Figure1. The minimal sequence difference of each human BCR germline κ light chain gene pair at full sequence length. The colors indicate the number of nucleotide differences between genes (the vertical bar on right gives the scale). The comparison begins at the 3' end of each full-length gene. The numbers in blue circles represent the numbers of mismatched nucleotides in the most similar pairs. These pairs of VH genes cannot be discriminated from each other at almost any length or mutation frequency.

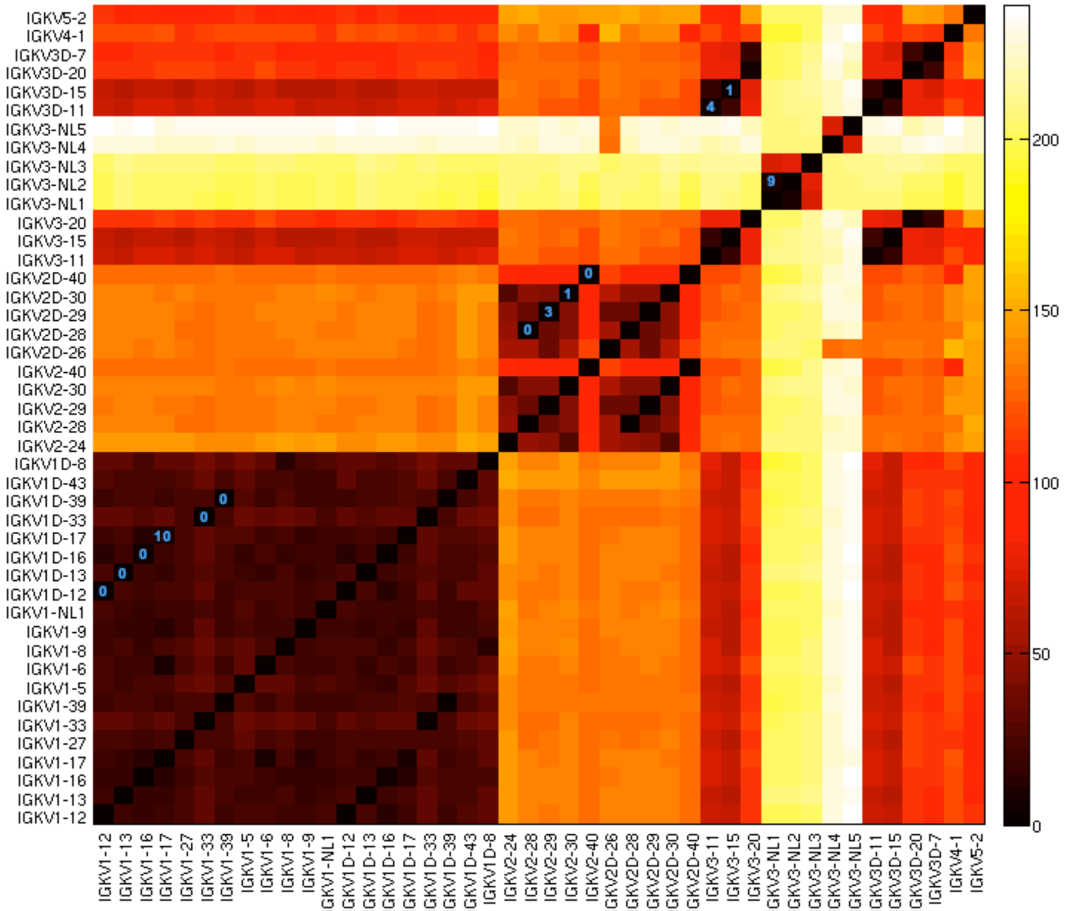


Figure 2. The minimal sequence difference of each human BCR germline λ light chain gene pair at full sequence length. The colors indicate the number of nucleotide differences between genes (the vertical bar on right gives the scale). The comparison begins at the 3' end of each full-length gene. The numbers in blue circles represent the numbers of mismatched nucleotides in the most similar pairs. These pairs of VH genes cannot be discriminated from each other at almost any length or mutation frequency.

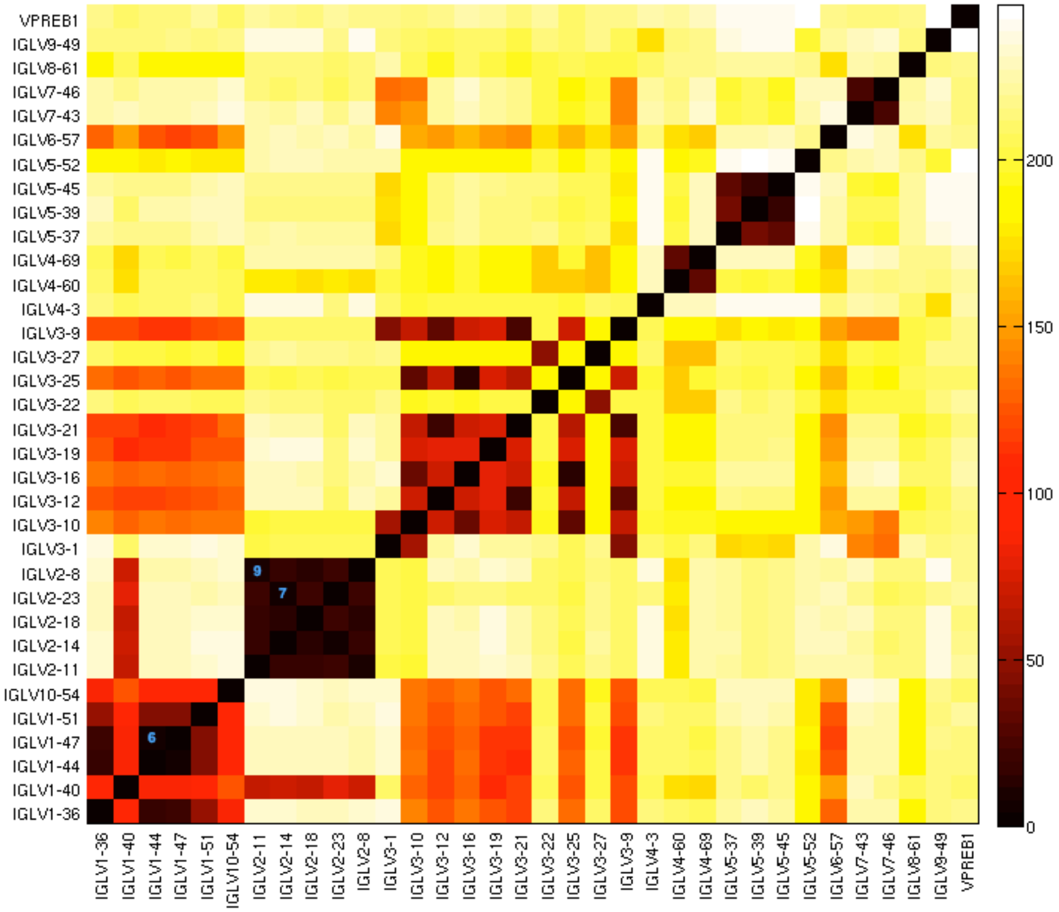


Figure 3. The minimal sequence difference of each human TCR germline α chain gene pair at full sequence length. The colors indicate the number of nucleotide differences between genes (the vertical bar on right gives the scale). The comparison begins at the 3' end of each full-length gene. The numbers in blue circles represent the numbers of mismatched nucleotides in the most similar pairs. These pairs of VH genes cannot be discriminated from each other at almost any length or mutation frequency.

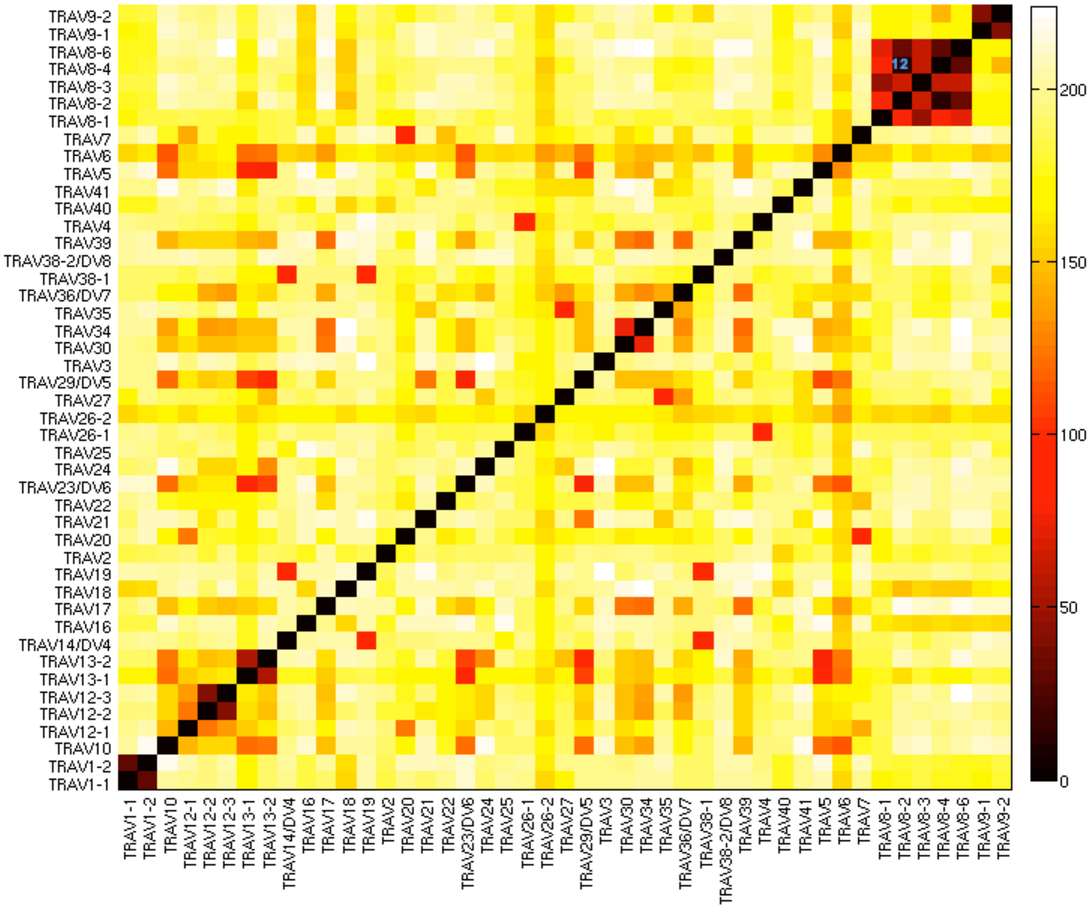


Figure 4. The minimal sequence difference of each human TCR germline β chain gene pair at full sequence length. The colors indicate the number of nucleotide differences between genes (the vertical bar on right gives the scale). The comparison begins at the 3' end of each full-length gene. The numbers in blue circles represent the numbers of mismatched nucleotides in the most similar pairs. These pairs of VH genes cannot be discriminated from each other at almost any length or mutation frequency.

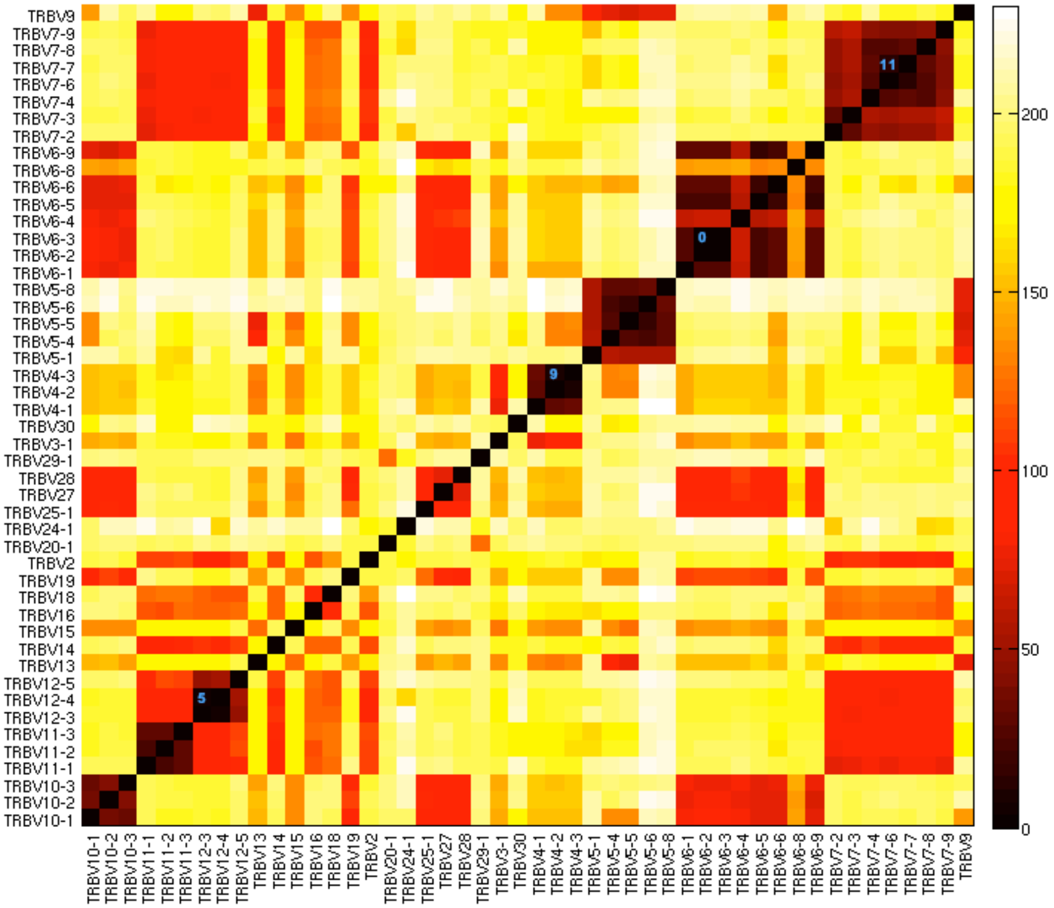


Figure 5. Comparison of human VH identification results using three different sequence identification methods at n=100 nucleotide read length and 0.05 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

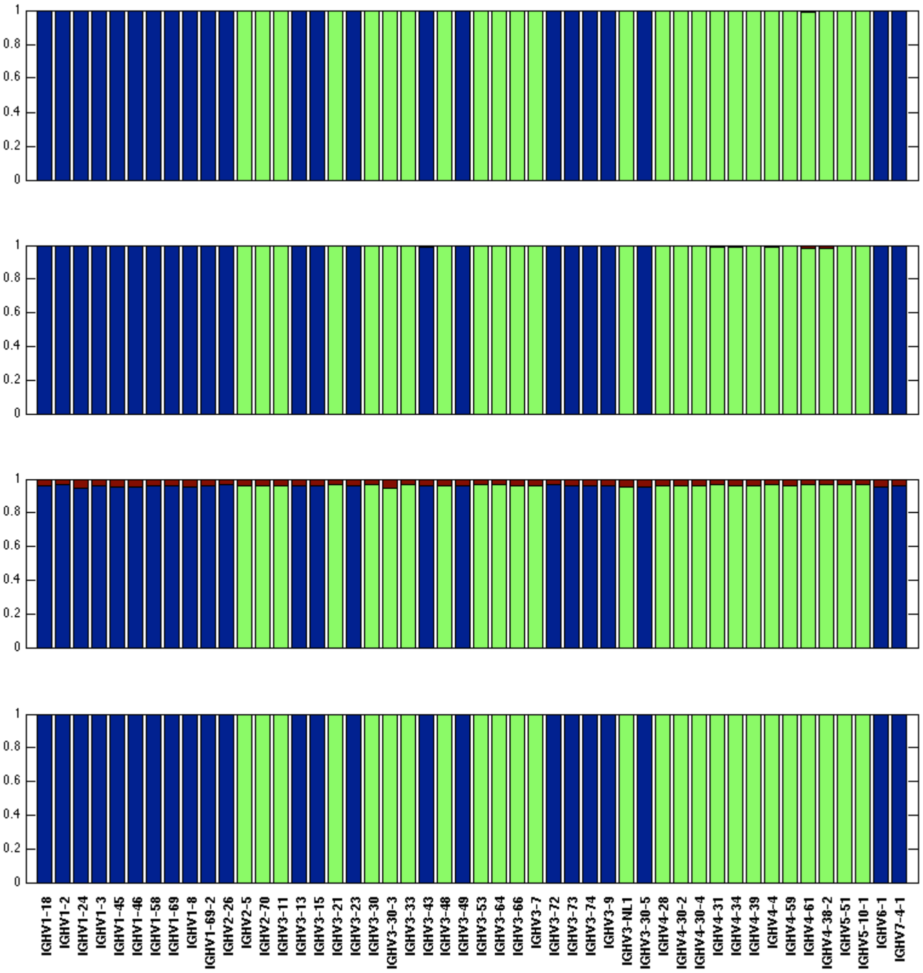


Figure 6. Comparison of human VH identification results using three different sequence identification methods at n=100 nucleotide read length and 0.15 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

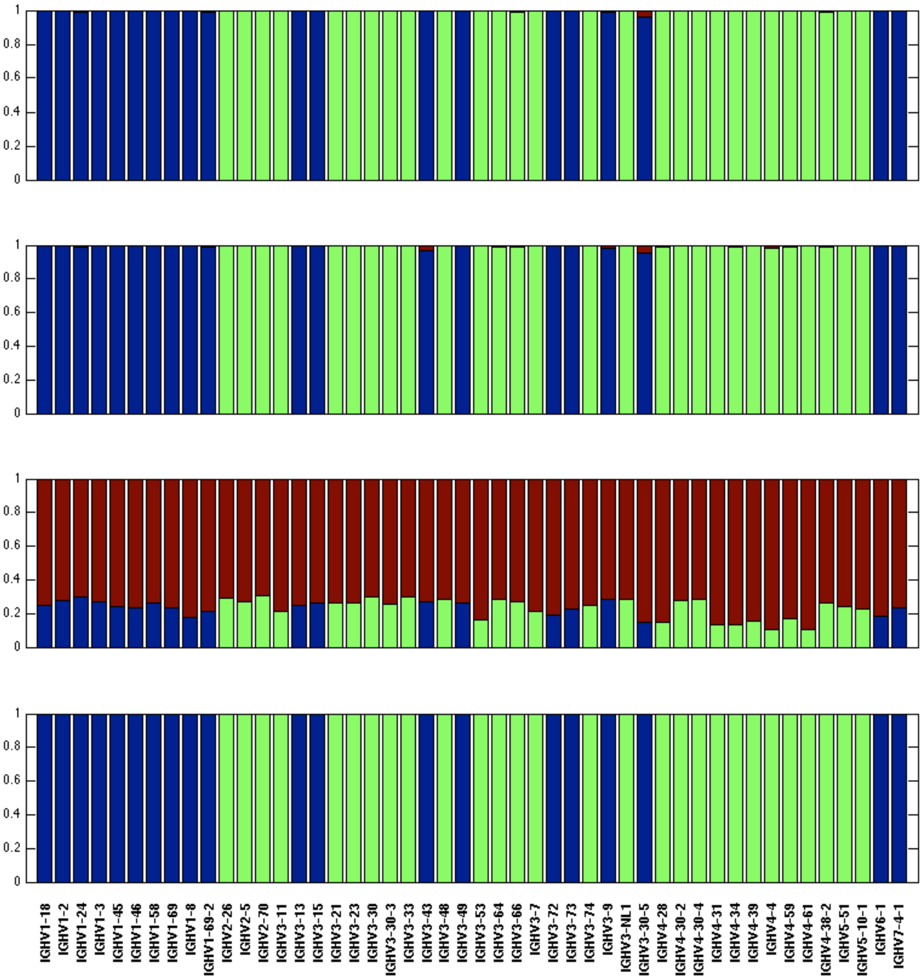


Figure 7. Comparison of human VH identification results using three different sequence identification methods at n=100 nucleotide read length and 0.30 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

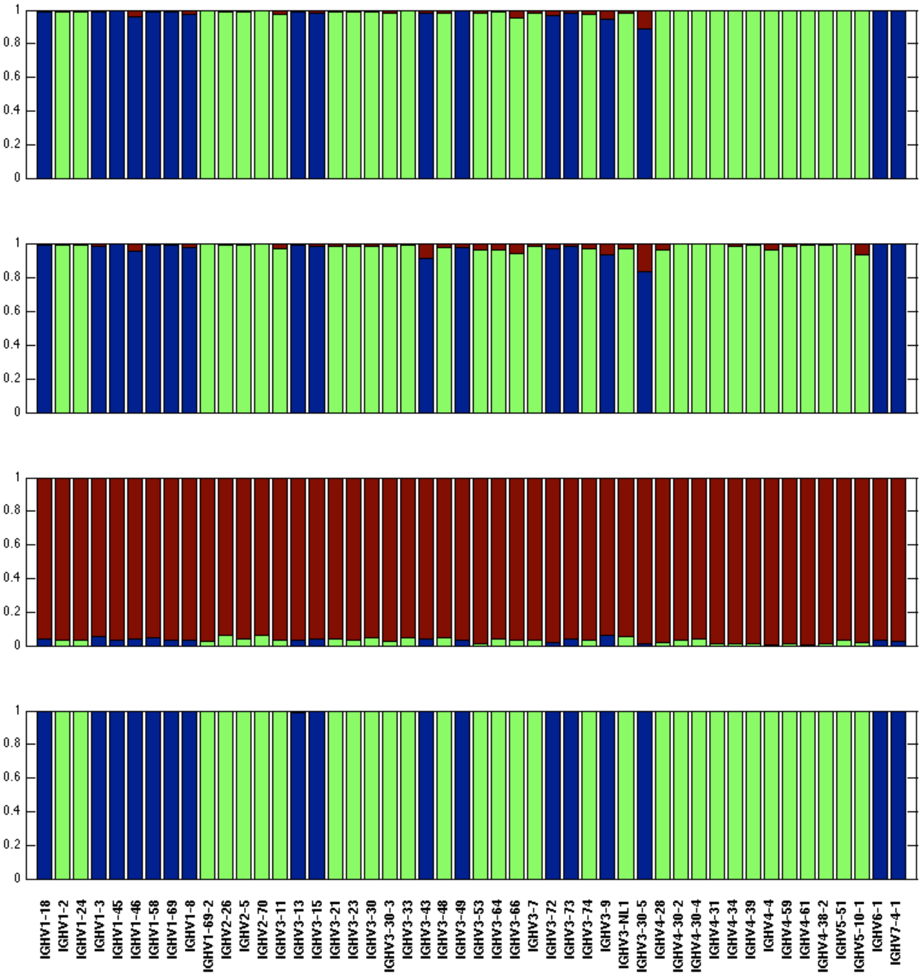


Figure 8. Comparison of human VH identification results using three different sequence identification methods at n=150 nucleotide read length and 0.05 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

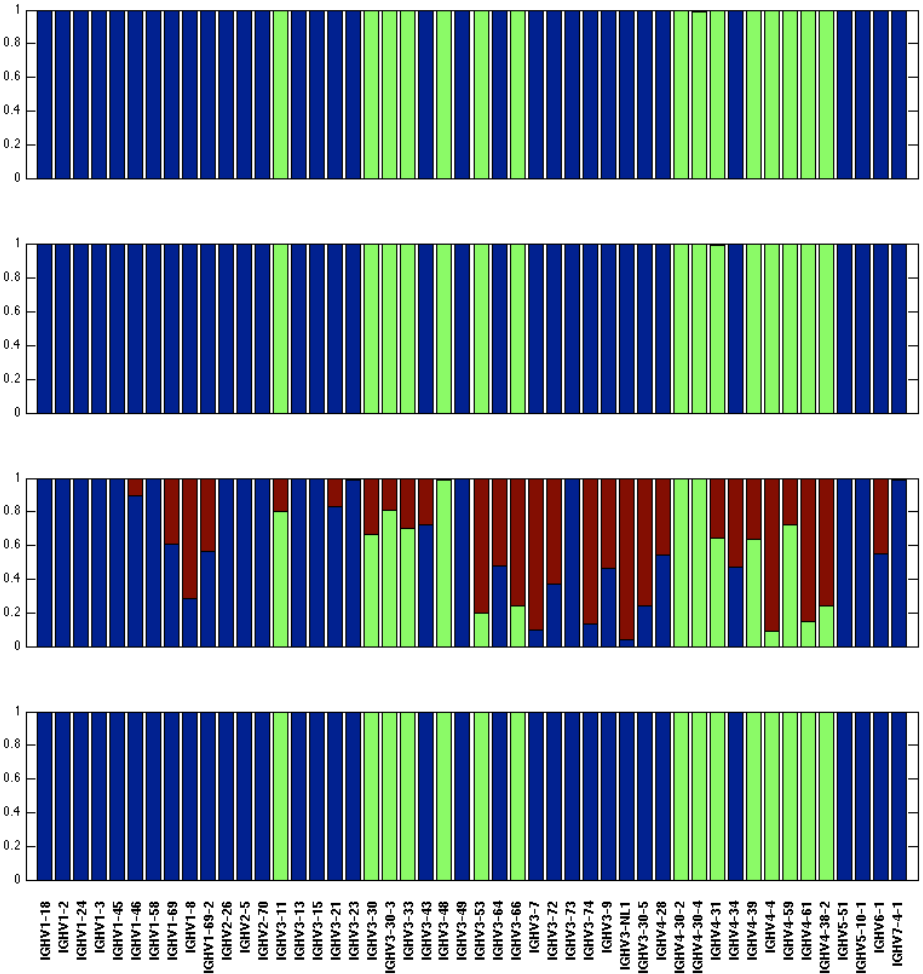


Figure 9. Comparison of human VH identification results using three different sequence identification methods at n=150 nucleotide read length and 0.30 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

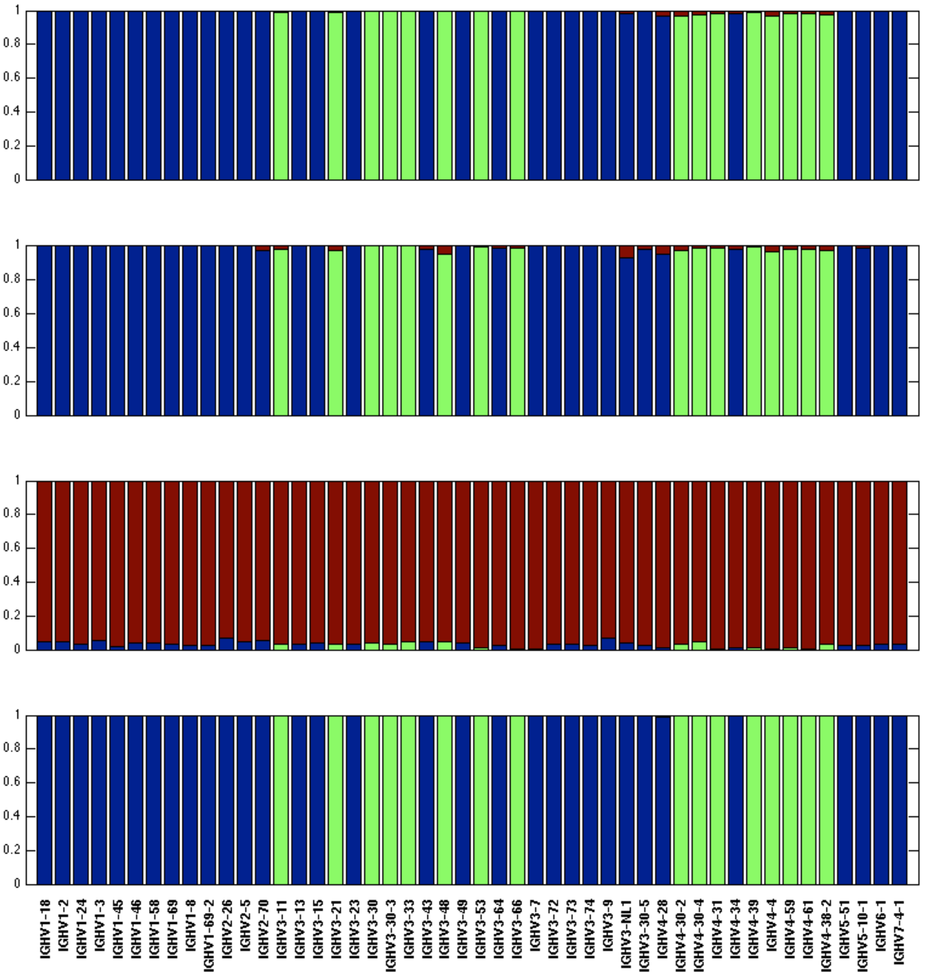


Figure 10. Comparison of human VH identification results using three different sequence identification methods at n=200 nucleotide read length and 0.05 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

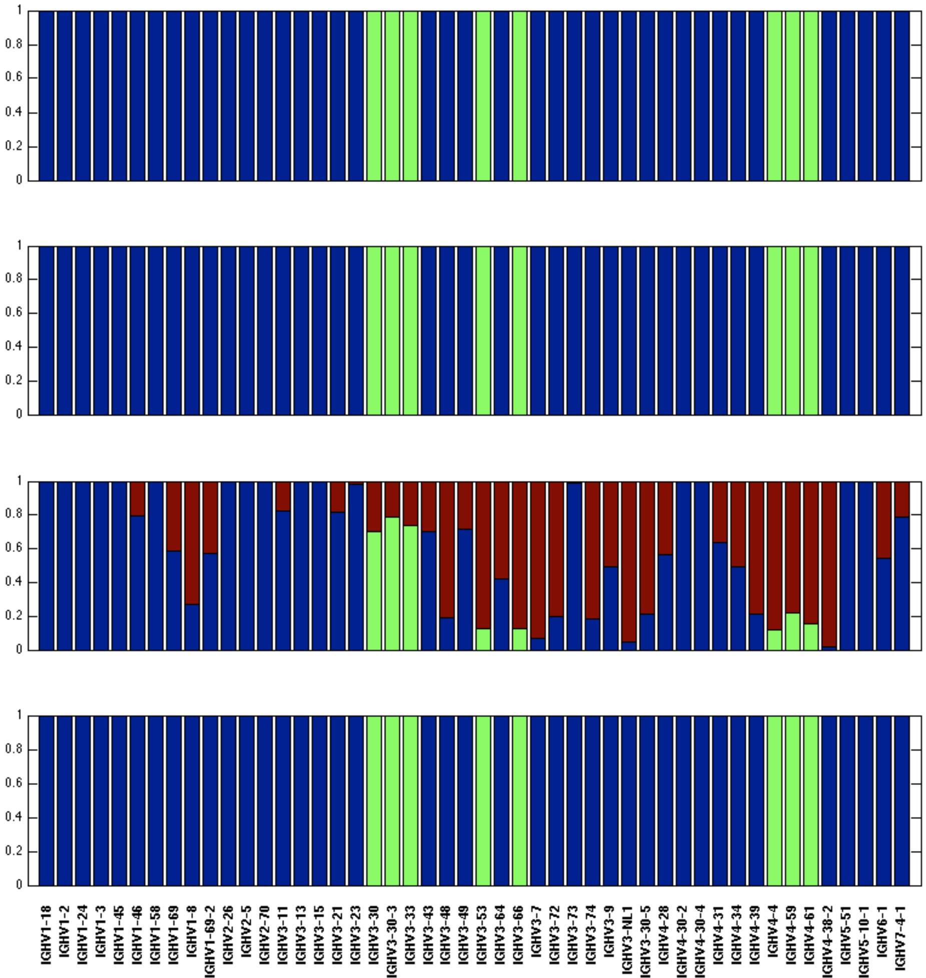


Figure 11. Comparison of human VH identification results using three different sequence identification methods at n=200 nucleotide read length and 0.15 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

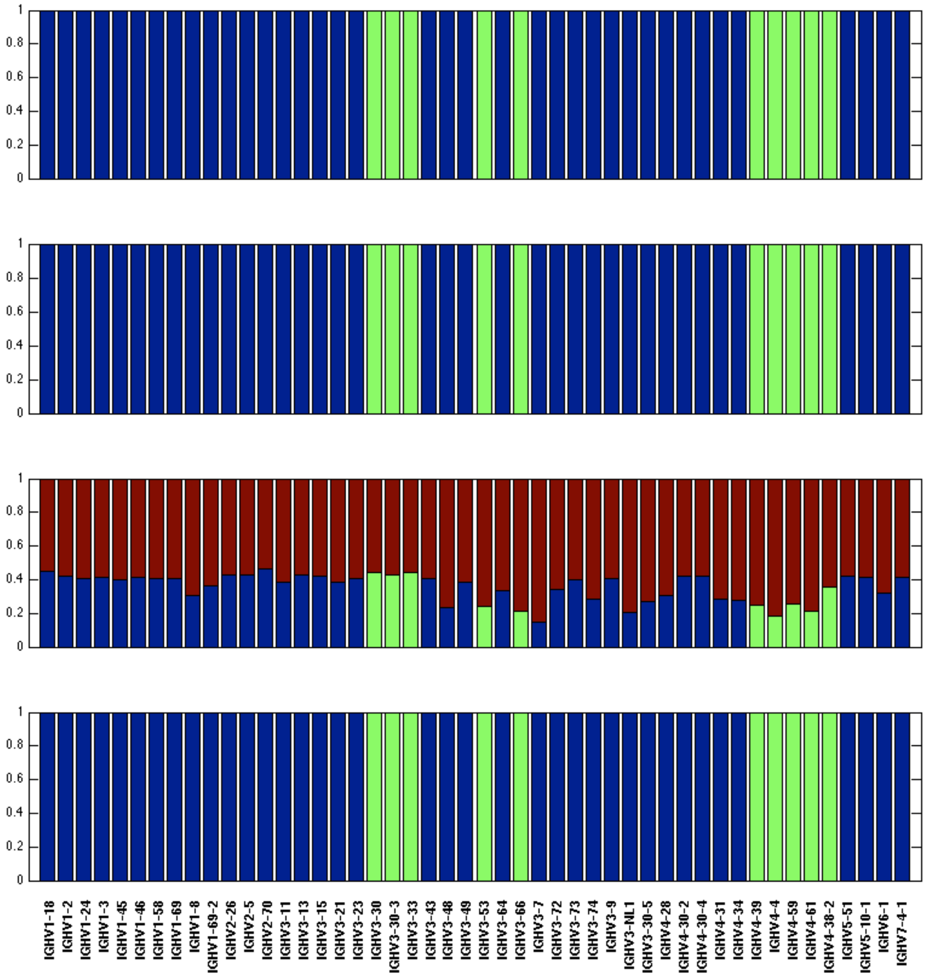


Figure 12. Comparison of human VH identification results using three different sequence identification methods at n=200 nucleotide read length and 0.30 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

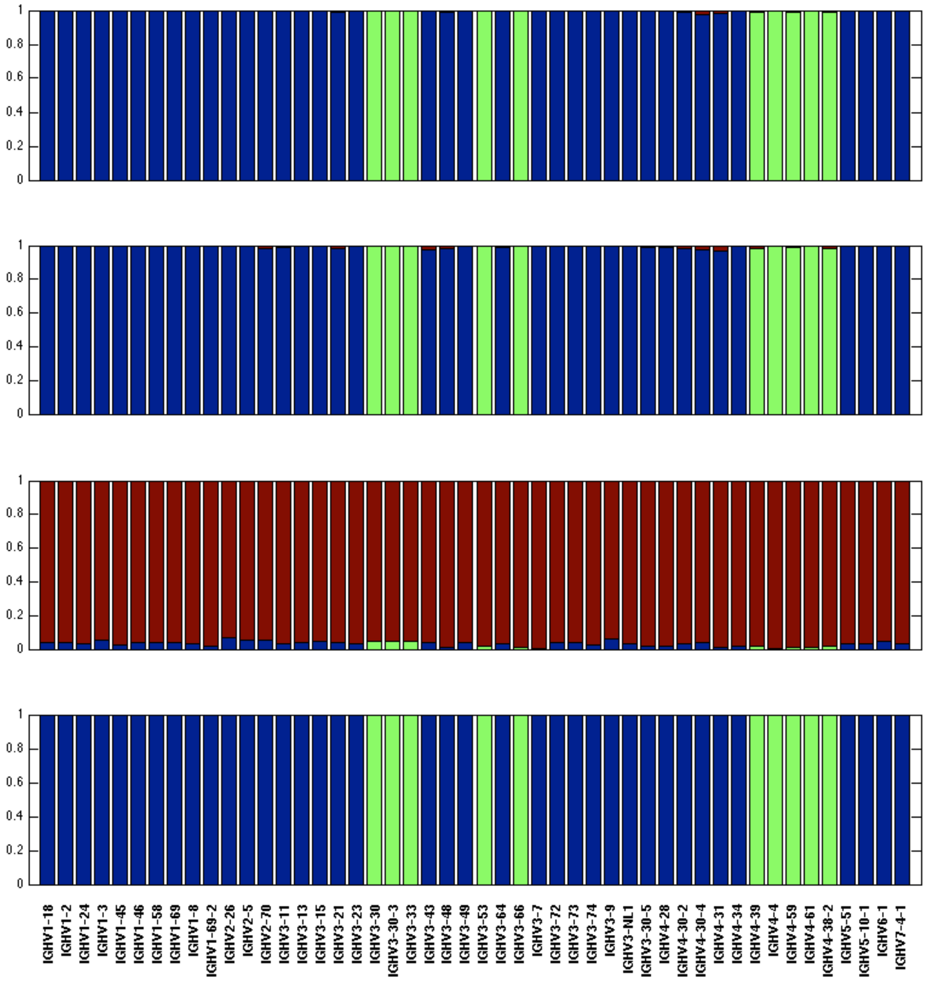


Figure 13. Comparison of human VH identification results using three different sequence identification methods at full read length and 0.05 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

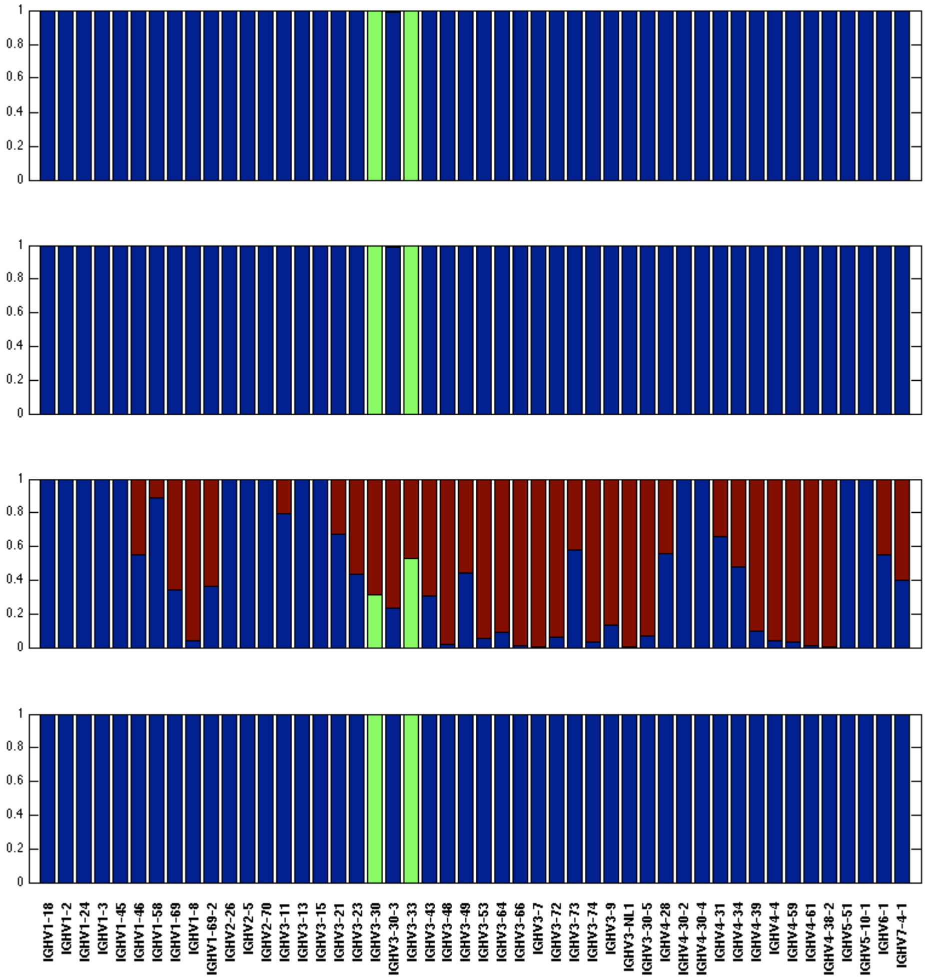


Figure 14. Comparison of human VH identification results using three different sequence identification methods at full read length and 0.15 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.

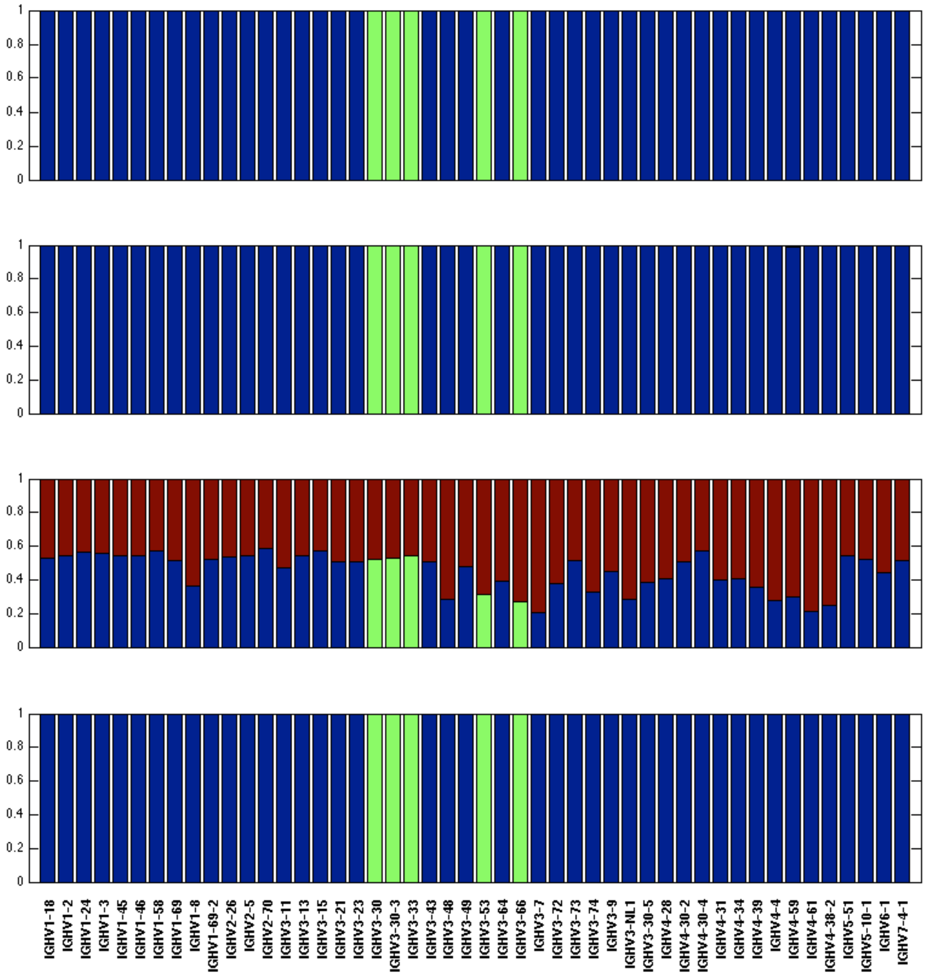
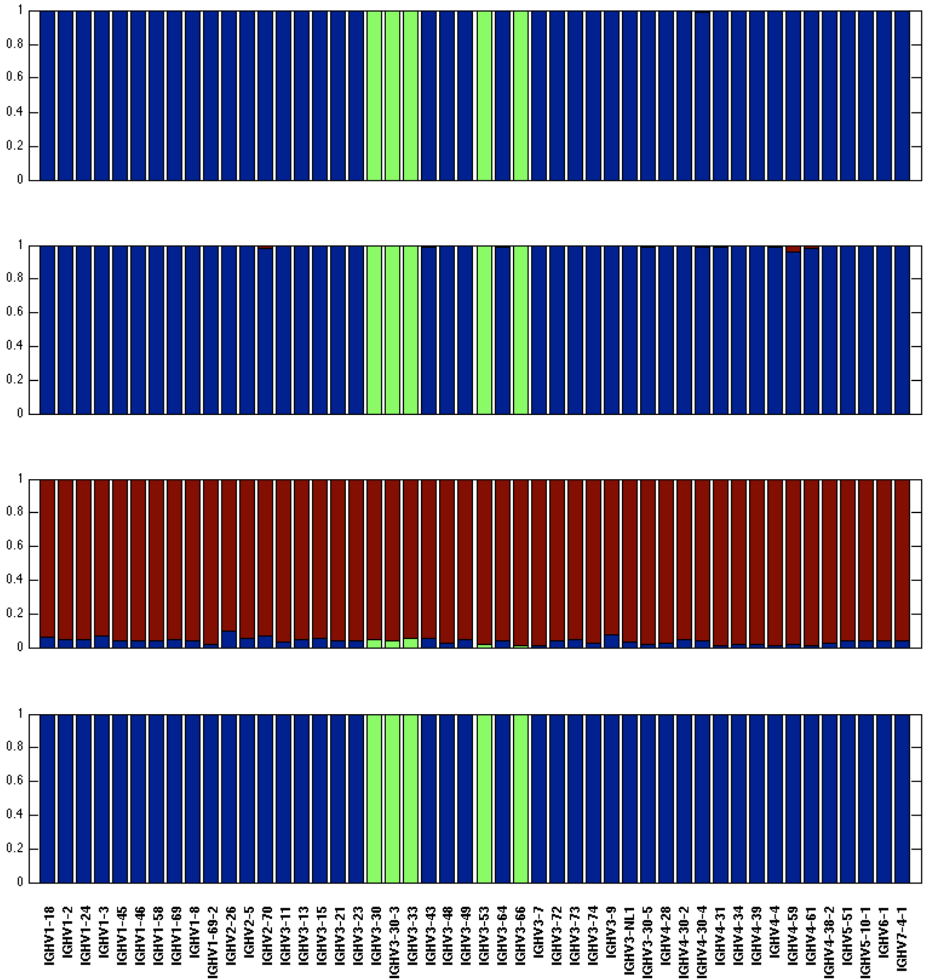


Figure 15. Comparison of human VH identification results using three different sequence identification methods at full read length and 0.30 mutation frequency. From top to bottom (A) The Anchor method (this paper); (B) High V-Quest; (C) IgBLAST using the default word length (9 characters); (D) IgBLAST using the minimal word length (4 characters). In all cases the blue color represents the fraction of correctly identified sequences in which a VH gene is uniquely identifiable; green is the fraction of identifications when a gene is not distinguishable from at least one other gene (using the calculation described in the Calcualtion and Methods section) and either the real gene or the one we predict to confuse it are identified; red is the fraction of incorrect identifications.


IGHV1-18
IGHV1-2
IGHV1-24
IGHV1-3
IGHV1-45
IGHV1-46
IGHV1-58
IGHV1-69
IGHV1-8
IGHV1-69-2
IGHV2-26
IGHV2-5
IGHV2-70
IGHV3-11
IGHV3-13
IGHV3-15
IGHV3-21
IGHV3-23
IGHV3-30
IGHV3-30-3
IGHV3-33
IGHV3-43
IGHV3-48
IGHV3-49
IGHV3-53
IGHV3-64
IGHV3-66
IGHV3-7
IGHV3-72
IGHV3-73
IGHV3-74
IGHV3-9
IGHV3-NL1
IGHV3-30-5
IGHV4-28
IGHV4-30-2
IGHV4-30-4
IGHV4-31
IGHV4-34
IGHV4-39
IGHV4-4
IGHV4-59
IGHV4-61
IGHV4-38-2
IGHV5-51
IGHV5-10-1
IGHV6-1
IGHV7-4-1