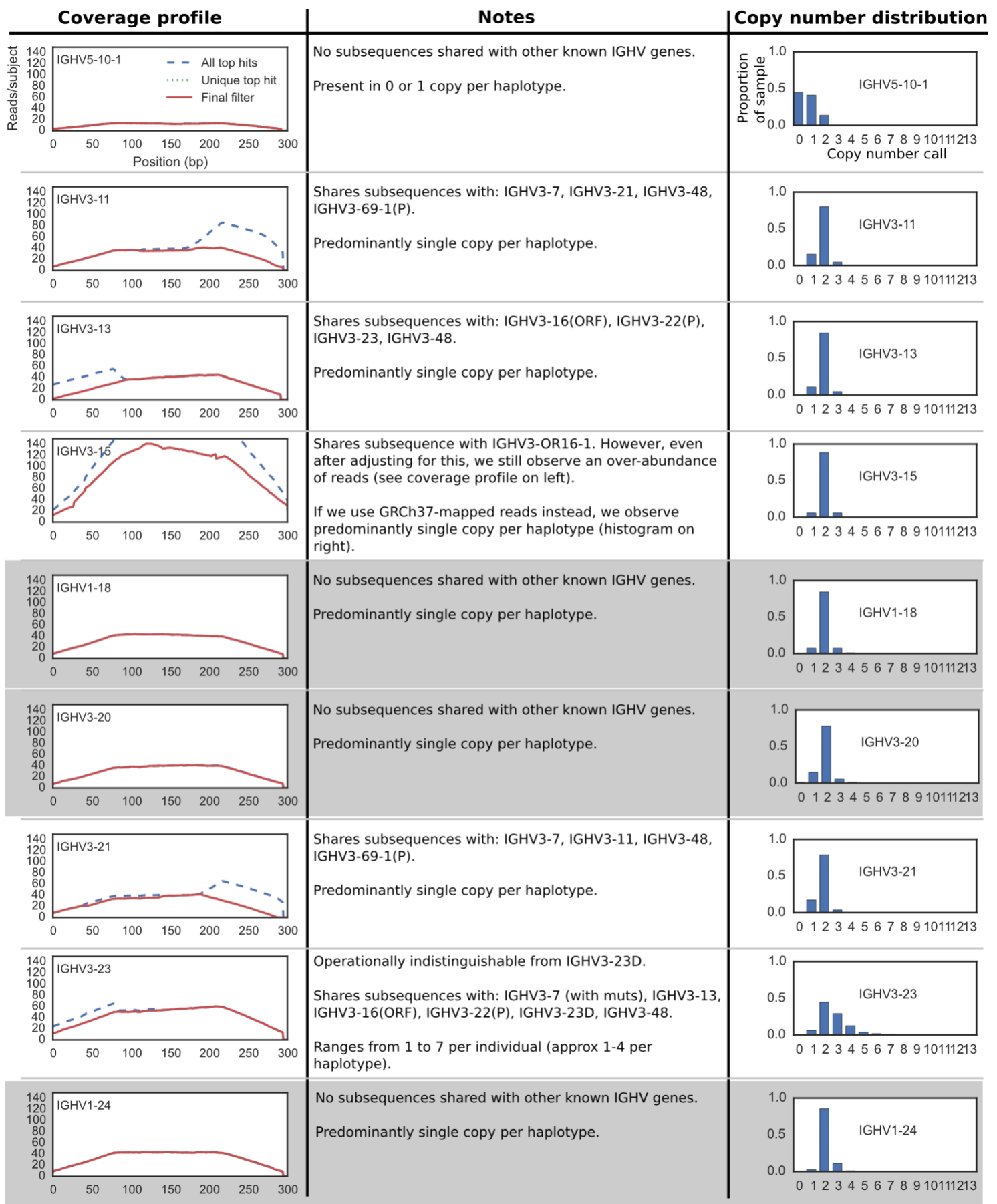


**Supplementary Information Figure 1 (part 1 of 5).** Detailed characteristics of each operationally distinguishable IGHV gene segment. Each row corresponds to a segment. The first column shows the per-base read coverage of the segment, averaged over 109 subjects (from blood/saliva samples). Blue dashed line is the coverage if we count all reads for which the segment is a top hit (an over estimate of actual coverage). Green dotted line is the coverage if we only count reads for which the segment is the unique top hit (an underestimate of coverage). Red solid line is the coverage after we apply the appropriate filter, customized for that segment (details in the Supplementary Text, Read filtering). A segment which does not share subsequences with any other segment should display a trapezoidal shape. The second column notes any operationally indistinguishable segments, whether there are any segments for which subsequences are shared (this is defined to be at least 75 bp of identical subsequence), and what the typical copy number of the segment is. If there are segments with which subsequences with a few mutations are shared, these are listed with “with muts” in parentheses. The third column displays the distribution of copy number for that segment. The two-copy segments that are used for allele and single nucleotide calls are shaded in grey.



Supplementary Information Figure 1 (part 2 of 5)

Coverage profile	Notes	Copy number distribution
	<p>No subsequences shared with other known IGHV genes.</p> <p>Predominantly single copy per haplotype.</p>	
	<p>Shares subsequence with IGHV4-4, IGHV4-55(P), IGHV4-59.</p> <p>Predominantly single copy per haplotype.</p>	
	<p>Operationally indistinguishable from IGHV3-30-3, IGHV3-30-5, IGHV3-33.</p> <p>Shares subsequence with IGHV3-66, IGHV3-NL1.</p> <p>Ranges from 1-13 per individual (approx 1-7 per haplotype).</p>	
	<p>Shares subsequence with IGHV4-4, IGHV4-30-4, IGHV4-38-2, IGHV4-39, IGHV4-59, IGHV4-61.</p> <p>High similarity with other functional genes makes calling copy number problematic.</p>	
	<p>Shares subsequence with IGHV4-30-2, IGHV4-31, IGHV4-59, IGHV4-61.</p> <p>High similarity with other functional genes makes calling copy number problematic.</p>	
	<p>Shares subsequence with: IGHV4-30-4, IGHV4-34, IGHV4-59, IGHV4-61.</p> <p>High similarity with other functional genes makes calling copy number problematic.</p>	
	<p>Shares subsequence with IGHV4-31, IGHV4-39, IGHV4-55(P), IGHV4-59, IGHV4-61.</p> <p>Predominantly single copy per haplotype.</p>	
	<p>Shares subsequence with IGHV4-4, IGHV4-30-2, IGHV4-39, IGHV4-55(P) (with muts), IGHV4-59, IGHV4-61.</p> <p>Present in 0 or 1 copy per haplotype.</p>	
	<p>Shares subsequence with IGHV4-4, IGHV4-30-2, IGHV4-34, IGHV4-38-2, IGHV4-59, IGHV4-61.</p> <p>Predominantly single copy per haplotype.</p>	

Supplementary Information Figure 1 (part 3 of 5)

Coverage profile	Notes	Copy number distribution
	<p>Operationally indistinguishable from IGHV3-43D.</p> <p>No subsequences shared with other known IGHV genes.</p> <p>Ranges from 2-6 per individual (approx 1-3 per haplotype).</p>	
	<p>No subsequences shared with other known IGHV genes.</p> <p>Predominantly single copy per haplotype.</p>	
	<p>Shares subsequences with: IGHV1-3 (with muts), IGHV1-OR15-2 (with muts)</p> <p>Predominantly single copy per haplotype.</p>	
	<p>Shares subsequences with: IGHV3-7, IGHV3-11, IGHV3-13, IGHV3-16, IGHV3-21, IGHV3-22(P), IGHV3-23, IGHV3-69-1(P)</p> <p>Predominantly single copy per haplotype.</p>	
	<p>No subsequences shared with other known IGHV genes, but inflated read coverage in 175-275 bp region.</p> <p>Predominantly single copy per haplotype (after adjusting for inflated coverage).</p>	
	<p>No subsequences shared with other known IGHV genes.</p> <p>Predominantly single copy per haplotype.</p>	
	<p>Operationally indistinguishable from IGHV3-66.</p> <p>Shares subsequences with: IGHV3-7, IGHV3-30, IGHV3-64, IGHV3-71.</p> <p>Though the histogram on the right suggests 3-4 copies per individual, we see only one main cluster in scaled read coverage data, so we believe it is actually predominantly 2 copies per haplotype.</p>	
	<p>No subsequences shared with other known IGHV genes.</p> <p>Predominantly single copy per haplotype.</p>	
	<p>Shares subsequence with: IGHV4-4, IGHV4-28, IGHV4-30-2, IGHV4-31, IGHV4-34, IGHV4-38-2, IGHV4-39, IGHV4-61.</p> <p>High similarity with other functional genes makes calling copy number problematic.</p>	

Supplementary Information Figure 1 (part 4 of 5)

Coverage profile	Notes	Copy number distribution
	<p>Shares subsequence with: IGHV4-4, IGHV4-30-2, IGHV4-30-4, IGHV4-31, IGHV4-34, IGHV4-38-2, IGHV4-39, IGHV4-59, IGHV4-OR15-8.</p> <p>High similarity with other functional genes makes calling copy number problematic.</p>	
	<p>Operationally indistinguishable from IGHV3-64D.</p> <p>Shares subsequences with: IGHV3-7, IGHV3-62(P), IGHV3-66, IGHV3-71, IGHV3-OR16-10 (with muts)</p> <p>Ranges from 2-4 per individual (approx 1-2 per haplotype).</p>	
	<p>Operationally indistinguishable from IGHV1-69D.</p> <p>No subsequences shared with other known IGHV genes.</p> <p>Ranges from 2-4 per individual (approx 1-2 per haplotype).</p>	
	<p>No subsequences shared with other known IGHV genes, but inflated read coverage in 75-175 bp region.</p> <p>Present in 0 or 1 copy per haplotype (after adjusting for inflated coverage).</p>	
	<p>Operationally indistinguishable from IGHV2-70D.</p> <p>Shares subsequences with: IGHV2-5.</p> <p>Ranges from 2-4 per individual (approx 1-2 per haplotype).</p>	
	<p>No subsequences shared with other known IGHV genes.</p> <p>Predominantly single copy per haplotype.</p>	
	<p>No subsequences shared with other known IGHV genes.</p> <p>Predominantly single copy per haplotype.</p>	
	<p>Shares subsequence with: IGHV3-OR16-13.</p> <p>Predominantly single copy per haplotype.</p>	
	<p>Shares subsequences with: IGHV3-30, IGHV3-30-3.</p> <p>Final filter in coverage profile (left, red solid line) is for the sole individual for which IGHV3-NL1 is present (as one copy on one haplotype).</p>	

Supplementary Information Figure 1 (part 5 of 5)