Factor VIII gene inversions causing severe hemophilia A originate almost exclusively in male germ cells

Judith Pratt Rossiter, Michele Young[1], Michelle L.Kimberland, Pierre Hutter[1], Rhett P.Ketterling[2], Jane Gitschier[3], Jürgen Horst[4], Michael A.Morris[1], Daniel J.Schaid[5], Philippe de Moerloose[6], Steve S.Sommer[2], Haig H.Kazazian, Jr.* and Stylianos E.Antonarakis[1],*

Center for Medical Genetics, The Johns Hopkins University School of Medicine, 600 N.Wolfe Street, Baltimore, MD 21287, USA, [1]Division of Medical Genetics, University of Geneva Medical School & Cantonal Hospital, Geneva, Switzerland, [2]Department of Biochemistry and Molecular Biology, Mayo Clinic/Foundation, Rochester, MN, [3]Howard Hughes Medical Institute and Department of Medicine, University of California, San Francisco, CA, USA, [4]Institut für Humangenetik der Universität, Münster, Germany, [5]Section of Biostatistics, Department of Health Sciences Research, Mayo Clinic/Foundation, Rochester, MN, USA and [6]Hemostasis and Thrombosis Unit, Cantonal Hospital, Geneva, Switzerland.

The factor VIII gene, which is defective in hemophilia A, is located in the last megabase of the long arm of the X chromosome. Inversions due to intrachromosomal homologous recombination between mispaired copies of gene A located within intron 22 of the gene and about 500 kb telomeric to it account for nearly half of all cases of severe hemophilia A. We hypothesized that pairing of Xq with its homolog inhibits the inversion process, and that, therefore, the event originates predominantly in male germ cells. In all 20 informative cases in which the inversion originated in a maternal grandparent, DNA polymorphism analysis determined that it occurred in the male germline. In addition, all but one of 50 mothers of sporadic cases due to an inversion were carriers. Thus, these data support the hypothesis and indicate that factor VIII gene inversions leading to severe hemophilia A occur almost exclusively in male germ cells.

INTRODUCTION

Hemophilia A is a common X-linked coagulation disorder, with an incidence of about 1 in 10,000 males before the onset of modern medical therapy (1). The disease is due to a large variety of mutations in the factor VIII gene, which is located in the last megabase of Xq (2). When an effort was made to characterize all the mutations in a series of patients with hemophilia A, an unexpected finding was obtained. Higuchi *et al.* were able to identify mutations in almost every case of mild-to-moderate hemophilia A (3), but were unable to identify the mutation in about half of the severely affected patients (4). Subsequently, Naylor *et al.* using RT−PCR of illegitimate transcripts of the factor VIII gene could not detect an amplification product between exons 22 and 23 of the gene in about 40% of severely affected patients (5). Recently Lakich *et al.* (6) and Naylor *et al.* (7) found that these patients have an inversion involving the factor VIII gene.

The inversion results from homologous recombination between gene A, a small, intronless gene within intron 22 of the factor VIII gene, and one of two additional copies of gene A located approximately 500 kb 5′ and telomeric to the factor VIII gene (6). Intron 22, the largest intron in the gene, contains two

transcribed sequences, gene A (8) and gene B (9). Interestingly, gene B is transcribed in the same orientation as factor VIII and uses exons 23 to 26 of the gene (9). In contrast, the copy of gene A within intron 22 is transcribed in the opposite direction to factor VIII, while the two upstream copies are thought to be transcribed in the same direction as the factor VIII gene (6). The postulated mechanism for the inversion involves flipping of the tip of the X chromosome, allowing homologous pairing between one of the telomeric copies of gene A and the copy in intron 22 (6). A single crossing over event causes a DNA inversion of roughly 600 kb, resulting in the factor VIII gene being divided into two parts, exons 1−22 and exons 23−26, which are widely separated and have opposite orientations. This inversion can be detected by Southern blot analysis (6), which allows direct detection of the factor VIII lesion in more than 25% of patients with hemophilia A.

Depending upon which extragenic copy of gene A is involved in the crossing over event, two main types of inversions are recognized. A type 1 inversion affects the distal copy, while a type 2 inversion affects the proximal copy (6). In addition, a rare Southern blot pattern, type 3, is thought to result from an
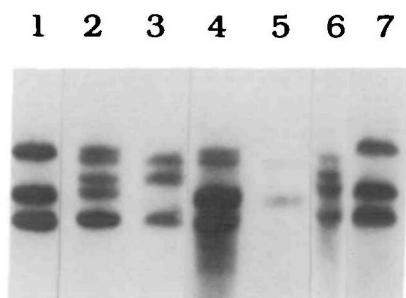
Figure 1. Southern blot demonstrating the various inversion mutation types. Genomic DNA from control individuals (lanes 1 and 7), carrier mothers (lanes 2 and 4), and affected individuals (lanes 3, 5 and 6) digested with *Bcl*I and hybridized to the gene A probe. Lanes 1 and 7 show the normal fragment pattern. Lane 2 shows a type 1 carrier, lane 3 a type 1 proband, lane 4 a type 2 carrier, lane 5 a type 2 proband, and lane 6 a type 3 proband.

Table 1. Inversion mutations in patients with severe hemophilia A

Patients studied	135
Inversion present	62 (46%)
Type 1	48 (77% of inversions)
Type 2	12 (19% of inversions)
Type 3	2 (3% of inversions)
Parent of origin of the mutation (defined from 26 sporadic cases[a])	
Maternal Grandfather	20
Maternal Grandmother	0
Mother	1
Estimated male:female ratio of mutations	302:1[b]

[a]Five sporadic families had a carrier maternal grandmother.
[b]The 95% confidence interval for this value is 22 – 12,115.

inversion occurring in an individual carrying more than two extragenic copies of gene A (Figure 1).

To our knowledge, inversions of this type due to homologous recombination have not been described previously in humans and may be unique to the factor VIII gene region, due to the intragenic and extragenic copies of gene A, the proximity of this region to the telomere, and its location on the long arm of the X chromosome. In this study, we hypothesized that the X chromosome plays an important role in the rearrangement, i.e., that the inversion may occur predominantly in male meiosis when Xq is unpaired with a homolog. In female meiosis the inversion process would be inhibited by pairing between the two X chromosomes. Our data now support the hypothesis and show that this common inversion originates almost exclusively in male germ cells. Since the mutation occurs very rarely in female germ cells, nearly all mothers of patients with inversions are carriers of factor VIII deficiency.

RESULTS

Frequency and type of inversion in severe hemophilia A

Southern blot analysis of genomic DNA from 135 patients known to have severe hemophilia A revealed an inversion of the factor VIII gene in 62, or 46%. Of these inversions, 48 (77%) were type 1, 12 (19%) were type 2 and 2 (3%) were type 3 (Table 1). Included in the type 1 group were two females with severe hemophilia A who were heterozygous for the normal pattern and the type 1 inversion.
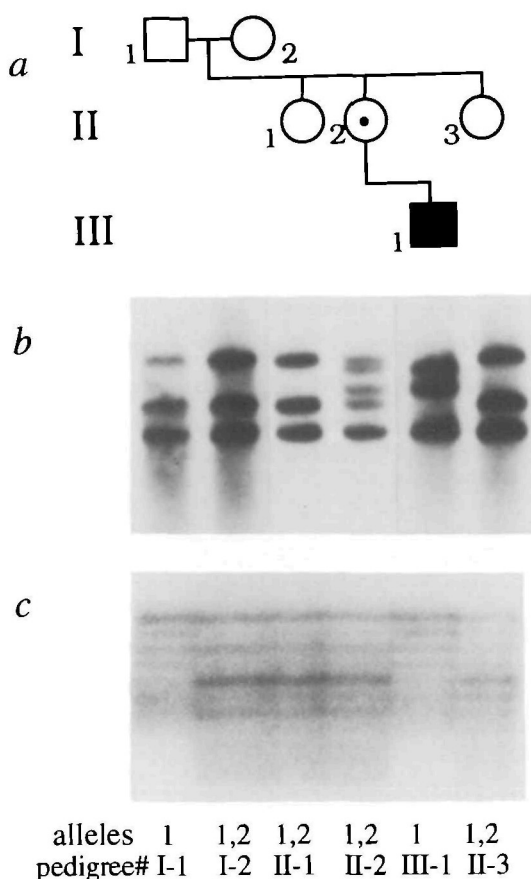

Figure 2. Pedigree, Southern blot analysis and polymorphism analysis of a family with a type 1 inversion mutation. (a) shows the pedigree, (b) shows the *Bcl*I digest hybridized to the gene A probe for all available family members. The grandfather (I-1), grandmother (I-2), and both maternal aunts (II-1 and II-3) have a normal pattern. The mother (II-2) is a carrier of the type 1 inversion mutation, demonstrated by the normal bands (21.5 kb, 16 kb, and 14 kb) as well as two abnormal bands (20 kb and 17.5 kb). The affected individual (III-1) has the type 1 inversion mutation, demonstrated by the abnormal 20 kb and 17.5 kb bands and the normal 14 kb band (which represents the uninvolved proximal extragenic copy of gene A). (c) shows polymorphism analysis using the $(CA)_n$ microsatellite repeat in intron 13 of the factor VIII gene, demonstrating that the affected individual inherited the X chromosome from his maternal grandfather.

Additional case samples were ascertained through the presentation of an individual for carrier detection or prenatal diagnosis because of a family history of hemophilia A. In these cases, the severity of the proband's hemophilia A could not always be determined. Therefore, this group was not included in the estimation of the frequency of the inversion mutation among patients with severe hemophilia A. An additional 33 patients with a factor VIII gene inversion were found in this group with undefined severity, 27 with type 1 and 6 with type 2. Thus, of a total of 95 inversions observed, 75 (79%) were type 1, 18 (19%) were type 2, and 2 (2%) were type 3.

Parental origin of the inversion

To test the hypothesis that the inversion is more likely to occur during male meiosis, family studies were carried out on 35 families with potentially new inversion mutations based on the absence of a family history of hemophilia A prior to the proband.

The analysis included 1) Southern blot for detection of the inversion in the patient, his mother, and his maternal grandparents, when available, and 2) determination of haplotypes in the factor VIII gene region in the same individuals to infer the parental origin of the inversion (Figure 2). In five families, both the mother and maternal grandmother were found to be carriers of the inversion despite the negative family history. In nine other families, the mothers were carriers, but the maternal grandmothers were not available for study. In addition to these 35 families, the mothers of 15 individuals with the inversion mutation were studied by Southern blot analysis, and in all cases these mothers were found to be carriers. No other family members were available for study in these 15 families.

In 20 informative families, the mother was found to be a carrier while the inversion was not present in either maternal grandparent (Table 1). In all of these 20 cases, polymorphic markers demonstrated that the X chromosome carrying the inversion originated in the germ cells of the unaffected maternal grandfather. The informative markers were intragenic $(CA)_n$ repeats and biallelic restriction fragment length polymorphisms in 16 cases, the extragenic DXS52 polymorphisms in 2, and the extragenic DXS15 in 2 cases. The error rate for the inference of origin is negligible when the intragenic markers are used and roughly 2-3% for the extragenic DXS52 and DXS15.

The age of the maternal grandfathers at the birth of their carrier daughters was available in 13 cases. The mean age was 29.9 years (range 24 to 38 years). When each grandparental age was compared to the distribution of male reproductive age in the general population (10), the mean age of males at conception of the individual with the *de novo* inversion was at the 54th percentile. None of the grandparental ages were in the 90th percentile or greater (11).

In only one family (HA26), a noncarrier mother had an affected son with a type 1 inversion. Since this was the only case of a maternal origin of the inversion, nonmaternity and sample mixup was excluded by analysis of 7 microsatellite polymorphisms (D6S292, D8S133, D21S167, D14S81, D18S59, D18S53, and CSF1R) which showed that mother and child shared one allele at each locus. Although a low level of germinal mosaicism in the mother cannot be excluded, it is unlikely because she also had 10 normal sons. Therefore, the mutation in this family probably occurred in maternal germ cells. Interestingly, the mother was in the 97th percentile for female reproductive age when she conceived her affected son.

Among the 95 unrelated patients with inversions, samples from their mothers were available in 50. Of these 50 mothers, 49 (98%) were carriers of the inversion. The only exception was that mentioned above.

The male:female sex ratio for the origin of the mutation was determined by Bayesian analysis (10). Since the sex ratio is very high for the inversion, it was helpful to include families in which the maternal grandmother is a carrier (details of calculation available on request). The male:female sex ratio is estimated at 302:1 and the 95% confidence interval is 22 to 12,115. The upper bound of the 95% confidence interval is so large because the female in family HA26 had so many normal sons. Even if the true sex ratio were 302:1, the mutation would originate in a female in 85% of families with this pedigree structure. If the mother of family HA26 had no normal sons, the estimated sex ratio would have been reduced to 55 and the 95% confidence interval would be 11 to 975.

DISCUSSION

We estimate that inversions of the factor VIII gene occur much more frequently in male germ cells (302:1). We hypothesize that mispairing leading to inversion is inhibited by pairing of the X chromosomes in female germ cells. The factor VIII gene and the gene A copies all reside at the telomere of Xq, which is generally unpaired in males. However, a pseudoautosomal region of Xq and Yq has recently been identified which extends approximately 400 kb from their telomeres (12). This region is distal to the telomeric gene A copies (13). DNA polymorphism analysis suggested the possibility of exchange between these segments of the X and Y chromosomes (12); however, the biological significance of this pairing in relation to the factor VIII inversion events is unknown. Because pairing of the distal Xq region differs in male and female meiosis, this region may be predisposed to flipping onto itself in male meiosis, allowing homologous recombination between copies of gene A. We estimate the frequency of the inversion event as roughly 1 in 240,000 male meioses (10^{-4} disease frequency $\times 16.6\%$ frequency of sporadic disease (14) $\times 25\%$ frequency of inversions) and 1 in 70,000,000 female meioses.

It is generally believed that all regions of the human genome are paired in meiosis with the exception of the majority of the sex chromosomes in male meiosis. Therefore, it is likely that the type of inversions observed in the factor VIII gene could be generated primarily in the unpaired regions of the long arms of the X and Y chromosomes. The inversions affecting the factor VIII gene are mediated by repeated sequences which are 5-10 kb in size (6), one copy of which lies within the gene, and the inversions result in splitting of the gene and a severe disease phenotype. It is possible that other long repeated sequences, such as LINE1 elements, mediate similar inversions involving other genes at the telomeres of Xq and Yq. These inversions could either be deleterious if one copy of the repeated sequence lies within an intron of a gene, or could merely lead to polymorphic rearrangement of marker order if the repeated sequence lies outside gene sequences. Although this mechanism should be more common on Xq and Yq, our finding the origin of the inversion in a female meiosis suggests that inversions of a similar type could occur at low frequency at the telomeres of autosomes.

Few mutations show a marked sex bias in their origin. To date the most pronounced sex difference is in transitions at CpG dinucleotides in the factor IX gene which originate 11 times more frequently in males than in females (10). The estimated male:female ratio of all point mutations was 3.5:1 in the factor IX gene (10). In the fragile X syndrome, expansion of the trinucleotide repeat in the FMR1 gene from the premutation to the full mutation occurs exclusively when the mutation is inherited from the female (15). However, since the full mutation is absent from sperm of fragile X males, these expansions may be mitotic, occurring in the somatic cells of the affected offspring (16). Expansion of trinucleotide repeats in other disorders is also influenced by the sex of origin. In Huntington disease, the largest increases in size of the $(CAG)_n$ repeat occur predominantly in male transmissions (17); while in myotonic dystrophy female transmissions tend to result in larger expansions (18). Another mutation with a bias in the sex of origin is the 1.5 Mb duplication within 17p11.2 that causes Charcot-Marie-Tooth disease type 1A, which was shown by Palau, *et al.* to originate in male meiosis in all 9 cases studied (19). Our data for the origin of the

inversion mutation in the factor VIII gene indicate the highest male:female sex ratio observed to date.

If the inversion occurred during DNA replication or mitosis, one would expect to find advanced grandpaternal age, since the number of cell divisions increases linearly with age after puberty. On the other hand, if the inversion occurred during the one meiosis that all germ cells undergo, one would expect to find the observed absence of advanced grandpaternal age.

The carrier frequency in mothers of patients with hemophilia A is theoretically 0.67 (20). One study of 28 isolated cases of hemophilia A in which the factor VIII gene mutations were identified found that 13 of the 28 mutations occurred *de novo* within two generations (21). This number is similar to that expected on a theoretical basis, 15 of 28. While the carrier frequency of 0.67 remains valid among mothers of individuals with mild-to-moderate hemophilia A and severe disease without an inversion, nearly all mothers of patients with the inversion (49 of 50) are carriers of factor VIII deficiency. Thus, the empiric risk of a mother being a carrier given that her son has severe hemophilia A can be calculated as $0.5 + (2/3 \times 0.5) = 0.83$ (approximately half of severe hemophilia A is caused by an inversion and nearly all of these cases have a carrier mother (thus 0.5) and the remaining half maintain the 0.67 carrier frequency). As a result, counselling of couples with a single child with severe hemophilia A must emphasize that there is a greater risk that the mother is a carrier than previously suspected. However, carrier detection and prenatal diagnosis can be quite straightforward using Southern blot analysis, and all mothers of sons with isolated cases of severe hemophilia A should be offered testing for the presence of an inversion mutation.

MATERIALS AND METHODS

Samples

Blood was collected from patients and family members after informed consent was obtained. Samples were received from the following hemophilia clinics: Vanderbilt University, Children's Orthopedic Hospital of Los Angeles, Mayo Clinic, University of Münster, Cantonal Hospital of Geneva, and the gene diagnostic laboratory of the Johns Hopkins Hospital. DNA was isolated by phenol-chloroform extraction (22) and stored at −20°C prior to use.

Southern blots

Genomic DNA (5 μg per reaction) was digested with *BclI* for 16 h at 50°C, the digest was subjected to electrophoresis for 24 h at 50 V on 0.7% agarose gels and the DNA fragments were transferred to nitrocellulose membranes. The membranes were probed with the 0.9 kb *EcoRI/SstI* fragment from intron 22 containing gene A. The fragment was prepared from plasmid p482.6 (ATCC catalogue no. 57203) as previously described (8). Southern blots were prehybridized and hybridized in QuikHyb hybridization solution (Stratagene).

Polymorphic markers

The following polymorphic markers were used to determine the parental origin of the factor VIII gene inversion: *BclI* from intron 18 (23), (CA)ₙ microsatellite repeat in intron 13 (24), (CA)ₙ microsatellite repeat in intron 22 (25), *XbaI* polymorphic site in intron 22 (26), *BglII* site of locus DXS15 (27), and the VNTR polymorphisms at loci DXS52 detected by *TaqI* (28). Information concerning the oligonucleotide primers or the hybridization probes for the detection of these polymorphisms can be found in the Genome Data Base. Microsatellite detection was performed as previously described (29).

Analysis of sex ratio

The ratio of mutation rates, r, for male:female was estimated by the method of maximum likelihood. First, the probability of observing each pedigree was calculated. The calculation of this probability allowed for four possible events for each pedigree: 1) the mother was the source of the mutation, 2) the grandfather was the source of the mutation, 3) the grandmother was the source of the mutation, or 4) the grandmother was a carrier. Under each event, the probability of the

pedigree was determined using Bayesian calculations (10), and these four probabilities sum to one. The likelihood of all 26 pedigrees was then calculated by multiplying the probabilities of the observed events across all pedigrees. The resulting likelihood depends only on r, allowing calculations of the maximum likelihood estimate of r. The likelihood ratio statistic was used to test if r was significantly different from 1, and to construct a 95% confidence interval for r.

Analysis of parental age

The age percentile for each family was determined using tables published in *Statistical Abstracts of the United States* or *Vital Statistics of the United States* (11). Birth rate per 1000 individuals in age groups of five years were multiplied by the total Caucasian male or female U.S. population for that age group. Once an exact natality number was calculated for each age group, the number was divided in five equal percentages to estimate the total natality for each individual age. Percentiles were calculated by adding all nativity numbers from age 15 to the age in question and dividing by the total natality number for that year. Prior to 1937, only maternal ages are listed. Thus, the relatively constant ratio of maternal to paternal age was utilized to estimate natality versus paternal age prior to 1937.

ABBREVIATIONS

RT−PCR, reverse transcript polymerase chain reaction; VNTR, variable number of tandem repeats

REFERENCES

1. Hoyer, L.W. (1994) *N. Engl. J. Med.* **330**, 38−47.
2. Tuddenham, E.G.D., Cooper, D.N., Gitschier, J., Higuchi, M., Hoyer, L., Yoshioka, A., Peake, I., Schwaab, R., Olek, K., Kazazian, H., Lavergne, J.M., Gianelli, F., Antonarakis, S. (1991) *Nucl. Acids Res.* **19**, 4821−4833.
3. Higuchi, M., Antonarakis, S., Kasch, L., Oldenburg, J., Economou-Petersen, E., Olek, K., Arai, M., Inaba, H., Kazazian, H.H. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 8307−8311.
4. Higuchi, M., Kazazian, H.H., Kasch, L., Warren, T.C., McGinniss, M.J., Phillips, J.A., Kasper, C., Janco, R., Antonarakis, S.E. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7405−7409.
5. Naylor, J.A., Green, P.M., Rizza, C.R., Giannelli, F. (1993) *Hum. Molec. Genet.* **2**, 11−17.
6. Lakich, D., Kazazian, H.H., Antonarakis, S.E., Gitschier, J. (1993) *Nature Genet.* **5**, 236−241.
7. Naylor, J., Brinke, A., Hassock, S., Green, P.M., Giannelli, F. (1993) *Hum. Molec. Genet.* **2**, 1773−1778.
8. Levinson, B., Kenwrick, S., Lakich, D., Hammonds, G., Gitschier, J. (1990) *Genomics* **7**, 1−11.
9. Levinson, B., Kenwrick, S., Gamel, P., Fisher, K., Gitschier, J. (1992) *Genomics* **14**, 585−589.
10. Ketterling, R.P., Vielhaber, E., Bottema, C.D.K., Schaid, D.J., Cohen, M.P., Sexauer, C.L., Sommer, S.S. (1993) *Am. J. Hum. Genet.* **52**, 152−166.
11. U.S. Bureau of the Census (1918−1989). Vital statistics of the United States. U.S. Government Printing Office, Washington, DC.
12. Freije, D., Helms, C., Watson, M.S., Donis-Keller, H. (1992) *Science* **258**, 1784−1787.
13. Freije, D. and Schlessinger, D. (1992) *Am. J. Hum. Genet.* **51**, 66−80.
14. Barrai, I., Cann, H.M., Cavalli-Sforza, L.L., Barbujani, G., DeNicola, P. (1992) *Am. J. Hum. Genet.* **37**, 680−699.
15. Rousseau, F., Heitz, D., Mandel, J.L. (1992) *Hum. Mutation* **1**, 91−96 .
16. Reyniers, E., Vits, L., DeBoulle, K., VanRoy, B., VanVelzen, D., deGraaff, E., Verkerk, A.J.M.H., Jorens, H.Z.J., Darby, J.K., Oostra, B., Willems, P.J. (1993) *Nature Genet.* **4**, 143−146.
17. Duyao, M., Ambrose, C., Myers, R., Novelletto, A., Persichetti, F., Frontali, M., Folstein, S., Ross, C., Franz, M., Abbott, M., Gray, J., Conneally, P., Young, A., Penney, J., Hollingsworth, Z., Shoulson, I., Lazzarini, A., Falek, A., Koroshetz, W., Sax, D., Bird, E., Vonsattel, J., Bonilla, E.,

Alvir, J., Bickham Conde, J., Cha, J.H., Dure, L., Gomez, F., Ramos, M., Sanchez-Ramos, J., Snodgrass, S., deYoung, M., Wexler, N., Moscowitz, C., Penchaszadeh, G., MacFarlane, H., Anderson, M., Jenkins, B., Srinidhi, J., Barnes, G., Gusella, J., MacDonald, M. (1993) *Nature Genet.* **4**, 387–392.

18. Redman, J.B., Fenwick, R.G., Fu, Y.H., Pizzuti, A., Caskey, C.T. (1993) *J. Am. Med. Assoc.* **269**, 1960–1965.

19. Palau, F., Löfgren, A., DeJonghe, P., Bort, S., Nelis, E., Sevilla, T., Martin, J.J., Vilchez, J., Prieto, F., Van Broeckhoven, C. (1993) *Hum. Molec. Genet.* **2**, 2031–2035.

20. Haldane, J.B.S. (1935) *J. Genet.* **31**, 317–326.

21. Antonarakis, S.E., Kazazian, H.H. (1988) *Trends in Genet.* **4**, 233–237.

22. Kunkel, L.M., Smith, K.D., Boyer, S.H., Borgaonkar, D.S., Wachtel, S.S., Miller, O.J., Breg, W.R., Jones, H.W., Rary, J.M. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 1245–1249.

23. Gitschier, J.D., Drayna, D., Tuddenham, E.G.D., White, R.L., Lawn, R.M. (1985) *Nature* **314**, 738–740.

24. Lalloz, M.R.A., McVey, J.H., Pattinson, J.K., Tuddenham, E.G.D. (1991) *Lancet* **338**, 207–211.

25. Lalloz, M.R.A., McVey, J.H., Michaelides, K., Tuddenham, E.G.D. (1992) *Br. J. Haematol.* **80**, 3a.

26. Wion, K.L., Tuddenham, E.G.D., Lawn, R.M. (1986) *Nucl. Acids Res.* **14**, 4535–4542.

27. Harper, K., Winter, R.M., Pembrey, M.E., Hartley, D., Davies, K.E., Tuddenham, E.G.D. (1984) *Lancet* ii, 6–8.

28. Oberle, I., Camerino, G., Heilig, R., Grunebaum, L., Cazenave, J.P., Crapanzano, C., Manucci, P., Mandel, J.L. (1985) *N. Engl. J. Med.* **312**, 682–686.

29. Weber, J.L., May, P.E. (1989) *Am. J. Hum. Genet.* **44**, 388–396.