Human Molecular Genetics, 2001, Vol. 10, No. 22 2557–2567

Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the *NEMO* and *LAGE2* genes

Swaroop Aradhya, Tiziana Bardaro[1], Petra Galgóczy[2], Takanori Yamagata[3], Teresa Esposito[1], Henry Patlan, Alfredo Ciccodicola[1], Arnold Munnich[4], Sue Kenwrick[5], Matthias Platzer[2], Michele D'Urso[1] and David L. Nelson*

Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza 902E, Houston, TX 77030, USA, [1]International Institute of Genetics and Biophysics (IIGB), Via G. Marconi 10, 80125 Naples, Italy, [2]Department of Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse 11, 07445 Jena, Germany, [3]Department of Pediatrics, Jichi Medical School, 3311-1 Yakushiji, Minamikawachi-machi, Tochigi 329-0433, Japan, [4]Department of Genetics, Unité des Recherches sur les Handicaps Génétiques de l'Enfant INSERM-393, Hopital Necker-Enfants Malades, 75015 Paris, France and [5]Wellcome Trust Centre for Molecular Mechanisms of Disease and University of Cambridge Department of Medicine, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2XY, UK

Received July 10, 2001; Revised and Accepted August 27, 2001

DDBJ/EMBL/GenBank accession nos+

The X-linked dominant and male-lethal disorder incontinentia pigmenti (IP) is caused by mutations in a gene called *NEMO* (*IKK-γ*). We recently reported the structure of *NEMO* and demonstrated that most IP patients carry an identical deletion that arises due to misalignment between repeats. Affected male abortuses with the IP deletion had provided clues that a second, incomplete copy of *NEMO* was present in the genome. We have now identified clones containing this truncated copy (Δ*NEMO*) and incorporated them into a previously constructed physical contig in distal Xq28. Δ*NEMO* maps 22 kb distal to *NEMO* and only contains exons 3–10, confirming our proposed model. A sequence of 26 kb 3′ of the *NEMO* coding sequence is also present in the same position relative to the Δ*NEMO* locus, bringing the total length of the duplication to 35.5 kb. The *LAGE2* gene is also located within this duplicated region, and a similar but unique *LAGE1* gene is located just distal to the duplicated loci. Mapping and sequence information indicated that the duplicated regions are in opposite orientation. Analysis of the great apes suggested that the *NEMO/LAGE2* duplication occurred after divergence of the lineage leading to present day humans, chimpanzees and gorillas, ~10–15 million years ago. Intriguingly, despite this substantial evolutionary history, only 22 single nucleotide differences exist between the two copies over the entire 35.5 kb, making the duplications >99% identical. This high sequence identity and the inverted orientations of the two copies, along with duplications of smaller internal sections within each copy, predispose this region to various genomic alterations. We detected four rearrangements that involved *NEMO*, Δ*NEMO* or *LAGE1* and *LAGE2*. The high sequence similarity between the two *NEMO/LAGE2* copies may be due to frequent gene conversion, as we have detected evidence of sequence transfer between them. Together, these data describe an unusual and complex genomic region that is susceptible to various types of pathogenic and polymorphic rearrangements, including the recurrent lethal deletion associated with IP.

INTRODUCTION

Mutations in *NEMO* (*IKBKG*, *IKK-γ*) cause the X-linked dominant disorder, incontinentia pigmenti (IP) (1,2). This disorder is typically lethal in male individuals but female patients survive because cells expressing the mutant X chromosome are selectively eliminated. Thus, skewed X-inactivation is a characteristic of this disorder (3,4). As a regulatory component of IκB kinase, NEMO is responsible for downstream activation of the NF-κB transcription factor. By inducing the transcription of various target genes, the NF-κB signaling pathway regulates immune and inflammatory reactions and prevents apoptosis (5,6). Disruption of NEMO or NF-κB renders cells susceptible to apoptosis, leading to the IP-associated male lethality and skewed X-inactivation in female patients (2). Nearly 70–80% of IP mutations are accounted for by an identical deletion within *NEMO*, which eliminates exons 4–10 (7). This mutation arises due to misalignment between two identical MER67B

*To whom correspondence should be addressed. Tel: +1 713 798 4787; Fax: +1 713 798 5386; Email: nelson@bcm.tmc.edu

+AF277315, AL596249 and AJ271718

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors

sequences (termed *int3h* repeats); one copy is located in intron 3 and another ~4 kb distal to the last exon of *NEMO*.

When the recurrent IP deletion was first identified due to an aberrant fragment on a Southern blot, fragments of normal size were also present (2). This led us to propose that a second copy of *NEMO* (Δ*NEMO*) existed in the genome. In addition, PCR analysis of DNA samples from spontaneous male abortuses with the IP deletion yielded the expected amplification products from exon 2 to 3 and exon 3 to 4, but failed to amplify from exon 2 to 4. These observations supported a model that the second copy of *NEMO* was truncated, lacking the first four exons. The human genome contains numerous examples of gene duplications, several of which are involved in genomic disorders (8). Thus, it was conceivable that rearrangements could occur between the two *NEMO* copies and that such events may have a role in the genetics underlying IP or another human disease due to disruption of genes between them. Rearrangements would be especially likely if the two *NEMO* copies share significant homology, and a preliminary analysis of exons suggested complete identity between *NEMO* and Δ*NEMO*.

We recently constructed a high-density bacterial- and P1-artificial chromosome (BAC and PAC) contig to study the region between *G6PD* and Xqter (9). After another group mapped *NEMO* to Xq28, we sequenced the entire gene from a BAC clone in the contig and showed that it lies head-to-head with *G6PD* and is transcribed in the centromeric to telomeric direction (Fig. 1A and B) (2,10). The 23 kb *NEMO* gene contains 12 exons with three alternative primary exons that independently splice into exon 2, where the initiating ATG codon lies (GenBank accession no. AJ271718). In our initial BAC/PAC contig, a gap existed just distal to *NEMO* and efforts to close it with flanking probes had failed repeatedly. When the idea of a second copy of *NEMO* was proposed, we decided to search the gap region since duplicated copies of genes are often located close to the parent copy, as exemplified by the *IDS* gene in Xq28 (11).

This report describes how we identified Δ*NEMO*, when and how it originated, its current structure and the homology it shares with *NEMO*. The duplication boundaries were cloned and the entire region containing *NEMO* and Δ*NEMO* was sequenced. Interestingly, *NEMO* was part of a larger duplication that originated during evolution of the great apes. The sequence, structure and evolution of this duplication provide significant insight into how the human genome evolves through mechanisms of structural alteration as well as sequence preservation.

RESULTS

Isolation of clones containing *NEMO* and Δ*NEMO*

We previously sequenced the *NEMO* gene (GenBank accession no. AJ271718) from BAC clone RP11-211L10 (Fig. 1A and B), including a 5 kb region downstream that contained the distal *int3h* copy, involved in mediating the recurrent IP deletion (2). When the existence of a second copy of *NEMO* was proposed, there were no clues to its genomic location. BAC and PAC clones containing *NEMO* had been previously isolated from the RP11 male BAC library, and the RP6 female and RP5 male PAC libraries. The clones from the RP11 and RP6 libraries were part of a larger physical contig (9), in which a gap existed between *NEMO* and *DKC1*. Screening the RP11 BAC library with the *NEMO* cDNA identified clones that contained Δ*NEMO* and that closed the gap (Fig. 1A). Multiple RP11 BAC clones (66O13, 196H18 and 515D14) contained Δ*NEMO*, and only one clone (103M23) included both *NEMO* and Δ*NEMO*. The two RP5 clones (865E18 and 1087L19) were isolated separately for large-scale sequencing of the *NEMO* region and together they covered both *NEMO* and Δ*NEMO* (12).

Cloning the duplication boundaries

The model that Δ*NEMO* contained only exons 3–10 suggested that the 5′ boundary of the duplication would be within intron 2 (Fig. 1B and C). Thus, the 5′ breakpoint was detected by hybridizing the *NEMO* cDNA probe on a *Hin*dIII digest of clones containing *NEMO*, Δ*NEMO* or both (Fig. 1D). Examination of the 5′ boundary clone from Δ*NEMO* showed that one end of the clone was within intron 7, as expected, and the other end was upstream of a gene called *LAGE1* (GenBank accession no. AJ223093) (Fig. 1C). This indicated that *LAGE1* was telomeric to Δ*NEMO*, and that Δ*NEMO* was in an inverse orientation relative to *NEMO* (Fig. 1A and C). The 3′ boundary was more difficult to detect because initially there was no sequence information 3′ of *NEMO*. Therefore, we walked distally from *NEMO* by cloning overlapping fragments and using them as probes to identify the 3′ duplication breakpoint. The 3′ boundary was in a 17 kb *Spe*I band in BAC clone RP11-211L10, and in a 28 kb *Spe*I fragment in clones containing Δ*NEMO* (Fig. 1D). Restriction analysis of clones from the probe walking experiments suggested a length of 35 kb for the duplication. We concurrently sequenced RP5 clones 865E18 and 1087L19 (GenBank accession no. AF277315), and RP11 clones 211L10 and 196H18 (GenBank accession no. AL596249). The sequence in this region was initially misassembled with the Δ*NEMO* sequence overlapping that of *NEMO* and leaving a gap at the Δ*NEMO* locus. Once this assembly was rectified, the sequence data confirmed the boundary cloning results and the opposite orientations of the *NEMO* copies. The sequence also revealed that the 5′ boundaries of both copies were located near *Alu* elements and that the 3′ boundaries were within LTR repeats.

Since there was initially a gap at the Δ*NEMO* locus in our physical contig, we suspected that this region might be unstable. Therefore, to ensure that the clones used in elucidating the structures and sequence of this region were faithful to the genome, and had not rearranged, a *Kpn*I digest was performed to detect two fragments that spanned the region between the two *NEMO/LAGE2* copies. Analysis of BAC clones and normal human genomic DNA showed the expected fragment sizes in all samples with the SA63F/SA64R probe (Table 1), indicating that the clones were intact (data not shown). However, it was interesting that clones containing either *NEMO* or Δ*NEMO* terminated in the same region, near the *66O13F/R* marker, suggesting that this region may indeed be unstable (Fig. 1A). The telomeric and centromeric end-sequences of RP11 clones 211L10 and 66O13, respectively, were located very close to each other between the duplicated regions.

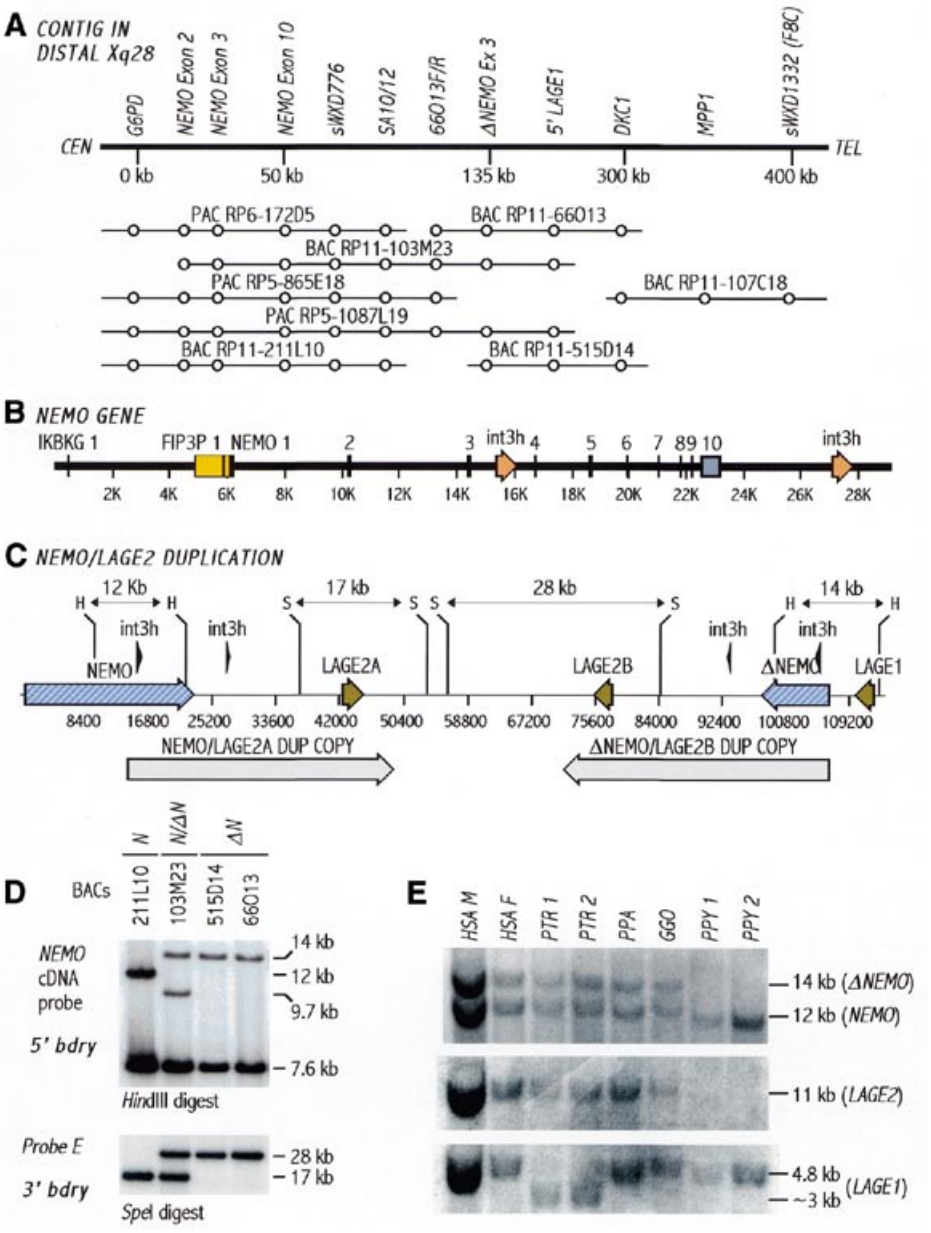

Figure 1. The *NEMO/LAGE2* duplication and its evolution. (**A**) Physical contig in the *NEMO/ΔNEMO* region spanning ~400 kb between *G6PD* and *F8C*. Note that several clones end in the region between *NEMO* and Δ*NEMO*, around the STS markers *SA10/12* and *66O13F/R*. (**B**) The gene structure of *NEMO* (GenBank accession no. AJ271718) with an additional 5 kb at the 3′ end. This was the information available before elucidating the duplication boundaries. The scale begins at nucleotide 22525 of GenBank entry AF277315. The blue boxes/black vertical lines correspond to exons, the yellow box represents a CpG island, and the *int3h* repeats are shown as orange arrows. There are two alternative upstream exons called 1a (IKBKG) and 1b (FIP3P), which independently splice into exon 2. (**C**) Structure of the *NEMO/LAGE2* duplication. The grey arrows below delineate the 35.5 kb duplicated regions. Note that Δ*NEMO* is smaller compared to *NEMO* and that its 5′ is at the telomeric end. The *int3h* copies that mediate the deletions in *NEMO* and Δ*NEMO* are shown as black arrows above the genes. (**D**) Southern blot analyses to identify the 5′ and 3′ boundaries of the *NEMO/LAGE2* duplications. Hybridization of a *Hin*dIII digest with *NEMO* cDNA showed that the 5′ duplication breakpoint in *NEMO* was 12 kb and that in Δ*NEMO* was larger at 14 kb. A 7.6 kb band, corresponding to the 3′ end of *NEMO*, was present in all clones, confirming that this region was also identical in Δ*NEMO*. Clone 103M23 has a shorter fragment of 9.65 kb instead of 12 kb because this clone does not contain the entire *NEMO* gene; the 5′ end-sequence (GenBank accession no. AQ311627) of this clone lies within intron 2 of *NEMO*. Hybridization of a Southern blot containing various digests with a *Kpn*I probe (probe E) located ~15 kb from *NEMO*, revealed the 3′ duplication boundaries. The 3′ boundary of the *NEMO*-containing copy was in a 17 kb *Spe*I fragment and that of the Δ*NEMO*-containing copy in a 28 kb *Spe*I fragment. *N*, *NEMO*; Δ*N*, Δ*NEMO*. (**E**) Southern blot analysis of *NEMO*, Δ*NEMO* and the *LAGE* genes in great apes. An exon 3 probe from *NEMO* on a *Hin*dIII digest detects the expected 12 kb fragment in human (*Homo sapiens*, *HSA*), common chimpanzee (*Pan troglodytes*, *PTR*), bonobo chimpanzee (*Pan paniscus*, *PPA*), gorilla (*Gorilla gorilla*, *GGO*) and orangutan (*Pongo pygmaeus*, *PPY*). The same probe fails to detect the 14 kb band corresponding to Δ*NEMO* in *PPY* while the other samples are positive. A *LAGE* probe that detects both the *LAGE1* and *LAGE2* genes hybridizes the expected 4.8 kb *Eco*RI *LAGE1* fragment in all samples, except in *PTR*, which appears to have a smaller, ~3 kb fragment. Only *PPY* lacks a signal corresponding to *LAGE2*, as indicated by the absence of the 11 kb *Eco*RI band in the middle panel.

Table 1. Primer sequences

Primer	Sequence (5′→3′)	Size (bp)	Location/purpose
LAGE1-F	GGATGGAGTATTGCAGCCTCTCGCC	529	Map location of *LAGE1*
LAGE1-R	GGTAGCTGGAACTGCATGTCTGGTGG		
SA15F	CTTGGCACATCACTTATCAG	5743 with NEMO-5R	Unique forward primer for *NEMO*
SA65F	GCAGAGTCAGCCGTGTGGAA	5261 with NEMO-5R	Unique forward primer for Δ*NEMO*
NEMO-5R	TGAAGGAGTCCTCAGAAGGCA		
NEMO3′-R1	ATGGTGCCTGACTGAGCTGG	3596 with SA65F	Verify Δ*NEMO int3h* deletion
SA25F	CCAGTGGCAGGGAAGAGC	995	Verify *int3h* duplication
INT3H-R2B	TGCAGATCAAGAGTGGACACTG		
SA63F	GCATTAGAGGAAGGCCGTGC	1711	Probe for *Kpn*I fragments between *NEMO* and Δ*NEMO*
SA64R	AGAGACTGGGCTGGGATTCTGT		
SA81F	AGAGGCTGCTGATTCCTGGA	1475	Verify *LAGE1–LAGE2A* inversion
SA82R	CTTGACCCTGTTGTCTCCGTC		
NEMO-10F	CGGCGGCTCCTGGTCTTACA	4576	Prepare probe A
SA25R	TCTTCTGTGCCAGTGGCAGG		
SA54F	CTCAAACATGCAGAGTGGTCGT	1000	Prepare probe telomeric to Δ*NEMO*
SA55R	AGCGATCTTCTTGCCTCGG		
LAGE-probeF	CAGAACAGAGACCTGGTACACCA	564	Hybridize *LAGE1* and *LAGE2* genes
LAGE-probeR	ATGTCCACAGTAGTTAATGCTGAATG		
hF50-F	CGGCGTGCTTATCATTAC	223	Isolate clones with *NEMO*/Δ*NEMO*
hF50-R	CCGAGAAAGTCCCAGTTT		
NEMO-3F	CCCAGCTCCCCTCCACTGT	293	Exon 3 probe for *NEMO*/Δ*NEMO*
NEMO-3R	ACACTGGCGTCACCGCGGGT		
NEMO-4R	CCGGTCTATCCTCATCAAGGA	2621 with NEMO-3F	Probe for *NEMO*/Δ*NEMO* intron 3

Structure and sequence of duplicated region

A BLAST search with the duplicated sequence identified the 3 kb *LAGE2* gene (GenBank accession no. AJ275978). Hence, exons 3–10 of *NEMO* and the entire *LAGE2* gene were both part of the duplication (Fig. 1C). The two copies derived from the duplication are hereafter termed '*NEMO/LAGE2*' copies. In each *NEMO/LAGE2* copy, the *NEMO* sequence occupied only 26% and *LAGE2* accounted for 8%; the remaining 66% of sequence consisted of non-coding DNA, repeats and regulatory regions. Sequence comparison of the two copies revealed near complete conservation. Examining the coding regions of *NEMO* and *LAGE2* between the two copies failed to reveal variations. However, outside the coding regions, 22 single nucleotide differences and two complex variations were found (Table 2). The complex polymorphisms included the presence or absence of multiple nucleotides within repeats of single- or 10-base units. Thus, the degree of sequence identity between the *NEMO/LAGE2* copies is >99%. Eight of the 22 single nucleotide differences were found either in RP5 clones or in RP11 clones but not in both libraries. Hence, these single nucleotide differences might be polymorphisms between individuals rather than between *NEMO* and Δ*NEMO* since the two libraries were prepared from different individuals. Eleven of the 22 single nucleotide differences were also within intronic or 3′-untranslated regions (3′-UTRs) of the *NEMO* sequence (Table 2). The duplications spanned a length of 35 470 bases, with the 3′ duplication junction located 26 306 bases from *NEMO* exon 10. The addition polymorphisms (nos 1, 2, 4, 6, 8, 9, 17, 19 and 23; Table 1) increased the length of the duplication to 35 518 bases. An intervening segment of 21 761 bases between the two copies did not appear to contain any genes.

Southern blot analysis of *NEMO* and *LAGE* genes in primates

DNA samples from humans and other great apes were analyzed by Southern blotting with probes for the *NEMO* and *LAGE* genes. All samples, except orangutan, showed the two expected bands that represent *NEMO* and Δ*NEMO* on a Southern blot hybridized with a *NEMO* exon 3 fragment (Fig. 1E). Orangutan only exhibited a fragment representing *NEMO* but not one corresponding to Δ*NEMO*. A second probe immediately telomeric to Δ*NEMO* also confirmed the expected band size in all samples except orangutan, which showed a smaller band (data not shown). A *LAGE* probe designed to detect both copies of *LAGE2* and the single copy of *LAGE1* hybridized the expected fragments in human, bonobo chimpanzee and gorilla (Fig. 1E). The common chimpanzee had a normal *LAGE2*-containing fragment but showed a smaller band for *LAGE1*. Orangutan did not exhibit a band corresponding to *LAGE2*.

Sequence exchange between the *NEMO/LAGE2* copies

Since the *NEMO/LAGE2* copies are in opposite orientation, we hypothesized that inversions might be responsible for their

Table 2. Single nucleotide differences between the two *NEMO/LAGE2* copies

Base	Base change with flanking sequence	Location within duplication[b]	Variable restriction sites
Single base differences between *NEMO/LAGE2* copies			
1[a]	GGTTCAGCCCTC G/A AGGCCTGCTTGC	*N/ΔN* intron 3	G-*Mnl*I, *Ava*I, *Tli*I
2[a]	TTAGGAGGCATT C/T TGGGGGCCCCGA	*N/ΔN* intron 4	C-*Bsm*I
3	CCCAGCACAGTA A/G GCGGTCAAGGTG	*N/ΔN* intron 4	None
4[a]	GCACTTGGGGCA G/C CCAGCAGGGCAG	*N/ΔN* intron 4	None
5	CCCCTTCCCCTG A/G CTTCCAGGTCTC	*N/ΔN* intron 4	G-*Bsl*I, *Bsa*JI; A-*Drd*I[c]
6[a]	GGCCGCACCGCA G/T GGTCTGTGGTTC	*N/ΔN* intron 5	G-*Alw*NI; T-*Nla*III
7	CACTGGGGCTCT –/+ (T) AGGGCTGGCCTT	*N/ΔN* intron 6	–(T)-*Bfa*I; +(T)-*Dde*I, *Mwo*I
8[a]	ATGCCGTGGTAG C/T GGCGGCTCCTGG	*N/ΔN* intron 9	C-*Mwo*I
9[a]	CCCGCCTGCCTA G/A CCCAGGATGAAG	*N/ΔN* exon 10 3′-UTR	G-*Mwo*I, *Bfa*I
10	GAGCTGGGTGGC A/C GCTCTTCCTCCC	*N/ΔN* exon 10 3′-UTR	C-*Aci*I, *Eae*I, *Hae*III; A-*Tse*I
11	CGACCCGCCCGC T/C GCTGTGCCCTGG	*N/ΔN* exon 10 3′-UTR	A-*Tse*I, *Alu*I; C-*Eae*I, *Hae*III
12	CACTGCAGCCTT C/G ACCTCCTGAGCT	Intergenic	None
13	TTACTGCTTTGA C/G TTTGGAGTCGTC	Intergenic	C-*Drd*I
14	TCCCCAGCACCC A/G GGCCTTCCTTCC	Intergenic	A-*Bst*NI, *Psp*GI; G-*Xma*I, *Hpa*II
15	GACCTTTCCCCT C/T– CTTCAAGCCAGG	Intergenic	C-*Mnl*I; T-*Mbo*II
16	TCACTGCAACTT A/C CGCCTCCAGGGT	Intergenic	C-*Aci*I, *Bsl*I
17[a]	GGATGGGGCGTG G/A GATGACGGTTCG	Intergenic	G-*Fok*I
18	GCAATGCTCTTA T/C GGCAGTGCCCCACCC	Intergenic	None
19[a]	AGGAATAATGGC C/T CTTCCTGCCGGC	Intergenic	C-*Hae*III; T-*Mbo*I, *Ear*I, *Sap*I
20	TAACACCAGATG C/T GGACTAGTGTGG	Intergenic	C-*Aci*I
21	CGAGGGAGTGGA A/G TAAGGTGGGAAT	Intergenic	None
22	AATTGGATTCGG C/T CAACCCTAGGCA	Distal to *LAGE2*	C-*Eae*I, *Hae*III; T-*Hin*cII
Complex polymorphisms between *NEMO/LAGE2* copies			
23[a]	GTGTGTGTGTGT –/+ (GTGT) ATTTTTTTTTTTTTTTTT –/+ (T) GAGACAGAGTTTTGCTCTTCT	Intergenic	None
24	CTTCTTTCCCTC –/+ (TCTTCGTTCCTTCCTCCCTTCCTTCCTTCCCTCCTTC-CTTTT) TTCCTTATTCCTTCCTTTCCT	Intergenic	None

Refers to bases in GenBank entry AF277315.

[a]Indicates that polymorphism may represent difference between two individuals rather than between the *NEMO/LAGE2* copies because the base alteration was found in a clone from one genomic library (RP5) but not in another clone from a different library (RP11). All of these base substitutions were found in the telomeric copy, except the 23 complex polymorphism, which was in the centromeric copy.

[b]Location indicates position where base alteration was discovered; *N*/Δ*N*, *NEMO* or Δ*NEMO*; intergenic, between *NEMO* coding sequence and the adjacent *LAGE2* gene sequence in each copy.

[c]*Drd*I polymorphism used to test for sequence transfer between *NEMO/LAGE2* copies (Fig. 2).

sequence homogeneity. Alternatively, gene conversion and recombination were considered contributing factors. To detect evidence for sequence exchange, we used a single nucleotide difference between the two copies (Table 1; *NEMO* intron 4), which created a *Drd*I site (Fig. 2). Analysis of RP5 and RP11 clones showed that *NEMO* contained the *Drd*I site and Δ*NEMO* did not. We hypothesized that if sequence exchange occurred between *NEMO* and Δ*NEMO* close to the 5′ duplication boundary, the *Drd*I site should shift back and forth between them. Moreover, observation of the same allele (presence or absence of the *Drd*I site) in both *NEMO* and Δ*NEMO* at a high frequency would support the case for gene conversion. Using genomic DNA from 10 normal male controls, we amplified fragments from both *NEMO* and Δ*NEMO* by PCR between intron 5 and unique sequences outside the *NEMO/LAGE2* copies. Digestion of these products with *Drd*I yielded three combinations (Fig. 2). Six samples had the *Drd*I polymorphism in both *NEMO* and Δ*NEMO*, one sample lacked the *Drd*I site in both copies, and three samples had the *Drd*I site in *NEMO* but not in Δ*NEMO*. We never observed the *Drd*I site in Δ*NEMO* only. This observation of the *Drd*I polymorphism in various combinations within *NEMO* and Δ*NEMO* indicated that sequence exchange occurred between them.

Δ*NEMO* internal (*int3h*-mediated) deletions

The most common mutation in IP patients is an *int3h*-mediated deletion within *NEMO* (2,7). Since Δ*NEMO* also contains two copies of *int3h*, we examined control individuals in an effort to find *int3h* deletions in Δ*NEMO*; these deletions were predicted to be non-pathogenic, and thus polymorphic, since Δ*NEMO* seemed to lack an obvious promoter and was considered non-functional.

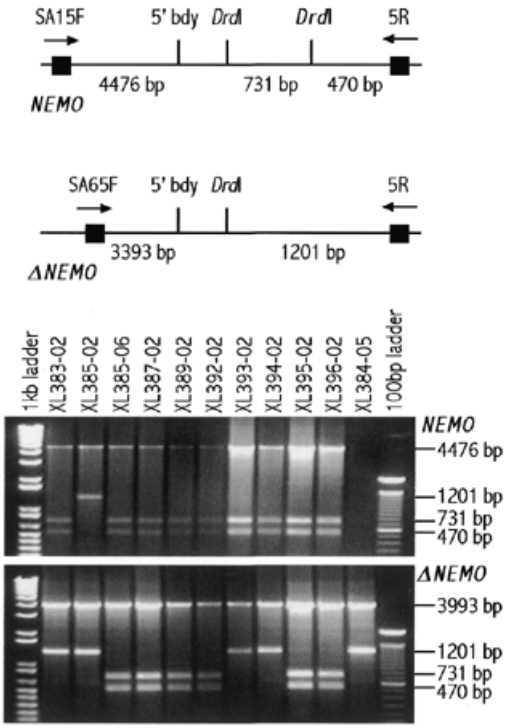

Figure 2. Sequence exchange between duplicated *NEMO/LAGE2* copies. Schematic shows relative location of the polymorphic *Drd*I restriction site in intron 4. Primer SA15F is unique to *NEMO*, and SA65F is located telomeric to Δ*NEMO*. Since both copies have exons 3–10, the NEMO-5R primer in intron 5 is used with the unique primers to generate specific fragments from each copy. Digestion of the resultant PCR products with *Drd*I yields a 5677 bp fragment in *NEMO* and 4504 bp band in Δ*NEMO*. This difference is because SA65F is closer to the 5′ duplication boundary (5′ bdy) than SA15F is. After digestion with *Drd*I, both loci produce a 1201 bp fragment unless the polymorphic *Drd*I site is present and cuts it into 731 and 470 bp pieces. The two panels show digestion of PCR products from *NEMO* (top) and Δ*NEMO* (bottom) in 10 control male individuals (10 unrelated X chromosomes). Note that three combinations are seen—six controls have the polymorphic *Drd*I site in both copies, three individuals have a *Drd*I site in *NEMO* only, and one person does not have the *Drd*I site in either copy. Control XL384-05 has the common IP deletion within *NEMO* and was used as PCR control.

A probe (SA54F/SA55R) was designed to look for alterations of a normal 13.8 kb *Hin*dIII fragment on a Southern blot to an ~10 kb fragment, indicative of a Δ*NEMO–int3h* deletion. Interestingly, analysis of 53 normal individuals (98 total X chromosomes) did not reveal this deletion (data not shown). However, while testing IP patients for the recurrent *NEMO–int3h* deletion with an intron 3 probe, we observed an extra 3.1 kb *Eco*RI band in DNA from two patients (XL203-01 and XL306-02), in addition to the aberrant 2.7 kb band representing the IP deletion (Fig. 3A). A third patient, XL233-01, also showed the 3.1 kb band but had a point mutation in *NEMO* instead of the IP deletion (7). Since the intron 3 probe detected both *NEMO* and Δ*NEMO*, and because an *int3h* deletion in Δ*NEMO* was expected to produce a 3.1 kb *Eco*RI fragment, we tried long-range PCR analysis on DNA from all three patients using primers SA65F and NEMO3′-R1 (Fig. 3B). Successful PCR amplification across the deletion junction and subsequent sequence verification of the PCR product (data not shown) confirmed that the 3.1 kb fragment was due to a Δ*NEMO–int3h* deletion, identical to the recurrent IP deletion in *NEMO*. Examination of over 100 unrelated female IP patients (more than 200 X chromosomes) revealed this rearrangement in only the three families mentioned above, emphasizing that it was rare. Notably, the aberrant band segregated with disease in all three IP families (data not shown), and there was no clinical variation between IP patients carrying only the *NEMO–int3h* deletion and those with both the *NEMO–int3h* and Δ*NEMO–int3h* deletions.

The *int3h* duplication

Given that the *int3h* repeats are oriented in the same direction and deletions are observed between them, we expected to find the reciprocal *int3h* duplications as well. An *Eco*RI digest probed with *NEMO* intron 3 revealed an extra ~26 kb band in one IP patient (XL206-03) who also demonstrated the 2.7 kb aberrant band representative of the recurrent IP deletion (Fig. 3A). The available sequence data suggested that duplication of one of the *int3h* copies would yield a 26 kb fragment. Thus, using PCR primers SA25F and INT3H-R2B, we successfully amplified across the duplication junction and confirmed the result by sequencing (data not shown). Since the *int3h* repeats in the normal *NEMO* gene are located in intron 3 and distal to the last exon (exon 10), the *int3h* duplication replicates exons 4–10 downstream of the normal *NEMO* exon 10 (Fig. 3B). Southern blot and PCR analyses both showed that patient XL206-03 inherited the duplication from her clinically normal father (XL206-01), indicating that this rearrangement was a polymorphism (Fig. 3C). Only patient XL206-03 and one control female individual exhibited this rearrangement from among more than 150 female IP and 48 normal unrelated individuals (approximately 400 total X chromosomes) tested, suggesting that it was very rare. We have not attempted to define whether the *int3h* duplication was within *NEMO* or Δ*NEMO* since the rearrangement would result in identical fragment lengths at both locations.

LAGE1–LAGE2A inversion

While examining 53 individuals (98 X chromosomes) for *int3h* deletions in Δ*NEMO*, DNA from one female IP patient (XL328-04) revealed an ~18 kb *Hin*dIII fragment, in addition to the normal 14 kb band on a Southern blot (Fig. 3A, right). Since the hybridizing probe also detected *LAGE1*, the available sequence supported a hypothesis that an inversion between the near-identical first exons of *LAGE1* and *LAGE2A* would produce an 18 kb *Hin*dIII fragment. A *LAGE1–LAGE2A* inversion would place *NEMO* adjacent to *LAGE1* at the centromeric inversion junction and place *LAGE2A* upstream of Δ*NEMO* at the telomeric end (Fig. 3B). Successful PCR amplification across the inversion junction with primers SA81F and SA82R, and confirmatory sequencing, in a DNA sample from patient XL328-04 confirmed this model (Fig. 3C). Unaffected relatives of IP patient XL328-04 also had the inversion, indicating that it represented a polymorphism. However, since it was found in only one out of 98 X chromosomes examined, it was rather rare.

DISCUSSION

When we initially identified the recurrent deletion in IP, we proposed that a second, truncated copy of *NEMO* was present in the genome (2). This work confirms our hypothesis that

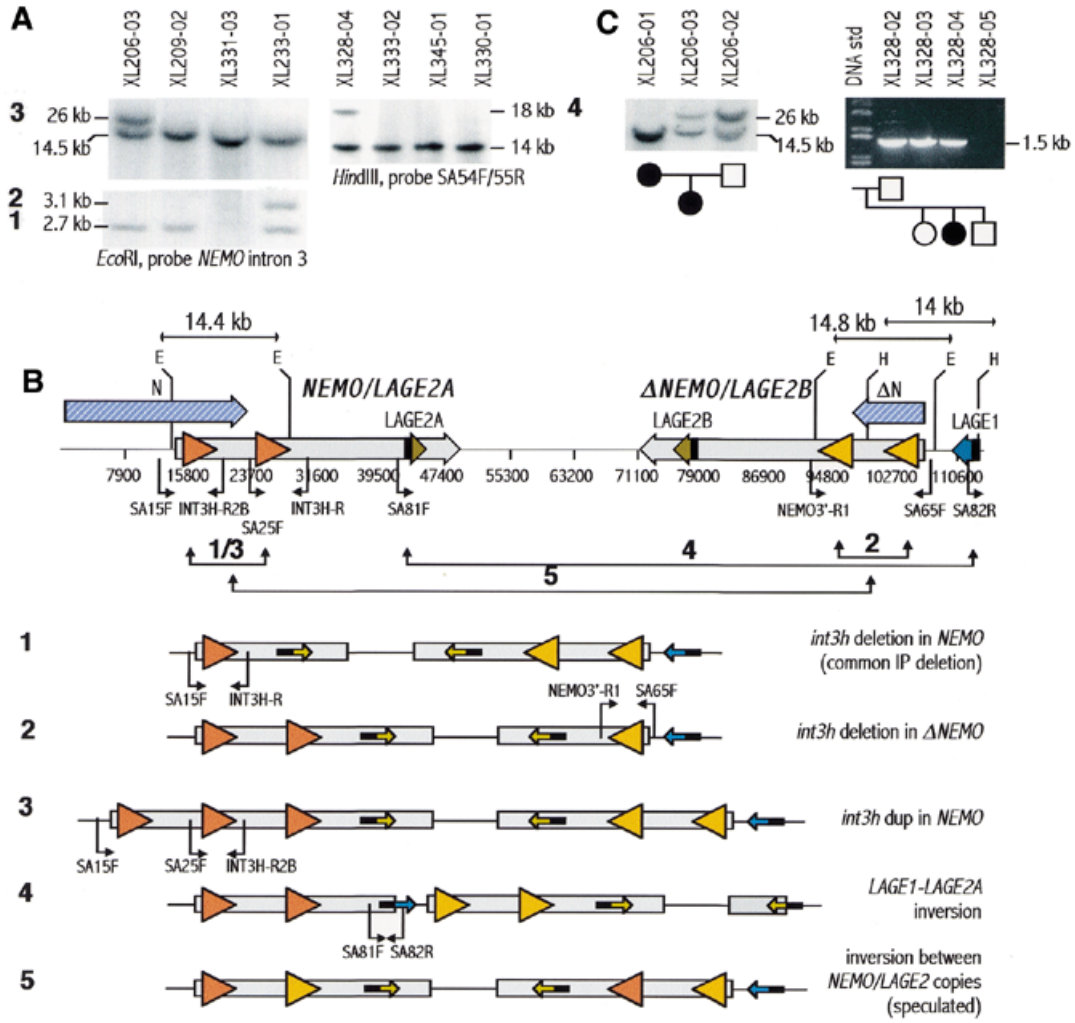

Figure 3. Various rearrangements at the *NEMO/LAGE2* loci (numbered 1–5 in bold). *NEMO* is centromeric to Δ*NEMO* and each copy contains identical *int3h* repeats (red/yellow triangles). Grey boxes mark the duplicated regions. The black portions of the *LAGE1* and *LAGE2* genes indicate sequence conservation between these genes. (**A** and **B**) Aberrant restriction fragments were observed due to genomic rearrangements in IP patients and normal individuals. Schematic representations of all rearrangements are shown below, with numbers corresponding to abnormal restriction fragments above. The common IP mutation, an *int3h* deletion in *NEMO* (1), created a 2.7 kb *Eco*RI band in three DNA samples. A 3.1 kb *Eco*RI fragment corresponded to an *int3h* deletion within Δ*NEMO* (2). The reciprocal rearrangement to the *int3h* deletion is the *int3h* duplication (3), represented by a 26 kb *Eco*RI fragment in DNA sample XL206-03. On the right, DNA sample XL328-04 exhibited an 18 kb *Hin*dIII fragment, representative of an inversion between *LAGE1* and *LAGE2A* (4). Although not experimentally proven yet, inversions between the two *NEMO/LAGE2* copies are likely (5) because of sequence homogeneity. (**C**) The *int3h* duplication and *LAGE1–LAGE2A* inversion are polymorphic, since they are found in normal individuals. Filled circles symbolize affected IP individuals. Left, Southern blot of DNA from family XL206, hybridized with *NEMO* intron 3 (same as *Eco*RI blot in A). Right, long-range PCR with primers SA81F and SA82R in IP family XL328-04 detects the inversion junction fragment.

Δ*NEMO* was located telomeric to *NEMO* and contained only exons 3–10. A section of *NEMO* was duplicated along with *LAGE2* ~10–15 million years ago. The two copies derived from the duplication are in opposite orientation, are nearly identical, and are separated by 22 kb of non-duplicated sequence. We have detected evidence for sequence exchange between the two copies, possibly pointing to a mechanism for maintaining the high sequence identity between them. In addition, various rearrangements arise due to sequence homology between and within the two copies. We detected four different rearrangements, including the recurrent and lethal IP-associated deletion (2).

The *NEMO/LAGE2* duplication was likely mediated by repeats since SINE and LINE sequences are located at its boundaries. These repeat sequences have been known to play a role in duplications and deletions of specific genomic regions (13–15). The *int3h* loci and both *LAGE* genes are also flanked by repeats. Our data suggest that the current structure of the *NEMO–LAGE1* genomic region evolved from two distinct events. The first event produced *LAGE2* from *LAGE1*, since orangutan only contains *LAGE1*. Sequence comparison clearly shows resemblance between the two *LAGE* genes, particularly within coding regions. This first event is somewhat reminiscent of the duplicative evolution of the related *MAGE* genes, also in Xq28 (16). The original *int3h* sequence likely replicated within the ancestral *NEMO* around the same time as, or before, the *LAGE1–LAGE2* duplication. The second event, mediated again by repeats, duplicated part of *NEMO* and the entire *LAGE2* gene together. Inversion of one of the copies may have occurred as a third step or taken place in conjunction with the

NEMO/LAGE2 duplication event itself. Insertion of the second copy, Δ*NEMO/LAGE2*, just telomeric to the first copy might have been facilitated by genomic instability in this region. The interval between the two *NEMO/LAGE2* loci appears to be unstable, since a high-density clone contig in this region initially contained a gap that was eventually closed but with relatively sparse coverage (9). Most of the clones in the map also end between the two copies, with very few containing both loci. It is not unprecedented to find genomic instability associated with gene rearrangements. For instance, the *XAP135* pseudogene is also inserted at an apparently unstable location near the *int22h-2* repeat, which lies ~200 kb from the *NEMO/LAGE2* loci and is involved in some hemophilia A inversions (9).

Many duplicated regions in the genome have originated during evolution of the great apes. For example, the Charcot–Marie tooth disease-linked CMT1A-REP repeats originated before the lineage leading to chimpanzees and humans (17). The *F8C*-associated *int22h* repeat was duplicated prior to orangutan speciation, but a third copy originated in the common ancestor to gorilla, chimpanzee and human (9). Similarly, the *NEMO/LAGE2* duplication occurred before the gorilla–chimpanzee–human lineage. Although both coding and non-coding regions of the genome have undergone duplication, it is interesting that many of them are associated with human diseases. Thus, while duplications predispose genomic regions to both pathogenic intrachromosomal and interchromosomal rearrangements, they have not been selected against, possibly because of some evolutionary advantage. Though functional divergence is often thought to be the outcome of genomic duplications (18), another previously unexplored possibility is that multiple copies of a gene offer a means to prevent sequence alterations of the parent copy through mechanisms such as gene conversion.

The remarkable sequence identity between the two copies of *NEMO/LAGE2* is an unusual occurrence in the genome. However, a few repeat loci are similar to the *NEMO/LAGE2* sequences (8,19), including the 9.5 kb *int22h* and the 24 kb CMT1A-REP loci, which have maintained significant (>98%) sequence homology between copies. The Hunter syndrome-associated *IDS* gene has also been partially duplicated nearby and both copies share >88% identity (11,20,21). Gene conversion has been proposed as a sequence conserving mechanism in all these cases, and supporting evidence has been presented for both the CMT1A-REP repeats and the *IDS* loci (15,21,22). Our finding that the *Drd*I polymorphism can be present in either copy of *NEMO* provides similar evidence for sequence exchange. However, we also considered that inversions between Δ*NEMO* and *NEMO* might help maintain sequence identity; base changes in Δ*NEMO* that are transferred to *NEMO* by inversion would likely be eliminated from the gene pool since most sequence alterations in *NEMO* are lethal (7). However, we had not anticipated that *NEMO* would only account for 26% of the entire duplication. Thus, inversions are less plausible but cannot be excluded; an inversion between the *NEMO/LAGE2* copies would likely be non-pathogenic since all coding regions would be reconstituted. This scenario is not unprecedented; two 11.3 kb, inversely oriented repeats with >99% homology flank the *FLN1* and *EMD* loci just upstream of *NEMO*, and they facilitate frequent inversion of the 48 kb intervening region (23).

Another reason for the sequence conservation between the two copies might be that the Δ*NEMO/LAGE2* copy has important biological significance. Since it lacks a promoter and the first four exons, we had presumed that Δ*NEMO* was not functional. Therefore, it was puzzling that we could not detect this potentially polymorphic Δ*NEMO–int3h* deletion in nearly 100 normal X chromosomes but found it in three unrelated IP patients and in all of their affected relatives. This suggested that Δ*NEMO* might have a role in the pathogenesis of IP. These findings have important implications for studying IP and Δ*NEMO*. Diagnostic testing for IP patients is complicated by the requirement of determining which *NEMO* copy contains the mutation. We currently use a Southern blot probe unique to *NEMO* to detect the recurrent deletion that accounts for the majority (70–80%) of IP mutations (7). The remaining patients have smaller exonic mutations for which long-range PCR has to be used to verify that they are at the *NEMO* locus and not at Δ*NEMO*. A few remaining patients have failed to exhibit mutations in *NEMO*, and these patients could potentially have alterations of Δ*NEMO* if this locus has any functional significance. We are currently investigating the function of Δ*NEMO*, although a preliminary database survey has failed to reveal evidence for an exclusive Δ*NEMO*-linked transcript, lacking exons 1 and 2. In addition, a previously described male IP patient has shown only an exon 10 mutation in *NEMO* by RT–PCR without the concurrent presence of a normal transcript sequence that might be derived from Δ*NEMO* (2,24,25). Thus, Δ*NEMO* does not appear to produce a transcript, but this possibility cannot be completely ruled out. We have tested the idea that Δ*NEMO* could be spliced onto the *LAGE1* transcript due to their relative positions to each other. However, RT–PCR on lymphoblast-derived RNA did not yield positive results, possibly because *LAGE1* is not expressed in this tissue. Therefore, other tissues that express this gene, such as breast, skin, placenta and uterus, need to be examined.

Unlike the *int3h* deletions in *NEMO* and Δ*NEMO*, the *LAGE1–LAGE2A* inversion appears to be non-pathogenic since it was found in a normal individual. This is not surprising since all of the genes involved are preserved. *LAGE1* is separated from its promoter but is apparently functional with a *LAGE2* promoter, probably because the two genes are evolutionarily related. The *LAGE* genes, similar to the *MAGE*, *BAGE* and *GAGE* genes, are expressed as antigens in various tumors and may cause disease phenotypes if disrupted (26–28). Similar to the *LAGE1–LAGE2A* inversion, the *int3h* duplication preserves normal gene sequences, although one might speculate that the duplicated exons 4–10 of *NEMO* might be spliced onto the end of an authentic *NEMO* transcript to create a larger-than-normal NEMO protein. However, it is not likely that the *int3h* duplication leads to an abnormal NEMO protein, since the C-terminal end of NEMO is known to be very sensitive to alterations and its disruption leads to either lethal or variant forms of IP (2,29–31). In support of this, we found the *int3h* duplication in an unaffected member of an IP family and in one normal female individual.

An interesting aspect of genomic rearrangements on the X chromosome is their parental origin. For instance, the inversion predominantly seen in hemophilia A patients occurs exclusively in the paternal germline (32,33). This is attributed to the fact that the X chromosome remains unpaired during male meiosis. We have previously reported that the common IP deletion also

shows a bias towards paternal origin but unlike the hemophilia A inversion, it is seen in female meioses as well (7,34). However, the relative distances between the repeats involved in each of these genomic disorders are very different. The *int3h* repeats associated with the IP deletion are separated by 11 kb, whereas the *int22h* repeats that predispose to the hemophilia A inversion are >100 kb apart. In this respect, it is likely that the *int3h* duplication also shows the same frequencies of parental origin as the *int3h* deletion. In contrast, the *LAGE1–LAGE2A* inversion occurs between identical sequences separated by nearly 68 kb and thus is more reminiscent of the hemophilia A inversion and might show a paternal origin bias. Unfortunately, the parental origin of the *LAGE1–LAGE2A* inversion could not be evaluated due to the lack of DNA samples from patient XL328-04's parents.

The *NEMO/LAGE2* duplication is complex compared to other reported genomic duplications because of additional internal sections of perfect homology within each copy (i.e. the *int3h* repeats and the *LAGE* genes). One might expect other types of alterations in addition to those reported here; therefore, a thorough analysis of the *NEMO/ΔNEMO* region in normal individuals would be interesting. Moreover, the non-pathogenic *int3h* duplication and the *LAGE1–LAGE2A* inversion could be transmitted to offspring in whom additional, novel rearrangements might occur. For example, the *LAGE1–LAGE2A* inversion places the two *NEMO/LAGE2* copies in tandem and in the same orientation, consequently predisposing one copy to deletion. As mentioned earlier, a non-pathogenic inversion between the *NEMO/LAGE2* copies would represent a fifth type of rearrangement, but we have not found this yet. If it exists, this rearrangement might be detectable by pulse-field gel electrophoresis, but it would be difficult because of the scarcity of appropriate restriction sites.

Finally, we have recently detected three novel restriction fragments on a diagnostic Southern blot intended to identify the common IP deletion. These unusual fragments are currently under investigation and likely represent new types of rearrangements at the *NEMO/LAGE2* duplication that will certainly provide greater insight into the dynamic nature of this region.

Collectively, this work suggests that genomic rearrangements may be more common than expected and may account for significant polymorphism in a general population. We currently lack efficient computational tools to detect large-scale sequence homologies that undergo benign rearrangements, which can appropriately be called genomic polymorphisms, in contrast to smaller single-nucleotide or short tandem repeat polymorphisms. With the completion of the Human Genome Project, sequence analysis tools will likely demonstrate that an appreciable proportion of our genome has evolved from duplication.

MATERIALS AND METHODS

Isolation of BAC clones containing *NEMO* and Δ*NEMO*

A contig had been constructed previously by screening the Genome Systems female BAC library, the RP11 male BAC library and the RP6 female PAC library with several probes in distal Xq28 (9). To enrich for clones near *NEMO*, the RP11 male BAC and RP6 female PAC libraries were screened with the *NEMO* cDNA and the *sWXD1332* marker. Additional clones were isolated from the RP5 male PAC library for sequencing the region between *NEMO* and *LAGE1* (12).

Probe walking to find duplication boundary

To detect the 3′ duplication boundary, DNA samples of RP11 BAC clones 211L10, 103M23, 66O13 and 515D14 were digested with various enzymes and transferred to a nylon membrane. The filter was prehybridized for 4 h and hybridized at 65°C with the appropriate probe overnight, starting with probe A (primers NEMO-10F and SA22R). The filter was washed to a final stringency of 1× SSC/0.1% SDS at 65°C and autoradiographed for 30 min to 2 h. To subclone overlapping fragments, a shotgun library was made with the appropriate enzyme from RP11-211L10 in a pBlueScript vector and hybridized with the relevant probe. Restriction digests confirmed the identity of positive colonies. The target bands from a restriction digest were gel-purified to extract successive probes from isolated clones.

Southern blot analysis

To analyze for the presence of Δ*NEMO* and determine the genomic structure between the duplicated copies, 5 µg of DNA was used for the following: human male and female (*Homo sapiens*, *HSA*), two male common chimpanzees (*Pan troglodytes*, *PTR*), one male bonobo chimpanzee (*Pan paniscus*, *PPA*), one male gorilla (*Gorilla gorilla*, *GGO*), and two male orangutans (*Pan pygmaeus*, *PPY*). DNA samples were digested with appropriate restriction enzymes overnight and electrophoresed on a 0.8% agarose gel for 24 h at 75 V. Following overnight transfer, the blots were hybridized with appropriate probes overnight at 65°C and washed to a stringency of 2× SSC/0.1% SDS.

Sequence generation and analysis

PAC clones RP5-865E18 and 1087L19 were sequenced as described previously (12). The GenBank accession no. is AF277315. The region between *NEMO* and *LAGE1* was also sequenced from BAC clones RP11-211L10 and 196H18 (GenBank accession no. AL596249). The sequences of both BAC clones were determined by a combined strategy of shotgun sequencing of M13 subclones and long-range genomic PCR products. BAC DNA was sonicated and the ends repaired with T4 DNA polymerase. Fragments of 1–2 kb were fractionated from the mixture by agarose gel electrophoresis and subcloned into M13mp18 vector prepared by digestion with *Sma*I. In total, 1800 subclones from BAC 196H18 and 1100 from BAC 211L10 were randomly selected and sequenced using Big-dye Terminator reactions on Applied Biosystems 3100 and 377 PRISM automated sequencers. Sequence traces were assembled using Applied Biosystem's FACTURA™ and INHERIT™ programs. These programs simultaneously assemble sequence files and facilitate subsequent editing to obtain a consensus sequence. Several smaller regions within the *NEMO/LAGE2* duplication that posed difficulties due to nucleotide differences between the two copies were sequenced directly from RP11-211L10 (containing *NEMO*) and from RP6-172D5 (containing Δ*NEMO*). PCR products were purified using a purification kit (Qiagen) and submitted to SeqWright (Houston, TX) for fluorescent sequencing. All sequences,

including GenBank entries, were analyzed with MacVector (Oxford Molecular Group, Cambridge, UK).

Sequence exchange detection—the *Drd*I assay

To detect sequence exchange between the two *NEMO/LAGE2* copies, a long-range PCR and digestion assay was used. A *Drd*I polymorphism (Table 1, no. 6) was present in intron 4 of *NEMO*. Specific PCR products of ~5 kb were amplified from *NEMO* with primer SA15F and separately, from Δ*NEMO* with primer SA65F. Both forward primers were used with the same reverse primer, NEMO-5R. The long-range PCR was done at 62°C annealing temperature in a 50 µl reaction with the EXPAND PCR kit (Roche). After verifying the correct product size, 25 µl of the PCR sample was digested with 1 U of *Drd*I for 4 h. The digest was electrophoresed on a 1% agarose gel and photographed under UV light.

Genomic DNA samples

Blood samples for IP patients were obtained through an IRB approved protocol, and DNA was extracted from these samples using conventional salt precipitation techniques. Genomic DNA samples from the three non-human great apes were isolated from lymphoblast cell lines maintained by D.L.Nelson.

ACKNOWLEDGEMENTS

We thank Kerry L.Wright for editing the manuscript. Evelyn Michaelis and Hella Ludewig provided excellent technical assistance. This work was supported by NIH grants 5 R01 HD35617 and 2 P30 HD24064 to D.L.N, Telethon-Italy grant E0927 to M.D. and German BMBF (BEO 0311108/0) and European Commission (BMH4-CT96-0338) grants to M.P.

NOTE ADDED IN PROOF

We recently re-evaluated the sequences of the two *NEMO/LAGE2* copies to a greater depth and found that the number of base differences between them is around 45 to 50 (instead of the 22 we have listed in this paper). However, the sequence identity between the two copies still exceeds 99%, and all of the points in this paper remain valid.

REFERENCES

1. Landy, S.J. and Donnai, D. (1993) Incontinentia pigmenti (Bloch–Sulzberger syndrome). *J. Med. Genet.*, **30**, 53–59.
2. International IP Consortium (2000) Genomic rearrangement in NEMO impairs NF-κB activation and is a cause of incontinentia pigmenti. *Nature*, **405**, 466–472.
3. Parrish, J.E., Scheuerle, A.E., Lewis, R.A., Levy, M.L. and Nelson, D.L. (1996) Selection against mutant alleles in blood leukocytes is a consistent feature in Incontinentia Pigmenti type 2. *Hum. Mol. Genet.*, **5**, 1777–1783.
4. Migeon, B.R., Axelman, J., de Beur, S.J., Valle, D., Mitchell, G.A. and Rosenbaum, K.N. (1989) Selection against lethal alleles in females heterozygous for incontinentia pigmenti. *Am. J. Hum. Genet.*, **44**, 100–106.
5. Israel, A. (2000) The IKK complex: an integrator of all signals that activate NF-κB? *Trends Cell Biol.*, **10**, 129–133.
6. Karin, M., Ben-Neriah, Y. (2000) Phosphorylation meets ubiquitination: the control of NF-[κ]B activity. *Annu. Rev. Immunol.*, **18**, 621–663.
7. International IP Consortium. A recurrent deletion in the ubiquitously expressed NEMO (IKK-γ) gene accounts for the vast majority of incontinentia pigmenti mutations. *Hum. Mol. Genet.*, in press.
8. Lupski, J.R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.*, **14**, 417–422.
9. Aradhya, S., Woffendin, H., Bonnen, P., Heiss, N.S., Esposito, T., Bardaro, T., Poustka, A., D'Urso, M., Kenwrick, S. and Nelson, D.L. Physical and genetic characterization reveals a pseudogene, an evolutionary junction, and unstable loci in distal Xq28. *Genomics*, in press.
10. Jin, D.Y. and Jeang, K.T. (1999) Isolation of full-length cDNA and chromosomal localization of human NF-κB modulator NEMO to Xq28. *J. Biomed. Sci.*, **6**, 115–120.
11. Timms, K.M., Lu, F., Shen, Y., Pierson, C.A., Muzny, D.M., Gu, Y., Nelson, D.L. and Gibbs, R.A. (1995) 130 kb of DNA sequence reveals two new genes and a regional duplication distal to the human iduronate-2-sulfate sulfatase locus. *Genome Res.*, **5**, 71–78.
12. Galgoczy, P., Rosenthal, A. and Platzer, M. (2001) Human–mouse comparative sequence analysis of the NEMO gene reveals an alternative promoter within the neighboring G6PD gene. *Gene*, **271**, 93–98.
13. Hu, X. and Worton, R.G. (1992) Partial gene duplication as a cause of human disease. *Hum. Mutat.*, **1**, 3–12.
14. Deininger, P.L. and Batzer, M.A. (1999) Alu repeats and human disease. *Mol. Genet. Metab.*, **67**, 183–193.
15. Boerkoel, C.F., Inoue, K., Reiter, L.T., Warner, L.E. and Lupski, J.R. (1999) Molecular mechanisms for CMT1A duplication and HNPP deletion. *Ann. N. Y. Acad. Sci.*, **883**, 22–35.
16. Rogner, U.C., Wilke, K., Steck, E., Korn, B. and Poustka, A. (1995) The melanoma antigen gene (MAGE) family is clustered in the chromosomal band Xq28. *Genomics*, **29**, 725–731.
17. Kiyosawa, H. and Chance, P.F. (1996) Primate origin of the CMT1A-REP repeat and analysis of a putative transposon-associated recombinational hotspot. *Hum. Mol. Genet.*, **5**, 745–753.
18. Gogarten, J.P. and Olendzenski, L. (1999) Orthologs, paralogs and genome comparisons. *Curr. Opin. Genet. Dev.*, **9**, 630–636.
19. Ji, Y., Eichler, E.E., Schwartz, S. and Nicholls, R.D. (2000) Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res.*, **10**, 597–610.
20. Bondeson, M.L., Dahl, N., Malmgren, H., Kleijer, W.J., Tonnesen, T., Carlberg, B.M. and Pettersson, U. (1995) Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. *Hum. Mol. Genet.*, **4**, 615–621.
21. Lagerstedt, K., Karsten, S.L., Carlberg, B.M., Kleijer, W.J., Tonnesen, T., Pettersson, U. and Bondeson, M.L. (1997) Double-strand breaks may initiate the inversion mutation causing the Hunter syndrome. *Hum. Mol. Genet.*, **6**, 627–633.
22. Timms, K.M., Bondeson, M.L., Ansari-Lari, M.A., Lagerstedt, K., Muzny, D.M., Dugan-Rocha, S.P., Nelson, D.L., Pettersson, U. and Gibbs, R.A. (1997) Molecular and phenotypic variation in patients with severe Hunter syndrome. *Hum. Mol. Genet.*, **6**, 479–486.
23. Small, K., Iber, J. and Warren, S.T. (1997) Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. *Nat. Genet.*, **16**, 96–99.
24. Mansour, S., Woffendin, H., Mitton, S., Jeffery, I., Jakins, T., Kenwrick, S. and Murday, V.A. (2001) Incontinentia pigmenti in a surviving male is accompanied by hypohidrotic ectodermal dysplasia and recurrent infection. *Am. J. Med. Genet.*, **99**, 172–177.
25. Doffinger, R., Smahi, A., Bessia, C., Geissmann, F., Feinberg, J., Durandy, A., Bodemer, C., Kenwrick, S., Dupuis-Girod, S., Blanche, S. *et al.* (2001) X-linked anhidrotic ectodermal dysplasia with immunodeficiency is caused by impaired NF-κB signaling. *Nat. Genet.*, **27**, 277–285.
26. Lethe, B., Lucas, S., Michaux, L., De Smet, C., Godelaine, D., Serrano, A., De Plaen, E. and Boon, T. (1998) LAGE-1, a new gene with tumor specificity. *Int. J. Cancer*, **76**, 903–908.
27. Chen, Y.T., Boyer, A.D., Viars, C.S., Tsang, S., Old, L.J. and Arden, K.C. (1997) Genomic cloning and localization of CTAG, a gene encoding an autoimmunogenic cancer-testis antigen NY-ESO-1, to human chromosome Xq28. *Cytogenet. Cell Genet.*, **79**, 237–240.
28. Castelli, C., Rivoltini, L., Andreola, G., Carrabba, M., Renkvist, N. and Parmiani, G. (2000) T-cell recognition of melanoma-associated antigens. *J. Cell Physiol.*, **182**, 323–331.
29. Yamaoka, S., Courtois, G., Bessia, C., Whiteside, S.T., Weil, R., Agou, F., Kirk, H.E., Kay, R.J. and Israel, A. (1998) Complementation cloning of NEMO, a component of the IκB kinase complex essential for NF-κB activation. *Cell*, **93**, 1231–1240.

30. Rothwarf, D.M., Zandi, E., Natoli, G. and Karin, M. (1998) IKK-γ is an essential regulatory subunit of the IκB kinase complex. *Nature*, **395**, 297–300.
31. Aradhya, S. and Nelson, D.L. (2001) NF-κB signaling and human disease. *Curr. Opin. Genet. Dev.*, **11**, 300–306.
32. Lakich, D., Kazazian, H.H.,Jr, Antonarakis, S.E. and Gitschier, J. (1993) Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nat. Genet.*, **5**, 236–241.
33. Rossiter, J.P., Young, M., Kimberland, M.L., Hutter, P., Ketterling, R.P., Gitschier, J., Horst, J., Morris, M.A., Schaid, D.J., de Moerloose, P. *et al.* (1994) Factor VIII gene inversions causing severe hemophilia A originate almost exclusively in male germ cells. *Hum. Mol. Genet.*, **3**, 1035–1039.
34. Aradhya, S., Woffendin, H., Jakins, T., Bardaro, T., Esposito, T., Smahi, A., Shaw, C., Levy, M., Munnich, A., D'Urso, M., Lewis, R.A., Kenwrick, S. and Nelson, D.L. (2001) A recurrent deletion in the ubiquitously expressed NEMO (IKK-gamma) gene accounts for the vast majority of incontinentia pigmenti mutations. *Hum. Mol. Genet.*, **10**, 2171–2179.