A p47-*phox* Pseudogene Carries the Most Common Mutation Causing p47-*phox*–deficient Chronic Granulomatous Disease

Agnes Görlach,* Pauline L. Lee,* Joachim Roesler,‡ Penelope J. Hopkins,§ Barbara Christensen,‖ Eric D. Green,¶ Stephen J. Chanock,‖ and John T. Curnutte‡*

**The Scripps Research Institute, Department of Molecular & Experimental Medicine, La Jolla, California 92037;* ‡*Genentech Inc., Department of Immunology, S. San Francisco, California 94080;* §*Sequana Therapeutics Inc., La Jolla, California 92037;* ‖*National Cancer Institute, Pediatric Branch, Bethesda, Maryland 20892; and* ¶*National Center for Human Genome Research, National Institutes of Health, Bethesda, Maryland 20892*

Abstract

The predominant genetic defect causing p47-*phox*–deficient chronic granulomatous disease (A47° CGD) is a GT deletion (ΔGT) at the beginning of exon 2. No explanation exists to account for the high incidence of this single mutation causing a rare disease in an unrelated, racially diverse population. In each of 34 consecutive unrelated normal individuals, both the normal and mutant ΔGT sequences were present in genomic DNA, suggesting that a p47-*phox* related sequence carrying ΔGT exists in the normal population. Screening of genomic bacteriophage and YAC libraries identified 13 p47-*phox* bacteriophage and 19 YAC clones. The GT deletion was found in 11 bacteriophage and 15 YAC clones. Only 5 exonic and 33 intronic differences distinguished all ΔGT clones from all wild-type clones. The most striking differences were a 30-bp deletion in intron 1 and a 20-bp duplication in intron 2. These results provide good evidence for the existence of at least one highly homologous p47-*phox* pseudogene containing the ΔGT mutation. The p47-*phox* gene and pseudogene(s) colocalize to chromosome 7q11.23. This close linkage, together with the presence within each gene of multiple recombination hot spots, suggests that the predominance of the ΔGT mutation in A47° CGD is caused by recombination events between the wild-type gene and the pseudogene(s). (*J. Clin. Invest.* 1997. 100: 1907–1918.) Key words: human chromosome 7 • neutrophils • gene conversion • NADPH oxidase • respiratory burst

Introduction

The phagocyte NADPH oxidase is a complex enzyme system that plays an important role in host defense. After stimulation with opsonized microorganisms or other activating agents, the oxygen consumption of these cells increases dramatically (respiratory burst) and they release a large amount of superoxide (1). Superoxide is then converted to more potent reactive oxygen species such as hydrogen peroxide, hydroxyl radical, and hypohalous acids, which are used by phagocytes to control microbial infections. The importance of this defense mechanism is made evident by a rare inherited syndrome, chronic granulomatous disease (CGD),[1] in which phagocytes fail to generate superoxide, rendering the patients highly susceptible to life-threatening microbial infections (2).

Characteristically, the components of the NADPH oxidase are found in different cellular compartments in the resting state. Two tightly linked membrane components, p22-*phox*[1] and gp91-*phox*, which form an unusual low potential b-type cytochrome (cytochrome b_{558}) contain the two redox centers of the oxidase (heme and flavin) as well as the NADPH-binding site (3–5). Upon proper stimulation, the cytosolic proteins p47-*phox* and p67-*phox* translocate to the membrane and associate with the cytochrome, thus allowing the electron transfer from NADPH to molecular oxygen (6–8). Both cytosolic components contain two SH3 (src homology region 3) domains that bind to proline-rich regions, and are probably sites of interaction between the NADPH oxidase components during activation (9, 10). In addition, a low molecular weight GTP-binding protein, Rac, is required for the function of the NADPH oxidase (11, 12). Recently, another cytosolic factor, p40-*phox*, has been identified (13, 14). The importance of this new factor for the function of NADPH oxidase, however, remains to be elucidated.

The genes encoding many of the NADPH oxidase components have been cloned, and their chromosomal localizations have been identified (13–20). Different forms of CGD are caused by mutations in the genes that encode p47-*phox*, p67-*phox*, p22-*phox*, and gp91-*phox* (for review see references 2, 21, and 22). Defects in the X-linked gp91-*phox* gene lead to the most common form of CGD. Various types of family-specific mutations have been detected in gp91-*phox*–deficient CGD, ranging from a decreased amount of protein and superoxide production to a total lack of protein and oxygen radical formation. Various genetic defects have also been described for the rare forms of p22-*phox* and p67-*phox*–deficient CGD. In contrast, studies from Europe, the United States, and Japan investigating mutations in the second most common form of CGD, the deficiency of p47-*phox* (A47°), reported the same mutation in 19 alleles in ten patients (23–25). At the beginning of exon 2 of this gene, a dinucleotide deletion within a GTGT repeat was found that predicts a frameshift and a premature stop codon at amino acid 51. One A47° CGD patient has been described as a compound heterozygote carrying the GT deletion on one allele, and a deletion of a G at bp 502 on the other

Address correspondence to John T. Curnutte, M.D., Ph.D., Department of Immunology, Mail Stop 34, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080. Phone: 650-225-8131; FAX: 650-225-8136; E-mail: curnutte.john@gene.com

Received for publication 4 October 1996 and accepted in revised form 23 June 1997.

The Journal of Clinical Investigation
Volume 100, Number 8, October 1997, 1907–1918
http://www.jci.org

1. *Abbreviations used in this paper:* CGD, chronic granulomatous disease; HMR, human minisatellite repeat; oxidese; YAC, yeast artificial chromosome.

allele (24). Considering that CGD affects about 1 in 500,000 individuals (21), and only 23% of these cases are caused by A47° CGD (2), it appears surprising that more than 90% of A47° CGD patients carry the same mutation, even though they are coming from unrelated and racially diverse populations. To date no explanation for the prevalence of the GT deletion in A47° CGD has been presented. It has been speculated that the high frequency of this mutation might be related to the presence of the dinucleotide repeat, which would lead to a tendency for the DNA strands to slip at this site, generating deletions during copying of this sequence by DNA polymerase (24, 25). It has also been hypothesized that the dinucleotide repeat could act as a site for unequal crossovers (23), such as in the case of one form of α-thalassaemia (hemoglobin H disease) where an AG deletion in an AG tandem repeat at the exon 1/intron 1 border causes a frameshift and a premature termination 25 amino acids further downstream (26).

Here we describe the structure and sequence of a previously unrecognized p47-*phox* pseudogene that carries the GT deletion identified in the majority of A47° CGD patients. This pseudogene colocalizes with the p47-*phox* gene at chromosome 7q11.23. Sequence analysis of the p47-*phox* gene and pseudogene showed a high degree of identity in exonic and intronic regions. The identification of multiple potential hot spots for recombination suggests that such events between the p47-*phox* wild-type gene and the pseudogene(s) might occur. It is proposed that recombination events might allow the transfer of the GT deletion from the pseudogene to the wild-type gene, resulting in mutant alleles. Such a mechanism might explain the unusually high number of p47-*phox*–deficient CGD patients carrying the ΔGT mutation.

Methods

Identification of genomic p47-phox clones. An EMBL3A bacteriophage library of human leukocyte DNA was screened with a full-length p47-*phox* cDNA probe spanning the entire coding region through the polyA tail. The cDNA probe was made by PCR with the primers ATGGGGGACACCTTCATCCGT and CACTCCAAGCAACATTTATTG. Positive clones were plaque-purified and identified by restriction digestion and Southern blot hybridization. Two out of three positive clones (L14 and L24) were characterized in detail in this analysis.

A human genomic P1 library (Genome Systems, St. Louis, MO) was screened by PCR amplification using two sets of primers. One set of primers (TTTTCCTTGTCCCTGCAGGT and GACTGGGTGGCCTCCAGTGCTCCCT) amplified a 212-bp fragment corresponding to the region immediately 5′ of the initiator methionine (termed STSA). Another set (AGACGCAGCGCTCTAAACCGCA and CTATAGAGCCTGGCGTCTGGA) amplified a 194-bp fragment including exon 11 and the flanking 3′ region of the p47-*phox* gene (termed STSB). Ten p47-*phox* P1 clones were identified with either primer pair. Six P1 clones (P38–P43) were investigated in detail in this study.

The inserts of the phage clones span about 15 kB as concluded from restriction analysis and direct sequencing. The phage clones extended ~ 2 kB from the start of translation. The average insert size of the P1 clones was about 80 kB.

Polymerase chain reaction. Genomic DNA was isolated from whole blood stored in EDTA using a DNA extractor (Applied Biosystems Inc., Foster City, CA). DNA from bacteriophage and P1 clones was isolated using standard protocols. Oligonucleotide primers were synthesized with a Model 394 DNA synthesizer (Applied Biosystems Inc.). Polymerase chain reaction was performed using a GeneAmp 9600 thermal cycler (Perkin Elmer Corp., Norwalk, CT). Typically 500–1,000 ng of genomic DNA (derived from healthy donors or p47-*phox*–deficient CGD patients), or 50–100 ng of cloned genomic DNA was amplified. Standard PCR reactions (50 μl) contained 2 μl 10 mM dNTP mix (dGTP was substituted by 1.5 mM 7-deaza-2′-dGTP and 0.5 mM dGTP) (Pharmacia LKB Biotechnology, Inc., Piscataway, NJ), 5 μl 10× PCR buffer (100 mM Tris-HCl, pH 8.3, 500 mM KCl, 1.5 mM $MgCl_2$) (Boehringer Mannheim Biochemicals, Indianapolis, IN) and 0.7 μl of each primer (100 ng/μl stock solution). DNA was denatured at 98°C for 3 min and kept at 80°C for the addition of 0.5 μl Taq polymerase (5,000 U/μl) (Boehringer Mannheim Biochemicals). Amplification conditions were as follows: 94°C for 1 min, 62°C for 1 min and 72°C for 2 min for 30 cycles, and a final extension for 7 min at 72°C. Total bacteriophage DNA was denatured at 72°C for 7 min before PCR. Longer fragments (> 2,000 bp) were amplified using the Expand Long Template PCR system (Boehringer Mannheim Biochemicals). After an initial denaturation step at 72°C for 7 min, 100 ng phage DNA was added to a 50-μl PCR reaction containing 5 μl 10× buffer I (500 mM Tris-HCl, pH 9.2, 160 mM $(NH_4)_2SO_4$, 17.5 mM $MgCl_2$), 1.75 μl of a 10 mM solution of each dNTP, and 1.5 μl of each primer (100 ng/μl stock solution). After an initial denaturation step at 94°C for 2 min, amplification was performed for 10 cycles at 94°C for 10 s, 60°C for 30 s, and 68°C for 2 min, followed by 15 cycles at 94°C for 10 s, 60°C for 30 s, and 68°C for 2 min with a cycle elongation of 20 s/cycle. A final extension was carried out at 68°C for 7 min. PCR fragments were purified using the Qiaquick PCR purification kit (QIAGEN Inc., Chatsworth, CA) and eluted with either H_2O or 10 mM Tris-HCl, pH 8.5. PCR products were analyzed on agarose gels to verify proper amplification.

Sequencing. Cycle sequencing of PCR-amplified DNA was performed using the fmol DNA cycle sequencing system (Promega Corp., Madison, WI). The sequencing primers (100 ng) were labeled in a 10-μl reaction containing 1 μl ^{32}P γ-ATP (6,000 Ci/mmol, Amersham Corp., Arlington Heights, IL), 1 μl 10× T4 polynucleotide kinase buffer (500 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine), and 1 μl T4 polynucleotide kinase (10 U) by incubation for 30 min at 37°C. The reaction was stopped at 80°C for 2 min. Labeled primers (1.5 μl) were mixed with 9.5 μl of PCR-amplified DNA (100–500 ng), 5 μl 5× sequencing buffer (250 mM Tris-HCl, pH 9, 10 mM $MgCl_2$) and 1 μl sequencing grade Taq polymerase (5,000 U/ml). Aliquots of this mixture (3.5 μl) were added to 2 μl of each d/ddNTP. After an initial denaturing step at 95°C for 1 min, 30 cycles were performed (95°C for 20 s, 42°C for 20 s, and 70°C for 30 s) followed by an extension at 72°C for 7 min. Stop solution (3.5 μl) provided with the kit was then added to each reaction. The samples were heated at 72°C for 2 min, immediately chilled on ice, and then loaded on an 8% denaturing polyacrylamide gel. Electrophoresis was performed at a constant power of 50 W in 1× Tris borate EDTA. The sequencing gels were vacuum dried and autoradiographed at −70°C overnight. Sequencing was also performed using the dye primer cycle sequencing method (27) with a Model 373A DNA sequencer (Applied Biosystems Inc.).

Reverse transcriptase-polymerase chain reaction (RT-PCR). Total RNA was isolated from lymphocytes/monocytes from healthy donors or from EBV-transformed B-cell lines derived from normal individuals (kindly provided by Dr. Ernest Beutler, The Scripps Research Institute) using the Trizol reagent (GIBCO BRL, Gaithersburg, MD) according to the manufacturer's instructions. RNA was dissolved in RNAse-free H_2O and stored at −70°C until used. RT-PCR was performed using the Superscript RT-PCR system (GIBCO BRL) according to the manufacturer's instructions. In a 12 μl reaction 1 μg RNA and 1 μl oligo(dT) (0.5 g/μl) were heated for 10 min at 70°C and immediately chilled on ice. The reaction was first preincubated for 5 min at 42°C with a mixture containing 2 μl of 10× buffer (200 mM Tris-HCl, pH 8.4, 500 mM KCl), 2 μl 25 mM $MgCl_2$, 1 μl dNTP mix (10 mM), and 2 μl of 0.1 M DTT. Superscript reverse transcriptase (200 U) was added, and the mixture was incubated at 42°C for 50 min and terminated by heating to 70°C for 15 min and chilling

on ice. *E. coli* RNaseH (2 U) was then added and incubated for 20 min at 37°C. Amplification was performed with 2 μl cDNA and exonic primers using the PCR conditions described above.

Restriction analysis. Restriction analysis with Bsp1407I (New England Biolabs Inc., Beverly, MA) was performed directly on 1 μg of bacteriophage DNA or on 500–1000 ng amplified DNA derived from cloned or genomic DNA. Restriction digestion was carried out in a 20-μl reaction containing 1.5 μl Bsp1407I (10 U/μl), 2 μl of the buffer supplied and 2 μl of bovine serum albumin (0.1 mg/ml) at 37°C overnight. Restriction digestion with DraIII (UUSSB Biologicals, Cleveland, OH) was performed with 5 μl of amplified cDNA in a 10-μl reaction containing 3 μl DraIII (3 U/μl), 1 μl of the buffer supplied, and 1 μl of BSA (0.1 mg/ml) for 2 h at 37°C. The reaction products were analyzed by electrophoresis on 3% Nusieve/1% agarose gels (FMC Bioproducts, Rockland, ME).

Ratio of pseudogene and wild-type gene copies. Five normal human DNA samples (50 ng) were used to amplify three genomic fragments encompassing the 5′ region of exon 2 using conditions as described above except that the primers were end-labeled with T4 polynucleotide kinase, and the reaction was amplified for only 25 cycles. The first primer pair I-1 (5′-TGCAATCCAGGACAACCGCAA-3′) and I-2RA (5′-TGTAATCAGAGAATCATGA-3′) amplified a 512-bp wild-type fragment and a 530-bp pseudogene fragment. The amplified products were separated on a 1.5% Trevigel 500 gel, visualized, and the bands were excised and counted in a beta scintillation counter. Alternatively, the amplified bands were digested with 15U of Bsp1407I which cuts only the wild-type PCR product into a 138-bp fragment and a 374-bp fragment. The second primer pair, I-1 and I-2R (5′-TGGAACTCGTAGATCTCG-3′), amplified a 216-bp or 218-bp fragment from the pseudogene and gene, respectively. Restriction digestion with Bsp1407I specifically cuts the wild-type fragment leading to two bands of 138-bp and 80-bp. The third primer pair, I-1 and I-2RB (5′-CTTCCCAAAGGGTGGAGCT-3′), amplified a 337- or 339-bp fragment from the pseudogene and gene, respectively. When digested with Bsp1407I, fragments of 138 bp and 201 bp were obtained from the wild-type amplification product.

Chromosomal localization. The chromosomal location of the p47-*phox* gene and related sequences were determined by amplification of DNA derived from a panel of human–hamster hybrid cell lines, each containing different human chromosomes (Corriell Institute for Medical Research, Camden, NJ). Two sets of primers were used: STSB, which amplifies exon 11 and the flanking 3′ region (see above), and STSC (TGCAATCCAGGACAACCGCAA and GCTCATGCCTGTAATCAGA) which amplifies a 451-bp fragment containing exon 2 and flanking intronic sequences. In addition, a set of yeast artificial chromosomes (YACs) highly enriched for human chromosome 7 DNA was also screened with the primer sets STSB and STSC as described previously (28). With this method a total of 19 positive YAC clones were identified.

Database analysis, sequence assembly, and editing. Assembly and analysis of DNA and protein sequences were performed with the software package of the University of Wisconsin Genetics Computer Group (29). Sequences were compared with the program FASTA (30).

Results

Sequence analysis of exon 2 in normal individuals and A47° CGD patients. In the course of mutation analysis of our series of A47° CGD patients, we also examined genomic DNA from normal donors. Exon 2 was amplified using the primers 2LB (CTTTCTGCAATCCAGGACAA), and 2RB (ATCACCTGGGCTAAGGTCCT), and the resulting DNA fragment was subjected to direct sequencing. Surprisingly, the (unrelated) control donors appeared to be heterozygous for the GT deletion frequently found in A47° CGD patients. To investigate further this unexpected presence of ΔGT in these normal individuals, we performed sequence analysis of exon 2 in genomic DNA from an additional 33 normals. Starting with the second GT of the GTGT tandem repeat at the beginning of exon 2, two different overlying sequences were observed in all of the normal individuals analyzed (Fig. 1). One was consistent with the published wild-type sequence containing the GTGT tandem repeat, while the other was identical to the sequence found in A47° CGD patients homozygous for the GT deletion. This result suggests that more than one p47-*phox*–related sequence might be present in normal genomic DNA.

To further confirm this finding, restriction digestion was performed with Bsp1407I, which cuts the wild-type sequence AGGTGTACA but cannot digest the ΔGT sequence (AGGTACA). No digestion was observed with amplified fragments containing exon 2 (using primers 2LB and 2RB) derived from ten A47° CGD patients homozygous for the GT deletion (Fig. 2 *a*). Bsp1407I digestion of the same fragment derived from a total of ten different normal subjects revealed three bands in all cases. One band was consistent with the undigested ΔGT-containing fragment, whereas the two smaller bands were consistent with the digestion products of the GTGT-containing fragment. This finding was confirmed by Bsp1407I digestion of the same amplification products derived from two genomic bacteriophage clones (to be further discussed below). One clone (L14) carried the GT deletion while the other clone (L24) carried the normal GTGT sequence. As expected, no digestion was observed in the fragment derived from the ΔGT clone L14, whereas the fragment from clone L24 carrying the GTGT sequence was completely digested into two fragments (Fig. 2 *a*).

Several explanations can be provided for this finding. First, all 34 normal individuals investigated here were carriers for A47° CGD. This is highly unlikely considering that the preva-

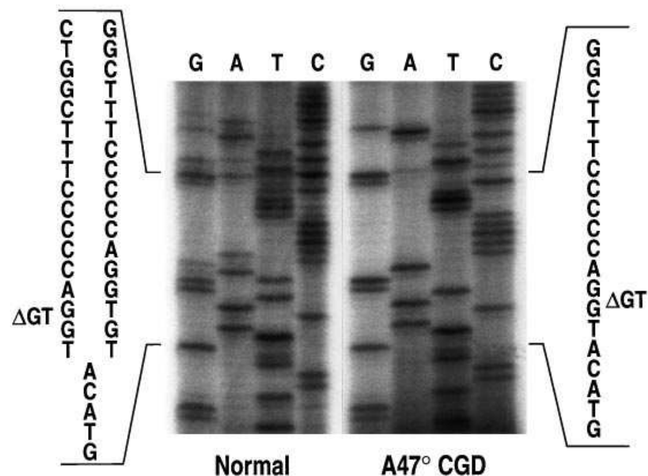

Figure 1. Sequence analysis of the intron 1/exon 2 border of p47-*phox*. Genomic DNA samples from a normal individual and a A47° CGD patient were amplified with primers 2LB (CTTTCTGCAATCCAGGACAA) and 2RB (ATCACCTGGGCTAAGGTCCT) and sequenced with the primer 2RB. The patient was homozygous for the GT deletion at the beginning of exon 2. The normal individual showed a sequence superimposed onto the wild-type sequence that was identical to that seen in a A47° CGD patient. This resulted in two overlying sequences starting with the beginning of exon 2. This sequence is representative of 34 normal individuals characterized and ten A47° CGD patients homozygous for the GT deletion.

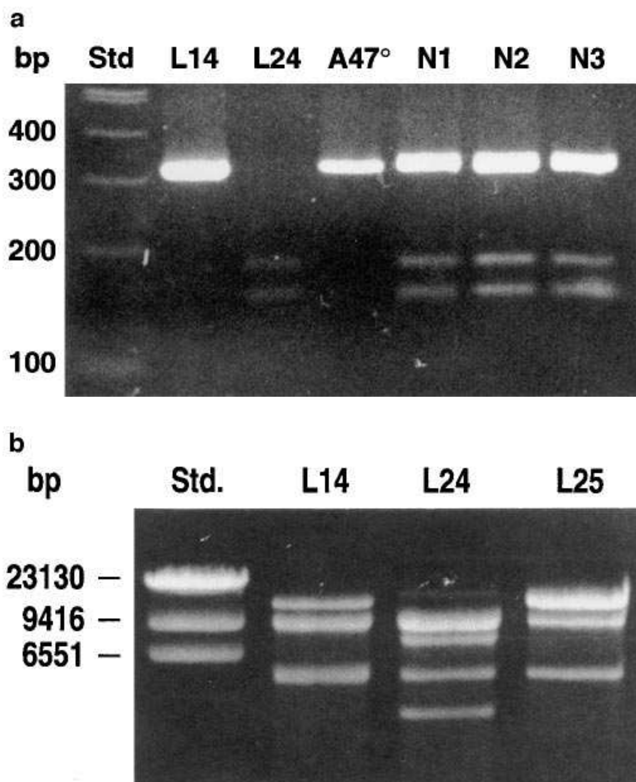

Figure 2. Restriction digestion with Bsp1407I. Restriction digestion was performed with Bsp1407I, which specifically cuts the wild-type sequence T/GTACA. (*a*) Genomic DNA from normal individuals (n = 10) and A47° CGD patients homozygous for the GT deletion (n = 10) was amplified with primers 2LB (CTTTCTGCAATCCAG-GACAA) and 2RB (ATCACCTGGGCTAAGGTCCT), and the amplification products (305 bp and 303 bp) were digested with Bsp1407I. Representative examples are shown here. The PCR product derived from a A47° patient was not digested, indicating that only sequence carrying the GT deletion was present. Bsp1407I digestion of PCR products derived from three normal individuals (N1–N3) revealed three DNA fragments: the fragments of 145 and 160 bp resulted from digestion of the wild-type product, and a 303-bp fragment corresponded to the uncut product. Bsp1407I digestion of the wild-type clone L24 showed complete digestion resulting in two fragments of 145 and 160 bp, whereas the fragment from the ΔGT clone L14 remained uncut. (*b*) Restriction digestion of unamplified bacteriophage clones L14 and L25 (pseudogene) as well as L24 (wild-type). The pseudogene clones were missing a restriction site, leaving the 12 kB fragment uncut, indicating that the GT deletion was not generated by a PCR or sequencing artifact.

lence of this autosomal recessive disease is probably about 1:2,000,000 (22) with an expected carrier frequency of approximately 1:700. Second, the GT deletion could have been consistently generated as an artifact by PCR amplification and/or the sequencing methods used. Third, a very similar gene (such as a pseudogene) carrying the GT deletion may be present in the normal population, and was therefore coamplified and cosequenced in these experiments.

Identification and characterization of p47-phox genomic clones. To examine further the presence of the GT deletion in the normal population, two different types of human genomic libraries (EMBL3A bacteriophage and P1) were screened with a p47-*phox* probe. Three bacteriophage clones and ten P1 clones were identified (Table I). Direct sequencing of exon 2 revealed that, of the three bacteriophage clones, two carried the GT deletion and one showed the normal GTGT sequence, thus giving a ΔGT:GTGT ratio of 2:1. In the ten P1 clones this ratio was 9:1. Since none of the clones showed both the GTGT and the ΔGT sequences, it appeared unlikely that the GT deletion was generated by PCR or sequencing artifacts. To confirm the presence of ΔGT by a method independent of PCR or sequencing, however, restriction digestion of unamplified bacteriophage clones was performed using Bsp1407I (Fig. 2 *b*). Whereas clone L24 contained four restriction fragments, only three fragments were observed in the clones L14 and L25, indicating that these two clones were lacking a restriction site compared to clone L24. This finding was consistent with the sequencing data indicating that L14 and L25 carried the GT deletion, whereas L24 showed the wild-type GTGT sequence. Thus, the GT deletion can be demonstrated in normal genomic DNA clones by a method independent of PCR. As the GT deletion predicts premature termination in the p47-phox protein, the most likely explanation for the large number of ΔGT clones obtained from the genomic DNA libraries is that there is at least one pseudogene. The identification of the ΔGT sequence in all normal subjects investigated would also be consistent with the presence of a ΔGT containing pseudogene.

Table I. Characterization of p47-phox Genomic Clones

Type of library	No. of clones	Exon 2	
		GTGT	ΔGT
EMBL3Aλ	3	1	2
P1	10	1	9
YAC	19	4	15
Σ	32	6	26

A total of 32 p47-*phox* genomic clones were isolated from two normal human genomic bacteriophage libraries (EMBL3Aλ and P1) as well as a normal human YAC library by screening with either hybridization using a p47-*phox* cDNA probe (EMBL3Aλ) or with PCR amplification using p47-*phox* primers (P1 and YAC). Sequence analysis of exon 2 of these clones revealed that only six clones showed the wild-type GTGT sequence at the beginning of exon 2, whereas 26 clones carried the GT deletion found in A47° CGD patients.

The high proportion of ΔGT-containing p47-*phox* clones identified suggested that more than one pseudogene might exist in the genome. Using three radiolabeled primer pairs encompassing the GTGT/ΔGT region of exon 2, quantitative PCR amplification was performed on genomic DNA from normal individuals. The amplified DNA was subsequently digested with Bsp1407I to distinguish between the gene (GTGT) and the pseudogene (ΔGT). Quantitative analysis of the corresponding fragments suggested a ratio of pseudogene to gene of 2:1 (Table II).

To characterize further the ΔGT containing p47-*phox* genomic clones, we analyzed two of the bacteriophage clones (L14 and L24) and six of the P1 clones (P38-P43) in greater detail. The six clones carrying the GT deletion (L14, P38, P39, P40, P41, P43) were designated p47-*phox* pseudogene clones, whereas the two clones carrying the GTGT sequence (L24, P42) were considered to be wild-type clones. The size and

Table II. Ratio of p47-phox Pseudogene and Wild-type Gene Copies

Primer pair	No. of DNA samples	No. of experiments	Ratio of pseudogene/wild-type
I-1 to I-2RA	5	2	2.2±0.5
I-1 to I-2R	5	4	2.1±0.3
I-1 to I-2RB	5	1	2.3

Genomic DNA samples were amplified with three different primer pairs encompassing the 5′ region of exon 2 as described in Methods. PCR products were either separated directly or after Bsp1407I digestion by agarose gel electrophoresis. Fragments derived from either the wild-type or pseudogene were quantitatively analyzed, and the ratios between pseudogene and wild-type fragments were calculated.

structure of the two wild-type and six pseudogene clones were determined by PCR amplification and sequence analysis using exonic primers. The wild-type clone P42 was found to be full-length containing 11 exons and 10 introns, consistent with the gene structure of the p47-*phox* gene (Chanock, et al., manuscript submitted for publication). The gene structures of the pseudogene clones P40, P41, and P43 were virtually identical to the wild-type clone in that they contained all 11 exons and ten introns (Fig. 3). All exon–intron borders followed the GT/AG rule (data not shown) (31). Each of the remaining four clones (L14, L24, P38 and P39) contained sequence from exon 1 to exon 8.

Comparison of the exonic regions of the p47-phox gene and pseudogene clones. The exonic regions of the six pseudogene clones were amplified with intronic primers, sequenced, and compared to the corresponding regions of both wild-type clones as well as to genomic DNA from a series of healthy donors. A total of nine single base pair substitutions, in addition to the GT deletion in exon 2, were found in the exonic regions of the pseudogene clones (Table III). Single base pair changes in exons 4, 6, and 9 (nucleotide positions 269, 558, 825, and 861) segregated between wild-type and pseudogene clones; i.e., all wild-type clones differed from all pseudogene clones at these four positions. For example, the wild-type clones L24 and P42 showed a G at position 269 in concordance with the published sequence, whereas all six pseudogene clones carried an A at this position (Table III). This G→A substitution predicts an Arg90→His exchange. At position 558, the pseudogene clones showed a G compared to an A in the wild-type clones. Interestingly, the identity of the nucleotide at this position in cDNA has been ambiguous in the literature, as either G or A has been seen by different investigators (17, 18, 32). If translated, this G→A transition would remain silent, as valine would be incorporated in both cases. Two segregating differences were identified in exon 9. Since clones L14, L24, P38, and P39 did not contain exons 9, 10, and 11 (see above), only one wild-type clone (P42) and three pseudogene clones (P40, P41, P43) could be analyzed in this region. At bp 825, a C was seen in all pseudogene clones compared to a T in the wild-type clone. At the cDNA level, this nucleotide position has also been questionable as both C and T have been reported (17, 18, 32). The C was considered to be a silent polymorphism (33). At bp 861, all pseudogene clones carried a T, while the wild-type clone contained a G. The published cDNA sequence shows a G at this position. Both nucleotide substitutions in exon 9 would also remain silent if translated.

In contrast to the nucleotide differences just described that segregated neatly between wild-type and pseudogene clones, there were five nucleotide positions in which only a subset of the pseudogene clones differed in their sequence from the other clones and the published wild-type sequence (Table III). At bp 295, for example, a G was found in two pseudogene clones (P40 and P41), whereas all other wild-type and pseudogene clones had an A at this position as in the published sequence. Such a nucleotide substitution would predict a Ser99→Gly exchange. A similar situation was found at bp 345. The pseudogene clones L14 and P43 contained a T at this position instead of a C, as seen in all of the other clones and in the published sequence. At bp 387, only one pseudogene clone (L14) showed an A for a G. The nucleotide at this position has also been questioned, as two reports described a G at this position in the normal cDNA (18, 32), whereas another study reported an A (17). Eventually, the G was considered to be a silent polymorphism (33). The pseudogene clone L14 also carried a G at bp 496, whereas all other clones showed an A. This nucleotide position remains ambiguous in the literature (17, 18, 32). In the corrected p47-*phox* consensus sequence (33), G was reported at this position. The A was considered to be a nonconservative polymorphism. If translated, the G would code for asparagine, whereas the A would code for aspartic acid. At bp 765, a C to A transversion was seen in five pseudogene clones, whereas one pseudogene clone (L14) and

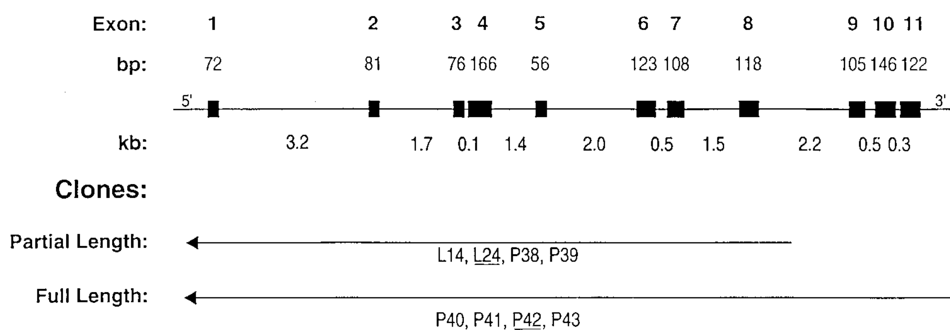

Figure 3. Structural organization of the human p47-*phox* gene. Exons are represented by boxes and numbered from 1 to 11; the size of each exon is given in basepairs. The sizes of the ten introns are given in kB. The arrows at the bottom indicate the clones investigated. The phage clones L14 and L24 and the P1 clones P38 and P39 were only partial length comprising sequence from the 5′ region to exon 8. The P1 clones P40 to P43 were full-length clones. Clones L24 and P42 (*underlined*) were wild-type; all other clones were pseudogene clones.

Table III. Exonic Differences Between p47-phox Wild-type and Pseudogene Clones

Exon	Nucleotide position	cDNA*	Wild-type clones		Pseudogene clones						Predicted AA‡		Normal gDNA§
			L24	P42‖	L14	P38	P39	P40	P41	P43	Wt	Ps	No. wt + ps/No. total
2	73/74	GTGT[1–4]	GTGT	GTGT	ΔGT	ΔGT	ΔGT	ΔGT	ΔGT	ΔGT	V[25]	V	34/34
4	269	G[1–4]	G	G	A	A	A	A	A	A	R[90]	H	12/14
	295	A[1–4]	A	A	A	A	A	G	G	A	S[99]	G	17/17
	345	C[1–4]	C	C	T	C	C	C	C	T	L[115]	L	17/17
	387	A[1,2]/G[3,4]	G	G	A	G	G	G	G	G	T[129]	T	4/4
6	496	G[1,2,4]/A[3]	A	A	G	A	A	A	A	A	N	D[166]	5/7
	558	G[1,2,4]/A[3]	A	A	G	G	G	G	G	G	V[186]	V	6/7
8	765	C[1–4]	C	C	C	A	A	A	A	A	V[255]	V	4/4
9	825	T[1,2,4]/C[3]	N/A	T	N/A	N/A	N/A	C	C	C	F[275]	F	6/6
	861	G[1–4]	N/A	G	N/A	N/A	N/A	T	T	T	V[287]	V	2/2

Differences in the exonic regions between p47-*phox* gene and pseudogene clones. All eleven exons of the two p47-*phox* wild-type clones (L24 and P42) and six pseudogene clones (L14, P38, P39, P40, P41, and P43) were sequenced. Clones L24, L14, P38, and P39 did not contain exons 9, 10, and 11 (N/A, not available). Sequence differences between wild-type and pseudogene clones were compared to the published cDNA sequences (third column). Direct sequence analysis was also performed on normal genomic DNA samples (last column). The data in this column indicate the number of samples found with both wild-type (wt) and pseudogene (ps) sequences compared to the total number of samples investigated at this nucleotide position. The nucleotide numbering starts with the first nucleotide of the initiator methionine. *Published p47-*phox* cDNA sequence (see references 1–4 below in legend). ‡Predicted amino acid (AA) for wild-type sequence (Wt) and pseudogene sequence (Ps). §No. of genomic DNA samples with wild-type and pseudogene sequence versus total number of genomic DNA samples. ‖The nucleotide sequences of the following clones have been submitted to the GenBank database: wild-type clone P42 (accession no. U57833-35); pseudogene clones L14 (U61224-25); P38 (U69639-43); P39 (U72356-61); P40 (U61238-40); P41 (U61241-44); and P43 (U60970-72). [1]See reference 33. [2]See reference 17. [3]See reference 18. [4]See reference 32.

the wild-type clones carried the published sequence. When viewed together, these segregating and nonsegregating nucleotide differences constitute a 0.5% to 0.75% divergence from the wild-type exonic sequence.

All ten of the exonic pseudogene differences (the ΔGT and the nine single base pair substitutions) were also seen in normal genomic DNA (Table III). The pseudogene nucleotides were found to be superimposed on the wild-type sequence in the majority of genomic DNA samples, indicating that both the wild-type and pseudogene were coamplified and cosequenced with the primers used. Finally, at two additional nucleotide positions (bp 849 and bp 936), all of the wild-type and pseudogene clones investigated contained the same sequence, but one that differed from the published cDNA sequence (33) (G [published]→A at bp 849, T→C at bp 936). Interestingly, at both positions, the sequences found in all of the clones have been reported as silent polymorphisms (33). In all of the normal genomic DNA samples investigated (five and six, respectively), bands for both nucleotides were observed (data not shown).

Comparison of the intronic regions of the p47-phox gene and pseudogene. Since comparison of the wild-type and pseudogene clones in the exonic regions indicated a remarkable degree of identity, we examined the intronic regions to identify additional distinguishing differences. Of the ∼ 13.5 kB of total intronic region present in the p47-*phox* gene (and pseudogene), more than 80% was sequenced in at least one wild-type clone, and in three pseudogene clones.

Interestingly, the major differences between the wild-type gene and pseudogene clones were identified in the regions flanking exon 2. At a position 580 bp upstream from the 5′ end of exon 2, a 30-bp duplication was found in all wild-type clones. All pseudogene clones, however, had only a single copy of this sequence. This 30-bp segment contained several tandem repeats (TCCCCTCCCCTCTCCTCCTGTCCCCTCCCT), and was incorporated within a larger 90-bp sequence that had a high content of C (62 bp) and was flanked on each site by an Alu repeat (see below). Amplification of DNA with primers Int1del (GTTTCACCATATTGGTCAGGCT) and 1RA (GTTGTCCTGGATTGCAGAAA) showed that a single 550-bp fragment was derived from the pseudogene clones, while two products were obtained from normal genomic DNA ($n = 4$) that were ∼ 550 and 580-bp in length (data not shown). All pseudogene clones tested also contained a 20-bp duplication in intron 2 (CAGGGTCTTGCTCTGTCACC), beginning 176 bp downstream from the 3′ end of exon 2; wild-type clones had only one copy of this sequence. Amplification of DNA with primers Int2.1 (GTTCCAGCTCCACCCTTTGGAA) and Int2dup (CAAAACCACCTAAAAGGCCGA) revealed a 207-bp fragment from the wild-type clones, a 227-bp fragment from the pseudogene clones, and both forms from normal genomic DNA. The duplication was also seen as an overlying sequence in all 34 normal genomic DNA samples investigated (data not shown).

Within the 11.1 kb of intronic regions analyzed, a total of 33 segregating differences, (including the intron 1 deletion and intron 2 duplication described above), between the wild-type and pseudogene clones were observed (Table IV). Two small deletions of 2 and 3 bp were found segregating in intron 1 as well as a CG→TG exchange 123 bp upstream from the 5′ end of exon 2. Both C and T were observed as overlying sequences in each of 34 normal genomic DNA samples analyzed. Only one difference was identified in intron 3, a segregating C→A exchange 64 bp downstream from the 3′ end of exon 3. Again,

Table IV. Intronic Differences Between Wild-type and Pseudogene p47-phox Clones

			Segregating differences				Nonsegregating differences			
Intron	Size	% sequenced	Total no.	Δbp	Insertions	Deletions	Total no.	Δbp	Insertions	Deletions
1	3194 bp	100	14	10	1	3	26	25	1	0
2	1733 bp	100	4	3	1	0	27	27	0	0
3	99 bp	100	1	1	0	0	0	0	0	0
4	1359 bp	100	0	0	0	0	9	8	0	1
5	2 kB	50	2	2	0	0	12	12	0	0
6	462 bp	100	0	0	0	0	8	8	0	0
7	1549 bp	100	5	4	1	0	14	14	0	0
8	2.2 kB	40	5	5	0	0	8	8	0	0
9	471 bp	100	0	0	0	0	1	1	0	0
10	334 bp	100	2	2	0	0	0	0	0	0
Σ	13.4 kB	83	33	27	3	3	107	105	1	1

87% of the intronic regions were sequenced in at lease one wild-type and three pseudogene clones. In the second column to the left, the size of the intron indicated in column 1 is given. Introns that have not been sequenced to completion are sized by their PCR products visualized on an agarose gel. The total number of segregating or nonsegregating differences (total no.) is given for each intron, as well as the number of single base pair changes (Δbp), insertions, and deletions.

both sequences were observed in normal genomic DNA (n = 15). This intron, which is present in wild-type and pseudogene clones, is of some potential interest since it contains 99 bp lacking a termination codon. A predicted translation product would be in frame with the flanking exons. The total number of segregating and nonsegregating intronic differences revealed a range of divergence from the wild-type sequence between 0.4 and 0.95% (Table V). The presence of nonsegregating differences in both the exonic and intronic regions could be due to allelic polymorphisms because of the lack of selective pressure or, alternatively, could represent differences between the two pseudogenes.

Sequence analysis of the 5′ upstream region. The high homology between the p47-*phox* gene and its pseudogenes raised the question as to whether the degree of conservation extended into the promoter region, thereby suggesting that the pseudogenes may be transcriptionally active. Therefore, the 5′ region of the wild-type clone L24 and the pseudogene clone L14 were amplified with an exon 1 antisense primer (Ex1-: TAGTGCTGGCTGGGTACGAAG), and the EMBL3A vector primers GAGTCTTGCAGACAAACTGCGCAA (for amplification of L14) or CTCGTCCGAGAATAACGAGTGGAT (for amplification of L24). Comparison of 874 bp of 5′ upstream region revealed an insertion of 2 nucleotides (AA) at −816 bp in the pseudogene clone. This difference is part of an A repeat flanking an Alu sequence that is located 820–570 bp upstream from the initiator methionine, and was found to be 76% identical to the Alu consensus sequence II (34).

Table V. Intronic Sequence Variations Between Genomic Bacteriophage Clones

Clone	No. base pair sequenced	L24	P42	L14	P38	P39	P40	P41	P43
		9341	11597	9341	5479	3230	11597	8801	11597
L24	9341	—	2	43	32	28	45	32	45
P42	11597	0.02	—	43	32	28	54	46	60
L14	9341	0.46	0.46	—	11	22	32	28	32
P38	5479	0.58	0.58	0.2	—	11	13	10	9
P39	3230	0.87	0.87	0.68	0.34	—	22	20	19
P40	11597	0.48	0.47	0.34	0.24	0.68	—	24	42
P41	8801	0.34	0.52	0.32	0.18	0.62	0.27	—	32
P43	11597	0.48	0.52	0.34	0.16	0.59	0.36	0.34	—

The total number of sequence differences between individual clones is given in the upper right half of the matrix above the diagonal axis. The part below the diagonal axis shows the degree of divergence between two clones in percent of divergence (no. of differences between two clones divided by number of basepairs sequenced in both clones). The second column and row show the number of basepairs sequenced in each clone.

Expression of the pseudogene(s). The virtual identity of the 5′ upstream region in the p47-*phox* gene and pseudogene clones suggested that the pseudogene(s) might be transcriptionally active. This was supported by the finding that many previously reported sequence discrepancies at the cDNA level were also observed in some or all of the genomic pseudogene clones.

Therefore, RNA from lymphocytes/monocytes (n = 7) or from EBV-transformed B-cell lines derived from healthy donors (n = 4) was reverse transcribed, the cDNA amplified with exonic primers (Ex1+: CACCTTCATCCCGTCACATCGCC, Ex3−: GATCCTGTTCTCTGGATTGA) and sequenced. At nucleotide position 75 (corresponding with the second GT of the tandem repeat at the beginning of exon 2) an overlying sequence identical to the pseudogene sequence carrying the GT deletion was observed (Fig. 4 *a*). The presence of ΔGT in cDNA was also confirmed independently by restriction digestion of these RT-PCR products with DraIII, which digests only wild-type cDNA. As shown in Fig. 4 *b*, three bands of 207, 147, and 60 bp were observed in normal RT-PCR samples, whereas an amplification product derived from a p47-*phox* wild-type cDNA clone with known wild-type GTGT sequence had only two digestion fragments of 147 and 60 bp.

Chromosomal location of the p47-phox gene and pseudogenes. The p47-*phox* gene has been previously mapped to

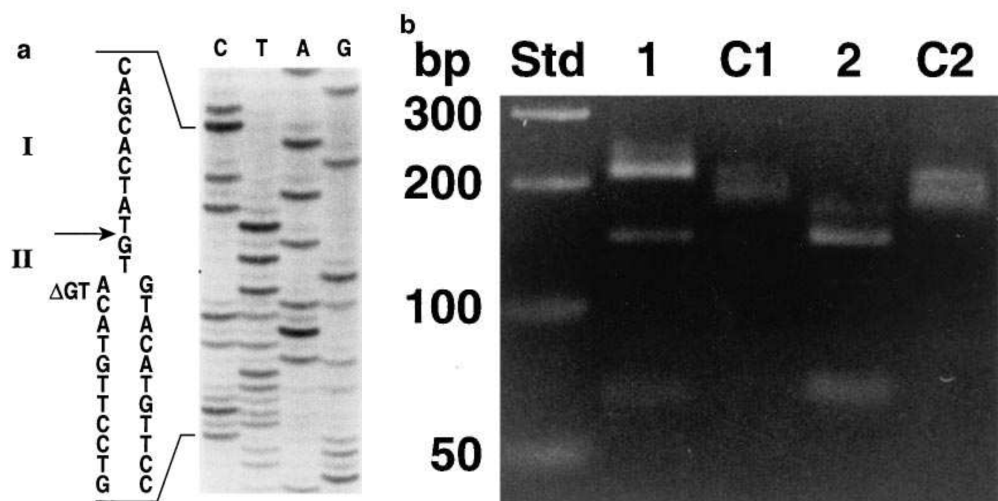

Figure 4. p47-*phox* cDNA analysis. (*a*) cDNA derived from normal lymphocytes/monocytes ($n = 7$) or normal EBV-transformed B-cell lines ($n = 4$) was PCR-amplified (Ex1+: CACCTTCATCCCGTCA-CATCGCC, Ex3−: GATCCT-GTTCTCTGGATTGA) and sequenced with primer Ex1+. Beginning with exon 2, a sequence containing the GT deletion was superimposed on the wild-type sequence. (*b*) PCR amplification (Ex1+/Ex3−) was performed with normal cDNA and a p47-*phox* cDNA clone with known wild-type sequence. The 207-bp products were digested with DraIII, which only cuts the wild-type sequence at CACNNN/GTG. The normal cDNA fragment (*1*) showed three bands: 207, 60, and 147 bp; indicating the presence of the GT deletion in the undigested cDNA fragment. The PCR product of the wild-type cDNA clone was digested to completion resulting in only two bands: 60 and 147 bp (*2*). Lanes C1 and C2 show the undigested PCR products.

chromosome 7q11.23 by Southern analysis of somatic cell hybrid lines, and by chromosomal in situ hybridization (35). To determine the chromosomal location of the pseudogenes, two approaches were used. First, DNA from a panel of human–hamster hybrid cell lines, each containing different human chromosomes, was amplified with two sets of primers: STSB, which generates a fragment containing exon 11 and the flanking 3′ end, and STSC, which amplifies exon 2 and the flanking intronic regions. This latter primer set would be able to amplify the region of the pseudogene, which contains the GT deletion and the 20-bp duplication. Appropriate PCR products were generated only from cell lines containing human chromosome 7, including two cell lines with chromosome 7 as their only human DNA. In hamster cell lines or cell lines that did not contain chromosome 7, no amplification could be observed, suggesting that only chromosome 7 contains p47-*phox* and p47-*phox*–related genes (data not shown).

In a second approach to determine the chromosomal location of the p47-*phox* pseudogenes, a YAC library highly enriched for chromosome 7 was screened using the same PCR assays (STSB and STSC) as described above. With this method, 19 p47-*phox*–positive YACs were identified. Sequence analysis of the fragment amplified with the primer set STSC, which contains exon 2 and flanking intronic regions, showed that 15 clones carried the GT deletion at the beginning of exon 2 and the 20-bp duplication in intron 2. The four remaining clones had the GTGT sequence at the beginning of exon 2. Since all 19 YACs have been localized to a single contig that maps to chromosomal band 7q11.23 (Green, E.D., unpublished data), the p47-*phox* wild-type and pseudogenes very likely colocalize to the same chromosomal region within 7q11.23.

Repetitive sequence elements and recombination hot spots. Sequence analysis of the p47-*phox* gene and pseudogene clones revealed a high number of repetitive sequence elements. Within the approximately 11.1 kB of intronic sequence obtained from wild-type and pseudogene clones, 17 complete Alu repeats and one Alu half site were found that had at least 70% identity with the Alu consensus sequence II (34) (Fig. 5). Most of the Alu motifs were flanked by short repeats. Intron 1 contained a cluster of six Alu sequences accounting for more than 50% of the total sequence of this intron. This intron also contained a large number of direct repeats. The entire intron (3194 bp) contained 94 repeats in a range between 10 and 42 bp, comprising a total of 1250 bp. Thus, more than 75% of intron 1 was represented by nonunique sequence. The 20-bp duplicated sequence in intron 2 characteristic for the pseudogene (described above) was repeated one other time in the same intron 600 bp further downstream in all of the wild-type and pseudogene clones analyzed. This intron (1733 bp) also contained 29 other repeats that ranged between 10 and 37 bp, and accounted for more than 400 bp. Of the remaining introns, only intron 10 showed a substantial number of repeated sequences that represented $\sim$ 45% of this intron. The largest segregating differences between wild-type and pseudogene clones involved repeated sequence elements.

The high number of sequence repeats suggested that the p47-*phox* gene might be susceptible to recombination events. We therefore searched for sequence motifs of potential recombination hot spots such as the Chi sequence (5′GCTGGTGG) (36) and the human minisatellite repeat (HMR) (5′GGG-CAGGAXG) (37). These motifs have been associated with recombination events between homologous genes. Three sequences that were identical to the Chi motif were found in intron 1 and exons 8 and 10 (the exon 10 sequence was in the antisense direction) (Fig. 5). Similarly, one sequence identical to the HMR sequence was identified in intron 10. There were an additional 12 Chi and 17 HMR motifs in the p47-*phox* gene that differed from the consensus sequences by only a single nucleotide. When viewed together, these potential recombination hot spots formed clusters located in introns 2, 4, and 7 (Fig. 5). Only introns 3 and 6 did not contain consensus motifs to either the Chi or HMR sequence.

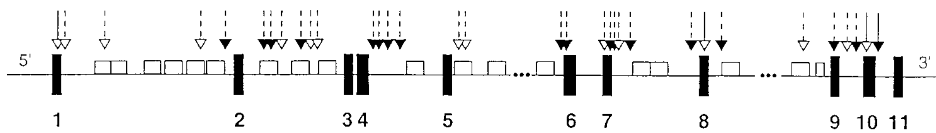

Figure 5. Localization of recombination hot spots in the p47-*phox* gene. Consensus sequences to the Chi sequence (5′GCTGGTGG) (*open arrowhead*) and the HMR (5′GGGCAGGAXG) (*full arrowhead*) are marked with *full arrowtails*; modified consensus motifs allowing one mismatch are indicated by *broken arrowtails.* Furthermore, the localization of Alu consensus sequences is indicated by *open boxes.* The numbers indicate the exons. The *dotted lines* in introns 5 and 8 indicate the size of sequence gaps.

Discussion

The original sequence analyses of the p47-*phox* gene revealed a considerable number of discrepancies between different cDNA clones and an apparently high number of polymorphisms (17, 18, 32, 33). Furthermore, mutation analysis of A47° CGD patients has shown a prevalence of a single mutation, a GT deletion at the beginning of exon 2 (23–25). In a recent evaluation of our A47° CGD patients, we found 20 out of 23 patients homozygous for this GT deletion (38). The prevalence of a single mutation in an unrelated, racially diverse population is in sharp contrast to the situation in the other three forms of CGD, where various family-specific mutations are found (for review see references 2 and 21). In this study, we show that the GT deletion is not only characteristic of A47° CGD, but is present in normal genomic DNA as well. Our data strongly suggest the presence of at least one pseudogene carrying the GT deletion found in A47° CGD patients. First, in all normal individuals investigated, the GT deletion could be identified at the genomic level, appearing as a superimposed sequence on that of the wild-type sequence. Second, screening of three different types of genomic libraries for p47-*phox* revealed (based on the size of each library and the corresponding coverage provided) an approximately two to three times higher number of positive clones than one would expect for a single copy gene. Third, only two bacteriophage clones and four YACs showed the wild-type sequence in exon 2, whereas eleven bacteriophage clones and 15 YAC clones carried the GT deletion at the beginning of exon 2. Sequence analysis of the clones demonstrated that the GT deletion was not introduced artificially by PCR or the sequencing methods used. The data strongly support the existence of at least one p47-*phox* pseudogene which carries the GT deletion in exon 2.

Characterization of two wild-type clones and six pseudogene clones showed a remarkable degree of identity between the p47-*phox* wild-type and pseudogene(s). We were able to demonstrate that the pseudogene clones analyzed in this study had the same gene structure as the p47-*phox* wild-type gene (Chanock et al., manuscript submitted for publication). In the exonic regions, four sequence differences (in addition to the GT deletion) were identified in all pseudogene clones, whereas at five other nucleotide positions, only a subset of pseudogene clones showed sequence differences. The divergence in the exonic regions between the wild-type clones and individual pseudogene clones ranged from 0.5 to 0.75%. Interestingly, all five of the nucleotide substitutions that varied among pseudogene clones were also observed as superimposed sequences in the majority of normal genomic DNA samples. This suggests that these differences represent either polymorphisms within a single pseudogene or the presence of more than one pseudogene. The latter would be consistent with the ratio of 26:6 pseudogene to wild-type clones that we observed in three different types of genomic libraries. Preliminary results using quantitative PCR suggest that there are two p47-*phox* pseudogenes present in the genome. Since PCR amplification is subject to technical artefacts, including the possibility that not all the pseudogenes may be amplified by the selected primer pairs, stretch FISH analysis is being performed to determine the actual number of p47-*phox* genes in the genome and to estimate the physical distance between them.

Whereas several of the exonic sequence differences seen in the pseudogene clones have been reported previously at the cDNA level (17, 18, 32, 33), the polymorphisms described at bp 555 (32) and at bp 621 (33) were not observed in either wild-type or pseudogene clones, suggesting the presence of true polymorphic sites in the wild-type sequence. Alternatively, it is possible that these nucleotide substitutions may be present on pseudogene isoforms that have not yet been identified and sequenced.

Our data show that the pseudogene(s) are transcribed in normal individuals. The pseudogene cDNA would predict a translation product of 50 amino acids, of which only 20 amino acids would be identical to the corresponding region of the normal p47-*phox* protein. Whereas truncated proteins derived from the translation of pseudogenes have been observed (e.g., the leukocyte interferon pseudogenes [39]), the presence of a p47-*phox* pseudogene polypeptide has not been demonstrated. It is possible that such a truncated protein is unstable and is degraded rapidly.

Similar to the findings with the exonic sequences, the intronic regions of the pseudogene clones were strikingly conserved in relation to the wild-type gene. Of the 83% of intronic sequence analyzed, only 33 segregating and 107 nonsegregating differences between wild-type and pseudogene(s) were identified, indicating a divergence of just 0.4% to 0.9%. The most striking differences were found in the flanking regions of exon 2: a 30-bp deletion in intron 1 and a 20-bp duplication in intron 2. Together with the GT deletion, these two differences serve as useful markers for the pseudogenes. The overall degree of divergence (exons plus introns) from the wild-type clones was found to be in the range 0.4% to 0.8%, thus rendering the p47-*phox* pseudogenes among the most conserved unprocessed pseudogenes known. Examples of other highly homologous unprocessed pseudogenes are the 21-hydroxylase pseudogene (CYP21A) (98% homology in exons and 96% homology in introns) (40), the von Willebrand factor pseudogene (96.9% overall homology) (41) and the β-glucocerebrosidase pseudogene (96% overall homology) (42).

The high degree of homology between the p47-*phox* gene and its pseudogenes suggests that the pseudogenes arose from gene duplication, as has been described for other unprocessed pseudogenes such as the 21-hydroxylase pseudogene (40) and

the α-globin gene (43). The p47-*phox* gene maps to chromosome 7q11.23 (35), and using cell-hybrid studies we have also localized the pseudogene to chromosome 7. Furthermore, all YACs identified as being positive for either the wild-type or the pseudogenes map to chromosomal band 7q11.23, indicating that both are confined to this region of chromosome 7. Interestingly, the q11.23 band of chromosome 7 appears to contain large block(s) of duplicated (i.e., nonunique) DNA segments (E.D. Green, unpublished observation).

The high level of sequence homology between p47-*phox* wild-type and pseudogenes together with the colocalization of these genes to the same chromosomal band suggests that these genes might be susceptible to recombination events such as gene conversion or crossing over. Whereas crossing over leads to reciprocal exchange of DNA, gene conversion was originally defined as nonreciprocal exchange of homologous genetic information (44). As the products of crossing over and gene conversion events are virtually indistinguishable at the DNA level in higher eukaryotes, however, such events are commonly referred to as gene conversions (45). Gene conversion events between homologous genes and their pseudogenes have been described in the pathogenesis of several genetic disorders such as 21-hydroxylase deficiency (40), von Willebrand disease (46), and Gaucher disease (42), and have also been postulated for many clustered gene families (47). The presence of at least one p47-*phox* pseudogene carrying the GT deletion suggests that such events might also be responsible for the transfer of the GT deletion from the p47-*phox* pseudogene to the wild-type gene, thus leading to A47° CGD.

Gene conversion events most frequently occur between stretches of sequence identity (48, 49). In addition, specific sequence features such as recombination hot spots and repetitive elements (for example, Alu sequences) may facilitate such recombination events (50). It has been clearly demonstrated in *Escherichia coli* that gene conversion frequently occurs at recombination hot spots such as the Chi sequence. There is now growing evidence that this sequence might be a recombination hot spot in mammals as well (36). In addition, the HMR has been described as another hot spot for recombination in humans (37). The high incidence of recombination events between genes and pseudogenes for 21-hydroxylase or for von Willebrand factor has been attributed to the presence of Chi and HMR sequences in these genes (46, 51, 52). The presence of three Chi consensus sequences, one of them in intron 1, and one HMR consensus motif, might also be related to recombination events between p47-*phox* wild-type and pseudogene. In addition, 12 modified Chi and 17 modified HMR motifs containing one mismatch were identified, with six of them clustered in intron 2. It has been reported that sequences similar but not identical to the Chi consensus also stimulate recombination (53). Thus, the high number of potential recombination hot spots might facilitate recombination events between the p47-*phox* gene and pseudogene.

Furthermore, an unusually high density of Alu repeats is present in the p47-*phox* gene. Since the average distance between these elements in the human genome is about 4 kB (34), the identification of 17 complete Alu sequences within the 11.1 kB of intronic sequence analyzed indicates that these repetitive elements are overrepresented in the p47-*phox* gene. Within 5 kB of sequence in introns 1 and 2, nine Alu repeats were identified. The presence of repetitive sequence elements such as Alu repeats has been associated with recombination-related mutational events in the LDL receptor gene (50) and the complement component C1 inhibitor gene (54). In human growth hormone genes it has been suggested that Alu repeats define breakpoints of gene duplication (55). It has also been proposed that gene conversion is promoted by palindromic sequences, as they have been observed near recombination breakpoints (56). The GTGT tandem repeat is located within a palindromic sequence, suggesting that the immediate flanking region might be susceptible to breakage and recombination.

For the fetal γ-globin and the 21-hydroxylase genes and pseudogenes it has been suggested that regions containing clusters of differences between wild-type and pseudogene sequences might represent recombination hot spots (57, 58). A large number of sequence differences between p47-*phox* wild-type and pseudogene is found in introns 1 and 2, among them the 30-bp deletion in intron 1 and the 20-bp duplication in intron 2. This concentration of sequence differences might constitute similar recombination hot spots in the introns flanking the GT deletion, and might be responsible in part for the predominance of the GT deletion in A47° CGD.

The prevalence of a single mutation in causing p47-*phox* deficiency may also occur because the GT deletion is the only deleterious mutation in the pseudogene, in that it leads to a premature termination of p47-*phox* translation. In contrast, the 21-hydroxylase and von Willebrand factor pseudogenes contain a substantial number of potentially deleterious mutations (41, 58). In the 21-hydroxylase pseudogene, for example, an 8-bp deletion, a single-bp insertion and a nonsense mutation (as well as six additional mutations) notably cause loss of the function of the wild-type protein when transferred to the wild-type gene. Each of these deleterious mutations accounts for a proportion of 21-hydroxylase-deficient patients (40). Out of the nine single base pair substitutions identified in the coding region of the p47-*phox* pseudogene, however, only three (R→H, S→G, D→N, Table III) predict amino acid changes. Whether these changes would alter the function of p47-*phox* is not known.

Another possible mechanism explaining the high incidence of the GT deletion in A47°CGD patients would be deletion of the wild-type gene, particularly in the promoter region, leaving only the pseudogene intact for transcription. Such a gene deletion in the wild-type promoter region might have occurred in the few A47° CGD patients who are heterozygous for the GT deletion, and do not carry other apparent mutations in the coding region of the p47-*phox* gene.

Taken together, the high degree of homology between the p47-*phox* gene and pseudogenes, the presence of many consensus motifs likely to be recombination hot spots, and the close chromosomal localization of gene and pseudogenes suggest the involvement of recombination events, such as gene conversion, between the p47-*phox* wild-type gene and ΔGT-containing pseudogenes in the pathomechanism of A47° CGD. Further studies evaluating genomic DNA from A47° CGD patients and their families are currently underway and should, with the help of the sequence analysis of the p47-*phox* gene and its pseudogene provided here, further elucidate the pathogenesis of A47° CGD.

Acknowledgments

The authors wish to thank the CORE facility of the Department of Molecular and Experimental Medicine at The Scripps Research Insti-

tute for extensive help with primer syntheses and sequence analyses, Drs. Andrew Cross and Paul Heyworth for helpful suggestions and reading the manuscript, and Mrs. Valerie Moreau for help with editing the manuscript.

This study was supported by National Institutes of Health grants RO1 AI24838 (J.T. Curnutte) and RR00833 (Scripps General Clinical Research Center), the Stein Endowment Fund, an Otto-Hahn-fellowship from the Max-Planck Society, Germany (A. Görlach), and a grant from the Deutsche Forschungsgemeinschaft (J. Roesler).

References

1. Babior, B.M., R.S. Kipnes, and J.T. Curnutte. 1973. Biological defense mechanisms: the production by leukocytes of superoxide, a potential bactericidal agent. *J. Clin. Invest.* 52:741–744.

2. Curnutte, J.T. 1993. Chronic granulomatous disease: the solving of a clinical riddle at the molecular level. *Clin. Immunol. Immunopathol.* 67:2s–15s.

3. Segal, A.W., I. West, F. Wientjes, J.H.A. Nugent, A.J. Chavan, B. Haley, R.C. Garcia, H. Rosen, and G. Scrace. 1992. Cytochrome b_{-245} is a flavocytochrome containing FAD and the NADPH-binding site of the microbicidal oxidase of phagocytes. *Biochem. J.* 284:781–788.

4. Rotrosen, D., C.L. Yeung, T.L. Leto, H.L. Malech, and C.H. Kwong. 1992. Cytochrome b_{558}: the flavin-binding component of the phagocyte NADPH oxidase. *Science (Wash. DC).* 256:1459–1462.

5. Doussiere, J., G. Brandolin, V. Derrien, and P.V. Vignais. 1993. Critical assessment of the presence of an NADPH binding site on neutrophil cytochrome b_{558} by photoaffinity and immunochemical labeling. *Biochemistry.* 32: 8880–8887.

6. Clark, R.A., B.D. Volpp, K.G. Leidal, and W.M. Nauseef. 1990. Two cytosolic components of the human neutrophil respiratory burst oxidase translocate to the plasma membrane during cell activation. *J. Clin. Invest.* 85:714–721.

7. Heyworth, P.G., J.T. Curnutte, W.M. Nauseef, B.D. Volpp, D.W. Pearson, H. Rosen, and R.A. Clark. 1991. Neutrophil nicotinamide adenine dinucleotide phosphate (reduced form) oxidase assembly: translocation of p47-*phox* and p67-*phox* requires interaction between p47-*phox* and cytochrome b_{558}. *J. Clin. Invest.* 87:352–356.

8. Cross, A.R., and J.T. Curnutte. 1995. The cytosolic activating factors p47-*phox* and p67-*phox* have distinct roles in the regulation of electron flow in NADPH oxidase. *J. Biol. Chem.* 270:6543–6548.

9. Leto, T.L., A.G. Adams, and I. De Mendez. 1994. Assembly of the phagocyte NADPH oxidase: binding of src homology 3 domains to proline-rich targets. *Proc. Natl. Acad. Sci. USA.* 91:10650–10654.

10. McPhail, L.C. 1994. SH3-dependent assembly of the phagocyte NADPH oxidase. *J. Exp. Med.* 180:2011–2015.

11. Knaus, U.G., P.G. Heyworth, T. Evans, J.T. Curnutte, and G.M. Bokoch. 1991. Rac 2 is a regulator of phagocyte oxygen radical production. *Science (Wash. DC).* 254:1512–1515.

12. Abo, A., E. Pick, A. Hall, N. Totty, C. Teahan, and A.W. Segal. 1991. Activation of the NADPH oxidase involves the small GTP-binding protein p21rac1. *Nature (Lond.).* 353:668–670.

13. Wientjes, F.B., J.J. Hsuan, N.F. Totty, and A.W. Segal. 1993. p40-*phox*, a third cytosolic component of the activation complex of the NADPH oxidase to contain src homology 3 domains. 1993. *Biochem. J.* 296:557–611.

14. Tsunawaki, S., H. Miyzunari, M. Nagata, O. Tatsuzawa, and T. Kuratsuji. 1994. A novel cytosolic component, p40-*phox*, of respiratory burst oxidase associates with p67-*phox* and is absent in patients with chronic granulomatous disease. *Biochem. Biophys. Res. Commun.* 199:1378–1382.

15. Parkos C.A., M.C. Dinauer, L.E. Walker, R.A. Allen, A.J. Jesaitis, and S.H. Orkin. 1988. Primary structure and unique expression of the 22 kilodalton light chain of human neutrophil cytochrome b. *Proc. Natl. Acad. Sci. USA.* 85: 3319–3323.

16. Dinauer, M.C., E.A. Pierce, G.A.P. Bruns, J.T. Curnutte, and S.H. Orkin. 1990. Human neutrophil cytochrome-b light chain (p22-*phox*): gene structure, chromosomal location, and mutations in cytochrome negative autosomal recessive chronic granulomatous disease. *J. Clin. Invest.* 86:1729–1737.

17. Volpp, B.D., W.N. Nauseef, J.E. Donelson, D.R. Moser, and R.A. Clark. 1989. Cloning of the cDNA and functional expression of the 47-kilodalton cytosolic component of human neutrophil respiratory burst oxidase. *Proc. Natl. Acad. Sci. USA.* 86:7195–7199.

18. Lomax K.J., T.L. Leto, H. Nunoi, J.I. Gallin, and H.L. Malech. 1989. Recombinant 47-kilodalton cytosol factor restores NADPH oxidase in chronic granulomatous disease. *Science (Wash. DC).* 245:409–412.

19. Leto, T.L., K.J. Lomax, B.D. Volpp, H. Nunoi, J.M.G. Sechler, W.M. Nauseef, R.A. Clark, J.I. Gallin, and H.L. Malech. 1990. Cloning of a 67-kD neutrophil oxidase factor with similarity to a non-catalytic region of p60c-src. *Science (Wash. DC).* 248:727–730.

20. Royer-Pokora, B., L.M. Kunkel, A.P. Monaco, S.C Goff, P.E. Newburger, R.L. Baehner, F.S. Cole, J.T. Curnutte, and S.H. Orkin. 1986. Cloning the gene for an inherited human disorder—chronic granulomatous disease—on the basis of its chromosomal location. *Nature (Lond.).* 322:32–38.

21. Roos, D., M. de Boer, F. Kuribayashi, C. Meischl, R.S. Weening, A.W. Segal, A. Ahlin, K. Nemet, J.P. Hossle, E. Bernatowska-Matuszkiewicz, and H. Middleton-Price. 1996. Mutations in the X-linked and autosomal recessive forms of chronic granulomatous disease. *Blood.* 87:1663–1681.

22. Thrasher, A.J., N.H. Keep, F. Wientjes, and A.W. Segal. 1994. Chronic granulomatous disease. *Biochem. Biophys. Acta.* 1227:1–24.

23. Casimir, C.M., H.N. Bu-Ghanim, A.R.F. Rodaway, D.L. Bentley, P. Rowe, and A.W. Segal. 1991. Autosomal recessive chronic granulomatous disease caused by deletion at a dinucleotide repeat. *Proc. Natl. Acad. Sci. USA.* 88: 2753–2757.

24. Volpp, B.D., and Y. Lin. 1993. In vitro reconstitution of the respiratory burst in B lymphoblasts from p47-*phox*–deficient chronic granulomatous disease. *J. Clin. Invest.* 91:210–207.

25. Iwata, M., H. Nunoi, H. Yamazaki, T. Nakano, H. Niwa, S. Tsuruta, S. Ohga, S. Ohmi, S. Kanegasaki, and I. Matsuda. 1994. Homologous dinucleotide (GT or TG) deletion in Japanese patients with chronic granulomatous disease with p47-phox deficiency. *Biochem. Biophys. Res. Comm.* 199:1372–1377.

26. Safaya, S., and R.F. Rieder.1988. Dysfunctional alpha-globin gene in hemoglobin H disease in blacks. A dinucleotide deletion produces a frameshift and a termination codon. *J. Biol. Chem.* 263:4328–4332.

27. Chan, E.K.L., S. Takano, L.E.C. Andrade, J.C. Hamel, and A.G. Matera. 1994. Structure, expression, and chromosomal localization of human p80-coilin gene. *Nucleic Acids Res.* 22:4462–4469.

28. Green, E.D., V.V. Braden, R.S. Fulton, R. Lim, M.S. Ueltzen, D.C. Peluso, R.M. Mohr-Tidwell, J.R. Idol, L.M. Smith, I. Chumakov, et al. 1995. A human chromosome 7 yeast artificial chromosome (YAC) resource: construction, characterization and screening. *Genomics.* 25:170–183.

29. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387–395.

30. Pearson, W.R., and D. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl.Acad. Sci. USA.* 85:2444–2448.

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

32. Rodaway, A.R.F., C.G. Teahan, C.M. Casimir, A.W. Segal, and D.L. Bentley. 1990. Characterization of the 47-kilodalton autosomal chronic granulomatous disease protein: tissue-specific expression and transcriptional control by retinoic acid. *Mol. Cell. Biol.* 10:5388–5396.

33. Volpp, B.D., W.N. Nauseef, J.E. Donelson, D.R. Moser, and R.A. Clark. 1989. Cloning of the cDNA and functional expression of the 47-kilodalton cytosolic component of human neutrophil respiratory burst oxidase. (correction). *Proc. Natl. Acad. Sci. USA.* 86:9563.

34. Britten, R.J., W.F. Baron, D.B. Stout, and E.H. Davidson. 1988. Sources and evolution of human Alu repeated sequences. *Proc. Natl. Acad. Sci. USA.* 85:4770–4774.

35. Francke, U., C.-L. Hsieh, B.E. Foellmer, K.J. Lomax, H.L. Malech, and T.L. Leto. 1990. Genes for two autosomal recessive forms of chronic granulomatous disease assigned to 1q25 (NCF2) and 7q11.23 (NCF1). *Am. J. Hum. Genet.* 47:483–492.

36. Smith, G.R. 1994. Hotspots of homologous recombination. 1994. *Experientia (Basel).* 50:234–241.

37. Jeffreys, A.J., V. Wilson, and S.L. Thein. 1985. Hypervariable minisatellite regions in human DNA. *Nature (Lond.).* 314:67–73.

38. Roesler, J., A. Görlach, J. Rae, P.J. Hopkins, P. Patino, P. Lee, J.T. Curnutte, and S.J. Chanock. 1995. Recombination events between the normal p47-*phox* gene and a highly homologous pseudogene are the main cause of autosomal recessive chronic granulomatous disease (CGD). *Blood.* 86:260a (Abstr.).

39. Goeddel, D.V., D.W. Leung, T.J. Dull, M. Gross, R.M. Lawn, R. McCandliss, P.H. Seeburg, A. Ullrich, E. Yelverton, and P.W. Gray. 1981. The structure of eight distinct cloned human leukocyte interferon cDNAs. *Nature (Lond.).* 290:20–26.

40. Helmberg, A. 1993. Twin genes and endocrine disease: CYP21 and CYP11B genes. *Acta Endocrinol.* 129:97–108.

41. Mancuso, D.J., E.A. Tuley, L.A. Westfield, T.L. Lester-Mancuso, M.M. Le Beau, J.M. Sorace, and J.E. Sadler. 1991. Human von Willebrand Factor gene and pseudogene: structural analysis and differentiation by polymerase chain reaction. *Biochemistry.* 30:253–269.

42. Horowitz, M., S. Wilder, Z. Horowitz, O. Reiner, T. Gelbart, and E. Beutler. 1989. The human glucocerebrosidase gene and pseudogene: structure and evolution. *Genomics.* 4:87–96.

43. Wilde, C.D. 1986. Pseudogenes. *Crit. Rev. Biochem.* 19:323–352.

44. Kourilsky, P. 1986. Molecular mechanisms for gene conversion in higher cells. *Trends Genet.* 2:60–63.

45. Maeda N., and O. Smithies. 1986. The evolution of multigene families: human haptoglobin genes. *Ann. Rev. Genet.* 20:81–108.

46. Eikenboom, J.C.J., T. Vink, E. Briet, J.J. Sixma, and P.H. Reitsma. 1994. Multiple substitutions in the von Willebrand factor gene that mimic the pseudogene sequence. *Proc. Natl. Acad. Sci. USA.* 91:2221–2224.

47. Collier, S., M. Tassabehij, and T. Strachnan. 1993. A de novo pathological point mutation at the 21-hydroxylase locus: implications for gene conversion in the human genome. *Nat. Genet.* 3:260–265.

48. Metzenberg, A.B., G. Wurzer, T.H.J. Huisman, and O. Smithies. 1991.

Homology requirements for unequal crossing over in humans. *Genetics.* 128: 143–161.

49. Liskay, R.M., A. Letsou, and J.L. Stachelek. 1987. Homology requirement for efficient gene conversion between duplicated chromosomal segments in mammalian cells. *Genetics.* 115:161–167.

50. Lehrman, M.A., D.W. Russell, J.L. Goldstein, and M.S. Brown. 1987. Alu-Alu recombination deletes splice acceptor cites and produces secreted low density lipoprotein receptor in a subject with familial hypercholesterolemia. *J. Biol. Chem.* 262:3354–3361.

51. Amor, M., K.L. Parker, H. Globermann, M.I. New, and P.C. White. 1988. Mutation in the CYB21B gene (Ile-172→Asn) causes steroid 21-hydroxylase deficiency. *Proc. Natl. Acad. Sci. USA.* 85:1600–1604.

52. Urabe, K., A. Kimura, F. Harada, T. Iwanaga, and T. Sasazuki. 1990. Gene conversion in steroid 21-hydroxylase genes. *Am. J. Hum. Genet.* 46:1178–1186.

53. Cheng, K.C., and G.R. Smith. 1987. Cutting of Chi-like sequences by the RecBCD enzyme of Escherichia coli. *J. Mol. Biol.* 194:747–750.

54. Stoppa-Lyonnet, D, P.E. Carter, T. Meo, and M. Tosi. 1990. Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. *Proc. Natl. Acad. Sci. USA.* 87:1551–1555.

55. Hirt, H., J. Kimelman, M.J. Birnbaum, E.Y. Chen, P.H. Seeburg, N.L. Eberhardt, and A. Barta. 1987. The human growth hormone locus: structure, evolution, and allelic variations. *DNA (New York).* 6:59–70.

56. Krawinkel, U., G. Zoebelein, and A.L.M. Bothwell. 1986. Palindromic sequences are associated with sites of DNA breakage during gene conversion. *Nucl. Acids Res.* 4:3871–3882.

57. Slightom, J.L., A.E. Blechl, and O. Smithies. 1980. Human fetal G- and A-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell.* 21:627–638.

58. Higashi Y., H. Yoshioka, M. Yamane, O. Gotoh, and Y. Fuji-Kuriyama. 1986. Complete nucleotide sequence of two steroid 21-hydroxylase genes tandemly arranged in human chromosome: a pseudogene and a genuine gene. *Proc. Natl. Acad. Sci. USA.* 83:2841–2845.