Am. J. Hum. Genet. 59:825–833, 1996

Genetic Recombination at the Human *RH* Locus: A Family Study of the Red-Cell Evans Phenotype Reveals a Transfer of Exons 2–6 from the *RHD* to the *RHCE* Gene

C.-H. Huang,[1] Y. Chen,[1] M. Reid,[1] and S. Ghosh[2]

[1]Lindsley F. Kimball Research Institute, New York Blood Center, New York; and [2]East of Scotland Blood Transfusion Service, Ninewells Hospital, Dundee

Summary

The human *RH* locus appears to consist of two structural genes, *D* and *CE*, which map on the short arm p34-36 of chromosome 1 and specify a most complex system of blood-group genetic polymorphisms. Here we describe a family study of the Evans (also known as "D- -") phenotype, a codominant trait associated with both qualitative and quantitative changes in D-antigen expression. A cataract-causing mutation was also inherited in this family and was apparently cotransmitted with Evans, suggesting a chromosomal linkage of these two otherwise unrelated traits. Southern blot analysis and allele-specific PCR showed the linkage of Evans with a *Sph*I RFLP marker and the presence of a hybrid gene in the *RH* locus. To delineate the pattern of gene expression, the composition and structure of Rh-polypeptide transcripts were characterized by reverse transcriptase–PCR and nucleotide sequencing. This resulted in the identification of a novel Rh transcript expressed only in the Evans-positive erythroid cells. Sequence analysis showed that the transcript maintained a normal open reading frame but occurred as a CE-D-CE composite in which exons 2–6 of the *CE* gene were replaced by the homologous counterpart of the *D* gene. This hybrid gene was predicted to encode a CE-D-CE fusion protein whose surface expression correlates with the Evans phenotype. The mode and consequence of such a recombination event suggest the occurrence, in the *RH* locus, of a segmental DNA transfer via the mechanism of gene conversion.

Introduction

The Rh polypeptides of human erythrocytes are of considerable interest, owing to their polymorphic nature and probable role in membrane structure and physiology (Agre and Cartron 1991; Anstee and Tanner 1993; Cartron and Agre 1995). With regard to their polymorphic nature, these transmembrane proteins represent a natural model system for studies of human genetic diversity, since they are associated with the occurrence, in various ethnic groups, of a large number of Rh antigens, including the common D, C/c, and E/e series and related allelic variants (Race and Sanger 1975). With regard to their physiologic role, the Rh proteins appear to function as finely tuned regulators in the maintenance of membrane integrity, since Rh-deficiency syndrome, characterized by chronic but moderate hemolytic anemia and stomatocytosis, may develop in their absence (Agre and Cartron 1991). Nevertheless, despite such pathological correlations, little is known about how the Rh proteins are involved in the erythrocyte-membrane processes. In addition, the question of how Rh-protein deficiency leads to the multiple facets of membrane abnormality still requires rigorous investigations.

As are most potent immunogens, the Rh antigens are clinically important, because they can cause severe transfusion reactions and hemolytic disease in the newborn, in the case of blood-type incompatibility (Mollison et al. 1993). The locus controlling the expression of these antigens appears to consist of two structural genes, *D* and *CE* (the latter including *ce, cE, Ce,* and *CE* alleles), which map on the short arm p34-36 of chromosome 1 (Cherif-Zahar et al. 1991; MacGeoch et al. 1992). The deduced primary structures of D and CE Rh polypeptides bear a high degree of sequence identities (Avent et al. 1990; Cherif-Zahar et al. 1990; Le Van Kim et al. 1992; Arce et al. 1993) and show a correlation with the Rh-antigenic phenotypes (Mouro et al. 1993). As suggested by hydropathy-plot analysis, these proteins share an overall similar, if not identical, membranous organization that is characterized by 12 spans with short peptides being exposed on either side of the lipid bilayer. However, such a topological model provides no evidence for the immunopotency of Rh antigens, because of lack of molecular detail concerning the size, location, and structure of the particular epitopes.

In order to gain insight into the basis for Rh-antigen

Received April 5, 1996; accepted for publication July 25, 1996.

Address for correspondence and reprints: Dr. Cheng-Han Huang, Lindsley F. Kimball Research Institute, New York Blood Center, 310 East 67th Street, New York, NY 10021.

0002-9297/96/5904-0012$02.00

Table 1

Rh Phenotyping and Association of Evans and Cataract Traits in the A.T. Family

Subject (Sex)	Relation	Rh Antigen Status[a]					Evans Status[a]	Cataract Association[a]
		D	C	E	c	e		
C1[b]		+	+	−	−	+	−	−
C2[c]		−	−	−	+	+	−	−
I-1 (F)	Grandmother	+	−	−	+	+	+	+
I-2 (M)	Grandfather	+	−	+	+	+	−	−
II-1[d] (M)	Son	+	−	+	+	−	+	+
II-2 (M)	Son	−	−	−	+	+	−	−
II-3 (M)	Son	+	−	+	+	+	−	−
II-4 (F)	Daughter	+	−	−	+	+	+	+
II-5 (M)	Son	+	−	+	+	−	+	+
III-1 (F)	Granddaughter	+	+	−	−	+	+	+

[a] A plus sign (+) denotes presence; and a minus sign (−) denotes absence.
[b] Control 1, Rh positive (genotype *DCe/DCe*).
[c] Control 2, Rh negative (genotype *dce/dce*).
[d] Propositus.

expression, molecular genetic approaches are being used to define the underlying mechanisms for the different Rh phenotypes. Here we describe the identification of a three-generation family in which two unrelated traits, the red-cell Evans phenotype (also known as "D· ·") and the inheritable form of cataract, are cotransmitted. We report the genomic structure and transcript expression of *Evans*, a gene that specifies the codominant trait associated with both qualitative and quantitative changes in D-antigen expression. Our results show that this trait arose from a genetic recombination event at the human *RH* locus, in which exons 2–6 of the *CE* gene have been lost and replaced by a segmental transfer of the homologous counterpart of the *D* gene.

Subjects, Material, and Methods

Blood Samples and Rh Phenotyping

Ten individuals were examined in this study, and their blood-group phenotypes were determined by standard serological tests. Two samples, one each obtained from unrelated human blood donors with the Rh-positive and Rh-negative phenotypes, were used as controls (Huang et al. 1996). The remaining eight samples were collected from members of a three-generation family (designated "A.T.") of caucasoid origin. In this family, the red-cell Evans-positive phenotype (also known as "D· · phenotype") and a congenital form of cataract were cotransmitted (table 1).

Isolation of Nucleic Acids

Erythroid total RNAs were prepared from hemolysates by preferential lysis of peripheral reticulocytes in NH_4Cl solution, followed by acetic acid titration of the recovered supernatant (Goossens and Kan 1981). The resultant RNA pellet was then solubilized and extracted with Trizol reagent, as recommended by the manufacturer (Bethesda Research Laboratories). The leukocytes pelleted in NH_4Cl solution were washed several times with TBS and were used to isolate high-molecular-weight genomic DNAs, as described elsewhere (Huang et al. 1991).

DNA Probes and Synthetic Primers

The genomic and cDNA probes specific for the Rh-polypeptide genes were as described elsewhere (Huang et al. 1995, 1996), and their nucleotide positions were numbered according to the distance from the first base of the initiation codon ATG. The three Rh cDNA probes encompassed the 5′ (exons 1–3, nt 1–480), the middle (exons 4–7, nt 515–1073), and the 3′ (exons 8–10 plus 3′ UTR, nt 1074–1456) portions. Probes for the particular exons were generated by specific amplification of Rh cDNAs and then were purified by native PAGE. Synthetic oligonucleotide primers (see below) used in genomic amplification and reverse transcriptase–PCR (RT-PCR) were synthesized on a 380A automated DNA synthesizer and were purified by denaturing PAGE.

Mapping of the RH *Locus by Southern Blot Analysis*

Genomic DNAs from control individuals and Evans family members were digested with the various restriction endonucleases, were fractionated by agarose gel electrophoresis, and were transferred onto nylon membranes by vacuum blotting. Blots were hybridized in a Bellco Autoblot apparatus with the various ^{32}P-Rh probes that were labeled according to the random

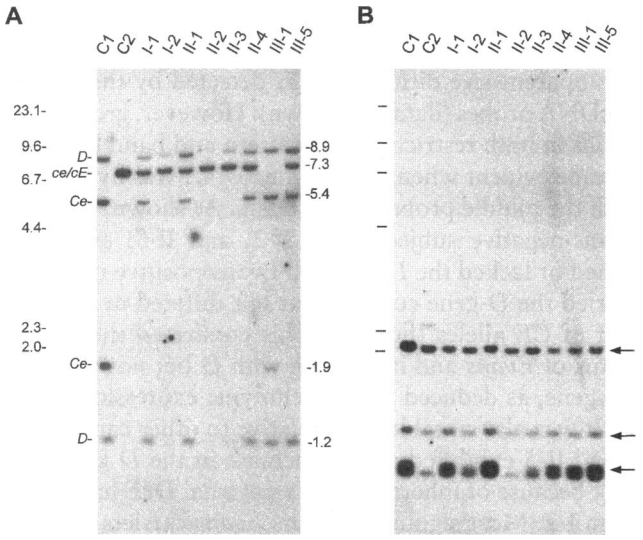

Figure 1 Genomic analysis of the A.T. family by Southern blot hybridization. Genomic DNAs isolated from two control individuals and eight members of the A.T. family were digested with restriction enzymes and hybridized with the Rh cDNA probes. *A, Sph* I blot hybridized with the cDNA probe encompassing exons 4–7, the middle part of the *RH* gene (nt 515–1073). The size marker of λ DNA *Hin*dIII bands and the origin of the various *RH* gene fragments are indicated at the left margin. Note that (1) the 1.9-kb band, but not the 5.4-kb band, is missing in all Evans-positive members except III-1; and (2) the 8.9-kb band, but not the 1.2-kb band, of *D* shows a change of intensity among Evans-positive members. *B, Msp*I blot hybridized with the same Rh probe as above. The bands showing increased and decreased intensities in the *D* and *CE* genes, respectively, are marked by arrows at the right margin. The upper and middle arrows point to the comigrating bands from both *D* and *CE*, whereas the lower one denotes the *D*-specific band.

primer-extension technique (Feinberg and Vogelstein 1984).

Genomic Amplification by Allele-Specific PCR and Restriction Analyses

RH genomic sequences were amplified by PCR using the thermostable DNA polymerase *Taq* (Saiki et al. 1988) and then were analyzed by restriction digestion. As template DNAs, the 5.4-kb *Sph*I fragments (fig. 1*A*) were recovered by electroelution, as described elsewhere (Huang and Blumenfeld 1991). PCR was run for 30 cycles in 50 μl of reaction volume containing the forward and reverse primers (for sequence and specification of the two primer sets, see the legend to fig. 2). The time and temperature segments were programmed as follows: DNA denaturation at 94°C for 60 s, primer annealing at 55°C for 45 s, and chain extension at 72°C for 30 s. The last polymerization step at 72°C was prolonged to 7 min.

Reverse Transcription and Amplification of Rh cDNAs by RT-PCR

Rh polypeptide mRNAs were converted into single-stranded cDNAs and then were amplified by PCR, as described elsewhere (Huang et al. 1995). In brief, 2–3 μg of erythroid total RNA, together with 25–30 ng of the cDNA synthesis primer, were incubated at 65°C for 5 min and were allowed to anneal on ice for 5 min. The reaction volume was brought to 20 μl by addition of 15 units of AMV reverse transcriptase, 1 × cDNA synthesis buffer, 20 units of RNase inhibitor, and water. Reverse transcription of cDNA was carried out at 42°C for 75 min and then the enzyme was inactivated at 72°C for 10 min. One fourth, or 5 μl, of the cDNA solution was amplified for 35 cycles after its addition to 45 μl of a premade PCR mixture containing 1 × reaction buffer, 2.5 units of *Taq* DNA polymerase (AmpliTaq; Perkin-Elmer), and 250 ng each of the amplimers. The cycling conditions were similar to those described above, and the amplification products were analyzed by electrophoresis and restriction digestion. The two primers for cDNA synthesis, *q* and *r*, were located in the 3′ UTR downstream of the homology breakpoint nt 1358 and were specific for *D* and *CE* genes, respectively (Avent et al 1990; Le Van Kim et al. 1992). The nucleotide sequence of *q* is 5′-GTATTCTACAGTGCATAATAA-ATGGTG-3′ (antisense, nt 1432–1458), whereas that of *r* is 5′-CTGTCTCTGACCTTGTTTCATTATAC-3′ (antisense, nt 1363–1388). The primer sets for cDNA amplification were upstream of the aforementioned

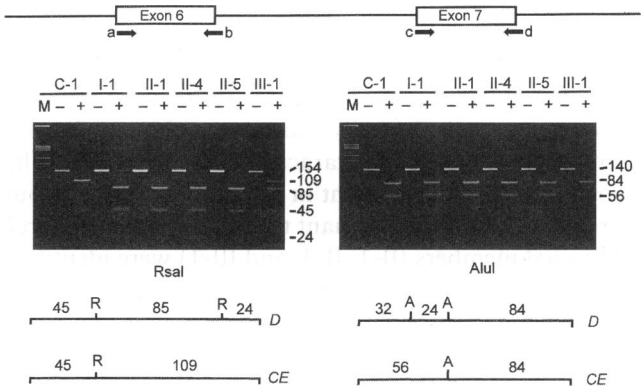

Figure 2 Identification of the 5.4-kb fragments as a *D-CE* hybrid by allele-specific PCR and restriction analysis. *Top,* 5.4-kb *Sph*I fragment spanning exons 6 and 7. The electroeluted 5.4-kb bands were used as templates to amplify the exons with the following primer sets. Set 1 (for exon 6): a, 5′-ACTTATGTGCACAGTGCGGTG-TTGG-3′ (sense) and b, 5′-TG TCTAGT TTCTTACCCGGCAGG-3′ (antisense). Set 2 (for exon 7): c, 5′-GGATTCACCACATCTCCG TCATGCA-3′ (sense) and d, 5′-GTGACCCACATGCCATTGCCG-TTC-3′ (antisense). *Middle,* 7% PAGE analysis of the amplified DNA products cleaved with *Rsa*I and *Alu*I. Lane M is the size marker of *Hae*III-cleaved ϕ174 DNA. A minus sign (−) denotes uncut; and a plus sign (+) denotes cut. The size of the various fragments (in bp) is indicated at the right margin. *Bottom,* Diagram for the distribution of *Rsa*I and *Alu*I cleavage sites expected of the *D* and *CE* exons. Note that the exon 6–containing fragments showed three patterns in *Rsa*I digestion, while the exon 7–containing fragments were cut only once by *Alu*I in all individuals.

cDNA synthesis primers but common to *D* and *CE* genes: set Ia, 5′-ATGAGCTCTAAGTACCCGCGG-TCTG-3′ (sense, nt 1–25 of exon 1), and Ib, 5′-TGG-CCAGAACATCCACAAGAAGAG-3′ (antisense, nt 640–663 of exon 5); set IIa, 5′-CCAAAATAGGCT-GCGAACACGTAGA-3′ (sense, nt 515–539 of exon 4), and IIb, 5′-TTAAAATCCAACAGCCAAATGAGG-AAA-3′ (antisense, nt 1228–1254 of exon 10); and set IIIa, 5′-ATCGCTCCCTCAAGCCCTCAAGTA-3′ (sense, nt −86 to −64 of exon 1) and IIIb, 5′-CTTTTG-ATCCTCTAAGGAAGCGTCATA-3′ (antisense, nt 100–126 of exon 1).

Nucleotide Sequence Determination by Cycle Sequencing

Amplified Rh cDNA and genomic DNA products were separated by native 5% PAGE and eluted with 1 × TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). The purified templates were cycle-sequenced for both strands on a 373A automated DNA sequencer (Applied Biosystem) using fluorescent dye tags as chain terminators. The deduced amino acid sequence of the Evans protein was analyzed by hydropathy plot to assess the arrangement of transmembrane segments, as described elsewhere (Kyte and Doolittle 1982).

Results

Rh Phenotyping and Cotransmission of Evans and Cataract Traits in the Family

Table 1 summarizes the family data on Rh typing and the status of Evans and cataract traits. It became evident from the pattern of family inheritance that subject I-1 carried both Evans and cataract-causing mutations. The two mutations were present in a heterozygous state but were expressed as codominant traits. In the family, three additional members (II-1, II-4, and III-1) were identified as exhibiting both Evans phenotype and congenital cataract (table 1). Subject II-5 also showed the two same phenotypes, although his cataract trait was late onset and developed in the course of chemotherapy. With regard to its relation to Rh antigens, Evans cosegregated with D but not with CcEe (table 1). This suggested that *Evans* is linked with *D* but is situated at the *CE* locus, producing no CcEe antigens. These findings, together with the cotransmission of the *Evans* and *cataract* genes, suggested that both congenital mutations could be localized to chromosome 1. Nevertheless, the character of this cataract mutation and its identity with the lamellar form that maps to chromosome 1 are not known (McKusick 1992, pp. 196–199).

Identification of Genomic Alteration by Southern Blot Analysis

To determine whether the *RH* locus harbored structural changes, genomic DNAs from the A.T. family members were analyzed by Southern blot hybridization. Except for a varied intensity in some restriction bands, no apparent size difference was detected by the 5′ and 3′ cDNA probes (data not shown). However, gross alterations in both restriction-site pattern and band intensity became evident when the genomic DNA was hybridized with the middle probe of Rh cDNA. As shown, whereas Evans-negative subjects (I-2, II-2, and II-3) either retained or lacked the *D* gene, all Evans-positive members carried the *D*-gene complement but differed in the context of *CE* alleles (fig. 1*A*). This confirmed the carrier status of *Evans* and its linkage with *D* but not with the *CE* gene, as deduced from phenotypic expression (table 1). It was also notable that, relative to other carriers, II-1 and II-5 exhibited a dose increase in the *D* gene (fig. 1*B*), because of inheritance of a paternal *DcE* haplotype from I-2. Most significantly, the *Evans* carriers differed in the banding pattern of 5.4- and 1.9-kb *Sph*I fragments previously assigned to the *Ce* or *ce* allele (Huang et al. 1996). Whereas both bands were seen in III-1, who inherited from her father a *Ce* gene, they segregated in Evans carriers of the first and second generations, who retained the 5.4-kb band but lacked the 1.9-kb band (fig. 1*A*). Together, these findings suggested that the locus responsible for *Evans* is associated with the occurrence of a partial and/or internal deletion in one *CE* gene.

Delineation of the Alteration by Exon Mapping and PCR Amplification

To further dissect the structural alteration, genomic DNAs from Evans carriers were mapped by individual exon-specific probes. In this way, the exon content of the 1.9- and 5.4-kb *Sph*I fragments was delineated. The missing 1.9-kb band contained the coding sequence of exons 4 and 5, whereas the 5.4-kb fragment encompassed exons 6 and 7 (data not shown). To establish the identity of the latter two exons in Evans carriers, the 5.4-kb band was electroeluted, amplified by PCR, and digested with the diagnostic enzymes (fig. 2). With regard to the *Rsa*I digestion of exon 6–containing products, three banding patterns were observed (fig. 2, *left panel*). Compared with the C-1 control, four Evans carriers showed an opposite pattern that was unique to the *D* gene. However, the pattern for carrier III-1 was intermediate, owing to a coexistence of the 5.4-kb fragments that originated from both *Evans* and *Ce* genes (see below). With regard to the exon 7 product, the pattern of *Alu*I cleavage was identical in all individuals, confirming its origin from the *CE* gene (fig. 2, *right panel*). These results pointed to the *D-CE* composite nature of the 5.4-kb band and identified *Evans* as a hybrid *RH* gene rather than a simple DNA deletion.

Characterization of Rh Transcripts Expressed in Erythroid Cells

To investigate the expression of *RH* genes in the A.T. family, the *D* or *CE* cDNA was reverse-transcribed from

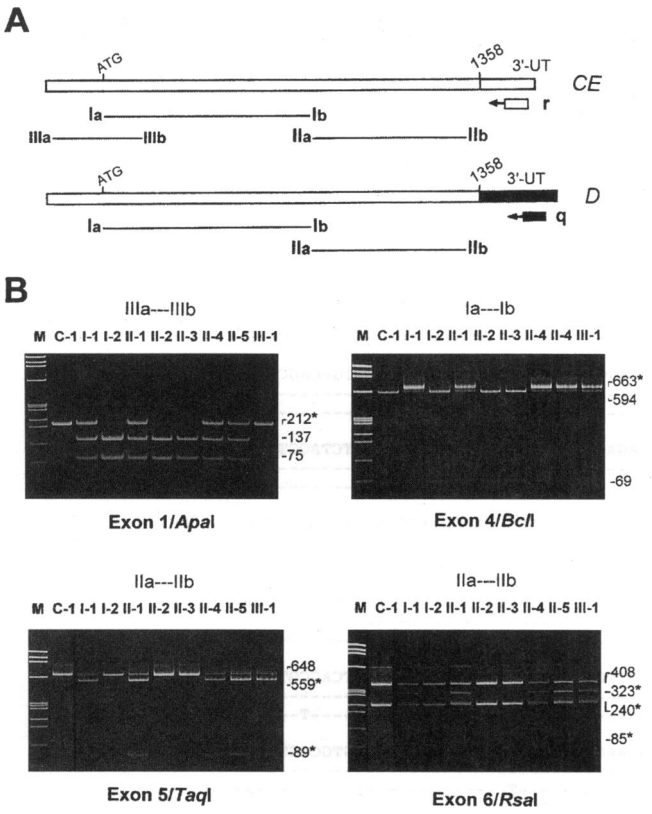

Figure 3 RT-PCR and restriction analysis of Rh transcripts expressed in erythroid cells from the A.T. family members. *A*, Diagram for synthesis and amplification of Rh cDNAs by RT-PCR. Rh mRNAs were converted into *CE* and *D* cDNAs, respectively, by reverse transcription with gene-specific primers that are located in the nonhomologous 3′ UTR (marked by nt 1358). The cDNA products were then amplified with three sets of primers common to *D* and *CE*. *B*, Restriction analysis of the *CE* primer–generated cDNA segments. The cDNA digests were electrophoresed on 5% PAGE and stained by ethidium bromide. The restriction enzyme used and its exon specificity are indicated at the bottom of each gel panel. The size of various fragments is denoted (in bp) and Evans-specific bands are indicated by stars. Note that Evans carrier III-1 showed no *Apa*I cleavage because of the lack of a *ce* or *cE* transcript.

erythroid total RNA with the 3′ UTR primer and then was amplified in three overlapping segments (fig. 3*A*). The composition and identity of the various cDNA forms were analyzed by restriction digestion (fig. 3*B*). All family members except subject II-2 expressed the *D* transcript in their erythroid cells (data not shown). In contrast, the cDNA forms obtained with primer *r* showed a pattern that correlated with the status of Evans and with the allelic expression of CcEe antigens (fig. 3*B*). With regard to the cleavage of exons 1 and 4–6, Evans-negative subjects all showed the same pattern, a pattern that conformed to the origin of their cDNA forms from the *CE* gene. In Evans-positive members, however, the expression of two distinct Rh transcripts was readily shown by the presence of two types of exons, one each specific for *D* and *CE*. These results indicated that Evans is encoded by a unique hybrid mRNA species whose internal part most likely is derived from the *D* gene.

Nucleotide and Deduced Amino Acid Sequences of the Putative Evans *Gene*

To determine the structure of Rh transcripts, both *D* and *CE* cDNA products were cycle-sequenced after gel purification. Figure 4*A* shows the nucleotide sequence of the putative *Evans* gene and compares it with the *D* and *Ce* transcripts. Although *Evans* occurred as the product of a *CE-D-CE* composite gene, it maintained the same open reading frame as did the normal genes. In *Evans,* the partial *D* sequence was flanked by two specific nucleotides, nt 48C and nt 941T, that are characteristic of *Ce* (fig. 4*A*). Apparently, *Evans* gained its exons 1 and 7–10 from a *Ce* allele but gained its internal portion spanning exons 2–6 from the *D* gene. This assignment is in full agreement with the linkage of *Evans* to the *D* locus (table 1 and fig. 1*A*), its cDNA cleavage pattern (fig. 3*B*), and its placement at the *CE* locus, without changing the gene number on the resident chromosome (fig. 5). Together with the finding of a hybrid genomic fragment (fig. 2), these results established the occurrence of *Evans* as a *CE-D-CE* hybrid gene in the *RH* locus, with the formation of two hybrid exon-exon junctions, 5′ *Ce*1-*D*2 and 3′ *D*6-*Ce*7, that differed from either parent *D* or parent *CE* (fig. 6*A*).

Figure 4*B* shows the deduced amino acid sequence of the Evans protein that has the same total number of residues as do the common Rh polypeptides. As suggested by hydropathy-plot analysis, the Cys residue at position 16 encoded in the 5′ *Ce*1-*D*2 junction would be embedded in the membrane, whereas the signature sequence, 308GAKYLPVCCN317, specified by the 3′ *D*6-*Ce*7 junction would be exposed on the fifth extracellular loop of the Evans protein (fig. 6*B*). Because no other structural changes were found in the Rh transcripts from either Evans-positive members or Evans-negative members, the functional expression of such a fusion protein most likely is the direct cause of the Evans phenotype. Interestingly, Evans carried neither C nor e antigenicity (table 1), although its amino terminus from residues 1–120 was identical to that of Ce protein (Mouro et al. 1993). This could be attributed to both altered protein conformation and lost linear sequences that are required for epitope exhibition.

Genotypic Status of the RH *Locus in Members of the A.T. Family*

Owing to the complexity of Rh antigens and uncertainty in Evans linkage, the assortment of *RH* haplotypes in the family could not be deduced from the results of phenotyping. Nevertheless, the combined data of ge-

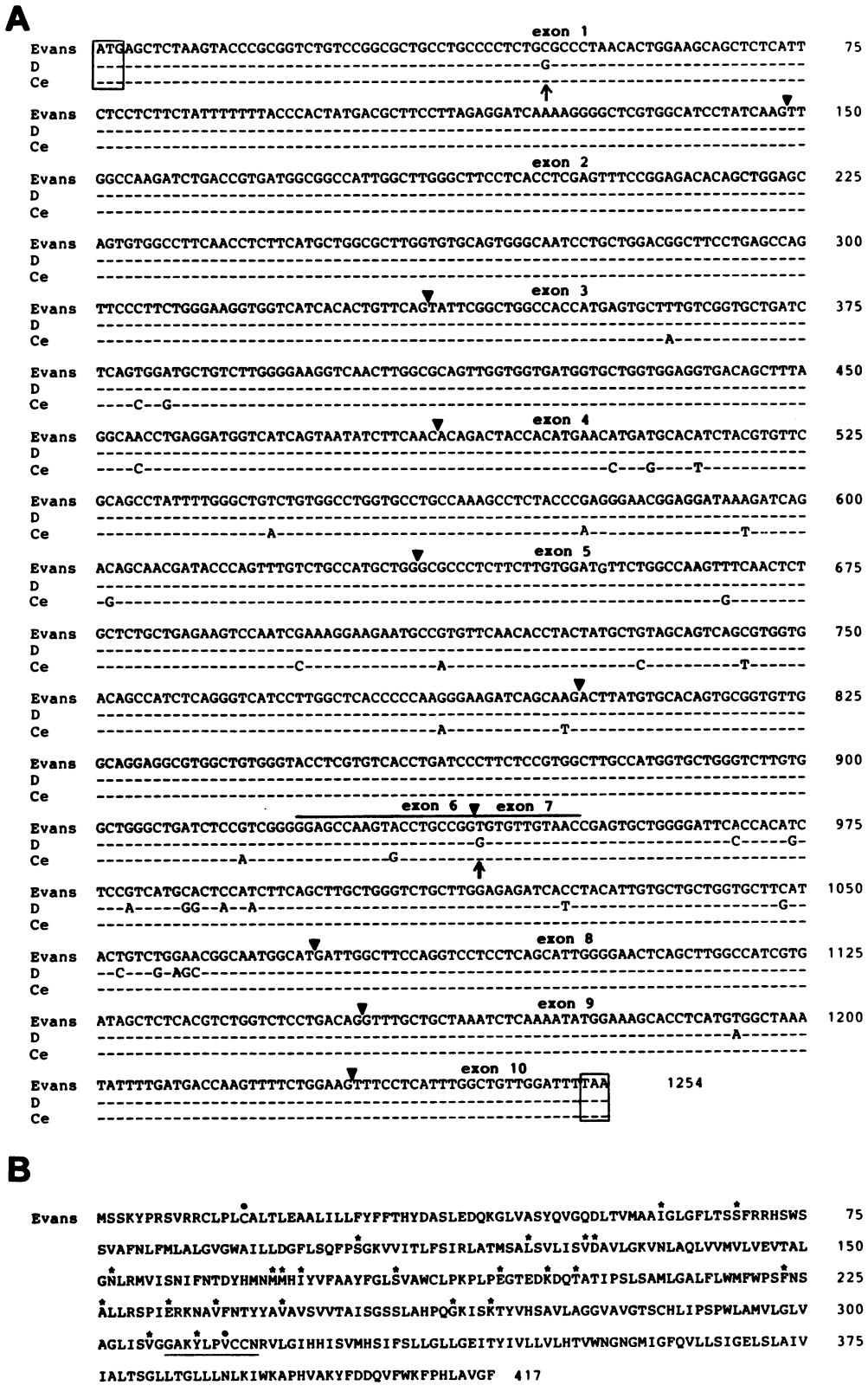

Figure 4 Nucleotide sequence of *Evans* and primary structure of Evans protein. *A*, Alignment of nucleotide sequences of Evans (Ev), D, and Ce cDNAs. The initiation and termination codons are boxed. Exon-exon boundaries are denoted by triangles. Different nucleotides are spelled out, whereas identical ones are indicated by dashes. The two nucleotides characteristic of *Ce*, nt 48C in exon 1 and nt 941T in exon 7, are marked by vertical arrows. The exon 6–exon 7 hybrid junction of *Evans* that differs from either *D* or *Ce* is overlined. *B*, Deduced amino acid sequence for Evans protein. Note that the length of Evans protein is identical with that of the D or Ce protein (not shown). The amino acid residues in the middle portion of Evans that are characteristic of the D protein are indicated by stars. The Evans signature sequence from residues 308–317 for the hybrid junction is underlined. Cys16 and Val314 are dotted to mark the amino and carboxy terminal borders with respect to the Ce protein.

nomic and transcript analyses, shown above, allowed an unequivocal determination of the *RH* genotypic status on all family members. Figure 5 provides a molecular account of the family inheritance, based on the transmission of three haplotypes—*DEvans* (*DEv*), *dce*, and *DcE*—from the first generation to their offspring. This model was supported by two key findings, the linkage of *Evans* (*Ev*) with *D* in the form of a hybrid *CE-D-CE* locus and the detection of subject II-2 as a Rh-negative homozygote (*dec/dce*) lacking both the *D* gene and the *Evans* gene. The recognition of two haplotype combinations—*DEv/dce* for I-1 and *DcE/dce* for I-2, explained the *Sph*I RFLP segregation pattern (fig. 1*A*) and correlated with the expression of Rh phenotypes in family members (table 1).

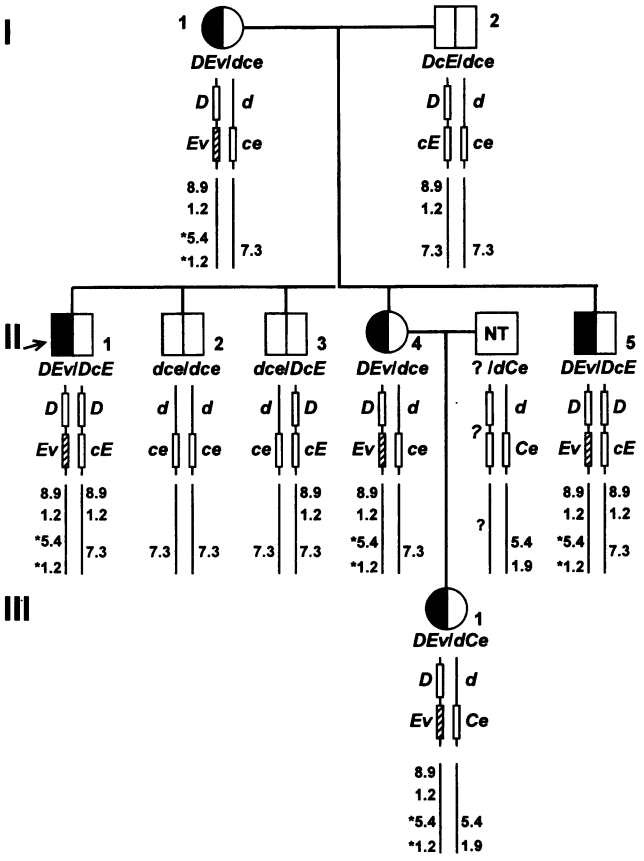

Figure 5 Genotypic status of *RH* locus and transmission of *Evans* in A.T. family members. Evans carriers who also carry the cataract mutation are indicated by half-filled circles (for female) or squares (for males). An arrow indicates propositus, i.e., carrier II-1. For each family member, the *RH* structure and linkage on the haploid genome are illustrated by two vertical bars. The italicized letters *D, ce, cE, Ce,* and *Ev* (*Evans*) stand for the respective genes. The *Evans* gene that occupies the *CE* locus is shaded. The *D*-negative status caused by a deletion of the *D* gene is denoted by *d* along with a straight line. Below the diagram of each genotype is also shown the distribution of *Sph*I RFLPs encompassing exons 4–7 of the *RH* genes. The *Evans*-specific 5.4- and 1.2-kb bands are marked by stars. NT, subject not tested; one of his haplotypes was unknown (?), but the other haplotype could be inferred from carrier III-1.

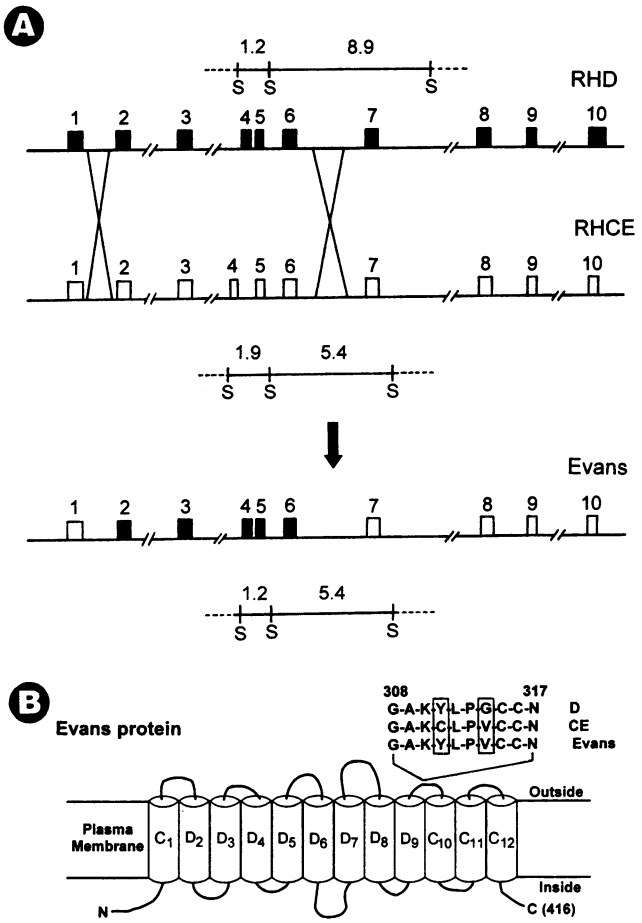

Figure 6 Model for the Evans phenotype: genetic recombination and protein structure. *A*, Homologous alignment of the donor *D* gene and the recipient *CE* gene. Introns not drawn to scale are denoted by broken lines. The *Sph*I fragments (in kb) characteristic of each gene are illustrated. A segmental transfer from *D* to *CE* covering exons 2–6 is indicated by two crosses localized to introns 1 and 6. This event results in a gene bearing the *CE-D-CE* hybrid structure and relocation of the *Sph*I cleavage sites. The *Evans* gene gained the 1.2-kb *D*-specific band and the 5.4-kb *D-CE* hybrid fragment, as the *Sph*I site downstream of exon 7 is unevenly positioned in the parent genes. *B*, Topological representation of the Evans protein. The signature sequence from amino acid residues 308–317 encoded by the hybrid junction of *D* exon 6 and *CE* exon 7 is illustrated. This sequence is predicted to reside in the fifth extracellular loop. The origin of each membrane-spanning segment is shown. The carboxy terminal amino acid is numbered 416, as the first methionine residue is cleaved off from mature protein.

Discussion

In this paper, we have described studies of a Caucasoid family in which two unrelated but inheritable codominant traits—the red-cell Evans phenotype and a congenital form of cataracts—are transmitted. The mutations

underlying these two traits were coexpressed in five members, and both of them were present in a heterozygous or carrier state, as inferred from the pattern of family inheritance. With regard to the congenital cataracts, they are highly heterogeneous and occur mostly with no known etiology (McKusick 1992, pp. 196–199); nevertheless, several types have been molecularly defined. For example, one is closely linked to the mutations of γ-crystallin genes in the 2q33-36 region (Lubsen et al. 1987), and another is associated with point mutations of the galactokinase gene in the 17q24 region (Stambolian et al. 1995). For the family studied here, both the definition of *Evans* as a hybrid *RH* gene and its cosegregation with the cataract mutation suggest that the latter can be localized to chromosome 1. One congenital form of cataracts has already been linked to the Duffy-antigen locus (Renwick and Lawler 1963), which resides in the 1q22-23 region (Mathew et al. 1994). However, such a linkage in the present family was not informative (data not shown), thus requiring further studies to verify the genetic nature of this cataract mutation.

Our study focused on the molecular genetic aspect of the red-cell Evans phenotype, a low incidence Rh antigen that was thought to result from a deleted *RH* locus similar but not identical to the *D*- - complex (Contreras et al. 1978). Here the molecular basis for the Evans phenotype is defined, showing that it is specified by a *CE-D-CE* hybrid gene located in the *CE* locus and is not associated with *RH*-locus deletion that has been described for the *D*- - variants (Blunt et al. 1994; Huang et al. 1995). With detailed analysis of gene expression, a novel Rh hybrid transcript specifically expressed in the Evans-positive erythroid cells has been identified and characterized. Because the transcript maintains an open reading frame identical with that of the parent *D* or *CE* transcript, it is predicted to be actively translated into a fusion polypeptide, bearing the same total number of amino acid residues as the common Rh proteins. Thus, our study establishes a correlation of the fusion protein with the expression of Evans antigen on the red-cell membrane.

The identification of *Evans* as a hybrid gene clearly points to the occurrence of a genetic recombination event at the *RH* locus. On the basis of the *Sph*I map for *RH* (Huang et al. 1996) and the structure of the Evans transcript (fig. 4), a hypothetical model can be proposed to account for the origin of the Evans phenotype (fig. 6). In forming *Evans*, the transfer of exons 2–6 from *D* to *CE* appeared to have involved two sites for DNA breakage and reunification. Consistent with this, the 5′ and 3′ breakpoints were localized to introns 1 and 6, respectively, by mapping the *Sph*I fragments of *RH* genes. It becomes evident that, by the transfer, the *Sph*I sites of the recombining parent genes were exchanged and relocated in the *Evans* gene (fig. 6*A*). This explains both the nature of the 5.4-kb band as a *D-CE* hybrid fragment (fig. 2) and its segregation from the *CE*-specific 1.9-kb band (fig. 1*A*). Despite the ultimate replacement of *CE* exons 2–6 by the *D* counterpart, such a recombination event was accompanied by no increase or decrease of gene number on the *Evans* chromosome. This implies that *Evans* arose by a directional DNA transfer via the mechanism of gene conversion, although unequal homologous recombination through double crossover could not be excluded formally (Huang and Blumenfeld 1995). Recently, similar events have been described as the molecular basis for the D-category variants (Mouro et al. 1994; Rouillac et al. 1995), which occur as D-CE-D hybrids opposite yet nonreciprocal to Evans. The nucleotide sequence and linkage status of *Evans* (figs. 4 and 5) indicate further that the recipient involved in the recombination event originated from a *DCe* haplotype. Whether *Evans* in other subjects has occurred on the genetic background of different haplotypes remains to be investigated.

At the level of primary structures, Evans protein differs from the D polypeptide mainly in the Cys16Trp replacement and in the sequence encoded by exon 7 (fig. 4). This feature, together with the presence of two hybrid junctions, may well explain the phenotypic association of Evans with D but the dissociation from the CcEe antigens. During the formation of *Evans*, the transfer of a large segment of *D* was expected to bring in some sequences coding for the D antigen–associated epitopes. Such a sequence exchange defines Evans as a partial D variant and explains its association with exalted D expression. Moreover, the exon fusions of *Ce1-D2* and *D6-Ce7* have given rise to new amino acid sequences (fig. 4*B*), which in turn could alter the spatial relationship and cis interaction of the adjacent peptide segments in the plasma membrane. Indeed, the Ce1-D2 fusion sequence makes the Evans protein identical with the Ce polypeptide in the first 120 residues (Mouro et al. 1993), yet in this family no C antigen was detectable on the red cells of Evans carriers from the first and second generations. With regard to the Evans antigen, both the D6-Ce7 fusion sequence and the inherent conformational perturbation may involve its epitope structure and contribute to its qualitative and quantitative changes associated with the D antigen. This hypothesis can be tested by transfection studies with Evans cDNA constructs in an ex vivo cell system (Smythe et al. 1996). It is also of interest to note that the D6-Ce7 fusion sequence of Evans may result from other types of DNA rearrangement, such as either a single unequal crossover via the sixth introns of *D* and *CE* or a segmental transfer of exon 7 from *CE* to *D*. Identification of these potential recombination events should provide additional tests of the

hypothesis and reveal molecular heterogeneities underlying the Evans phenotypic expression.

Acknowledgment

We thank Dr. Olga Blumenfeld, of the Albert Einstein College of Medicine, Bronx, for critical reading and helpful discussion of the manuscript. We are indebted to J. Kay, East of Scotland Blood Transfusion Service, for collecting and typing the blood samples. We are grateful to Bob Ratner, Tellervo Huima-Byron, and Yelena Oskov for photographs. This work was supported by the NIH grant HL54459.

References

Agre P, Cartron JP (1991) Molecular biology of the Rh antigens. Blood 78:551–563

Anstee DJ, Tanner MJA (1993) Biochemical aspects of the blood group Rh (rhesus) antigens. Baillieres Clin Haematol 6:401–422

Arce MA, Thompson ES, Wagner S, Coyne KE, Ferdman BA, Lublin DM (1993) Molecular cloning of RhD cDNA derived from a gene present in RhD-positive, but not RhD-negative individuals. Blood 82:651–655

Avent ND, Ridgwell K, Tanner MJA, Anstee DJ (1990) cDNA cloning of a 30 kDa erythrocyte membrane protein associated with Rh (Rhesus)-blood group-antigen expression. Biochem J 271:821–825

Blunt T, Steers F, Daniels G, Carritt B (1994) Lack of RH C/E expression in the Rhesus D- - phenotype is the result of a gene deletion. Ann Hum Genet 58:19–24

Cartron JP, Agre P (1995) RH blood groups and Rh-deficiency syndrome. In: Cartron JP, Rouger P (eds) Blood cell biochemistry. Vol 6: Molecular basis of major human blood group antigens. Plenum, New York, pp 189–225

Cherif-Zahar B, Bloy C, Le Van Kim C, Blanchard D, Bailly P, Hermand P, Salmon C, et al (1990) Molecular cloning and protein structure of a human blood group Rh polypeptide. Proc Natl Acad Sci USA 87:6234–6247

Cherif-Zahar B, Mattei MG, Le Van Kim C, Bailly P, Cartron JP, Colin Y (1991) Localization of the human Rh blood group gene structure to chromosome region 1p34.3-1p36.1 by in situ hybridization. Hum Genet 86:398–400

Contreras M, Stebbing B, Blessing M, Gavin J (1978) The Rh antigen Evans. Vox Sang 34:208–211

Feinberg AP, Vogelstein B (1984) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 137:266–267

Goossens M, Kan YW (1981) DNA analysis in the diagnosis of hemoglobin disorders. Methods Enzymol 76:805–817

Huang C-H, Blumenfeld OO (1991) Molecular genetics of human erythrocyte MiIII and MiVI glycophorins. J Biol Chem 266:7248–7255

——— (1995) MNSs blood groups and major glycophorins: molecular basis for allelic variation. In: Cartron JP, Rouger P (eds) Blood cell biochemistry. Vol 6: Molecular basis of major human blood group antigens. Plenum, New York, pp 153–183

Huang C-H, Guizzo ML, McCreary J, Leigh EM, Blumenfeld OO (1991) Typing of MNSs blood group specific sequences in the human genome and characterization of a restriction fragment tightly linked to S-s- alleles. Blood 77:381–386

Huang C-H, Reid ME, Chen Y (1995) Identification of a partial internal deletion in the RH locus causing the human erythrocyte D- - phenotype. Blood 86:784–790

Huang C-H, Reid ME, Chen Y, Coghlan G, Okubo Y (1996) Molecular definition of red cell Rh haplotypes by tightly linked *Sph*I RFLPs. Am J Hum Genet 58:133–142

Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132

Le Van Kim C, Mouro I, Cherif-Zahar B, Raynal V, Cherrier C, Cartron JP, Colin Y (1992) Molecular cloning and primary structure of the human blood group RhD polypeptide. Proc Natl Acad Sci USA 89:10925–10929

Lubsen NH, Renwick JH, Tui L-C, Breitman ML, Schoenmakers JGG (1987) A locus for a human hereditary cataract is closely linked to the γ-crystallin gene family. Proc Natl Acad Sci USA 84:489–492

MacGeoch C, Mitchell CJ, Carritt B, Avent ND, Ridgwell K, Tanner MJ, Spurr NK (1992) Assignment of the chromosomal locus of the human 30-kDa Rh (rhesus) blood-group-antigen-related protein (Rh30A) to chromosome region 1p.36.13-p34. Cytogenet Cell Genet 59:261–263

Mathew S, Chaudhuri A, Murty VVVS, Pogo AO (1994) Confirmation of Duffy blood group antigen locus (FY) at 1q22-q23 by fluorescence in situ hybridization. Cytogenet Cell Genet 67:68

McKusick VA (1992) Mendelian inheritance in man, 10th ed. John Hopkins University Press, Baltimore

Mollison PL, Engelfriet CP, Contreras M (1993) Blood transfusion in clinical medicine. Blackwell, Oxford

Mouro I, Colin Y, Cherif-Zahar B, Cartron JP, Le Van Kim C (1993) Molecular genetic basis of the human Rhesus blood group system. Nat Genet 5:62–65

Mouro I, Le Van Kim C, Rouillac C, van Rhenen DJ, Le Pennec PY, Bailly P, Cartron JP, et al (1994) Rearrangements of the blood group RhD gene associated with the D^{VI} category phenotype. Blood 83:1129–1135

Race RR, Sanger R (1975) Blood groups in man. Blackwell, Oxford

Renwick JH, Lawler SD (1963) Probable linkage between a congenital cataract locus and the Duffy blood group locus. Ann Hum Genet 27:67–84

Rouillac C, Colin Y, Hughes-Jones NC, Boelet M, D'Ambrosio AM, Cartron JP, Le Van Kim (1995) Transcript analysis of D category phenotypes predicts hybrid Rh D-CE-D proteins associated with alteration of D epitopes. Blood 85:2937–2945

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Smythe JS, Avent ND, Judson PA, Parsons SF, Martin PG, Anstee DJ (1996) Expression of RHD and RHCE gene products using retroviral transduction of K562 cells establishes the molecular basis of Rh blood group antigens. Blood 87: 2968–2973

Stambolian D, Ai Y, Sidjanin D, Nesburn K, Sathe G, Rosenberg M, Bergsma DJ (1995) Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts. Nat Genet 10:307–312