Am. J. Hum. Genet. 42:446–451, 1988

Analysis of Genetic Variation Reveals Human Immunoglobulin VH-Region Gene Organization

Michael A. Walter* and Diane W. Cox*,†,‡

*Research Institute, Hospital for Sick Children, and Departments of †Paediatrics and ‡Medical Genetics, University of Toronto, Toronto

Summary

We have investigated the extent of genetic variation and the number of germ-line heavy-chain-variable (VH) genes to obtain information on the organization and repertoire of the VH genes. Our studies revealed extensive genetic variation in this region, indicated by restriction-endonuclease site polymorphisms. Analysis of the distribution of selected polymorphic loci revealed evidence of linkage disequilibrium, particularly between VH2 and VH3 subclass loci, indicating that the subclasses are interdispersed in the human germ-line chromosome. Absolute correlation was detected between alleles of a VH2 locus and the alleles of three VH3 loci, evidence for an extra set of VH genes, which are present in 48% of the Caucasian population. A preliminary estimate of the number of VH genes, approximately 50, indicates a smaller number of VH genes than suggested by the amount of protein variation. The extensive genetic variation we have observed may be associated with genetic differences in the immune response and potentially with variable susceptibility to autoimmune disorders.

Introduction

Immunoglobulins are composed of two types of polypeptide chain, the heavy (H) and the light (L), each of which contains a variable (V) region encoding the NH_2 terminus attached to a COOH terminal constant (C) region. The antigen specificity of an antibody molecule is determined by the amino acid sequence of the variable region. The variable region of the heavy-chain peptide of the antibody is coded by the sequences of three "minigenes"—the VH region sequences, diversity (D) region sequences, and joining (J) region sequences, each selected from a pool of such genes (for review, see Tonegawa 1983; Yancopoulos and Alt 1986).

The VH genes have been subdivided into three subclasses based upon protein and DNA homologies—subclasses VH1, VH2, and VH3 (Kabat et al. 1983). Genes within a VH family generally share >80% sequence homology while the sequence homology between VH subclasses is generally <70% (Brodeur and Riblet 1984). While protein sequence data suggest that the VH repertoire may be large, perhaps more than 140 genes (Kabat et al. 1983), the organization of the VH gene region and the exact number of germ-line VH genes are unknown.

We searched for polymorphisms in the VH3 and VH2 subclasses to characterize VH genetic variation and as an approach to the study of VH gene organization. The distributions of the alleles of six selected VH polymorphisms were studied in 10 3-generation normal Canadian Caucasian families and 10 2-generation families. We present here a map of these VH loci based upon the degree of correlation found between the alleles of these loci. Second, we have obtained an estimate of the extent of the VH gene repertoire, at the DNA level, of the VH1, VH2, and VH3 subclasses.

Material and Methods

DNA was obtained from peripheral blood as described elsewhere (Linsley et al. 1983). (Leukocytes

Received August 3, 1987; revision received October 7, 1987.

Address for correspondence and reprints: Diane W. Cox, Ph.D., Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada, M5G 1X8.

0002-9297/88/4203-0006$02.00

were an appropriate source of DNA for the study of unrearranged VH genes because leukocyte DNA and fibroblast DNA digested with restriction endonucleases and hybridized with VH gene probes showed identical hybridizing fragments). Three micrograms of DNA was digested with 3U restriction endonuclease/μg in manufacturer's recommended buffer overnight and electrophoresed in 0.8% agarose gels (with 0.8 μg/ml ethidium bromide) for 1,200–1,600 volt/hours. Gels were then treated with 0.25 N HCl for 15 min and transferred to Hybond® nylon membrane according to manufacturer's recommendations. Filters were baked for 2 h, prehybridized for ≥1 h, and hybridized at 42 C for 18–24 h with 2–5 × 10^6 cpm of nick-translated probe/ml of hybridization solution. Prehybridization and hybridization solutions were as decribed elsewhere (Linsley et al. 1983). Filters were washed in 2 × SSC for ½ h and twice for 1 h in 0.1 × SSC, 0.1% SDS at 50 C. Autoradiography was done at −70 C overnight.

DNA probes used in this study were as follows: VH26-8, a VH3 gene probe that was a gift of T. Rabbitts (Matthyssens and Rabbitts 1980); VHE2.2, a 2.2-kb *Eco*RI fragment flanking VH26-8 and derived from λ VH26, the latter also a gift of T. Rabbitts (Matthyssens and Rabbitts, 1980); V_{CE-1}, a cloned VH2 subclass gene that was a gift of T. Honjo (Takahashi et al. 1984); VH2.EB1.2, a 1.2-kb *Eco*RI/*Bam*HI fragment derived from V_{CE-1} that no longer contains JH sequences; and V_{E3-D10}, a cloned VH1 subclass gene that was a gift of T. Honjo (Noma et al. 1984).

Results

VH3 Polymorphisms

DNA extracted from leukocytes of six random Canadian Caucasian individuals was digested with 10 restriction endonucleases to detect VH-region polymorphic loci. VHE2.2 detected polymorphisms with the following restriction enzymes: *Eco*RI, *Sac*I (discussed below), *Apa*I (Walter et al. 1987), *Hin*dIII, *Ava*II, *Rsa*I, *Taq*I, and *Bgl*II. Variation in intensity between bands was observed and can be explained by VH genes having differing degrees of homology with the probe. The greatest variation of VH3 genes was observed with *Eco*RI, which was therefore selected for further study. The three *Eco*RI polymorphic loci were designated loci 1, 2, and 3 (fig. 1). Locus 1 has a 6.7-kb fragment with no allelic counterpart. Locus 2

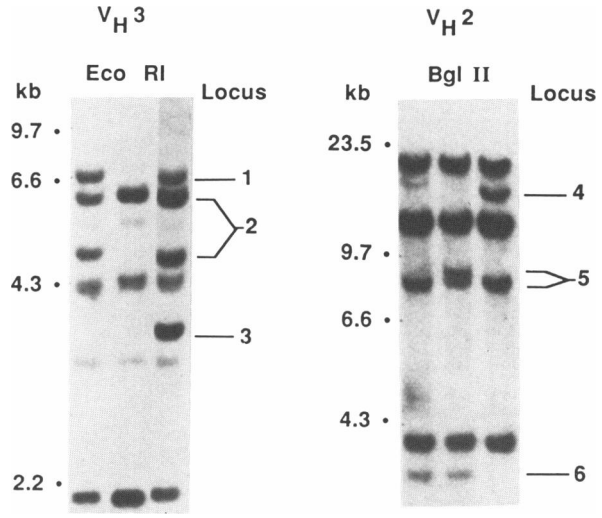

Figure 1 Autoradiograph of polymorphic human VH region loci. Lanes labeled "VH3" indicate *Eco*RI-digested DNA probed with VHE2.2. Lanes labeled "VH2" indicate *Bgl*II-digested DNA probed with V_{CE-1}.

has two alleles–a 6.0- and a 4.6-kb fragment. Locus 3 was scored as the presence or absence of a 3.5-kb fragment. No variation was observed in digests with *Xba*I or *Bam*HI. The high degree of polymorphism observed in the VH region may be due in part to several enzymes detecting the same difference in gene number or to the large number of genes a probe detects.

VH2 Polymorphisms

V_{CE-1} was used to probe similar panels of random individuals' DNA digested with 12 different enzymes. The probe detected polymorphisms with *Bgl*II, *Eco*RI, *Apa*I, *Hin*dIII, *Xba*I, *Msp*I, *Rsa*I, *Ava*II, and *Taq*I. *Bgl*II revealed the greatest variability with three VH2 subclass polymorphic loci, as described elsewhere (Johnson et al. 1984). These were designated here as loci 4, 5, and 6 (fig. 1) and were selected for further analysis. Locus 4 has a 12.0-kb fragment with no allelic counterpart. Locus 5 has two alleles—a 7.5- and a 7.0-kb form. Locus 6 has one detectable allele—a 3.4-kb fragment. No VH2 polymorphic loci were detected in digests with *Bam*HI, *Pst*I, or *Sac*I.

Analysis of the distribution of these six loci in 10 3-generation Canadian Caucasian families and five 2-generation families, yielding 134 independent chromosomes, confirmed allelic designations. Allelic frequencies are shown in table 1. Segregation of the

Table 1

Allelic Designations and Frequencies of Six VH-Region Polymorphic Loci

Enzyme (Subclass) and Locus[a]	Alleles[b] (kb)	Average ± SE Frequency
*Eco*RI (VH3):		
1	6.7	.28 ± .04
	0	.72 ± .04
2	6.0	.83 ± .03
	4.6	.17 ± .03
3	3.5	.87 ± .03
	0	.13 ± .03
*Bgl*II (VH2):		
4	12.0	.28 ± .04
	0	.72 ± .04
5	7.5	.94 ± .02
	7.0	.06 ± .02
6	3.4	.45 ± .04
	0	.55 ± .04

[a] Loci 1–3 are polymorphisms detected in *Eco*RI-digested DNA with probe VHE2.2; loci 4–6 are polymorphisms detected in *Bgl*II-digested DNA with probe V_{CE-1}.

[b] Fragment sizes were determined from measurements of a minimum of five similarly probed filters. A zero indicates that an allelic counterpart to the polymorphic band could not be found.

allelic combinations (haplotypes) was compatible with a lack of recombination between VH polymorphic loci in all families studied. The alleles of all six loci were in Hardy-Weinberg equilibrium. The allele frequencies of locus 6 were found to be statistically different from those elsewhere reported for this locus (Johnson et al. 1984). This discrepancy could be due to different sample populations or to technical differences, since we found that consistent scoring of allelic dosage was possible only with 50 C washes, not with the 65 C washes used elsewhere (Johnson et al. 1984). Any results determined by allele dosage were obtained by comparison of the relevant polymorphic band with constant bands within the same lane.

VH Linkage Disequilibrium

The distributions of the alleles of the six VH3 and VH2 polymorphic loci on the 134 chromosomes generated from our family data were analyzed for the presence of linkage disequilibrium as revealed by tests of significant correlations. χ^2 Analysis of all pairwise combinations of the six loci in 2 × 2 contingency tables was carried out, and the uncorrected correlations (calculated as $r = [\chi^2/n]^{1/2}$) are summarized in table 2. Statistical significance was considered as $P < .05$, determined by χ^2 analysis or by Fisher's exact method when appropriate (Steel and Torrie, pp. 504–506). Linkage disequilibrium was found in the VH region, as in the heavy-chain constant region (Bech-Hansen et al. 1983). Six correlations were significant (table 2). Four involve a VH2 locus in linkage disequilibrium with a VH3 locus.

Table 2

Correlation Coefficients and Associated Probabilities of Pairwise Comparisons of Six Polymorphic Loci

	VH3 Loci			VH2 Loci		
	1	2	3	4	5	6
VH3:						
1		.21[a]	.17[a]	1.00[a]	.16	.06
2	.02		.11	.21[a]	.22[a]	.06
3	.04	.13		.17[a]	.02	.03
VH2:						
4	6×10^{-34}	.02	.04		.16	.06
5	.07	.02	.80	.07		.04
6	.50	.50	.70	.50	.60	

NOTE.—Numbers in the upper right are correlation coefficients, numbers in the lower left are the probability that these values could occur randomly.

[a] Significant at $P < .05$.

Other VH Polymorphisms

Two other VH3 polymorphisms—a 6.6-kb *Sac*I fragment detected with VHE2.2 and a 2.7-kb *Hin*dIII fragment detected with VH26-8 (fig. 2)—also were analyzed through family studies (Walter and Cox 1986). A preliminary analysis of allelic distributions in more than 25 independent chromosomes revealed that these polymorphisms also show a correlation of 1.00 at $P < .05$ with loci 1 and 4.

VH Gene Number

Our data were also used to estimate the number of germ-line VH genes. A number of assumptions were necessary for our approach. As VH genes are small (<0.5 kb) (Matthyssens and Rabbitts 1980) and the regions between genes are large (on average 17 kb [Kodaira et al. 1986]), hexameric restriction endonucleases should usually recognize sites between, rather than within, genes. The number of bands detected in the DNA of six random individuals averaged for several hexameric restriction enzymes should, therefore, equal the number of genes detected by the given probe. This assumes that the probes used are specific for one subclass only and detect all members of that subclass at low stringency conditions. Intensity differences between bands presumably re-

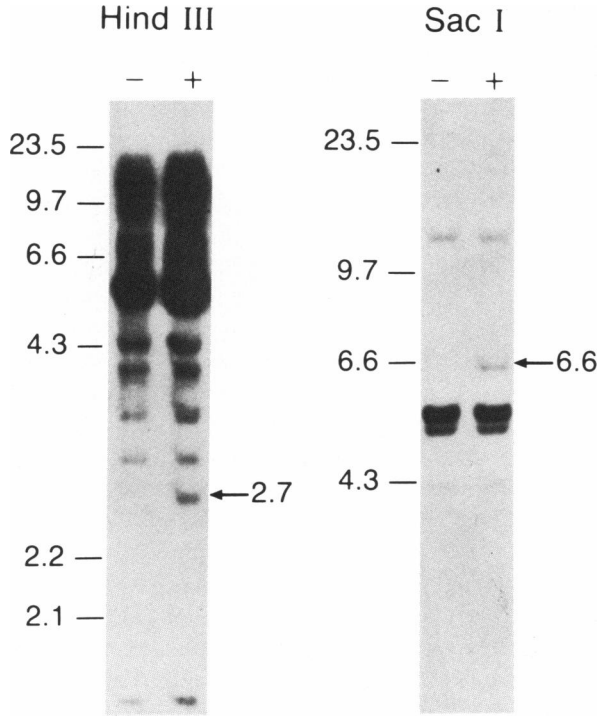

Figure 2 Autoradiograph of VH3 polymorphic loci in linkage disequilibrium with loci 1 and 4. Lanes labeled "Hind III" are DNA digested with restriction enzyme *Hin*dIII and probed with VH26-8. Lanes labeled "Sac I" are DNA samples digested with enzyme *Sac*I and probed with VHE2.2. Plus signs (+) and minus signs (−) refer to the presence and absence of the polymorphic fragments, respectively.

flect differing degrees of homology to the probe, as born out in experiments where fainter bands are usually lost first as the stringency is increased (data not shown). As polymorphisms are known to exist from our studies, the number obtained from this method will overestimate the actual VH gene number.

The number of bands observed for a series of enzymes for all three VH subclasses is shown in table 3. $V_{E3\text{-}D10}$ was used as the VH1 subclass probe; VH2.EB1.2 was used as the VH2 subclass probe; VH26-8 was used as the VH3 subclass probe. None of these probes contained JH sequences. As estimated by this method, there are ~17 VH1 genes, 7 VH2 genes, and 24 VH3 genes, totaling 48.4 ± 4.3 VH genes of these VH subclasses.

Discussion

Our analysis of VH polymorphic loci demonstrated that linkage disequilibrium occurs in the VH region. In four cases, an association was observed between VH3 and VH2 subclass polymorphic loci. A likely explanation is that the loci involved are physically close or adjacent, suggesting that the genes of the VH2 and VH3 subgroups are interdispersed, as are the genes for the human VK subclasses (Pech and Zachau 1984). This explanation is consistent with the recent discovery of genes of different subclasses interdispersed in clusters, on overlapping cosmid clones (Kodaira et al. 1986). This organization is different than that found in the mouse, where genes within the 7–10 VH subclasses (Brodeur and Riblet 1984) are distinctly clustered (Kemp et al. 1981; Rechavi et al. 1982; Yancopoulos et al. 1984).

The presence of linkage disequilibrium in the VH region yields insight into the organization of the VH-gene cluster. Based upon the degree of correlation observed between VH polymorphic loci, a "correlation map" of some of the polymorphic loci has been constructed (fig. 3). Similar studies have been used for CH gene order, which proved consistent with cosmid cloning data (Bech-Hansen et al. 1983). To determine whether the degree of correlation between loci on this map is an indication of the physical relation of these loci, the distances between these loci are being determined by pulsed-field gel electrophoresis.

Table 3

Germ-Line Number of VH Genes

Subclass[a]	Restriction Enzymes	Average ± SE No. of Genes[b]
VH1	*Bgl*II, *Bam*HI, *Eco*RI, *Hin*dIII, *Xba*I, *Sac*I, *Apa*I	17.4 ± 1.4
VH2	*Bgl*II, *Bam*HI, *Eco*RI, *Hin*dIII, *Xba*I, *Sac*I, *Apa*I, *Pst*I	6.8 ± 0.4
VH3	*Eco*RI, *Bam*HI, *Xba*I, *Sac*I, *Hin*dIII, *Apa*I	24.2 ± 0.9
Total		48.4 ± 4.3

[a] $V_{E3\text{-}D10}$ was the VH1 subclass probe; VH2.EB1.2 was the VH2 subclass probe; VH26-8 was the VH3 subclass probe.

[b] Bands were counted as single genes under the assumptions outlined in the text.

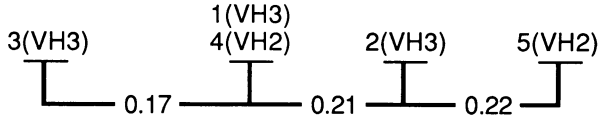

Figure 3 Correlation map of VH polymorphic loci. This figure only includes significant correlations; nonsignificant correlation values are given in table 2. The ordering shown was determined by the strength of the correlations between pairs of VH genes. The broadness of the confidence intervals of these correlation values allows other possible orderings.

Interestingly, the linkage disequilibrium between locus 1 (VH3) and locus 4 (VH2) is absolute, suggesting that these loci must be very close, perhaps adjacent. Experiments in which VHE2.2 was used to probe *Bgl*II-digested DNA and V_{CE-1} was used to probe *Eco*RI-digested DNA did not reveal cross homologies of the bands in question (data not shown). Thus the observations are not due to probe cross homologies. Furthermore, two other VH3 polymorphisms—the 6.6-kb *Sac*I fragment detected with VHE2.2 and a 2.7-kb *Hin*dIII fragment detected with VH26-8—also show total correlation with loci 1 and 4. Turnbull and co-workers (Turnbull et al. 1987) independently demonstrated that the 2.7-kb *Hin*dIII polymorphism and locus 1 (which they refer to as HE7.3) were in tight linkage disequilibrium. We could not find allelic bands for locus 1, locus 4, the 6.6-kb *Sac*I fragment, or the 2.7-kb *Hin*dIII fragment. We suggest that the occurrence of an extra set of VH genes, perhaps generated by gene duplication followed by divergence, is consistent with these findings. This extra set of genes, involving a VH2 gene and at least one VH3 gene, occurs in at least one chromosome in 48% of the normal Caucasian population. The implication that a subset of the normal Caucasian population could have "extra" VH genes, potentially expanding their antibody repertoire, could have practical consequences.

Our data were also used to estimate the number of germ-line VH genes. There are 32, 13, and 91 known protein variants of VH1, VH2, and VH3 subclasses, respectively (Kabat et al. 1983), or approximately 140 in total. As somatic mutation increases the number of protein variants (Tonegawa 1983) and has been estimated as 10,000-fold greater in B lymphocytes than the normal spontaneous mutation rate (Meyer et al. 1986), these numbers are thought to be overestimates of the actual germ-line gene number. Our experiments were conducted on unrearranged chromosomes to avoid this problem.

Our results suggest that there are approximately 17 VH1 genes, 7 VH2 genes, and 24 VH3 genes, totaling 48.4 ± 4.3 VH genes belonging to these VH subclasses. This is a smaller number than protein data indicate. The difference between these estimates of VH gene number is due not only to somatic mutation but also to the extent of genetic variation present in the population.

Our estimate of the number of VH genes in humans, while inexact, contrasts sharply with that in the mouse, where the number of genes in one of the murine VH subclasses has been estimated as approximately 1,000 (Livant et al. 1986). The number and organizational differences between these species suggests that the evolution of VH genes in mouse and man has apparently been very different.

Historically, analysis of association between immune disorders and potential genetic differences in antibody structure has relied on Gm types, genetic markers of the CH region, despite the fact that differences in VH repertoires are more likely to be involved in immunological disorders. This was due to the absence of suitable well-characterized and useful VH markers. We have used the VH polymorphic loci to generate insight into VH gene organization. However, the six polymorphic VH loci described in this paper are also excellent candidates for association studies. Our use of a probe flanking a VH3 gene simplifies the DNA hybridization patterns obtained, in contrast to the extremely complex ones obtained using a VH3 gene probe (Turnbull et al. 1987). Potential disease predispositions and susceptibilities can now be examined in relation to differences in the germ-line VH region of antibodies.

Acknowledgments

We thank P. S. Linsley for isolating the probe VHE2.2, T. H. Rabbitts for the VH probes λ VH26 and VH26-8, and T. Honjo for the VH gene probes V_{CE-1} and V_{E3-D10}. This work was funded by grants from the National Science and Engineering Research Council of Canada and by a studentship from the Medical Research Council of Canada.

References

Bech-Hansen, N. T., P. S. Linsley, and D. W. Cox. 1983. Restriction fragment length polymorphisms associated with immunoglobulin C γ genes reveal linkage disequilibrium and genomic organization. Proc. Natl. Acad. Sci. USA 80:6952–6956.

Brodeur, P. H., and R. Riblet. 1984. The immunoglobulin heavy chain variable region (Igh-V) locus in the mouse I: one hundred Igh-V genes comprise seven families of homologous genes. Eur. J. Immunol. 14:922–930.

Johnson, M. J., A. M. Natali, H. M. Cann, T. Honjo, and L. L. Cavalli-Sforza. 1984. Polymorphisms of a human variable heavy chain gene show linkage with constant heavy chain genes. Proc. Natl. Acad. Sci. USA 81:7840–7844.

Kabat, E. A., T. T. Wu, H. Bilofsky, M. Reid-Miller, and H. Perry. 1983. Sequences of proteins of immunological interest. U.S. Dept. of Health and Human Services, Washington, DC.

Kemp, D. J., B. Tyler, O. Bernard, N. Gough, S. Geron-

dakis, J. H. Adams, and S. Cory. 1981. Organization of genes and spacers within the mouse immunoglobulin VH locus. J. Mol. Appl. Genet. **1**:245–261.

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

Linsley, P. S., N. T. Bech-Hansen, L. Siminovitch and D. W. Cox. 1983. Analysis of a break in chromosome 14 mapping to the region of the immunoglobulin heavy chain locus. Proc. Natl. Acad. Sci. USA **80**:1997–2001.

Livant, D., C. Blatt, and L. Hood. 1986. One heavy chain variable region gene segment subfamily in the BALB/c mouse contains 500–1000 or more members. Cell **47**:461–470.

Matthyssens, G., and T. H. Rabbitts. 1980. Structure and multiplicity of genes for the human immunoglobulin heavy chain variable region. Proc. Natl. Acad. Sci. USA **77**:6561–6565.

Meyer, J., H. M. Jack, N. Ellis, and M. Wabl. 1986. High rate of somatic point mutation in vitro in and near the variable region segment of an immunoglobulin heavy chain gene. Proc. Natl. Acad. Sci. USA. **83**:6950–6953.

Noma, Y., Y. Yaoita, N. Matsunami, A. Rosen, G. Klein, and T. Honjo. 1984. Immunoglobulin gene organization of ultraviolet-illuminated human lymphoblastoid cell lines producing both IgM and IgG. Mol. Biol. Med. **2**:337–350.

Pech, M., and H. G. Zachau. 1984. Immunoglobulin genes of different subclasses are interdigitated within the VK locus. Nucleic Acids Res. **12**:9229–9236.

Rechavi, G., B. Bienz, D. Ram, Y. Ben-Neriah, J. B. Cohen, R. Zakut, and D. Givol. 1982. Organization and evolution of immunoglobulin V_H gene subgroups. Proc. Natl. Acad. Sci. USA **79**:4405–4409.

Steel, R. G. D., and J. H. Torrie. 1980. Principles and procedures of statistics: a biometrical approach. 2d ed. McGraw-Hill, New York.

Takahashi, N., T. Noma, and T. Honjo. 1984. Rearranged immunoglobulin heavy chain variable region (V_H) pseudogene that deletes the second complementarity-determining region. Proc. Natl. Acad. Sci. USA. **81**:5194–5198.

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

Turnbull, I. F., O. Bernard, K. S. Sriprakash, and J. D. Mathews. 1987. Human immunoglobulin variable region genes: a new V_H sequence used to detect polymorphism. Immunogenetics **25**:184–192.

Walter, M. A., and D. W. Cox. 1986. Evidence for a large common duplication in the human immunoglobulin VH region. Genet. Soc. Can. Bull. **17**:71.

Walter, M. A., P. S. Linsley, and D. W. Cox. 1987. Apa I polymorphism of a human immunoglobulin VH3 subclass locus. Nucleic Acids Res. **15**:4697.

Yancopoulos, G. D., and F. W. Alt. 1986. Regulation of the assembly and expression of variable-region genes. Annu. Rev. Immunol. **4**:339–368.

Yancopoulos, G. D., S. V. Desiderio, M. Paskind, J. F. Kearny, D. Baltimore, and F. W. Alt. 1984. Preferential utilization of the most J_H-proximal V_H gene segments in pre-B-cell lines. Nature **311**:727–733.