig. 1.
- **Association testing and replication:** Linear models tested total copy number and haplotype-resolved effects in two stratified cohorts with bidirectional replication, adjusting for diabetes, age, sex, and genetic ancestry.
- **Functional enrichment:** Genes overlapping associated CNVs were tested for Reactome enrichment.
- **Multi-omic additive effects:** Multiple regression assessed the heritability for selected species.

Results

- 846 CNVs across 482 patients were tested for association with 68 bacterial species.

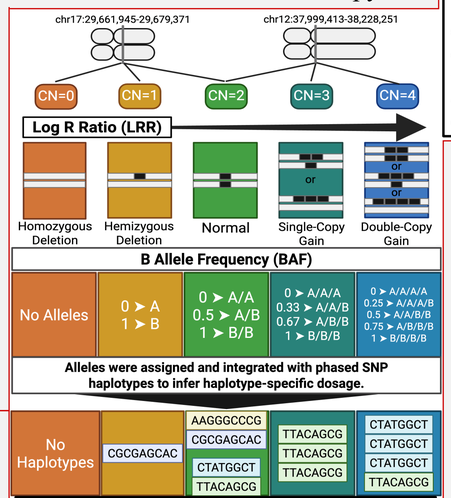

Figure 1. Haplotype assignment workflow. *Copy number states inferred from PennCNV were integrated with phased SNP haplotypes using Eagle to resolve haplotype-specific CNV dosage.*

- CNVs were most abundant and dense on chromosomes 19 and 22, and were predominantly short, low-incidence events that were most often hemizygous deletions (Fig. 2A, 2B, 2C, 2D).
- Shorter CNVs tended to occur at higher incidence, CNV regions with 3 to 4 haplotypes were most common, and longer CNVs exhibited greater haplotypic complexity overall (Fig. 2E, 2F, 2G).

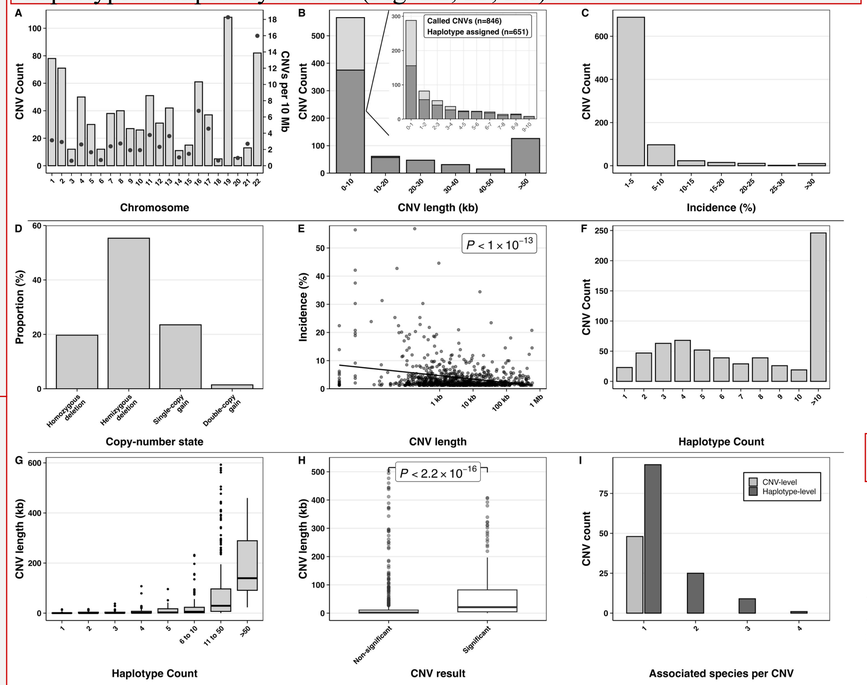

Figure 2. Overview of CNV genomic features, haplotypic complexity, and association patterns. *(A) CNV distribution across chromosomes, adjusted for chromosome size. (B) Length distribution for called CNVs and haplotype-assigned CNVs. (C) CNV incidence across samples. (D) Copy-number state distribution. (E) CNV length and incidence,* assessed by linear regression. *(F) Haplotype count distribution. (G) CNV length by haplotype count. (H) CNV length by significance status, assessed by Wilcoxon rank-sum test. (I) Number of species associations at the CNV and haplotype levels.*

- 169 CNVs were associated with 54 species (Fig. 3).
- Significant CNVs occurred in 1.3 to 34% of patients and were frequently located in telomeric or centromeric regions (Table 1).
- Associated CNVs were significantly larger than non-associated CNVs (median 21.1 kb vs 3.0 kb, respectively) (Fig. 2H).
- All 48 CNV-level associations were species-specific and involved 24 species, whereas haplotype-level associations were less species-specific, and linked 128 CNVs to 50 species (Fig. 2I).
- ~86% of CNVs overlapped with genes, and were enriched for metabolism, drug ADME, and disease, with additional species-level enrichment in signal transduction, developmental biology, autophagy, extracellular matrix organization, and immune function.

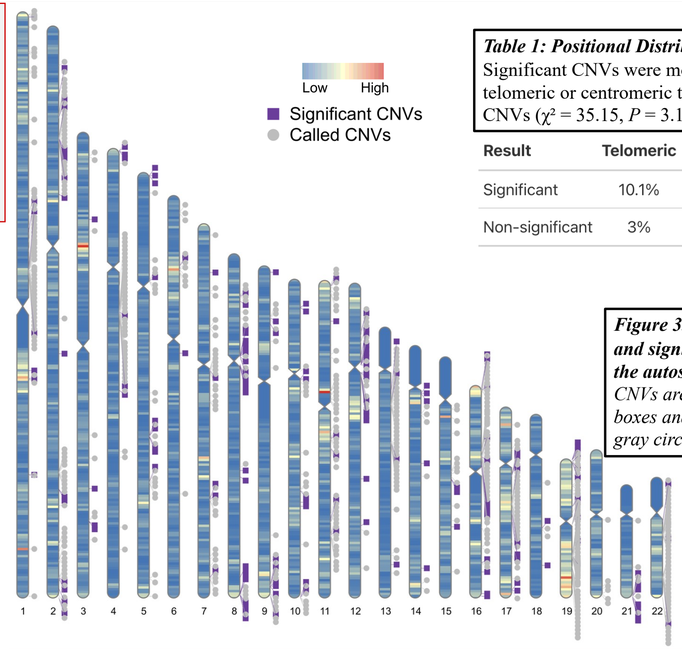

Table 1: Positional Distribution of CNVs. Significant CNVs were more frequently telomeric or centromeric than non-significant CNVs ($\chi^2 = 35.15$, $P = 3.15 \times 10^{-9}$).

Result	Telomeric	Centromeric	Internal
Significant	10.1%	17.2%	72.8%
Non-significant	3%	6.4%	90.7%

Figure 3. Ideogram of called and significant CNVs across the autosomes. *Significant CNVs are shown in purple boxes and all called CNVs in gray circles.*

- Multi-omic models explained up to 59% of species abundance (Fig. 4).

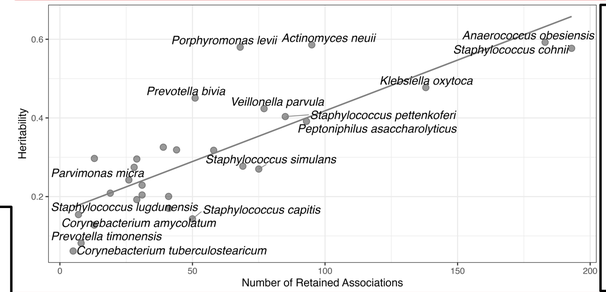

Figure 4. Relationship between adjusted R^2 and the number of significantly associated loci per species. *Points represent individual species. The number of associated loci was positively associated with model R^2 ($F = 49.97$, $df = 1, 27$, $P = 1.34 \times 10^{-7}$; $R^2 = 0.63$). Twenty-nine species were included. Some points are unlabeled due to overlap.*

Conclusions

- ~19% of CNVs were associated with chronic wound bacterial species in two cohorts, indicating frequent influence of host structural variation in shaping wound microbiomes.
- Associated CNVs were significantly larger and more often located in telomeric or centromeric regions, indicating that CNV size and location facilitate influence on wound-associated bacteria.
- Haplotype-resolved analysis suggests allelic composition at CNVs is important, emphasizing a genotype-dosage interaction in determining the human genome-wound microbiome relationship.
- Integrating CNVs with transcriptomic models improved explanation of microbial abundance, reinforcing structural variation as a meaningful contributor to inter-patient microbial differences in chronic wounds.