Automated Analysis of Whole Genomes to Interpret Complex and Uncommon MNS Alleles

Justin B.L. Halls*[1,2], Sunitha Vege*[3], Aline Floch[3,4,5], Helen H. Mah[1], Matthew S. Lebo[1,2,6,7], Prathik K. Vijay Kumar[7], Connie M. Westhoff[3], William J. Lane[1,2]

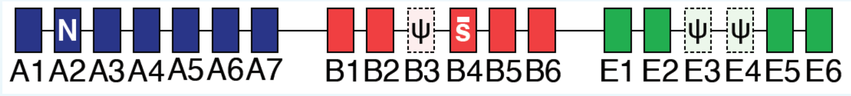
[1]Department of Pathology, Brigham and Women's Hospital, Boston, MA, [2]Harvard Medical School, Boston, MA, [3]New York Blood Center, New York, NY, [4]Univ Paris Est Creteil, INSERM, IMRB, Creteil, France, [5]Etablissement français du sang Ile-de-France, IMRB, Creteil, France, [6]Laboratory for Molecular Medicine, Boston, MA, [7]Partners Personalized Medicine, Boston, MA.

Purpose

Automated analysis of WGS data is an emerging technology with the capacity to enhance our understanding of blood group systems. Here we demonstrate the utility of automated analysis in identifying complex, rare, and novel alleles within the MNS system as well as delineating breakpoints of known alleles.

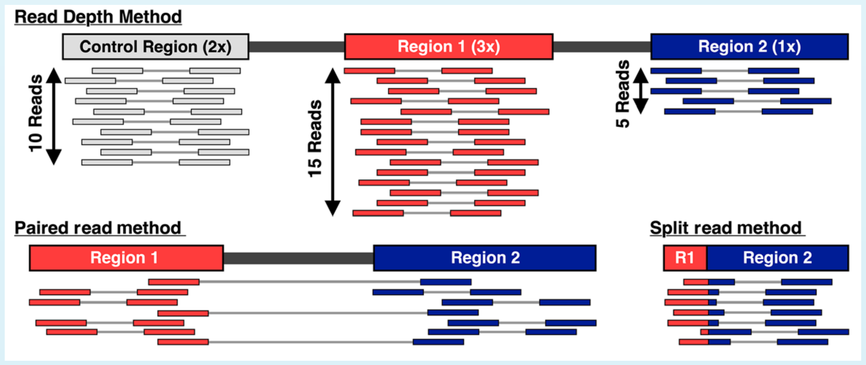
Introduction

The MNS blood group system is defined by three homologous genes: *GYPA*, *GYPB*, and *GYPE*. Many MNS alleles contain complex structural variations (SV) such as partial gene deletions and multi-step gene recombinations that form hybrid genes and represent a challenge for the development of WGS genotyping algorithms.



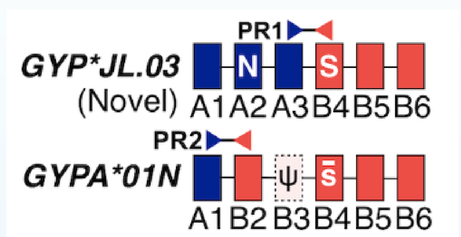
Methods

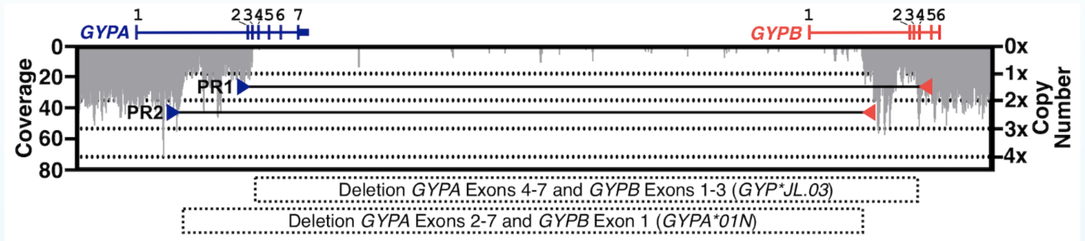
WGS was performed on 9 previously characterized MNS samples with diverse allele types including U+^{var} and different GYP hybrids including *GYP(A-B)*, *GYP(B-A)*, *GYP(B-A-B)*, and *GYP(B-E-B)*. The automated software, bloodTyper, identified single nucleotide variations (SNV) and utilized read depth, paired read, and split read methods to determine SV. Manual analysis was also performed to find interpretive gaps and update the bloodTyper software. This updated version was then used to call the MNS alleles in all 3,202 high coverage whole genomes from the 1000 Genomes Project (1000G).



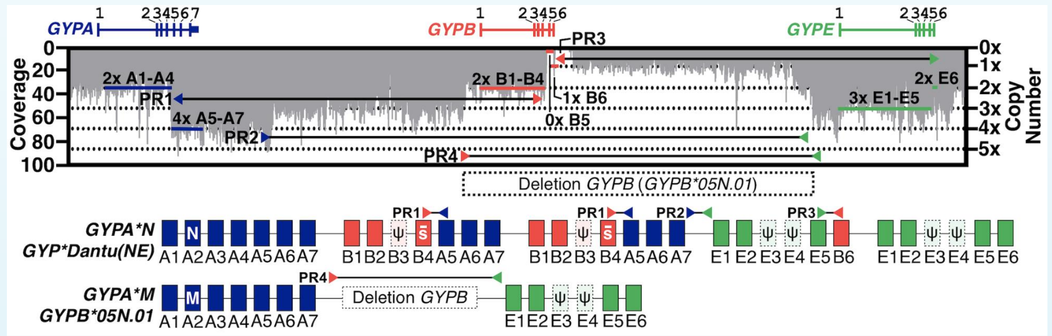
Results

Automated analysis of the 9 WGS samples allowed for description of unique breakpoints and characterization of three novel alleles: *GYP*Hil.02*, **JL.02*, **JL.03*, and confirmation of the recently described *GYP*Bun.02*. The *GYP*JL.03* sample was identified to be compound heterozygous with *GYPA*01N*, allowing for the first description of the exact breakpoint for this allele.



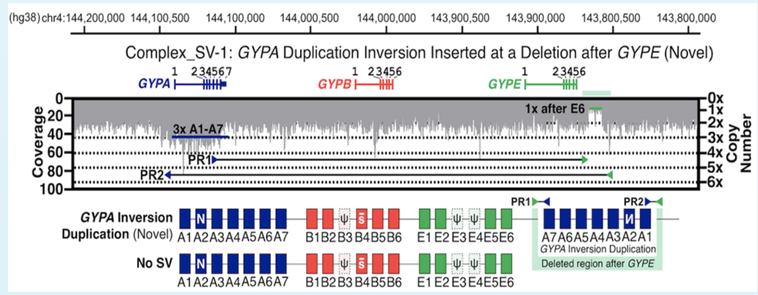


WGS analysis of *GYP*Dantu(NE)* found a duplication of *GYPE* that circumscribed a region of *GYPB* Exon 6 allowing clarification of the breakpoints. Analysis of the 1000G samples found *GYP*Hil*, **Sch* (with three different breakpoints), **Dantu(NE)* and several potentially novel alleles including two B-A hybrids, one E-A hybrid, and four complex SV, likely representing several recombination events.



Several uncommon and rare SNV alleles were found in 1000G including heterozygous *GYPB*He*, *GYPA*Vr*, *GYPA*Mta*, *GYPB*Mv*, *GYPB*sD*, *GYPB*Mit*, and 1 homozygous *GYP*Nob* [GYP(A-B-A)]. Of note, *GYPB*sD* has only one report in a Caucasian South African family but was found here in five East Asians.

Six samples demonstrated four different complex SVs (SV-1 to 4) likely due to multiple genetic events. SV-1 was found in three Europeans with 3x *GYPA* (duplication and inversion) and 1x in an intergenic region after *GYPE*. SV-2 was found in one African with 3x *GYPB* and 4x *GYPE*, which appears to be one partial duplication and inversion of *GYPB* and two full duplications of *GYPE* one with inversion. SV-3 was found in one South Asian in a duplication with 3x *GYPA* and 3x in the intergenic region after *GYPB*. SV-4 was found in one African with a duplication with 3x *GYPE* and 3x in intergenic region after *GYPA*.



Conclusions

This work enhances automated interpretation of blood group alleles and further develops genomic analysis of WGS data by determining exact breakpoints of known alleles, identifying rare and novel alleles, and identifying complex alleles.

No Acknowledgments or Disclosures