The Super-Information Age of Immunoglobulin Genetics

By Patrick C. Wilson and J. Donald Capra

From Microbial Immunogenetics, the Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73112

If the information age arose with the popularization of the personal computer, then the emergence of superpowered versions of these machines completely integrated into our lives and into the superstructure of society via the Internet has led us into the super-information age. Information is pouring into our lives and onto our desktops with unprecedented speed and volume. Genomics technology and the super-information age have led to a similar proliferation of information in the study of human genetics and immunobiology. In this issue of the Journal of Experimental Medicine, through innovation, persistence, and technological prowess, Matsuda et al. have completed sequencing the entire human \( V_H \) gene locus, including all of the immunoglobulin \( V_H \) genes as well as the hundreds of kilobases of intervening sequence (1). This landmark work and the efforts of several other groups of pioneering immunogeneticists in the study of the human \( V_H \) locus (2–5) have produced an extraordinary resource for the study of B cell biology and the important issues of immunoglobulin diversity. It appears that the study of immunoglobulin \( V_H \) genes in the human has also entered into the super-information era.

Tonegawa and the Information Age. With the final proof by Tonegawa (2) that the incredible diversity of immunoglobulin \( V \) gene sequences encoding the antibody repertoire were formed by the somatic recombination of relatively few genetic elements (for review see reference 6), the information age of \( V \) gene genetics was begun. In the almost twenty years since, a steady flux of monumental dimensions has entered the personal computer, then the emergence of superpowered versions of these machines completely integrated into our lives and into the superstructure of society via the Internet has led us into the super-information age. Information is pouring into our lives and onto our desktops with unprecedented speed and volume. Genomics technology and the super-information age have led to a similar proliferation of information in the study of human genetics and immunobiology. In this issue of the Journal of Experimental Medicine, through innovation, persistence, and technological prowess, Matsuda et al. have completed sequencing the entire human \( V_H \) gene locus, including all of the immunoglobulin \( V_H \) genes as well as the hundreds of kilobases of intervening sequence (1). This landmark work and the efforts of several other groups of pioneering immunogeneticists in the study of the human \( V_H \) locus (2–5) have produced an extraordinary resource for the study of B cell biology and the important issues of immunoglobulin diversity. It appears that the study of immunoglobulin \( V_H \) genes in the human has also entered into the super-information era.

With the final proof by Tonegawa (2) that the incredible diversity of immunoglobulin \( V \) gene sequences encoding the antibody repertoire were formed by the somatic recombination of relatively few genetic elements (for review see reference 6), the information age of \( V \) gene genetics was begun. In the almost twenty years since, a steady flux of monumental dimensions has entered the personal computer, then the emergence of superpowered versions of these machines completely integrated into our lives and into the superstructure of society via the Internet has led us into the super-information age. Information is pouring into our lives and onto our desktops with unprecedented speed and volume. Genomics technology and the super-information age have led to a similar proliferation of information in the study of human genetics and immunobiology. In this issue of the Journal of Experimental Medicine, through innovation, persistence, and technological prowess, Matsuda et al. have completed sequencing the entire human \( V_H \) gene locus, including all of the immunoglobulin \( V_H \) genes as well as the hundreds of kilobases of intervening sequence (1). This landmark work and the efforts of several other groups of pioneering immunogeneticists in the study of the human \( V_H \) locus (2–5) have produced an extraordinary resource for the study of B cell biology and the important issues of immunoglobulin diversity. It appears that the study of immunoglobulin \( V_H \) genes in the human has also entered into the super-information era.

With the addition of a second haplotype from isolated YAC clones, the Honjo lab constructed a map of the 3' end of the \( V_H \) locus in 1991 (13), which was extended in 1993 to 0.8 megabase encompassing the 64 most 3' \( V_H \) segments (4). The use of genomic Southern blot hybridization analyses provided an estimate at the time of the total compliment of \( V_H \) genes, determined to be between 60 and 200 genes (10, 14). During this period, the expressed \( V_H \) gene repertoire was being compiled and appreciated as many of the \( V_H \) genes recognized today were cloned from a number of sources. With these advances, it was recognized that the \( V_H \) genes consist of seven families based on homology of >80% among \( V_H \) genes within each family (9, 10, 15–18). In 1992, in a significant contribution to this effort, Tomlinson et al. used a PCR-based approach using various sets of \( V \) gene family-specific primers to sequence 74 \( V \) genes designated as DP-1 through DP-74 from a single individual (17). These sequences became instrumental in the efforts of this group to map the \( V_H \) locus (see discussion of Cook and colleagues, below), and provided a good approximation of the total \( V_H \) gene compliment of a single individual. The ongoing mapping efforts demonstrated that there is no significant clustering of the individual \( V_H \) gene families on the locus because they were found intermixed throughout. It was also recognized, primarily through cDNA sequence analysis of numerous autoantibodies performed in laboratories around the world, that although there is certainly allelism involving the individual human \( V_H \) genes (19–21) compared with other multigene families such as the HLA complex and to the mouse \( V_H \) genes, the human \( V_H \) locus exhibits relatively little polymorphism (22–24). This trait was an important factor in the efforts to sequence the human \( V_H \) locus. However, it should be realized that there are differences between people in the presence or absence of single or blocks of \( V_H \) genes due to insertion/deletion polymorphism of the \( V_H \) locus (11, 13, 25).
Finally, in 1994, Cook and colleagues completed the map of the human $V_H$ locus with the analysis of a second haplotype at the 3' end (25) and through an extension to the telomeric end of chromosome 14q32.3 (5). The final map of the $V_H$ locus at this time was $\sim$1,100 kb and included an estimated 95 $V_H$ genes varying depending on the haplotype, with $\sim$51 functional and the remainder pseudogenes. In addition, 24 orphan $V_H$ genes not believed to contribute to the production of functional antibodies were found on chromosomes 15 and 16 (26–28).

About a Million Basepairs. In this issue of *The Journal of Experimental Medicine*, Matsuda et al. report having independently mapped the telomeric end of the $V_H$ locus, and have completed sequencing the entire span of 957,090 bp (1). In their analysis, the $V_H$ locus contains 123 $V_H$ genes, including 39 functional genes known to produce heavy chains, 5 genes that appear functional but have not been reported as heavy chain proteins, and 79 pseudogenes. It is striking that approximately two-thirds of the $V_H$ genes in the human locus are nonfunctional. Again it should be appreciated that the human $V_H$ locus can contain insertional/deletional polymorphism depending on the particular haplotype (11, 13, 25). Previous estimates report that approximately half of the $V_H$ genes are functional, and we have found transcripts for 10–12 $V_H$ family genes in analyses of tonsils from five different individuals (Wilson, P.C., Y.J. Liu, J. Banchereau, V. Pascual, and J.D. Capra, unpublished results), compared with the seven transcribed $V_H$ genes reported by Matsuda et al. in this issue (1). The extent and importance of $V_H$ gene allelism between particular $V$ genes or the total compliment of $V_H$ genes in different individuals, disease states, or racial groups is an area of interest that should not be superceded. Of particular interest concerning the differential complexity of $V$ gene loci are the recent surprising findings of Green and Jakobovits that somatic gene segments. Indeed, the great debate of germline versus somatic is now fully laid to rest. Another generation of scientists is looking at promoters, enhancers, tissue-specific factors, chromosomal end points, and the like. All of this is now available thanks to this pioneering work by Matsuda et al. (1). The study of human $V_H$ genes has now entered the postgenomics era in which all human bioscience will be propelled in the near future via the human genome project.

Address correspondence to J. Donald Capra, Molecular Immunogenetics, The Oklahoma Medical Research Foundation, 825 N E 13th St., Oklahoma City, OK 73112. Phone: 405-271-7393; Fax: 405-271-8237; E-mail: the_capras@email.msn.com

Received for publication 30 September 1998.

References

1. Matsuda, F., K. Ishii, P. Bourvagnet, K.-i. Kuma, H. Hayashida, T. Miyata, and T. Honjo. 1998. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J. Exp. Med. 188:2151–2162.

2. Tonegawa, S. 1983. Somatic generation of antibody diversity. Nature 302:575–581.

3. Croce, C.M., M. Shander, J. Martinis, L. Cicurel, G.G. D’Ancona, T.W. Dolby, and H. Koprowski. 1979. Chromo-
somal location of the genes for human immunoglobulin heavy chains. Proc Natl Acad. Sci. USA. 76:3416–3419.

4. Matsuda, F., E.K. Shin, H. Nagaoa, M. Matsumura, M. Haino, Y. Fukita, S. Takaishi, T. Imai, J.H. Riley, R. Anand, and T. Honjo. 1993. Structure and physical map of 64 variable segments in the 3'0.8-megabase region of the human immunoglobulin heavy-chain locus. Nat. Genet. 3:88–94.

5. Cook, G.P., I.M. Tomlinson, G. Walter, H. Riethman, N.P. Carter, L. Buluwela, G. Winter, and T.H. Raabitts. 1994. A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q. Nat. Genet. 7:162–168.

6. Kindt, T.J., and J.D. Capra. 1984. The Antibody Enigma. Plenum Press, New York.

7. Cox, D.W., V.D. Markovic, and I.E. Teshima. 1982. Genes for immunoglobulin heavy chains and for alpha 1-antitrypsin are localized to specific regions of chromosome 14q. Nature 297:428–430.

8. Kirsch, I.R., C.C. Mortison, K. Nakahara, and P. Leder. 1982. Human immunoglobulin heavy chain genes map to a region of translocations in malignant B lymphocytes. Science 216:301–303.

9. Kodaira, M., T. Kinashi, I. Umemura, F. Matsuda, T. Noma, Y. Oono, and T. Honjo. 1986. Organization and evolution of variable region genes of the human immunoglobulin heavy chain. J. Mol. Biol. 190:529–541.

10. Berman, J.E., S.J. Mellis, R. Pollock, C.L. Smith, H. Suh, B. Heinke, C. Kowal, U. Surti, L. Chess, and C.R. Cantor. 1988. Content and organization of the human Ig VH locus: definition of three new VH families and linkage to the Ig CH locus. EMBO J. 7:727–738.

11. Walter, M.A., U. Surti, M.H. Hofker, and D.W. Cox. 1990. The physical organization of the human immunoglobulin heavy chain gene complex. EMBO J. 9:3303–3313.

12. van Dijk, K.W., L.A. Milner, and E.C. Milner. 1992. Mapping of human H chain V region genes (VH4) using deletion analysis and pulsed field gel electrophoresis. J. Immunol. 148:2923–2931.

13. Shin, E.K., F. Matsuda, H. Nagaoa, Y. Fukita, T. Imai, K. Yokoyama, E. Soeda, and T. Honjo. 1991. Physical map of the 3' region of the human immunoglobulin heavy chain locus: clustering of autoantibody-related variable segments in one haplotype. EMBO J. 10:3641–3645.

14. Matsuda, F., K.H. Lee, S. Nakai, T. Sato, M. Kodaira, S.Q. Zong, H. Oono, S. Fukuhara, and T. Honjo. 1988. Dispersion localized of D segments in the human immunoglobulin heavy-chain locus. EMBO J. 7:1047–1051.

15. Lee, K.H., F. Matsuda, T. Kinashi, M. Kodaira, and T. Honjo. 1987. A novel family of variable region genes of the human immunoglobulin heavy chain. J. Mol. Biol. 195:761–768.

16. Shen, A., C. Humphries, P. Tucker, and F. Blattner. 1987. Human heavy-chain variable region gene family nonrandomly rearranged in familial chronic lymphocytic leukemia.