Am. J. Hum. Genet. 46:795–800, 1990

Human Carboxypeptidase A Identifies a *Bgl*II RFLP and Maps to 7q31-qter

E. A. Stewart,* C. S. Craik,† L. Hake,* and A. M. Bowcock*

*Department of Genetics, Stanford University Medical Center, Stanford, CA; and †Department of Pharmaceutical Chemistry, University of California, San Francisco

Summary

A genomic clone for human carboxypeptidase has been isolated with a probe for rat CPA1 cDNA. A 1.7-kb *Hin*dIII/*Eco*RI fragment from the 3′ flanking region of human carboxypeptidase detects a DNA polymorphism with *Bgl*II. Multipoint linkage analysis with an established map of chromosome 7 markers shows that the most likely location of carboxypeptidase is at 7q31-qter, between D7S87 and D7S93. All other placements can be excluded with odds >100:1. These and other markers confirm that carboxypeptidase lies distal to the locus for cystic fibrosis, at a distance of approximately 12 centimorgans. The regions containing identity to the rat gene were sequenced and shown to be 82% identical to exons 9 and 10 of the rat gene. The presence of a codon for isoleucine at the residues corresponding to codon 255 of rat CPA1 cDNA strongly suggests that the A form of human carboxypeptidase has been isolated.

Introduction

Pancreatic carboxypeptidases are a family of zinc metalloenzymes that are required for degradation of proteins at their carboxytermini. Three members have been identified in the rat: CPA1, CPA2, and CPB. These proteolytic enzymes have similar amino acid sequences and catalyze the same reaction: degradation of polypeptides from the COOH terminus. Two members of this family, bovine carboxypeptidase A (CPA) and carboxypeptidase B (CPB) have been isolated and characterized extensively (Barrett and McDonald 1986). Recently, the genes and cDNAs for rat CPA1, CPA2, and CPB have also been isolated, and the corresponding proteins have been characterized (Clauser et al. 1988; Gardell et al. 1988). The different forms of rat carboxypeptidase differ in substrate preference. Both forms of rat CPA hydrolyze amide bonds in which the C-terminal amino acid has an aromatic or branched aliphatic side chain; however, rat CPA1 preferentially acts on smaller amino acids, and CPA2 acts on larger amino acids. CPB exhibits a substrate specificity for bonds in which the C-terminus is lysine or arginine. Substitutions of amino acids in the different carboxypeptidase family members at important positions within the substrate binding pocket may account for the differences in substrate specificity (Gardell et al. 1988).

The human carboxypeptidase gene had been assigned by somatic cell hybrids to 7q22-qter (Honey et al. 1986). We have isolated a human carboxypeptidase gene from a human genomic library. Its sequence predicts that it is CPA. RFLPs detectable with the CPA probe are described, and linkage analysis with the cystic fibrosis locus and other chromosome 7 markers is performed. Allelic frequencies for six populations are also determined.

Material and Methods

Isolation of Genomic CPA Clone

Total genomic DNA prepared from buffy coat was partially digested with *Mbo*I and ligated into the *Bam*HI cloning site of EMBL3. One million plaques were screened by transfer to nitrocellulose filters (Benton and Davis 1977), followed by hybridization to a rat CPA cDNA. Six positive plaques were picked and purified;

Received August 16, 1989; revision received December 12, 1989.

Address for correspondence and reprints: A. M. Bowcock, Department of Genetics, S-337, Stanford University Medical Center, Stanford, CA 94305.

0002-9297/90/4604-0018$02.00

all six clones had the same restriction map; therefore, only one clone (λHCPA) was characterized further.

Nucleotide Sequence Analysis

DNA from λHCPA was digested with *Sau*3A and cloned into the *Bam*HI site of M13mp18. Plaques were transferred to nitrocellulose filters as described above and hybridized to rat CPA cDNA. After two rounds of screening, six positive plaques were isolated. The human inserts were sequenced using the dideoxy-chain termination technique (Sanger et al. 1977) with Sequenase (USB) reagents. Sequence data were analyzed with the use of Genbank programs.

DNA Analysis

Genomic DNA digestions, Southern transfers, hybridizations, and autoradiography were done according to methods described by Feder et al. (1985). Probe DNA was ^{32}P labeled using the hexamer primer reaction described by Feinberg and Vogelstein (1983). In order to identify DNA polymorphisms, DNA for six unrelated caucasoid individuals was digested with a number of restriction enzymes and hybridized with subcloned probes from λHCPA.

Construction of a Restriction Map

λHCPA DNA was isolated, and single and double digestions were carried out. The digested DNA was analyzed by agarose gel electrophoresis. Gels were transferred to Zetabind by the method of Southern as described above and were probed with genomic DNA or rat CPA CDNA. In order to screen for RFLPs, *Eco*RI fragments from λHCPA were isolated and subcloned into pBR322.

Linkage of CPA to Chromosome 7 Markers

DNA samples from lymphoblastoid lines of 40 large reference families obtained from the Centre d'Etude du Polymorphisme Human (CEPH; Paris) were typed for the RFLP revealed by p184a. Linkage of the CPA gene to 16 chromosome 7 markers was evaluated using a data base of genotypes for the CEPH reference families (Donis-Keller et al. 1987). Linkage of the CPA gene to cystic fibrosis (CF) was also tested on 19 families each with at least two affected CF children (Bowcock et al. 1986).

Analysis of Population Allelic Frequencies

Allelic frequencies for λHCPA were determined in unrelated individuals from six populations. Five populations have been described elsewhere (Bowcock et al. 1987; Cavalli-Sforza et al. 1987). Japanese samples were obtained from individuals born in Japan and living in the San Francisco Bay area.

Results

Polymorphisms with the CPA Gene

Figure 1 shows the restriction map of the CPA genomic clone. Four *Eco*RI subclones were isolated: p184 (4.5 kb), p185 (0.9 kb), p190 (1.25 kb), and p191 (2.3 kb). p184 was then digested with *Hin*dIII to yield p184a (1.7 kb) and p184b (2.8 kb). p184a detected a two-allele polymorphism with the enzyme *Bgl*II, with alleles of 11 kb and 8.7–2.3 kb; all of the other clones had within them repetitive DNA sequences which could not be competed out with human DNA (Ardeshir et al. 1983) and so were not used for RFLP screening. An autoradiograph showing the p184a polymorphism can be seen in figure 2. The polymorphism is likely to be due to a single base-pair change within the *Bgl*II site, since it is not detected with any other enzyme. Other enzymes tested and found to be not polymorphic were *Aat*I, *Alu*I, *Apa*I, *Apa*LI, *Ava*I, *Bam*HI, *Ban*II, *Bcl*II, *Bsp*1286, *Bst*EII, *Bst*NI, *Dde*I, *Dra*I, *Eco*109, *Eco*RI, *Eco*RV, *Hae*II, *Hae*III, *Hin*dIII, *Hph*I, *Kpn*I, *Mbo*II, *Msp*I, *Nci*I, *Pst*I, *Pvu*II, *Rsa*I, *Sac*I, *Sau*961, *Sca*I, *Scr*FI, *Ssp*I, *Sty*I, *Xba*I, and *Xmn*I.

Mendelian Inheritance

Codominant segregation of alleles was shown in all families that were informative when tested: 19 CF families (94 individuals) and 23 CEPH families (378 individuals).

Linkage of CPA to Other Chromosome 7 Markers

Pairwise linkage with CPA and other chromosome 7 markers using the CEPH reference families gave lod scores ($\hat{Z}$) >3.0 with 13 markers (table 1). Location scores (Lathrop et al. 1984) were calculated using the

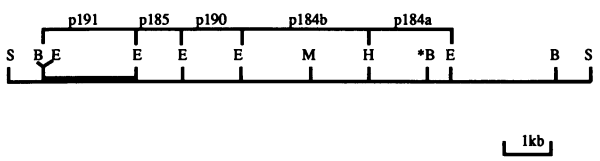

Figure 1 Restriction map of CPA genomic clone λHCPA, showing the locations of the subcloned probes. B = *Bgl*II; E = *Eco*RI; H = *Hin*dIII; M = *Bam*HI; S = *Sal*I (synthetic cloning sites). The thicker line indicates the region of the clone which hybridized to rat CPA cDNA. The location of the polymorphic *Bgl*II site is indicated by an asterisk.

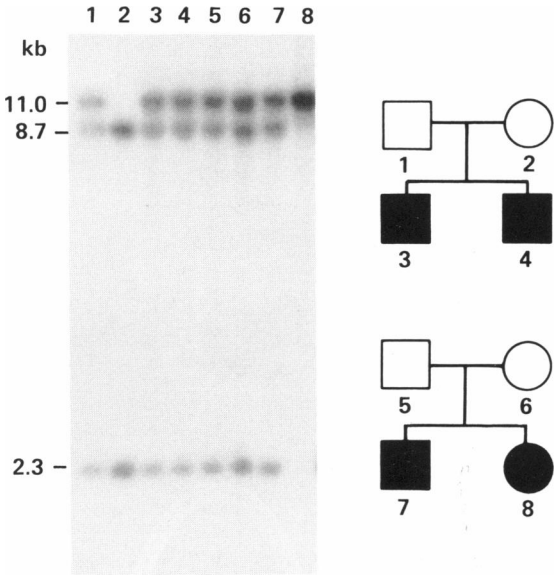

Figure 2 Autoradiograph of genomic blot containing DNA from two families (pedigrees shown on the right) that was digested with *BglII* and hybridized with p184a. Allele lengths (in kilobases) are shown.

previously established map of these and other markers in the region (Donis-Keller et al. 1987). A male-to-female ratio of 0.45 was assumed. The most likely location for the CPA gene is between D7S87 and D7S93 (fig. 3). The odds against the second most likely placement, between D7S107 and D7S93, were 718:1. Two other placements (between D7S72 and D7S107 and between D7S8 and D7S87) could be rejected with odds $>10^5$:1. All other placements could be rejected with odds $>10^6$:1. This confirms the chromosomal localization of human CPA to within the previously determined region of 7q22-qter (Honey et al. 1986). The locus for CF has been mapped between MET and D7S8 at θ = .004 and .003, respectively (Beaudet et al. 1986). Thus CPA lies distal to the CF locus at 7q31-qter. CPA was also found to be linked to CF in 19 multiply affected families; the $\hat{Z}$ 1.5 at $\hat{\theta}$ = .12 was not significant, because of insufficient data. The maximum-likelihood estimate of recombination between CPA and D7S93 is 3%, confirming a distance of approximately 12 centimorgans (cM) distal to the CF locus for CPA.

Nucleotide Sequence Analysis of CPA

Two M13mp18 CPA clones that hybridized to rat CPA were sequenced in order to confirm the genomic clone's identity. 175 bp were sequenced in M13 clone CPA6 (fig. 4). Nucleotides 1–84 had 82% identity with nucleotides 1296–1380 in the rat cDNA sequence (Quinto et al. 1982). Nucleotides 85–129 did not show any significant identity with the rat CPA cDNA sequence. The second clone, CPA5, was 117 bp long and also had 82% identity with the rat CPA cDNA sequence, at positions 1380–1497.

When the CPA6 sequence was compared with the rat CPA1 sequence (Clauser et al. 1988), nucleotides 1–84 were found to correspond to exon 9, with the splice site at nucleotide 84 and with a GT dinucleotide just downstream of position 84. No significant identity is seen within the intron. The CPA5 sequence corresponds to exon 10 in the rat CPA1 gene. The intron/exon boundary was at the cloning site, and therefore any upstream splicing sequences, such as the AG often seen at the 3′ end of an intron (Breathnach and Chambon 1981), were not cloned.

The sequence was translated and compared with the bovine CPA and CPB and with the rat CPA1, CPA2, and CPB amino acid sequences (fig. 4). The two human sequences covered the region from amino acid position 220 to position 286 and had 75% identity with the bovine CPA amino acid sequence, 60% identity with the bovine CPB sequence, 83% identity with the rat CPA1 sequence, 57% identity with the rat CPA2 sequence, and 57% identity with the rat CPB sequence. The presence of Ile at codon 255 confirms that the gene is CPA and not CPB; rat CPA1 and CPA2 and bovine CPA have Ile at position 255, but rat and bovine CPB have Asp at position 255.

Table 1

Maximum-Likelihood Estimates of $\hat{\theta}$ of CPA with Chromosome 7 Markers, $\hat{Z}$ and 1-Lod-Unit Confidence Intervals

Locus	Probe/Enzyme	$\hat{\theta}$	$\hat{Z}$	Confidence Interval
MET	METD/*TaqI*	.08	4.49	.015–.226
D7S8	pJ3.11/*MspI*	.12	6.82	.068–.210
D7S97	CRI-S158/*TaqI*	.06	5.40	.022–.118
D7S87	CRI-S94/*TaqI*	.08	6.78	.021–.186
D7S63	CRI-L1033/*MspI*	.12	3.28	.072–.200
D7S71	CRI-S2/*MspI*	.06	5.13	.024–.117
D7S93	CRI-S140/*MspI*	.03	7.32	.009–.086
D7S107	CRI-S201/*TaqI*	.03	7.34	.001–.120
D7S106	CRI-S199/*MspI*	.03	6.24	.001–.158
D7S91	CRI-S134/*HincII*	.08	4.39	.014–.221
D7S72	CRI-S3/*HindIII*	.12	4.74	.034–.269
D7S111	CRI-S241/*EcoRI*	.13	3.47	.050–.308
TCRβ	pJ2/*BglII*	.16	4.01	.094–.260

NOTE.—Only comparisons giving $\hat{Z} \geq 3.0$ are shown.

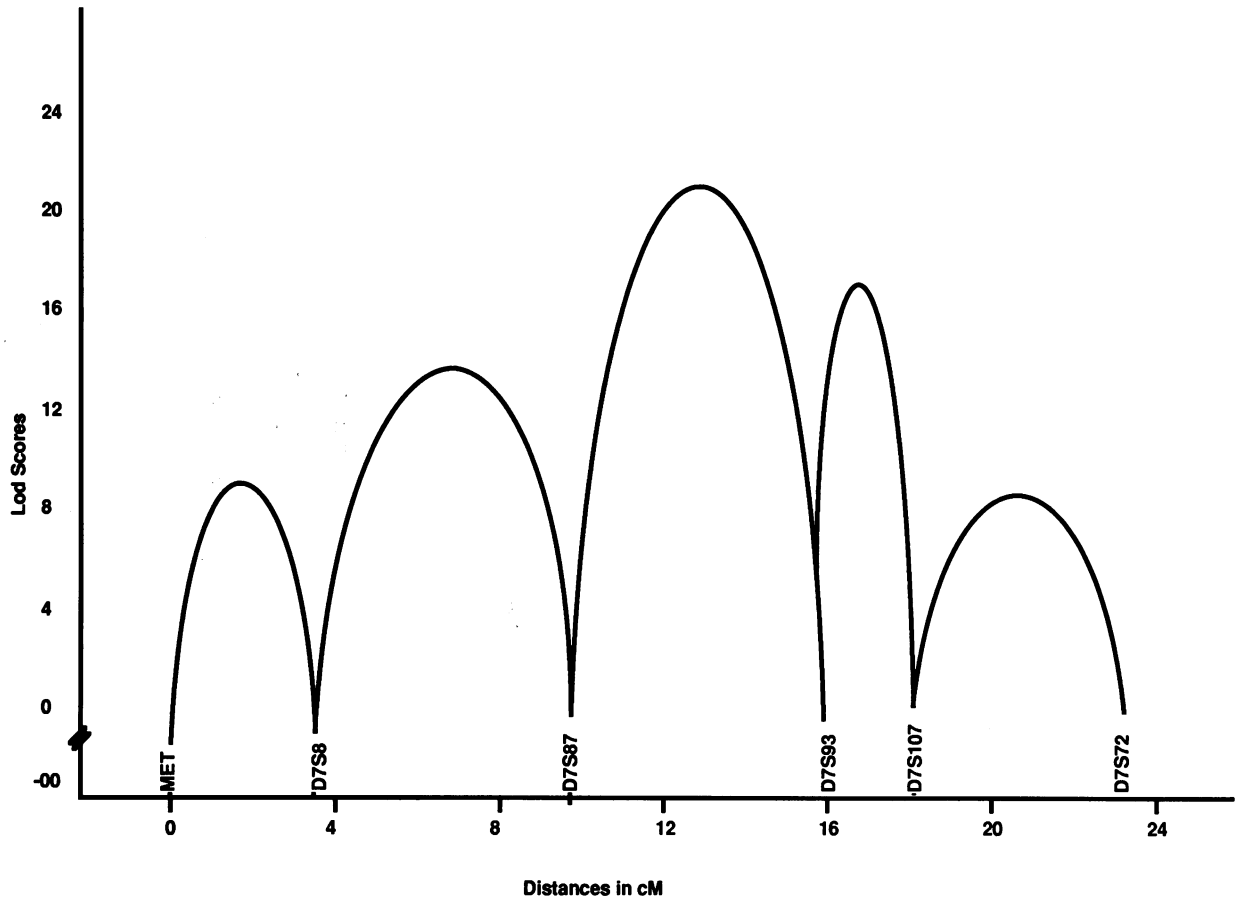

Figure 3 Support for the position of CPA with respect to chromosome 7 markers. The position of MET was arbitrarily set at 0, and the positions of the other loci were fixed according to the genetic map (Donis-Keller et al. 1987). The female/male distance ratio was set at 0.45.

Population Studies of CPA

Allelic frequencies are given in table 2. The polymorphism was in Hardy-Weinberg equilibrium in all populations studied. Allelic frequencies differed between populations; however, in all of the populations the + allele (presence of restriction site) with fragment sizes of 8.7 kb and 2.3 kb was the most common.

Discussion

In screening a human genomic library cloned into EMBL3, we isolated six clones that hybridize to rat CPA1 cDNA. On restriction analysis, these were shown to be identical. Sequence analysis of the genomic clone λHCPA showed that the regions that hybridized to rat CPA1 cDNA contained sequences from exons 9 and 10. Thus, only a small portion of the human gene has been isolated, but it was sufficient to allow confirmation that human carboxypeptidase has been cloned. The presence of isoleucine at the position identical to rat position 255 identified it as CPA. When rat cDNA was hybridized, at low stringency, to blots of human genomic digests (data not shown), a numbers of bands were observed that were not present in the genomic clone. These additional bands are likely to contain exons 1–8 and may also contain exons from additional forms of carboxypeptidase. Humans do have a B form of the gene (Geokas et al. 1975), and it is possible that the A form is present in two separate forms, A1 and A2 as in rat.

The *BglII* polymorphism mapped to the 3′ flanking region of the gene and is likely to be due to a base-pair substitution. It has a PIC of .33 (Botstein et al. 1980), which is not ideal in terms of informativeness but is sufficient to allow the gene to be localized on an established map of chromosome 7 (Donis-Keller et al. 1987) and to be positioned between the loci D7S87 and

```
        220
HCPA    gatcagctttccaaggctgctgtgacagccctggcctctctctacgggacc
HCPAA   D  Q  L  S  K  A  A  V  T  A  L  A  S  L  Y  G  T
RCPA1   D  Q  L  A  K  S  A  V  T  A  L  T  S  L  H  G  T
RCPA2   D  E  V  A  Q  K  A  A  Q  A  L  K  R  L  H  G  T
BCPA    N  Q  V  A  K  S  A  V  A  A  L  K  S  L  Y  G  T
RCPB    N  A  L  V  K  G  A  A  K  E  L  A  T  L  H  G  T
BCPB    N  T  L  A  K  G  A  V  K  K  L  A  S  L  H  G  T

HCPA    aagttcaactatggcagcatcatcaaggcaatttgtaagt
HCPAA   K  F  N  Y  G  S  I  I  K  A  I  Y
RCPA1   K  F  K  Y  G  S  I  I  D  T  I  Y
RCPA2   S  Y  K  V  G  P  I  C  S  V  I  Y
BCPA    S  Y  K  Y  G  S  I  I  T  T  I  Y
RCPB    K  Y  T  Y  G  P  G  A  T  T  I  Y
BCPB    T  Y  S  Y  G  P  G  A  T  T  I  Y

HCPA    tactccttcaccttcgagctccgggacactgggcgctatggcttcccgctg
HCPAA   Y  S  F  T  F  E  L  R  D  T  G  R  Y  G  F  L  V
RCPA1   Y  S  F  T  F  E  L  R  D  T  G  L  R  G  F  L  L
RCPA2   Y  S  F  A  F  E  L  R  D  T  A  F  Y  G  F  L  L
BCPA    Y  S  F  T  F  E  L  R  D  T  G  R  Y  G  F  L  L
RCP     Y  S  F  T  F  E  L  R  D  T  G  F  F  G  F  L  L
BCPB    Y  S  F  T  F  E  L  R  D  K  G  R  Y  G  F  V  L

        248
HCPA    gatcaagccagtggaagcactattgactggacctacagccagggcatcaag
HCPAA   Y  Q  A  S  G  S  T  I  D  W  T  Y  S  Q  G  I  K
RCPA1   Y  Q  A  S  G  S  T  I  D  W  T  Y  S  Q  G  I  K
RCPA2   Y  Q  A  S  G  G  S  I  D  W  A  Y  D  L  G  I  K
BCPA    Y  Q  A  S  G  G  S  I  D  W  S  Y  N  Q  G  I  K
RCPB    Y  P  A  A  G  G  S  D  D  W  S  Y  D  Q  G  I  K
BCPB    Y  P  A  S  G  G  S  D  D  W  A  Y  D  Q  G  I  K

                                                    286
HCPA    ccagcctcccagatc
HCPAA   P  A  S  Q  I
RCPA1   P  A  S  Q  I
RCPA2   P  A  K  Q  I
BCPA    P  A  S  Q  I
RCPB    P  E  S  Q  I
BCPB    P  E  S  Q  I
```

Figure 4 Nucleotide sequence of human CPA (HCPA), with comparison of the predicted amino acid sequences of human CPA (HCPAA), rat CPA1 (RCPA1), rat CPA2 (RCPA2), rat CPB (RCPB), and bovine CPA (BCPA) and CPB (BCPB). Numbers indicate the position of the amino acids according to the primary sequence of BCPA. Boldface letters indicate DNA sequences found within introns. Underlining indicates *Sau*3A cloning sites for HCPA. The tyrosine at position 248 is repeated because the codon (tat) is split at the cloning sites.

D7S107. This places CPA distal to the CF locus at a distance of 12 cM.

This polymorphism was preserved in all populations studied: pygmies from the Central African Republic and Zaire; caucasoids, Chinese, Japanese, and Melanesians. Heterozygosity ranged from .34 to .49; thus, for a two-allele system it was reasonably informative in all populations.

Table 2

Allelic Frequencies, Heterozygosity, and Hardy-Weinberg χ^2 Values for λ HCPA Alleles in Six Human Populations

Population	Allele Frequency		Heterozygosity	χ^2 with 1 df	No. of Chromosomes
	11.0 kb	8.7/2.3 kb			
Central African Republic pygmies	.43	.57	.48	.86	46
Zaire pygmies	.32	.68	.43	1.45	44
Caucasoids	.34	.66	.45	.01	70
Chinese	.31	.66	.43	.02	100
Japanese	.21	.79	.34	1.56	42
Melanesians	.43	.57	.49	.39	28

CPA is a digestive enzyme, and with the isolation of this gene it may now be possible to analyze the association of CPA with a number of digestive disorders.

Acknowledgments

We would like to acknowledge Luca Cavalli-Sforza for his support, Mary-Claire King for data on the cystic fibrosis families, and Judith R. Kidd for participating in the immortalization of cell lines for the Stanford-Yale collection of aboriginal populations. We wish to thank Laleh Daneshvar, Berta Young, and Joan Hebert for their help and encouragement and CEPH (Paris) for family DNA. This work was supported in part by NIH grant GM28428. C.S.C. is funded in part by NSF grant DMB 8608086.

References

Ardeshir F, Giulotto E, Zieg J, Brison O, Liao W, Stark G (1983) Structure of amplified DNA in different Syrian hamster cell lines resistant to N-(phosphonacetyl)-1-aspartate. Mol Cell Biol 3:2076–2088

Barrett A, McDonald J (1986) Mammalian proteases: a glossary and bibliography. Vol 2: Exopeptidases. Academic Press, London

Beaudet A, Bowcock A, Buchwald M, Cavalli-Sforza L, Farrall M, King M-C, Klinger K, et al (1986) Linkage of cystic fibrosis to two tightly linked DNA markers: joint report from a collaborative study. Am J Hum Genet 39:681–693

Benton W, Davis R (1977) Screening gt recombinant clones by hybridization to single plaques *in situ*. Science 196: 180–182

Botstein D, White R, Skolnick M, Davis R (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bowcock A, Bucci V, Hebert J, Kidd K, Friedlaender J, Cavalli-Sforza L (1987) Study of 47 DNA markers in five populations from four continents. Gene Geogr 1:47–64

Bowcock AM, Crandall J, Daneshvar L, Lee GM, Young B, Zunzunegui V, Craik C, et al (1986) Genetic analysis of cystic fibrosis: linkage of DNA and classical markers in multiplex families. Am J Hum Genet 39:699–706

Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 50:349–383

Cavalli-Sforza L, Kidd J, Kidd K, Bucci C, Bowcock A, Hewlett B, Friedlaender J (1987) DNA markers and genetic variation in the human species. CSH Quant Biol 51:411–417

Clauser E, Gardell S, Craik C, Macdonald R, Rutter W (1988) Structural characterization of the rat carboxypeptidase A1 and B genes: comparative analysis of the rat carboxypeptidase gene family. J Biol Chem 263:17837–17845

Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith T, et al (1987) A genetic linkage map of the human genome. Cell 51:319–337

Feder J, Yen L, Wijsman E, Wang L, Wilkins L, Schroder J, Spurr N, et al (1985) A systematic approach for detecting high-frequency restriction fragment length polymorphisms using large genomic probes. Am J Hum Gen 37: 635–649

Feinberg A, Voglestein B (1983) A technique for radiolabeling DNA restriction fragments to high specific activity. Anal Biochem 132:6–13

Gardell S, Craik C, Clauser E, Goldsmith E, Stewart C, Graf M, et al (1988) A novel rat carboxypeptidase CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. J Biol Chem 263:17828–17836

Geokas M, Largman C, Brodrick J, Raeburn S, Rinderknecht H (1975) Human pancreatic carboxypeptidase B. I. Isolation, purification, and characterization of fraction II. Biochim Biophys Acta 391:396–402

Honey N, Sakaguchi A, Lalley P, Quinto C, Rutter W, Naylor S (1986) Assignment of the gene for carboxypeptidase A to human chromosome 7q22-qter and to mouse chromosome 6. Hum Genet 72:27–31

Lathrop G, Lalouel J-M, Julier C, Ott J (1984) Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Proc Natl Acad Sci USA 81: 3443–3446

Quinto C, Quiroga M, Swain W, Nikovits W Jr, Stranding D, Pictet R, Valenzuela P, et al (1982) Rat precarboxypeptidase A: cDNA sequence and preliminary characterization of the gene. Proc Natl Acad Sci USA 79:31–35

Sanger F, Nicklen S, Coulson A (1977) DNA sequencing with the chain-termination inhibitors. Proc Natl Acad Sci USA 74:5463–5467