Cell Genomics

CellPress
OPEN ACCESS

Preview

Uncovering diversity in the immunoglobulin heavy chain locus

Jill A. Hollenbach[1,2,*]
[1]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA
[2]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA
*Correspondence: jill.hollenbach@ucsf.edu
https://doi.org/10.1016/j.xgen.2025.101134

The immunoglobulin heavy chain constant (IGHC) locus houses genetic determinates of antibody function and specificity. In this issue of *Cell Genomics*, Jana et al. use long-read sequencing to characterize extensive inter-individual diversity in the IGHC region across ancestrally diverse populations, highlighting potential functional consequences.

The adaptive immune response is characterized by the ability of effector cells to respond to an astonishing array of foreign antigens. Somatic recombination is fundamental to the generation of a legion of immune cells endowed with a fine degree of specificity for antigen and a wide variety of effector functions. During this process, variable genomic building blocks are rearranged in developing T or B cells to exponentially increase the potential diversity of their antigen receptors beyond what is encoded in the germline. In both T and B cells, these building blocks—the variable (V), diversity (D), and joining (J) segments—are spliced together in a process known as V(D)J recombination.[1] Significant insights into immune effector function have been gained in the past several years through examination of T cell receptor (TCR) diversity and its relationship to health and disease.[2] Like TCRs, B cell receptors (BCRs) and their related antibodies are the product of V(D)J recombination; here, these building blocks are housed within the immunoglobulin light chain and immunoglobulin heavy chain (IGH) region, alongside the IGH constant (IGHC) locus. The BCR is comprised of two identical light chains and two identical heavy chains. For the heavy chain, the product of V(D)J recombination is spliced to a specific IGHC gene. While the V, D, and J segments contribute to antigen specificity, it is the IGHC that determines the antibody class (IgG, IgM, IgD, IgA, IgE) and thus the effector function (e.g., complement activation, mucosal immunity, Fc receptor binding) of the antibodies produced by the B cell[3] (Figure 1). Upon activation, B cells can modulate these effector functions through a process known as class switching, in which further recombination events replace the constant region from the IGHC to produce a different class of antibodies, while maintaining antigen specificity.[4]

There is clear evidence of variation within the IGHC locus, which consists of nine annotated genes.[5] However, only limited studies have interrogated this region, and it is certain that variation in the IGHC is significantly greater than currently known. This understudied genomic region harbors structural variants representing gene duplications and deletions, as well as single nucleotide variation (SNV) polymorphism. While the functional consequences of IGHC polymorphism are not fully understood, variants resulting in protein changes may influence isotype distributions in antibody repertoires, serum antibody levels, Fc receptor affinity, and antibody-dependent cellular cytotoxicity (ADCC).[6] Understanding whether and how this variation is associated with immunity, health, and disease outcomes is reliant on first characterizing its full nature and extent.

In this issue of *Cell Genomics*, Jana et al..[7] describe application of long-read sequencing to fully resolve the IGHC region. Accurate representation of this variation is a critical tool underpinning large-scale studies aimed at more fully understanding how this variation is relevant for antibody structure and function and human health. Here, the authors present data that drastically increases the number of haplotype resolved assemblies for this region and bootstrap that work to develop a bioinformatic pipeline for long-read assemblies of the region to provide new data regarding both structural and SNV polymorphism across human populations.

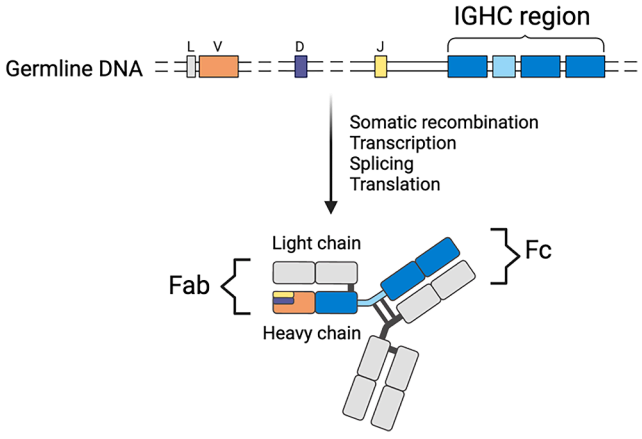

Figure 1. Variation in the immunoglobulin heavy chain constant region determines antibody specificity and effector functions
The B cell receptor is composed of identical pairs of light and heavy chains. The Fab region of the antibody is responsible for antigen specificity, while the Fc region determines effector function.

Importantly, the tools and resources developed in this work will enable other researchers to more accurately examine both germline and somatic variation in the context of health and population studies.

The cornerstone of the work presented here is the full resolution of 16 total haplotypes from eight donors, providing a set of validated reference haplotypes from highly accurate long-read sequencing methods. Eight haplotypes from four donors were resolved using a combination of overlapping fosmid clones and long-read assemblies, while eight additional haplotypes were resolved using haplotype-phased assemblies from parent-child trios. Building on these carefully curated haplotypes, the authors were able to validate a bioinformatic pipeline capable of assembling the IGHC locus from long-read sequencing data at scale. Having demonstrated the fidelity of these methods, application at high throughput in an additional 97 individuals allowed them to develop a catalog of IGHC variation from 105 individuals across a diverse set of ancestral backgrounds. In doing so, they identified several structural variants, including insertion-deletions and inversions. Likewise, they identified 2,000 SNVs comprising over 200 IGHC alleles, the vast majority of which had not been previously characterized. Remarkably, every individual examined harbored at least one novel allele, which were especially frequent in individuals with African ancestry. Similar to other highly variable immunogenomic regions, such as HLA,[8] the authors clearly demonstrate that this variation tracks closely with geographic ancestry.

One of the most striking results from this work is the finding that a large percentage of the variants identified are not represented in the dbSNP catalog[9] or 1000 Genomes[10] variant calls. This has important implications for genome-wide association studies (GWASs), for example, as it is likely that association signals in this region may have been missed or misinterpreted, including potential functional consequences of associated variation. Likewise, in the increasing number of whole genome sequencing studies, accurate and comprehensive reference sequence is critical to understanding whether a novel variant is rare or potentially deleterious, or likely due to sequencing artifact. The contribution by Jana et al. to the representation of IGHC variation will support work reliant on comprehensive variant catalogs.

While high-fidelity long-read sequencing and the accompanying validated bioinformatic pipeline were crucial to this work, extension to short-read sequencing data will be an important next step toward further characterization of IGHC variation. The potential to capitalize on the work presented by Jana et al. to provide a scaffold for both bioinformatic and targeted sequencing methods that take advantage of the lower cost and higher throughput characteristic of short read data is a tantalizing prospect. In the meantime, this work promises to catalyze research into the impact of genomic variation on antibody structure and function.

DECLARATION OF INTERESTS

The author declares no competing interests.

REFERENCES

1. Schatz, D.G., and Swanson, P.C. (2011). V(D)J recombination: mechanisms of initiation. Annu. Rev. Genet. *45*, 167–202.
2. Woodsworth, D.J., Castellarin, M., and Holt, R.A. (2013). Sequence analysis of T-cell repertoires in health and disease. Genome Med. *5*, 98.
3. Murphy, K.,W.C. (2016). Janeway's Immunobiology (Garland Science).
4. Stavnezer, J., Guikema, J.E.J., and Schrader, C.E. (2008). Mechanism and regulation of class switch recombination. Annu. Rev. Immunol. *26*, 261–292.
5. Lefranc, M.P. (2001). Nomenclature of the human immunoglobulin heavy (IGH) genes. Exp. Clin. Immunogenet. *18*, 100–116.
6. Rodriguez, O.L., Safonova, Y., Silver, C.A., Shields, K., Gibson, W.S., Kos, J.T., Tieri, D., Ke, H., Jackson, K.J.L., Boyd, S.D., et al. (2023). Genetic variation in the immunoglobulin heavy chain locus shapes the human antibody repertoire. Nat. Commun. *14*, 4419.
7. Jana, U., Rodriguez, O.L., Lees, W., Engelbrecht, E., Vanwinkle, Z., Peres, A., Gibson, W.S., Shields, K., Schultze, S., Dorgham, A., et al. (2025). The human IG heavy chain constant gene locus is enriched for large structural variants and coding polymorphisms that vary among human populations. Cell Genom. 101058. https://doi.org/10.1016/j.xgen.2025.101058.
8. Fernandez Vina, M.A., Hollenbach, J.A., Lyke, K.E., Sztein, M.B., Maiers, M., Klitz, W., Cano, P., Mack, S., Single, R., Brautbar, C., et al. (2012). Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. Philos. Trans. R. Soc. Lond. B Biol. Sci. *367*, 820–829.
9. Sherry, S.T., Ward, M., and Sirotkin, K. (1999). dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res. *9*, 677–679.
10. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.