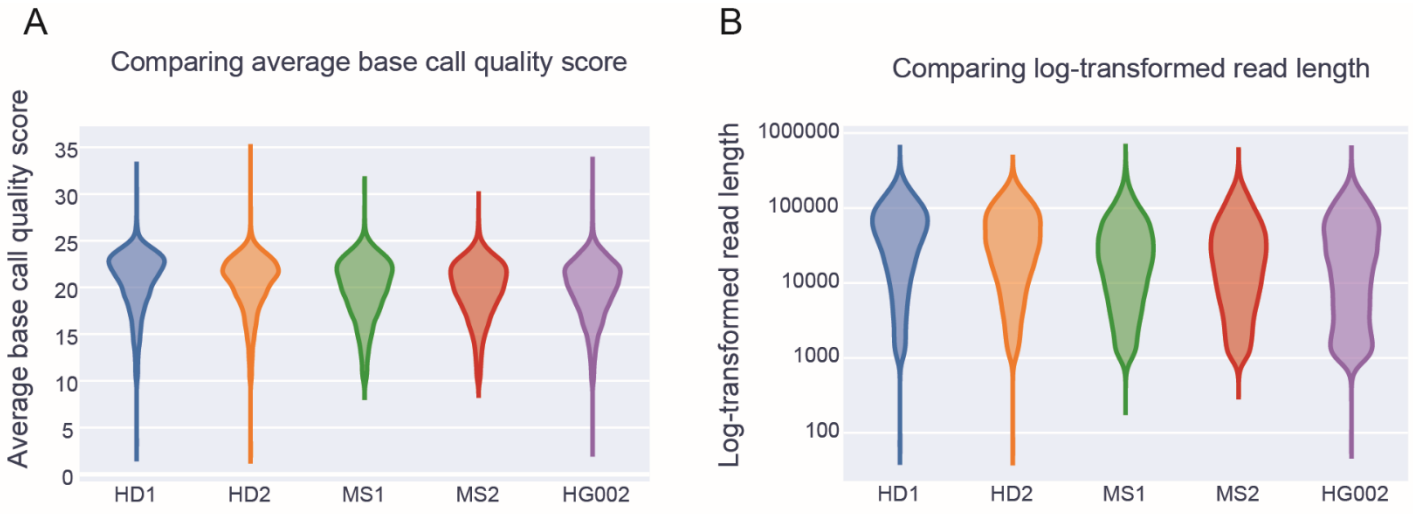
Supplemental Materials for "Ultra-long sequencing for contiguous haplotype resolution of the human immunoglobulin heavy chain locus" by Gornitzka et. al.

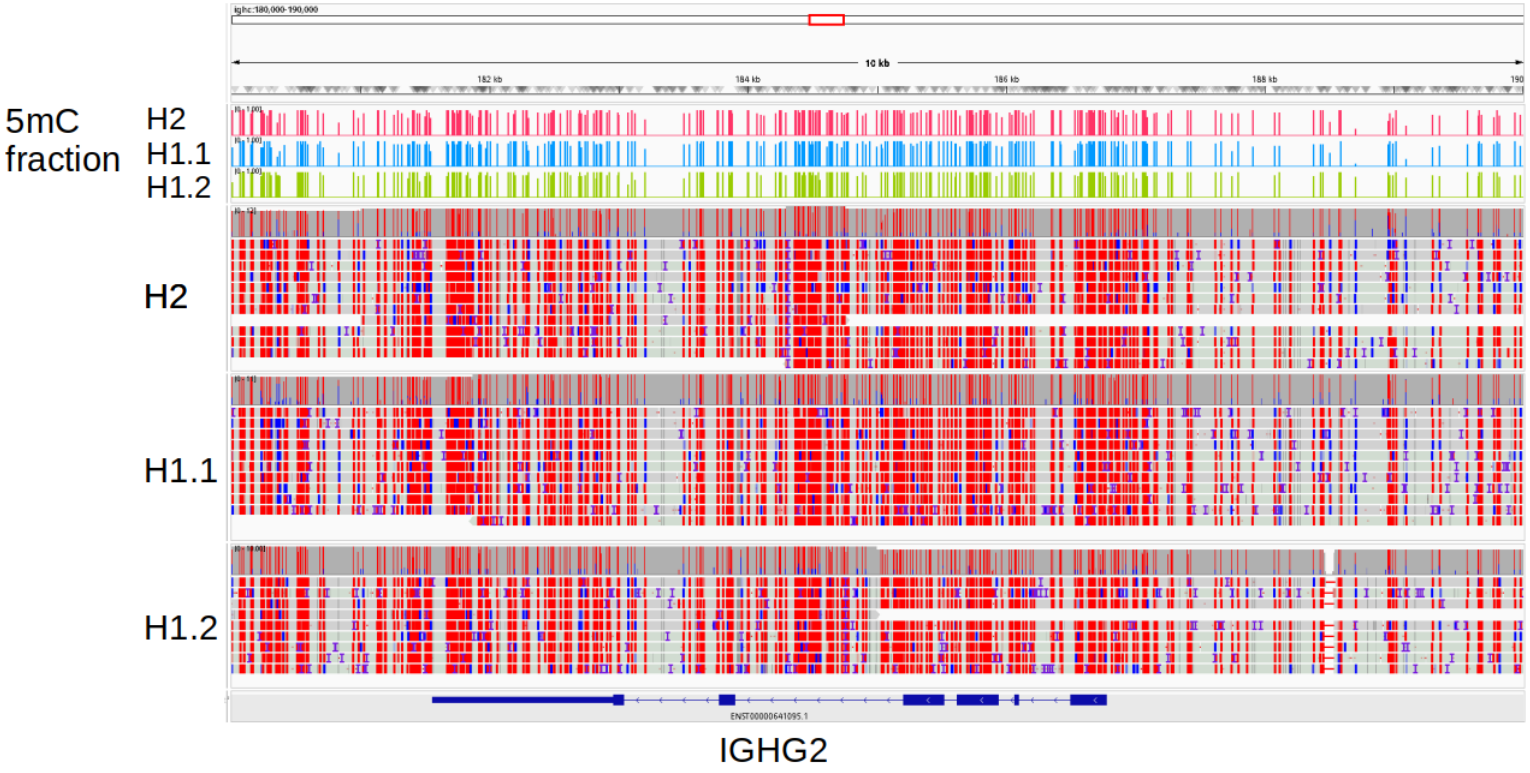
The Supplemental Materials include:

- Supplemental Figures S1-5
- Supplemental Text
- Supplemental Table S1, Quantification of IGH genes and alleles (in a separate Excel file)
- Supplemental Table S2, Immunoglobulin serum levels in MS2 (in a separate Excel file)



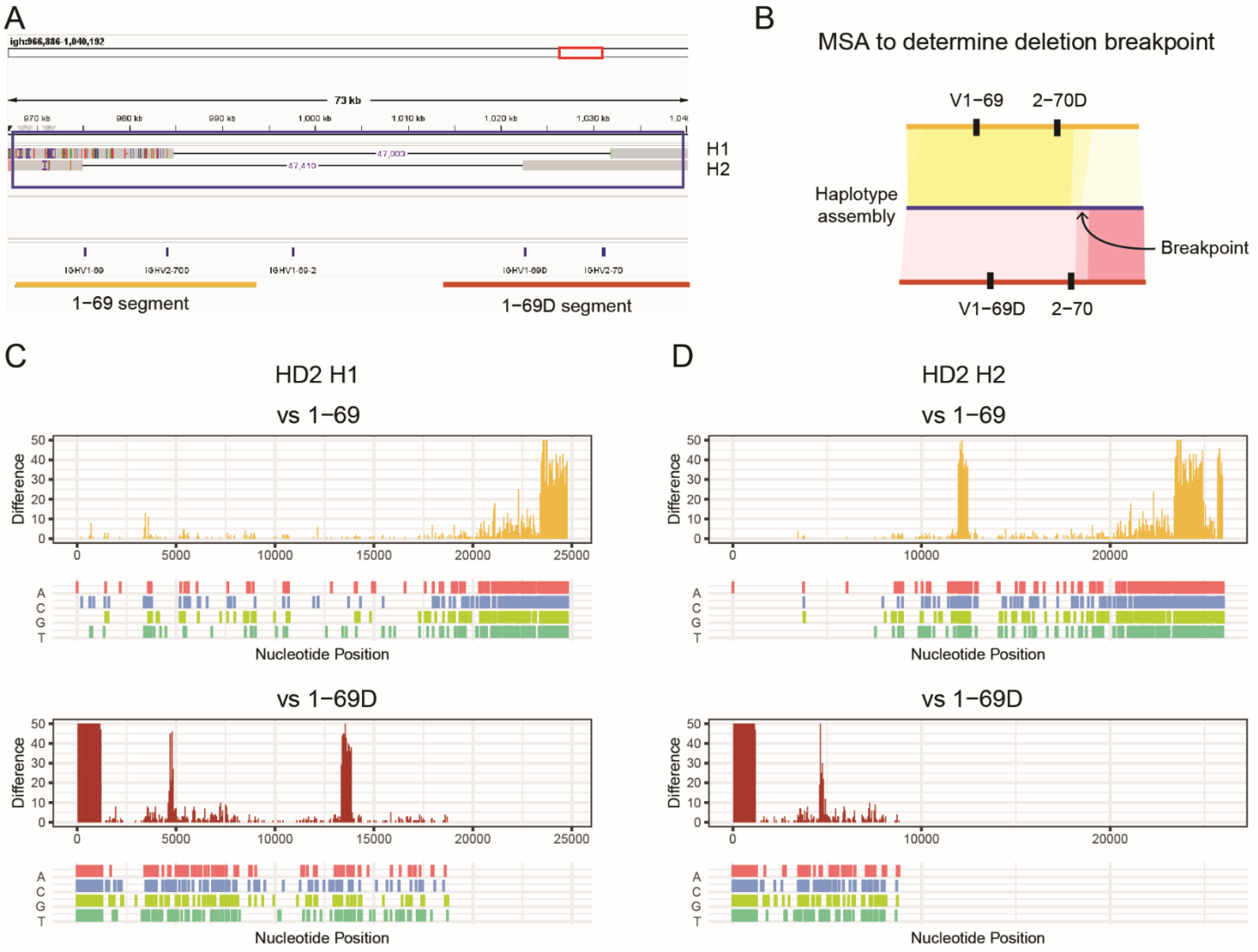
Supplemental Figure S1. Descriptive statistics on all rebasecalled sequencing reads, excluding those rejected by adaptive sampling.

A, Violin plot of base call quality scores from all donors and HG002. B, log-transformed violin plot of read lengths from all donors and HG002.



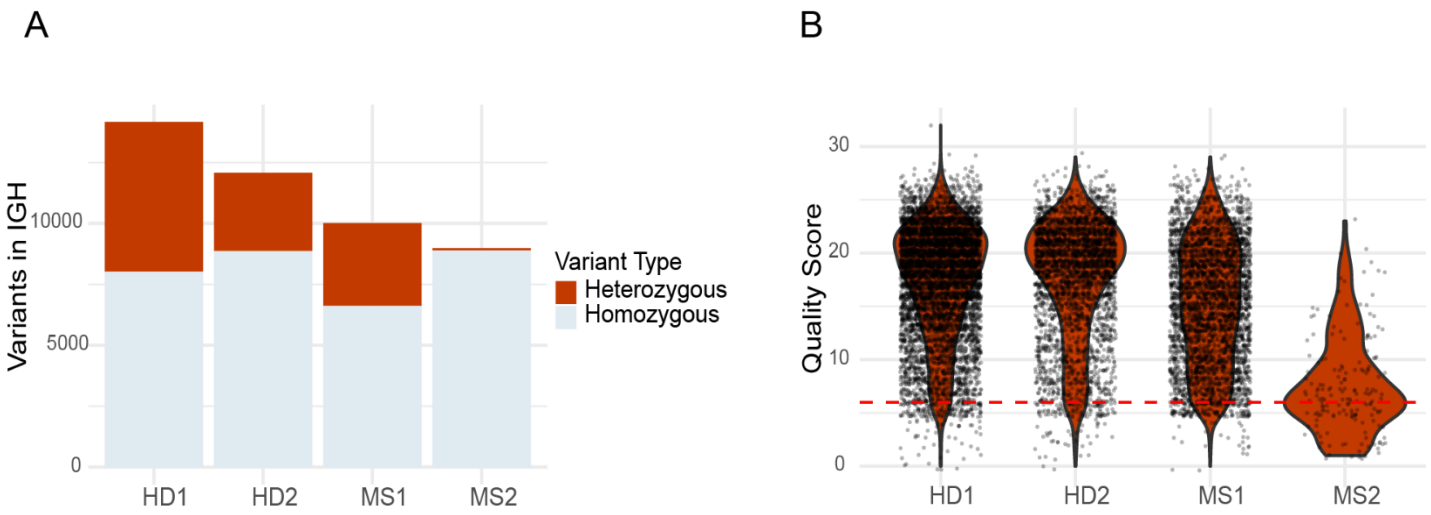
Supplemental Figure S2. IGV visualization of CpG methylation across the duplicated IGHG2 region in donor HD1.

Aligned nanopore reads are split by the three copies of the gene and methylation analysed separately. Top tracks are bedgraphs for 5mC methylation generated by modkit pileup, one per gene copy. Non-duplicated haplotype (H2, pink), centromeric copy on (H1.1, blue) and telomeric copy on H1 (H1.2, green). The read tracks are coloured by CpG methylation (red = methylated, blue = unmethylated).



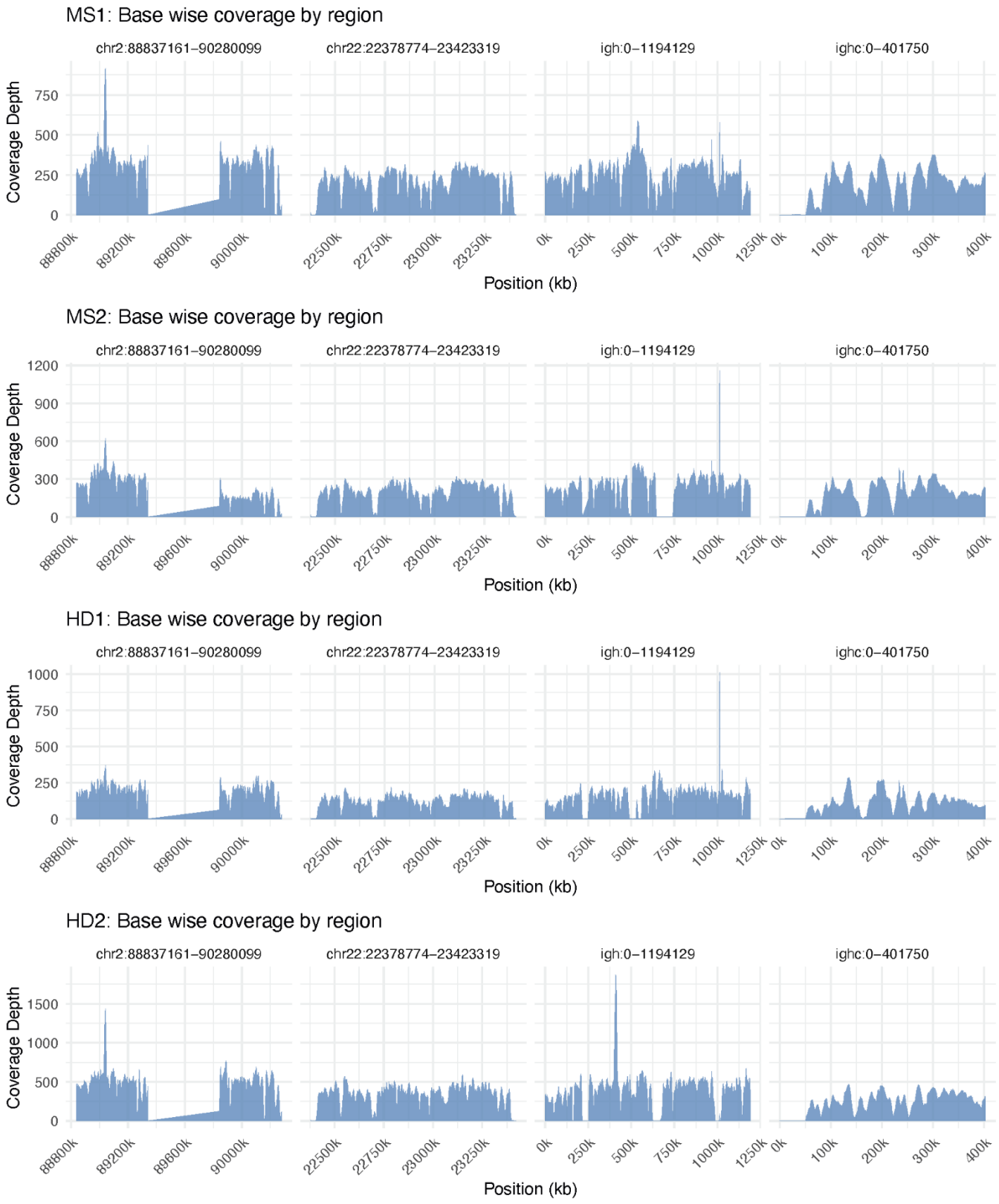
Supplemental Figure S3. Breakpoint analysis of deletion haplotypes of IGHV1-69/2-70 in HD2.

A, IGV window showing Minimap2 alignments of HD2 haplotype assemblies over the IGHV1-69/2-70 region. Blue box shows the segment of the haplotype 1 (H1) and 2 (H2) extracted for multiple sequence alignment (MSA) and breakpoint analysis. Coordinates of segments from IGH reference used for breakpoint analysis highlighted with yellow and red bars in bottom track. B, Schematic illustrating rationale of breakpoint analysis, higher opacity represents higher identity. Breakpoint is determined to be where the highest identity of HD2 switches from one V1-69/V2-70 segment to the other. C and D, Nucleotide differences between each V1-69/V2-70 segment and HD2 H1 (C) and H2 (D), shown both as accumulated differences and split by base.



Supplemental Figure S4. Analysis of heterozygous variants in IGH, and lack thereof in MS2.

A, Bar plot showing variants called between the donor's ONT reads and the custom IGH reference by Clair3 during phasing. Variants within the IGH locus that met a Phred quality (Q) score threshold of 6 are stratified by homozygous or heterozygous status. B, Distribution of variant quality scores for all heterozygous variants in IGH, as determined by Clair3. The stippled line represents the Q=6 threshold required for phasing.



Supplemental Figure S5. HiFi read depth across IGK, IGL and IGH.

Coverage analysis over different immune loci IGK (chr2), IGL (chr22) and IGH variable (igh) and IGH constant (ighc) region for the four donors MS1, MS2, HD1 and HD2.

Supplemental Text

To address the question of gene usage frequency of constant genes in HD1 with the IGHC duplication (main figure 5C), we investigated methylation patterns in the ONT reads. The POD5 files were basecalled with the 5mC model and quantified CpG methylation at:

(i) the duplicated IGHG2 segment in donor HD1, and

(ii) the entire IGH locus in all four donors.

All three IGHG2 copies, and the full IGH locus in every donor, are hypermethylated (≥ 70%). Specifically, 5mC levels were 82-86% in the constant regions and 74-79% in the variable regions. Such high methylation is expected in monocytes and bulk PBMC preparations like those used in our study, which are dominated by non-B cells whose IGH locus is transcriptionally silent. For illustration, we prepared an IGV snapshot with per-read 5mC tags, split by the three IGHG2 copies in HD1, showing no allele-specific differences (Supplemental Figure S2). Finally, we note that the constant region expression is primarily governed by class-switch recombination rather than DNA methylation. How DNA methylation may influence the switching-process itself remains unclear and lies beyond the scope of our dataset.