EYS (SPAM), an orthologue of Drosophila spacemaker, is mutated in autosomal recessive retinitis pigmentosa

Mai M Abd El-Aziz[1]*, Isabel Barragan[2, 3]*, Ciara A O'Driscoll[1]*, Leo Goodstadt[4], Elena Prigmore[5], Salud Borrego[2, 3], Marcela Mena[2, 3], Juan I Pieras[2, 3], Mohamed F El-Ashry[1], Leen Abu Safieh[1], Amna Shah[1], Michael E Cheetham[6], Nigel P Carter[5], Christina Chakarova[1], Chris P Ponting[4], Shomi S Bhattacharya[1, 7]& Guillermo Antinolo[2, 3]

Supplementary Methods

Human samples

All participating families have previously been linked to the RP25 locus [1, 2, 3]. Informed consent was obtained from all participants for clinical and molecular genetic studies and the study conformed to the tenets of the Declaration of Helsinki.

Mutation screening of the genes within the 100 Kb deleted interval

Fourteen primer pairs (primer details are available on request) were designed to ensure a total coverage of the predicted coding regions and the intronic flanking sequences of the 6 genes within the 100 Kb interval using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Polymerase chain reaction (PCR) and mutation screening were carried out as previously reported [4].

Expression studies

Expression analysis of the *RP25* gene was assessed by PCR amplification of human cDNAs from the retina, brain, kidney, liver, heart, skeletal muscle, pancreas, lung and placenta (Quick-Clone; Clontech) as well as from lymphoblast, Y79, ARPE19 and HELA cell lines. cDNA specific primers within exons 4 (forward) and 11 (reverse) of *RP25* were designed for this purpose (**Supplementary Table 1**). The ubiquitously expressed gene, *PGM1*, was used as an amplification control. PCR products were then electrophoresed on a 2% agarose gel and visualized by ethidium bromide staining.

Comparative genomics

Bioinformatics analysis of the genes within the deleted interval

The RP25 genomic interval was used as a query to search protein and translated nucleic acid sequence databases for homologs using the BLAST program at the U.S. National

Centre for Biotechnology Information website (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Further analysis of the protein sequence was performed using the tools available at the ExPASY (http://www.expasy.ch/tools/) and Molecular Biology Server (BCM) Search Launcher.

Validating gene sequence

The genomic extents for *RP25* were identified in the mouse, rat and draft dog, horse and rhesus macaque genome assemblies using the BLAT sequence alignment server at (http://genome.ucsc.edu) [5]. The full length coding sequences were predicted using exonerate (http://www.ebi.ac.uk/~guy/exonerate/) [6]. We confirmed splice site conservation manually.

Predicting domain architecture

The domain architecture of RP25 was predicted using the SMART protein domain database server [7]. The EGF, Ca^{2+}-binding EGF and Laminin-G domains predictions were further refined manually using seed alignments from SMART and hmmsearch from the HMMER suite of programs [8]. Functional EGF domains were indicated by conservation of the consensus cysteines required for disulphide bond formation. EGF-Ca^{2+} domains were identified if they retain consensus negatively-charged residues.

Demonstrating orthology

Fly and human homologues of *RP25* were obtained using BLASTP [9] and from the SMART protein domain database [7]. We selected all fly and human homologues whose sequence contained predicted LamG domains separated by one or two EGFs. The LamG domain sequences were manually aligned and the protein distances between them

calculated using protdist from PHYLIP [10]. A bootstrapped phylogenetic tree was then derived using neighbour joining (**neighbour** and **consense** from PHYLIP).

Mammalian lineage-specific loss of RP25

The orthologous sequence for the final coding exon was identified using BLAT [6] and BLAST [9] in the low coverage genome assemblies for bat, armadillo, chicken, platypus, opossum, pika, rabbit, tree shrew, marmoset and guinea pig, sheep and cattle genome sequences [11]. DNA sequences were multiply aligned guided by the translated peptide sequence. Putative disruptions to the coding sequence in unfinished genome sequences were corroborated by high quality sequence in multiple traces. Lineage-specific d_N/d_S values were calculated with the codeml programme [12] using the known species phylogeny and parameters model = 1, CodonFreq = 2.

RT-PCR and 5′ and 3′ RACE cDNA amplification

Identification of the *RP25* gene coding regions was carried out using the SMART™ RACE cDNA amplification kit (Clontech, Europe). Initially, first-strand retinal cDNA synthesis was performed which was then used directly in the 5′ and 3′-RACE PCR reactions, without the need for a second-strand synthesis and adaptor ligation. Gene specific primers (GSPs) were initially designed within exons 1 and 4 of the *Q9H557* gene. Subsequently, multiple GSPs were designed upstream and downstream of the first primer pairs to identify the full length gene (Primers details are available on request). The retinal cDNA was amplified using an Advantage 2 PCR kit (Clontech, Europe) enabling the use of a touchdown annealing temperature, which has significantly improved the specificity of BD SMART RACE amplification. Sequence analysis of the different PCR products was then carried out bi-directionally which was followed by alignment of the

multiple contigs to identify the transcript structure of the novel gene. Finally, we designed 2 pairs of GSPs (**Supplementary Table 1**) to PCR amplify the full length transcript of *RP25* in 2 overlapping fragments (**Supplementary Fig. 3).**

Mutation screening of the *RP25* gene

Forty eight primer pairs were designed using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) ensuring a total coverage of the entire coding region and the intronic flanking sequences of the *RP25* gene (**Supplementary Table 1**). Polymerase chain reaction (PCR) and mutation screening were carried out as previously described [4].

Cosegregation analysis

To follow the segregation of four of the identified mutations within families, a restriction digest analysis was performed. The PCR products obtained from amplifying exons 12, 17+18, 28 and 41 of *RP25* in their corresponding families were digested with 1 unit of *Spe*I, *TSP*45I, *Hph*I and *Pfl*MI enzymes, respectively (New England BioLabs, UK). The products were then analyzed on 2% agarose gel.

Array comparative genomic hybridization

Array comparative genomic hybrdization analysis (array-CGH) was performed on 10 DNA samples from 2 families (RP5 and RP73). Initially, a whole Genome Tiling Path (WGTP) BAC array comprising 26,574 large-insert clones with 93.7% coverage of the euchromatic portion of the human genome was used and the methods employed were as previously published [13]. Secondly, we have used a high resolution fine tiling array, NimbleGen custom designed oligo array, which was designed to cover a 2 Mb region spanning the genomic interval between 64753791 and 66753791 Mb on 6q and consisted

of 385K oligonucleotide probes, with a probe approximately every 5 bp. Selection and designing of the oligonucleotide probes, synthesis of microarrays and CGH experiments, including DNA labeling and hybridization were carried out by NimbleGen Systems Inc., as previously described [14]. Array data were analyzed using SignalMap software version 1.8 (NimbleGen Systems Inc).

Multiplex ligation dependent probe amplification (MLPA)

Seven synthetic probes for exons 12, 13 and 15-19 of *RP25* were designed according to the MRC Holland guidelines (**Supplementary Table 1**). Each probe sequence was checked for localization and putative polymorphisms using BLAT and BLAST. A set of control probes was included in each synthetic probe mix in order to provide invariable values. The reactions were carried out as previously described [15]. The MLPA analysis criteria were as follows: i) normal if the individual dosage quotient values are within 0.8-1.0; ii) deletions or duplications if the dosage quotient values are around 0.5 or 1.5, respectively; and iii) the mean standard deviation of all samples for each peak should be below 10%.

Generation of anti-spam antibody

Rabbit polyclonal antiserum was raised against one peptide derived from the predicted amino acid sequence by Eurogentec. The peptide (N-terminal residues 61-75, H2N-GVNTKIDTSGNQAVP-CONH2) was conjugated to hemocyanin through an N-terminal cysteine residue before immunization. Final bleed from the immunized rabbit was affinity purified against peptide by Eurogentec (Double XP program), and the affinity purified serum was used in all subsequent experiments.

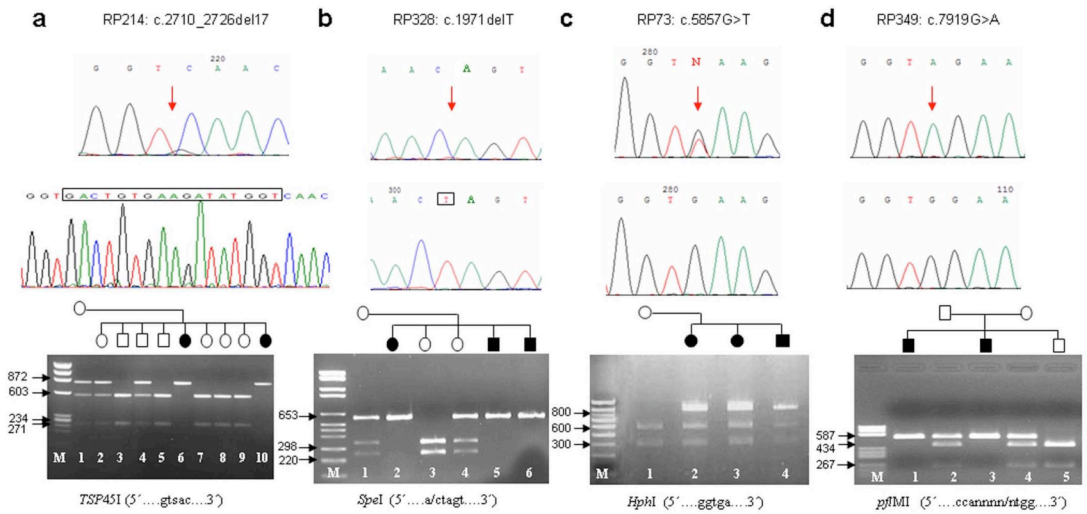
Immunohistochemistry

Adult pig eyes were isolated and fixed in 4% paraformaldehyde for overnight at 4 °C. The eyes were then rinsed in 1xPBS and then in sucrose gradient and finally left in 20% sucrose overnight at 4 °C. The sections were snap frozen in OCT and then stored in -80 °C. Cryosections (10 μm) were treated with 10% sodium borohydride in 1xPBS and then blocked with blocking solution (0.2% Goat serum and 0.1% TritonX in 1xPBS). Antibodies against rhodopsin (1D4, a gift from Professor R. Molday, UBC) [16] and spam were then diluted 1:1000 and 1:200, respectively, in 2% Goat serum and 0.1% TritonX in 1xPBS. Secondary antibodies, Cy3- conjugated AffiniPure Goat anti Mouse and Fluorescein (FITC)- conjugated AffiniPure Goat anti Rabbit (1:400, Jakson ImmunoResearch lab.) were also diluted in 2% Goat serum and 0.1% TritonX in 1xPBS and incubated for 1 h. Immunofluorescence was analysed with a LSM510 confocal laser scanning microscope (Zeiss) and images were processed with Adobe Photoshop CS2. The specificity of the outer segment spam immunostain was confirmed by staining retinal sections with the rabbit pre-immune serum or anti-spam antiserum preadsorbed with the peptide epitope (20 μg/ml). Retinal sections were also incubated with either the primary or secondary antibodies on their own to confirm that the signal observed was specific. In addition, the rabbit primary antibody was incubated with mouse secondary antibody and vice versa to verify that immunosignals did not arise from cross reaction between the antibodies in the double label procedure.

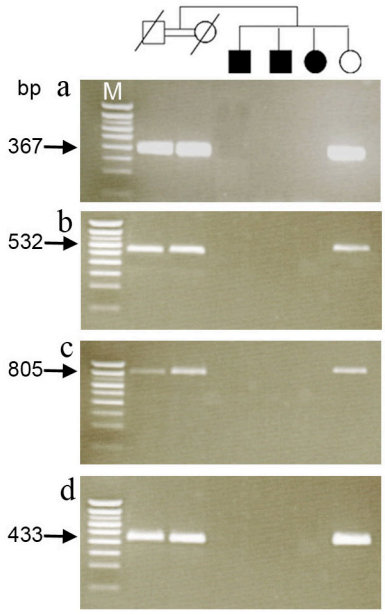
Supplementary Note

All mutations were homozygous except in family RP73 where a compound heterozygous change was detected in agreement with the genetic data [1]. It is also worth noting that the mutations in families RP214, RP349 and RP328 were homozygous even though the patients reported lack of consanguinity in their families. However, the identification of homozygous mutations in these families is explained by our previously published observation of homozygous haplotypes incorporating the *RP25* gene shared by the affected members in each family. Nevertheless these regions of shared haplotype were different between families hence explaining the finding of different mutations in each of these families [1, 3].

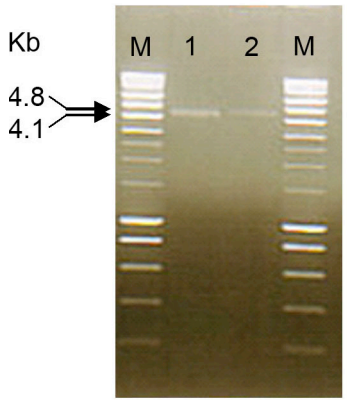
Supplementary Figures



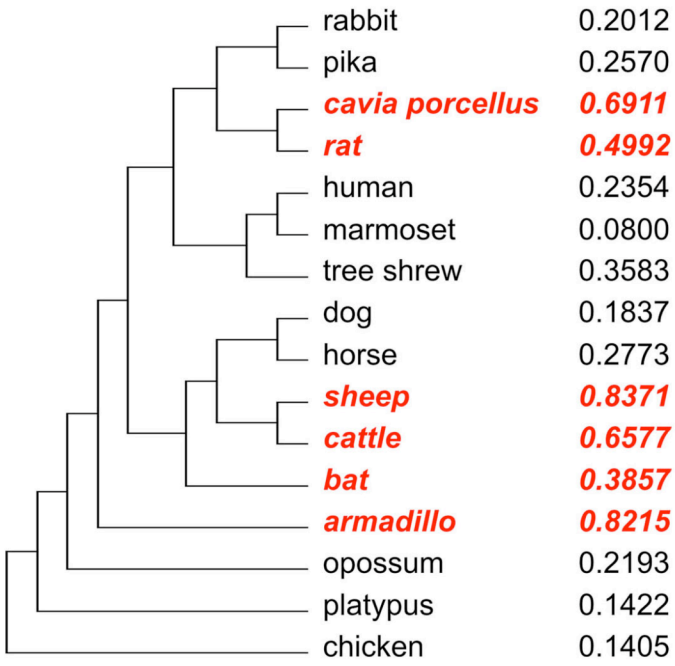
Supplementary Figure 1 Sequences of the small deletions and point mutations in *RP25* and its segregation in arRP families. (**a**) Homozygous c.2710_2726del17 in family RP214 (**b**) Homozygous c.1971delT in family RP328 (**c**) Heterozygous c.5857G>T in family RP73 (**d**) Homozygous c.7919G>A in family RP349. The position of mutations is indicated by red arrows and they all lead to loss of restriction enzyme sites as shown on the gel pictures in all affected members.



Supplementary Figure 2 PCR amplification of *Q5TEL3* and *Q9H557* genes in family RP5. Non amplification of *Q5TEL3* single exon (**a**) and *Q9H557* four exons (**b-d**) in all affected family members, respectively. M denotes a 1 Kb smart ladder.



Supplementary Figure 3 Amplification of the full length *RP25* transcript using the retinal cDNA as a template in two overlapping fragments. Lanes 1 and 2 show the corresponding size fragments of 4.1 and 4.8 Kb, respectively. M is a 10 Kb smart ladder.



Supplementary Figure 4 Phylogenetic tree showing the independent disruption of *eys* (spam) in each of the guinea pig, rat, sheep, cattle, bat and armadillo lineages (in red). High estimated d_N/d_S values for terminal branches show loss of constraint in lineages with pseudogenes. There are no traces of sequence correspondence to multiple regions of mouse genomic sequences in this region in the assembled genome which explains the absence of *eys* (spam) in the mouse.

Supplementary Table 1 Primers used for mutation screening, amplification, expression study and MLPA analysis of *RP25*

PRIMERS FOR MUTATION SCREENING OF THE *RP25* GENE

EXON	FORWARD PRIMER (5´ > 3´)	REVERSE RPIMER (3´ > 5´)
1(nc)	AGCTCCAGATGGTGATTTGC	TCCTTAAGCAAATGACAGAGAAAA
2(nc)	GGCAGGATACTGGCTAATCTG	GCCAAAAACGATAAAACCTGA
3(nc)	CCATCCTCTGGATTATCATAAAAG	CTTTGGGAAAGAAGGCCAAG
4a	GCTGCTGGTGACACTATCTTTG	GAAGTCCAAGCAGATGTTTTCTG
4b	GCCTGATGGTTTTTCACAGC	AAAATGGAGGCTGGCAATG
4c	GCAGTTCTGCCAGGAATCTC	GTGCTGGGATTACAGGTGTG
5	TTTCAATTTAAATGCATCATCG	TGAAAAGCATGTGAACTGTTG
6	TTTGCAAAAGTTACTGTAGAATTGC	GACCGTTCTTGTTCGTCTGAG
7	TGAGATGGGAGATGGTGTTG	CAACAATTAACCCAAAACATGC
8	GCTTTTTGGCTAAGATCACAGG	TGGCTAAGATTAATAAGAGCATTTG
9	GGCTTTTGAACATGGATATGAC	TCTCTTGCACCAAGTAGATTTCC
10	GGAACTTATTTTGTGGCAGATG	TGATTCTTCAAAATTTTTACTTTCC
11	CAAGCTTTGAACCCTTGTCC	CTTCTTCCCTCCTTTTATTGTGC
12	GCCAAGAATGGACACTTTAT	CCAAAGAAGCAATCCTATTATTCAA
13	TGGCATTCTTATCTAATAAATTTGG	TTGGTTGGTCACTTTAGAAGC
14	TTGGTTAAAAGTGAACTACATAAATGA	TGCTTGAGTTTCTGTTTTCTAACC
15	GAGATATCAAAATGGCCAGGAG	ATGATTGCGACACCATCTTG
16	TCACAGGAAATTAGGCAAACAA	GGATTTTTCCAACCCATTTT
17+18	CTGGCATGTTTTTATGCACAC	CATTTTGCTCAGGCACACAT
19	AAAATTTTGCAAGGAGAATTGC	ATTTTTGCCCTGTTTGCATC
20	TGTGCTTTGTTTTTGTCATTCAC	AACTGGCAGCATCTGTCATC
21	TTTCACCTGAACTAGGAAAGAAAAG	CAACAAGAGACAAAGAAAGAAAAGG
22	GAGGAAGGAAATGTCAAACAAG	TTGCAGAGTGCATTACTAGTGG
23	GTTGGAGCTCTGAAAACACG	TAGTATTGGTGGAGTGGATTGTC
24	CCACGGATAAGAGCTGAGAC	AGAGAAGGAGAGATGCGCTG
25	TGTCTATGGAAATGCAAATGGA	AAACAGGAGTCATAACCAATAATGC
26_1	AGTGCCAAAGTGGTTCGTTT	GGGGTCCTTGCTCTCCTATC
26_2	GCACCGGTTCAGTACAAAAG	AGTTTGTGAAGGGACAATGGA
26_3	ATCCTCATCCTTGGAAGAATCC	GACCATGACAGGCTCTTTCTTC
27	TTGAAAGAGGCAGGAAAGAGAC	AAGAGACATCCTGGTGGTGAGT
28	GCTTTTCTCAACCATGTTCCTC	GGGGATAGGGTCACCTTAAAAC
29	GATTAATCTGCTTCTGGCTTTG	TGGAAAAACAGACTGACATTGG
30	CCCATGTTCATTGCAGCATTAT	GAACGTAGGAATGTGAAGCAAA
31	AGGGTCATGTTATGTGGCTCAG	CAGCTGTTTCTTGTTTGTGC
32	TTCAGTCTTTTCCTCTGTACTGG	GCTCTGAAACATTTGCAGCAT
33	ACTCCTACCAACCCCCTAAATC	GTGGTGGTGCACATCTGTAGTC
34	TTCTGAAAGCATTCCATGTCC	TTTTCTGGTGCTTTGTTGAGAG
35	GCCAACAATAGCAACCTCTTTT	ACATGTGTGCATCATTTAGGT
36	GCATATGTGTTCATGCATGTGT	CTGCTTGGTGATCAGTCTCACT
37	AGATGCATCAGCAAAACGTAAC	GCATCTAGGCAAAGGTCTTTT
38	GAATAACAACAGCCAGTTGCAC	CTGTGAACTTCGTGGATGTAGG
39	TAACAGACACCAGCAGAGAAGC	TGTTCAAGTCTGAAAGCAATCC
40	TTCTCTGCGCATTTCTGTATTC	CTGTCCTCCCATCATGTAACAA
41	GACAAGTTAGCATCAGGGCAAT	GAAAAAGAGGACAGTGGATTTG
42	CTCACCTACAAGCAACTCTTGG	TACGCATACACTTGCAGTGACA
43_1	CTTATCCAACTTGGCCAGAAAC	TCAAACAGGACACAGACTGGTT
43_2	GGGTACAACACATGCAGAAATG	GTGGATCAATATCCTCGGAAAG
43_3	GGTCTCCATAATCAGACCTTGA	GATTCCCCGTAAGCAATGTATC

PRIMERS FOR AMPLIFICATION OF THE FULL LENGTH *RP25* GENE

EXON

4aF	CCAATTCCCAGGAATCCTTAACCACAAC
26R	GAATCGAGAAGAGGAAACATCTGCGG
26F	CCGCAGATGTTTCCTCTTCTCGATTC
43R	CCCAACCCAAAGTACACAGGCAACTG

PRIMERS FOR EXPRESSION ANALYSIS OF THE cDNA

EXON

4aF	CCAATTCCCAGGAATCCTTAACCACAAC
11R	CCAGATACATGTTGCCAGCCCATCTGAG

MLPA SYNTHETIC PROBES

EXON	5` PROBE SEQUENCE	3` PROBE SEQUENCE
12	TGAAGACTGCAAATCTGCGTCCTG	CAAAAATGGAACAACTAGTACACATTTAAGGGGA
13	TGGAGCCACCTGCATTGACCAACCT	GGTAATTACTTCTGCCAGTGTGTGCCT
15	TGTGAACAAGAATCCAATGAGTGTAAAATGAATCC	TTGCAAGAACAATTCCACCTGTACTGACCTTTAC
16	TACATCTGGATGGACTGGACAGAACTGTAGTGAA	GAAATAAATGAATGCGACTCTGATCCATGCATGA
17	TCTCATACTCTTTGCAGAATTTGAAGGTAAAAACTG	TGAAATTGATGTGAAAGACTGCCTCTTCCT
18	TGGGTTTTCTGGATCTCTGTGTGAAATTGAAATT	AATGAATGTTCCTCTGAACCTTGCAAAAATAATGG
19	TTGTGAACCTGAGTACCATGGGCCCTT	CTGTGAACTTGATGTAAATAAATGTAAAATCTCACCTT

Supplementary References

1. Ruiz, A. Borrego, S. Marcos, I. & Antinolo, G. *Am J Hum Genet* **62**, 1452-1459 (1998).

2. Barragán, I. Marcos, I. Borrego, S. & Antiñolo, G. *Ophthalmic Res* **37**, 89-93 (2005).

3. Barragan, I. *et al. Ann Hum Genet* online (2008).

4. Abd El-Aziz, M.M. *et al. Ann Hum Genet* **70**, 1-14 (2007).

5. Kent, W.J. *Genome Res* **12**, 656-64 (2002).

6. Slater, G.S. & Birneym, E. *BMC Bioinformatics* **6**, 31 (2005).

7. Letunic, I. *et al. Nucleic Acids Res* **32**, D142-4 (2004).

8. Eddy, S.R. *Bioinformatics* **14**, 755-63 (1998).

9. Altschul, S.F. *et al. Nucleic Acids Res* **25**, 3389-402 (1997).

10. Felsenstein, J. *Science* **246**, 941-942 (1989).

11. Green, P. *Genome Res* **17**, 1547-9 (2007).

12. Yang, Z. *Mol Biol Evol* **24**, 1586-91 (2007).

13. Redon, R. *et al. Nature* **444**, 444-454 (2006).

14. McCormick, M.R., Selzer, R.R. & Richmond, T.A. *Methods Mol Biol* **381**,189-211 (2007).

15. Schouten, J.P. *et al. Nucleic Acids Res* **30**, e57 (2002).

16. Spuy, J.V.D *et al. Mol Vision* **11**, 542-553 (2005).