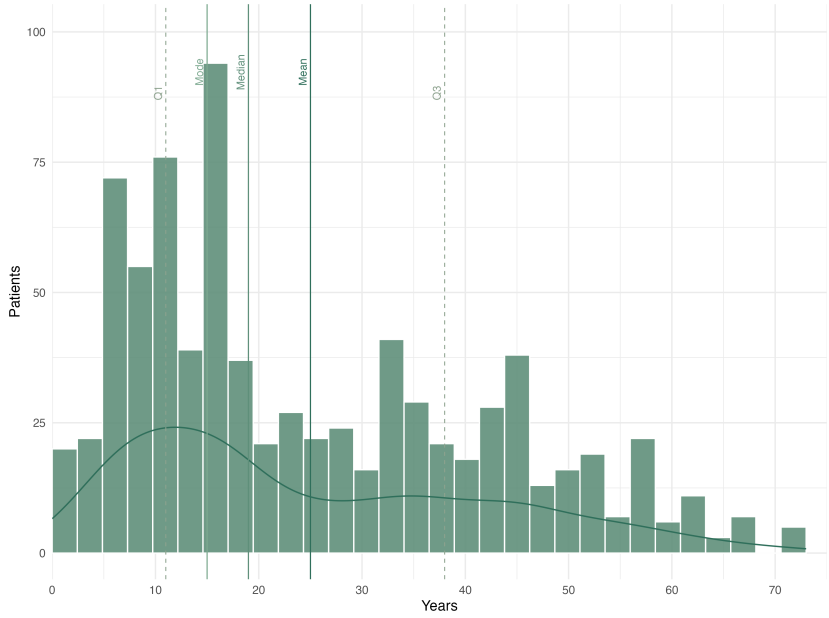
Genomic landscape of inherited retinal diseases for gene therapy and clinical trials in Argentina

Supplementary Information

Marcela Ciccioli; Jazmín Martínez; Ezequiel Fernández; Alejandra Antacle; Laura Echandi; Daniel Hidalgo; Mario Saravia; Rebeca Valero; Miquel Tuson; Gemma Marfany; Roser Gonzàlez-Duarte; Fabiana Arzuaga and Hernán Dopazo

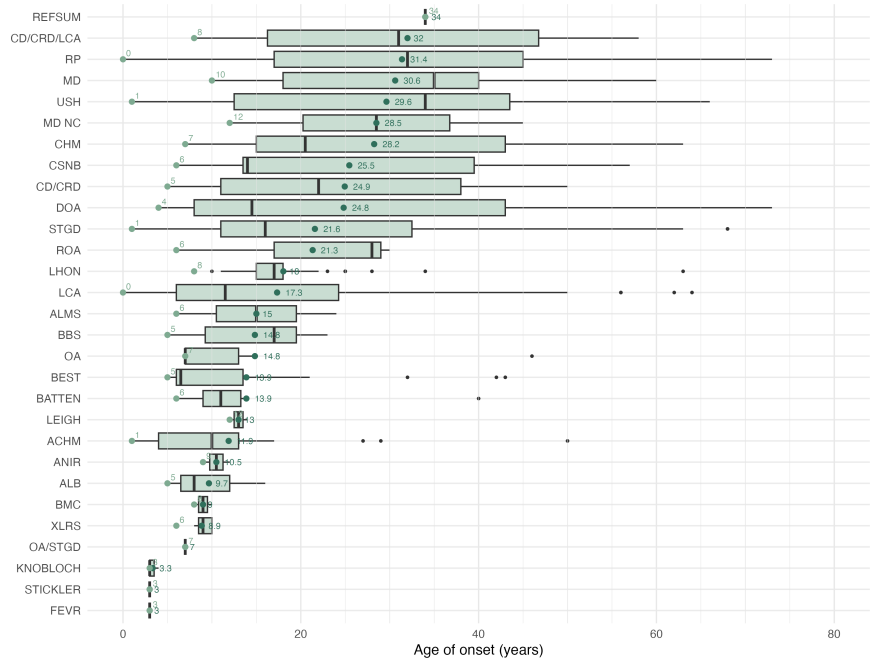
1 Supplementary Figures

1.1 Age at clinical onset



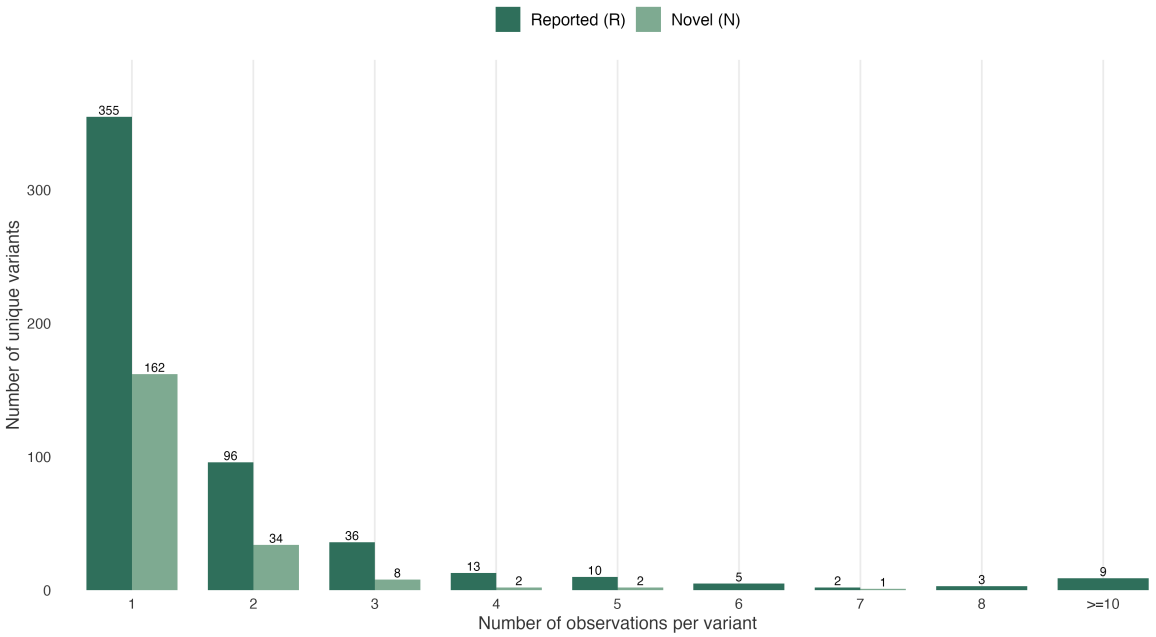
Supplementary Figure S1. Distribution of age of clinical onset in the SARA cohort. Histogram of age at clinical onset among all genetically resolved individuals. The distribution ranged from 1 month to 73 years and was right-skewed. The mode was 15 years, the median was 19 years, and the mean was 25.0 years. The interquartile range extended from 11 years (Q1) to 38 years (Q3), indicating that half of all cases developed symptoms within the first four decades of life.

1.2 Age of onset across IRD diagnoses



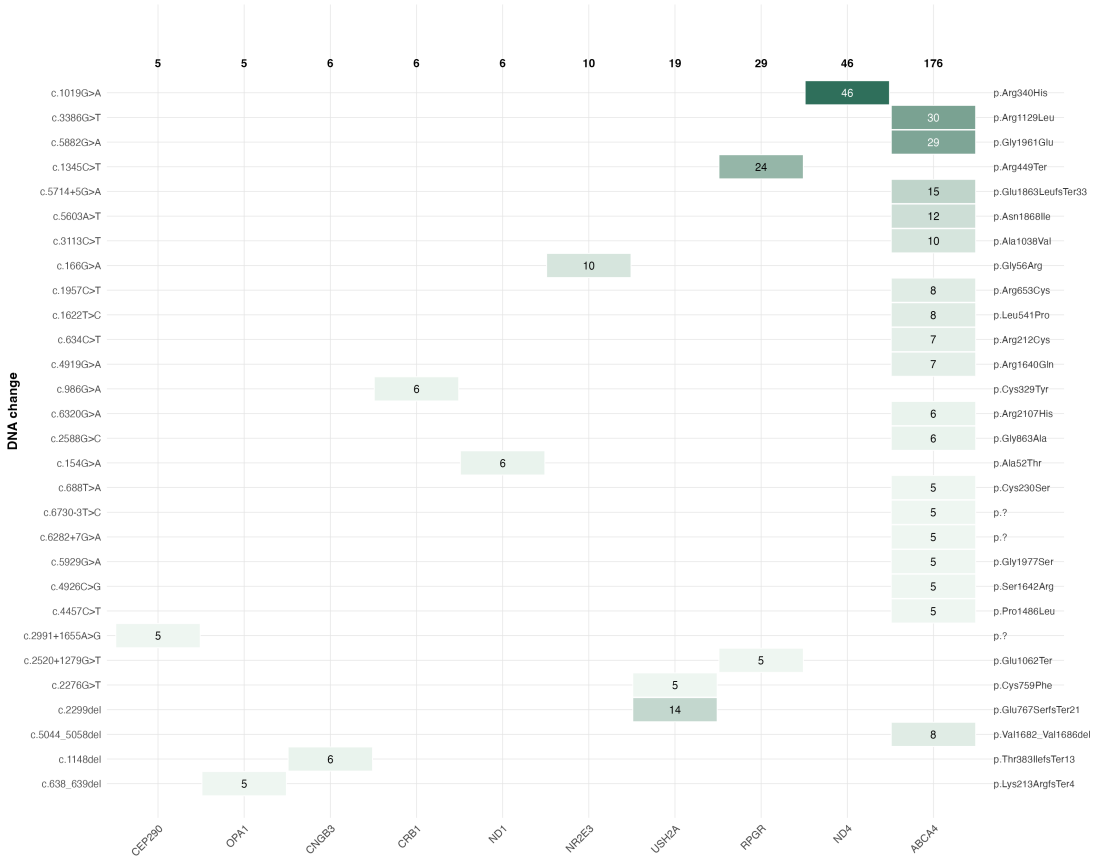
Supplementary Figure S2. Boxplots show the distribution of onset ages for each diagnosis, including minimum, maximum, and interquartile ranges. The vertical black line in each box represents the median age of onset, while blue dots indicate the mean, and red dots indicate the minimum observed onset for each diagnosis. Many IRDs exhibit early onset, typically in childhood or adolescence, whereas others—such as USH, macular dystrophy, STGD, DOA, and RP—display broader or later-onset distributions.

1.3 Recurrence spectrum



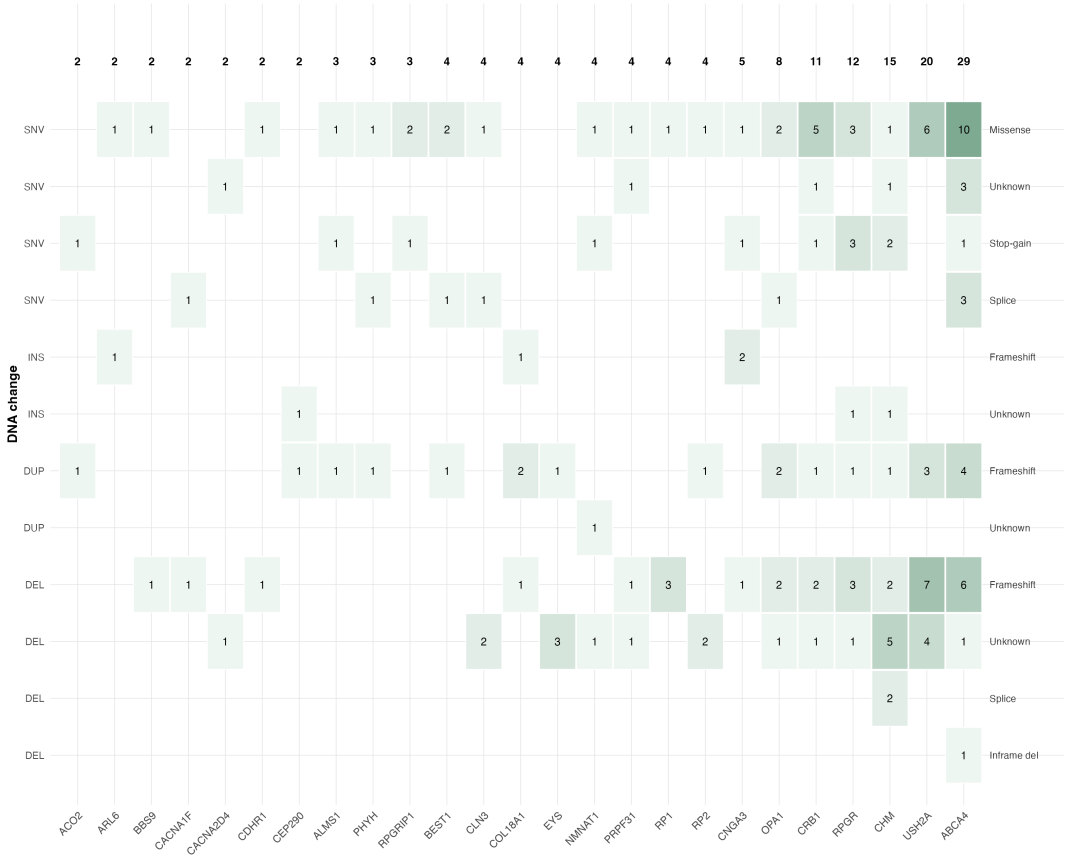
Supplementary Figure S3. Distribution of unique genetic variants by recurrence level in the SARA cohort, stratified by reported and novel status as defined by VEP/ClinVar public clinical annotation. Bars indicate the number of unique variants within each recurrence class; high-frequency classes (≥ 10 observations) are collapsed for clarity. Variants lacking public clinical annotation are largely confined to singleton and low-frequency classes, whereas reported variants span a broader recurrence spectrum, including highly recurrent alleles.

1.4 Matrix of reported variants



Supplementary Figure S4. Matrix representation of reported genetic variants observed five or more times in the SARA database. The figure highlights 29 unique recurrent variants distributed across 10 genes, annotated at both the DNA and protein levels.

1.5 Matrix of novel variants



Supplementary Figure S5. Distribution of variants lacking public clinical annotation across the most frequently affected genes. Matrix depicting novel genetic variants across the 25 genes with the highest burden of variants without public clinical annotation recovered by VEP/ClinVar in the SARA cohort.

2 Supplementary Tables

2.1 Phenotype summary

Supplementary Table S1. Disease acronym, clinical phenotype and number of patients ordered by their percentage contribution within the cohort.

N	Acronym	Phenotype	Patients	Pct
1	RP	Retinitis pigmentosa	328	40.6%
2	STGD	Stargardt disease	159	19.7%
3	LCA	Leber congenital amaurosis	60	7.4%
4	LHON	Leber hereditary optic neuropathy	56	6.9%
5	USH	Usher syndrome	30	3.7%
6	CHM	Choroideremia	24	3.0%
7	DOA	Dominant optic atrophy	22	2.7%
8	ACHM	Achromatopsia	17	2.1%
9	CD/CRD	Cone/cone-rod dystrophy	17	2.1%
10	BEST	Bestrophinopathy	16	2.0%
11	MD	Macular dystrophy	13	1.6%
12	CSNB	Congenital stationary night blindness	11	1.4%
13	BATTEN	Batten disease	8	1.0%
14	XLRS	X-linked retinoschisis	7	0.9%
15	BBS	Bardet-Biedl syndrome	6	0.7%
16	OA	Optic atrophy	6	0.7%
17	CD/CRD/LCA	Cone/cone-rod dystrophy / LCA	4	0.5%
18	ALB	Albinism	3	0.4%
19	KNOBLOCH	Knobloch syndrome	3	0.4%
20	ROA	Recessive optic atrophy	3	0.4%
21	ALMS	Alstrom syndrome	2	0.2%
22	ANIR	Aniridia	2	0.2%
23	BMC	Blue cone monochromacy	2	0.2%
24	LEIGH	Leigh syndrome	2	0.2%
25	MD NC	Macular dystrophy, North Carolina	2	0.2%
26	STICKLER	Stickler syndrome	2	0.2%
27	FEVR	Familial exudative vitreoretinopathy	1	0.1%
28	OA/STGD	Optic atrophy / Stargardt disease	1	0.1%
29	REFSUM	Refsum disease	1	0.1%

2.2 Rare genes below 1%

Supplementary Table S2. Rare genes in the SARA cohort (Disease-associated genes observed in fewer than 1% of genetically solved patients grouped by the number of patients in whom each gene was detected).

Patients per gene[a]	Number of genes[a]	Genes
8	2	CNGB3, PCARE
7	7	ADGRV1, CNGA3, IFT140, ND1, NMNAT1, PDE6B, RS1
6	1	GUCA1A
5	3	CDHR1, CERKL, IMPDH1
4	4	IMPG2, MFN2, NRL, PDE6A
3	14	ACO2, AIPL1, BBS5, CACNA1F, COL18A1, HK1, MYO7A, ND6, OPA3, RPGRIP1, RTN4IP1, SPATA7, TULP1, TYR
2	22	AGBL5, ALMS1, ATP6, CDH23, CNGB1, CRX, IQCB1, KLHL7, MERTK, NPHP1, NYX, OAT, OPN1MW, PAX6, PCDH15, PHYH, PRDM13, RBP4, ROM1, SCLT1, SNRNP200, UNC119
1	33	AHI1, ARL6, BBS2, BBS9, CACNA2D4, CC2D2A, CNGA1, COL11A1, COL11A2, CPLANE1, ELOVL4, FAM161A, INPP5E, LRP5, MFSD8, NPHP4, NR2F1, OTX2, PANK2, PDE6C, PITPNM3, PRPF3, RBP3, RDH11, RIMS1, RLBP1, RP9, TIMP3, TOPORS, TPP1, TTLL5, USH1C, ZNF408

[a] Patients per gene indicates the recurrence count for each gene, whereas Number of genes indicates how many distinct genes had that recurrence. For example, a row with "3 patients per gene" and "14 genes" indicates that 14 different genes were each observed in three patients. Most rare genes were observed in only one or two patients, whereas a small subset recurred in up to eight patients in our cohort.

2.3 Multigenic findings

Supplementary Table S3. Multigenic findings (Twenty-one cases carried variants in more than one IRD-associated gene, including secondary or non-primary gene findings).

Patient[a]	N	Genes	Phenotype
P0738	4	ADGRV1, IMPDH1, PHYH, RPGR	RP
P0145	2	NRL, RPGR	RP
P0147	2	NRL, RPGR	RP
P0227	2	KLHL7, NRL	RP
P0228	2	KLHL7, NRL	RP
P0234	2	EYS, ROM1	RP
P0256	2	IFT140, PROM1	RP
P0262	2	PRPF8, RPGR	RP
P0307	2	ABCA4, PROM1	STGD
P0317	2	ABCA4, ACO2	OA/STGD
P0357	2	CEP290, RPGR	LCA
P0358	2	CEP290, RPGR	LCA
P0469	2	NYX, RP9	CSNB
P0524	2	CPLANE1, PRPH2	RP
P0540	2	PRPH2, RHO	RP
P0541	2	RHO, RLBP1	RP
P0550	2	RIMS1, TOPORS	RP
P0733	2	PITPNM3, RPGR	RP
P0735	2	CNGA1, USH2A	USH
P0744	2	ABCA4, NPHP4	STGD
P0772	2	ABCA4, CHM	STGD

[a] For each patient, the implicated genes and clinical phenotype(s) are indicated.

2.4 IRD genotype-inheritance patterns

All patient–gene configurations were assigned to a descriptive IRD genotype–inheritance pattern before downstream aggregation. These categories define the structure of Figure 2 in the main manuscript and were used to summarise the Mendelian and mitochondrial architecture of the cohort. Classification was based on the curated inheritance mode of each gene, the number of distinct variants observed in that gene for a given patient, reported zygosity, and patient sex for X-linked loci.

Analytical note: case resolution was performed using a hierarchical rule-based framework evaluated at the gene level. The framework integrated inheritance mode, allelic configuration,

pathogenicity evidence, loss-of-function status, zygosity, sex for X-linked genes, population allele frequency and phenotype compatibility. Rules 1–15 capture canonical AR, AD, XL, AR/AD and mitochondrial configurations, whereas Rules 16–17 represent lower-confidence frequency-based extensions applied after stronger pathogenicity or loss-of-function criteria had failed. Resolved cases fulfilled complete inheritance-compatible genetic criteria; partially resolved cases contained plausible disease-associated variation but incomplete allelic, inheritance or evidence support.

Supplementary Table S4. IRD genotype-inheritance patterns observed in the SARA cohort

Item	Pattern	Operational definition	Classification rule
1	AR biallelic compound	Patient with two or more distinct variants in the same autosomal-recessive gene, with heterozygous calls consistent with a possible compound-heterozygous configuration.	Inheritance = AR; number of variants per patient-gene >=2; zygosity includes heterozygous
2	AR homozygous	Patient with a homozygous variant in an autosomal-recessive gene.	Inheritance = AR; zygosity includes homozygous
3	AR heterozygous carrier	Patient with a single heterozygous variant in an autosomal-recessive gene.	Inheritance = AR; number of variants per patient-gene = 1; zygosity includes heterozygous
4	AD hetero / homozygous	Patient with one or more variants in an autosomal-dominant gene, irrespective of heterozygous or homozygous reporting.	Inheritance = AD
5	XL hemizygous	Male patient with one or more variants in an X-linked gene.	Inheritance = XL; sex = male
6	XL female carrier	Female patient with one or more variants in an X-linked gene.	Inheritance = XL; sex = female
7	MT	Patient with a mitochondrial variant.	Inheritance = MT
8	AR/AD heterozygous	Patient with a single variant in a gene with both autosomal-recessive and autosomal-dominant disease associations.	Inheritance = AR/AD; number of variants per patient-gene = 1
9	AR/AD biallelic	Patient with two or more variants in a gene with both autosomal-recessive and autosomal-dominant disease associations.	Inheritance = AR/AD; number of variants per patient-gene >=2

2.5 Gene inheritance modes

Supplementary Table S5. Assigned modes of inheritance for the 109 IRD genes observed in the SARA cohort

Mode	N	Genes
AR	57	ABCA4, ADGRV1, AGBL5, AHI1, AIPL1, ALMS1, ARL6, BBS2, BBS5, BBS9, CACNA2D4, CC2D2A, CDH23, CDHR1, CEP290, CERKL, CLN3, CNGA1, CNGA3, CNGB1, CNGB3, COL18A1, CPLANE1, CRB1, EYS, FAM161A, IFT140, INPP5E, IQCB1, MERTK, MFSD8, MYO7A, NMNAT1, NPHP1, NPHP4, OAT, PANK2, PCARE, PCDH15, PDE6A, PDE6C, PHYH, RBP3, RDH11, RDH12, RLBP1, RPE65, RPGRIP1, RTN4IP1, SCLT1, SPATA7, TPP1, TTLL5, TULP1, TYR, USH1C, USH2A
AD	35	ACO2, BEST1, COL11A1, COL11A2, CRX, ELOVL4, GUCA1A, HK1, IMPDH1, KLHL7, LRP5, MFN2, NR2E3, NR2F1, NRL, OPA1, OPA3, OTX2, PAX6, PITPNM3, PRDM13, PRPF3, PRPF31, PRPF8, PRPH2, RHO, RIMS1, ROM1, RP1, RP1L1, RP9, SNRNP200, TIMP3, TOPORS, UNC119
XL	7	CACNA1F, CHM, NYX, OPN1MW, RP2, RPGR, RS1
AR/AD	6	GUCY2D, IMPG2, PDE6B, PROM1, RBP4, ZNF408
MT	4	ATP6, ND1, ND4, ND6

Analytical note: SARA inheritance assignments were compared with the curated inherited retinal disease classifications provided by the RetiGene Atlas. Of the 109 genes observed in the SARA cohort, 100 were found in the current RetiGene reference after gene-symbol harmonisation. Overall, inheritance assignments were highly concordant: 81 genes showed exact agreement, while 17 additional genes were considered compatible because RetiGene annotated them as dual-inheritance genes or because the SARA assignment represented a phenotype-specific subset of a broader RetiGene annotation. One formal difference was identified for **RBP4**, which was classified as AR/AD in SARA but AR in RetiGene; all SARA observations involving **RBP4** were biallelic and therefore compatible with recessive disease expression. Nine observed genes were not present in the current RetiGene reference used here (**COL11A2, CPLANE1, PITPNM3, PRDM13, RDH11, RIMS1, ROM1, RP9,** and **UNC119**) and should be interpreted cautiously as candidate, modifier, historical or non-primary annotations depending on gene and clinical context.

2.6 SARA–SPG et al., phenotype frequencies

Supplementary Table S6. Top 20 clinical phenotypes in the merged SARA + SPG et al. (2023) cohort, calculated at the patient level.[a]

Phenotype	SARA + SPG et al. (n, %)	SARA only (n, %)	SPG et al. only (n, %)[b]
RP	585 (59.9%)	328 (40.5%)	294 (62.3%)
STGD	182 (18.6%)	159 (19.7%)	34 (7.2%)
USH	64 (6.6%)	30 (3.7%)	35 (7.4%)
LCA	60 (6.1%)	60 (7.4%)	0 (0.0%)
LHON	56 (5.7%)	56 (6.9%)	0 (0.0%)
CHM	43 (4.4%)	24 (3.0%)	19 (4.0%)
MD	34 (3.5%)	13 (1.6%)	21 (4.4%)
CD/CRD	28 (2.9%)	17 (2.1%)	11 (2.3%)
ACHM	23 (2.4%)	17 (2.1%)	6 (1.3%)
DOA	22 (2.3%)	22 (2.7%)	0 (0.0%)
XLRS	17 (1.7%)	7 (0.9%)	10 (2.1%)
BEST	16 (1.6%)	16 (2.0%)	0 (0.0%)
OA	13 (1.3%)	6 (0.7%)	2 (0.4%)
CSNB	12 (1.2%)	11 (1.4%)	1 (0.2%)
BBS	11 (1.1%)	6 (0.7%)	5 (1.1%)
STICKLER	9 (0.9%)	2 (0.2%)	7 (1.5%)
BATTEN	8 (0.8%)	8 (1.0%)	0 (0.0%)
ANIR	5 (0.5%)	2 (0.2%)	3 (0.6%)
KNOBLOCH	5 (0.5%)	1 (0.1%)	4 (0.8%)
CD/CRD/LCA	4 (0.4%)	4 (0.5%)	0 (0.0%)

[a] Each patient was counted once per phenotype, regardless of the number of associated variants. Percentages are reported relative to the total number of patients with a defined clinical phenotype in each cohort (merged, SARA only, and SPG et al. only). The 20 most frequent phenotypes accounted for 97.3% of patients with a defined clinical phenotype in the merged cohort.

[b] The SPG et al. dataset was derived from the work of Schlottmann P.G. et al. (2023). Cases labelled as "EOSRD" in the SPG et al. dataset, representing a heterogeneous group of early-onset severe retinal disorders rather than a defined clinical phenotype, were excluded from phenotype frequency analyses and the corresponding percentage calculations.

2.7 SARA–SPG et al., gene frequencies

Supplementary Table S7. Top 20 disease-associated genes in the merged SARA–SPG et al. cohort, calculated at the patient level.[a]

Gene	SARA + SPG et al. (n, %)	SARA only (n, %)	SPG et al. only (n, %)[b]
ABCA4	189 (19.2%)	164 (20.3%)	37 (7.5%)
RPGR	128 (13.0%)	68 (8.4%)	61 (12.3%)
USH2A	117 (11.9%)	53 (6.6%)	64 (13.0%)
RHO	88 (8.9%)	17 (2.1%)	71 (14.4%)
ND4	46 (4.7%)	46 (5.7%)	0 (0.0%)
CHM	44 (4.5%)	25 (3.1%)	19 (3.8%)
CRB1	33 (3.4%)	21 (2.6%)	12 (2.4%)
PRPF31	31 (3.2%)	13 (1.6%)	18 (3.6%)
EYS	30 (3.0%)	15 (1.9%)	17 (3.4%)
RDH12	26 (2.6%)	14 (1.7%)	12 (2.4%)
PRPH2	24 (2.4%)	17 (2.1%)	7 (1.4%)
BEST1	22 (2.2%)	17 (2.1%)	5 (1.0%)
OPA1	21 (2.1%)	19 (2.3%)	2 (0.4%)
RP2	21 (2.1%)	13 (1.6%)	8 (1.6%)
RPE65	19 (1.9%)	18 (2.2%)	1 (0.2%)
RS1	19 (1.9%)	7 (0.9%)	12 (2.4%)
PROM1	18 (1.8%)	13 (1.6%)	5 (1.0%)
NR2E3	17 (1.7%)	14 (1.7%)	3 (0.6%)
CEP290	15 (1.5%)	14 (1.7%)	1 (0.2%)
GUCY2D	15 (1.5%)	9 (1.1%)	6 (1.2%)

[a] Each patient was counted once per gene, irrespective of the number of reported variants. Percentages are reported relative to the total number of patients with at least one reported gene in each cohort (merged, SARA only, and SPG et al. only). The 20 most frequently involved genes accounted for 71.6% of patients with at least one reported disease-associated gene.

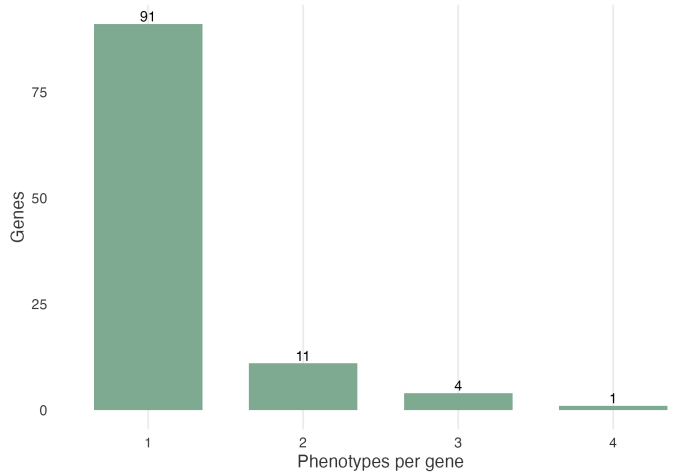
[b] The SPG et al. dataset was derived from the work of Schlottmann P.G. et al. (2023).

3 Other analyses: extended pleiotropy and phenotype–gene heterogeneity

Inherited retinal diseases (IRDs) exhibit both genetic pleiotropy and diagnosis-specific heterogeneity. While some genes are associated with multiple clinical phenotypes, some diagnoses can arise from many distinct disease-associated genes. To further explore this architecture in the SARA cohort, we analysed the distribution of phenotypes per gene and genes per phenotype, excluding non-primary (`NO-PRIM`) annotations from phenotype-level summaries.

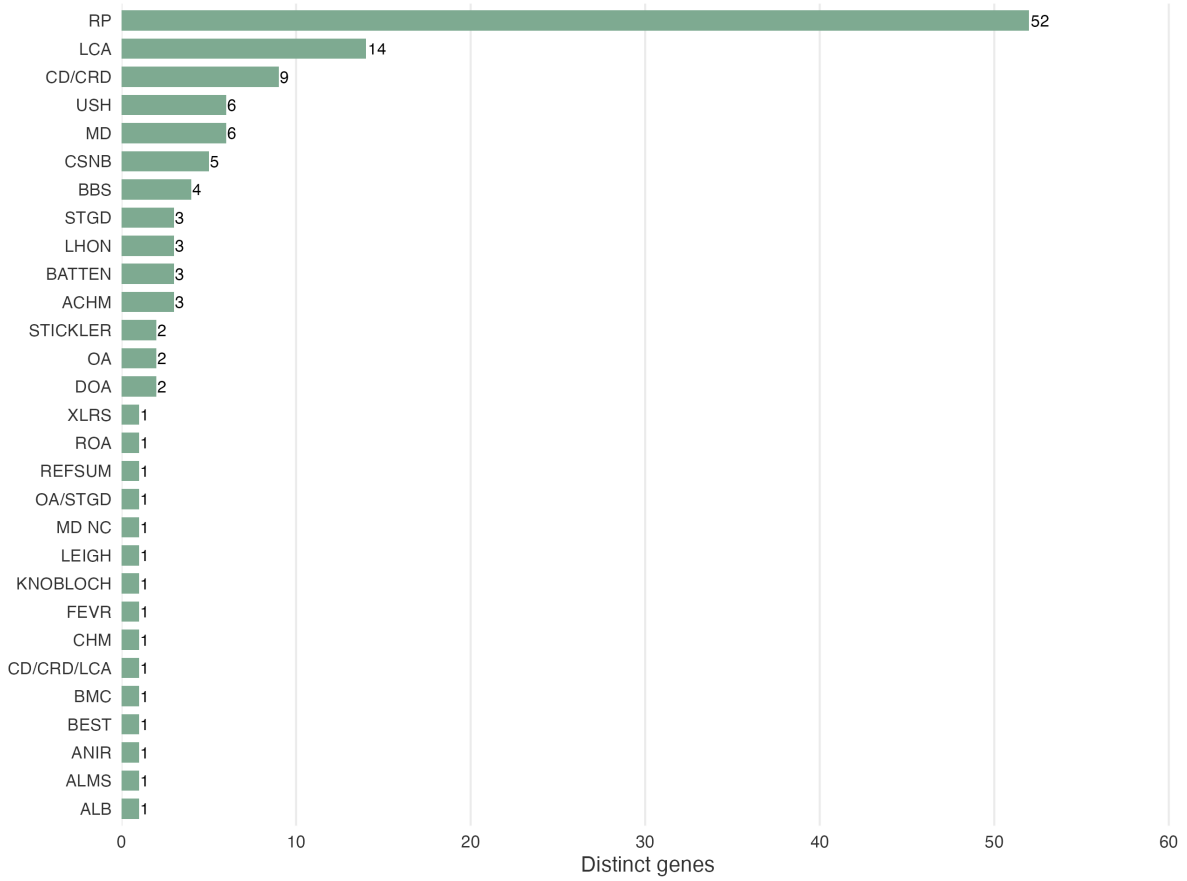
In addition, Phenotype–Gene associations were separated into recurrent (at least 1% of cases) and low-frequency (less than 1%) components to distinguish the compact recurrent core from the broader long-tail diversity of genotype–phenotype relationships across the cohort.

3.1 Distribution of phenotypes per gene (pleiotropy)



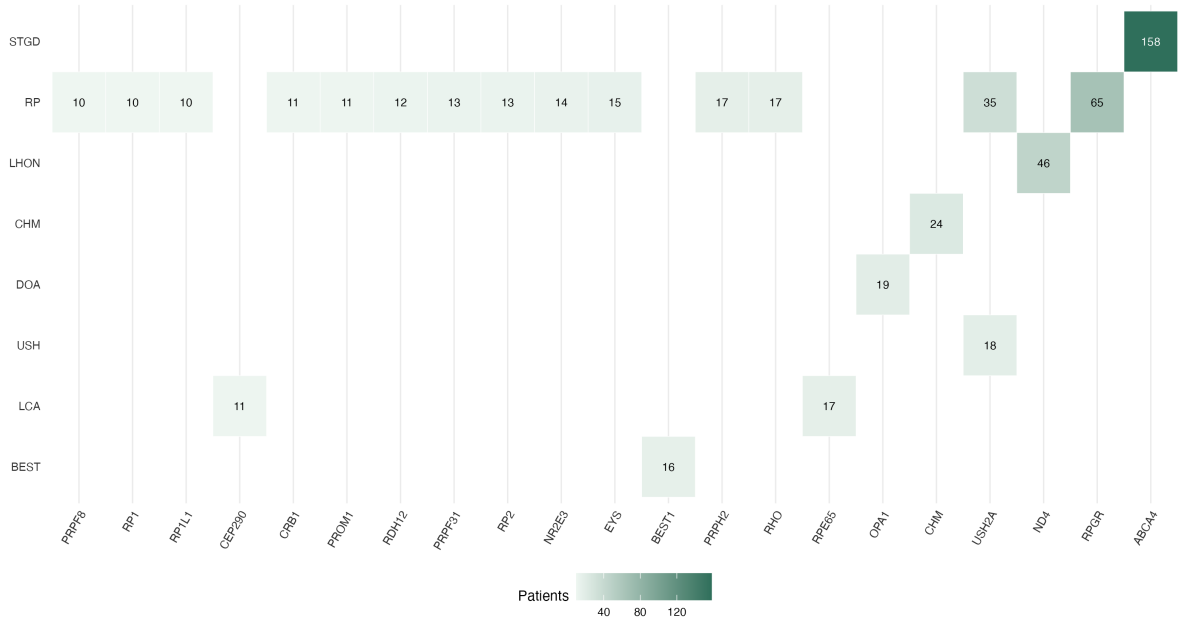
Supplementary Figure S6. Distribution of the number of phenotypes associated with each disease-associated gene in the SARA cohort. Among the 109 genes with at least one assigned clinical diagnosis, most were linked to a single phenotype (88 genes), whereas 16 genes were associated with two phenotypes, four genes with three phenotypes, and one gene with four phenotypes. This distribution illustrates a compact but detectable layer of genetic pleiotropy beyond high-frequency loci.

3.2 Genetic heterogeneity across IRD phenotypes



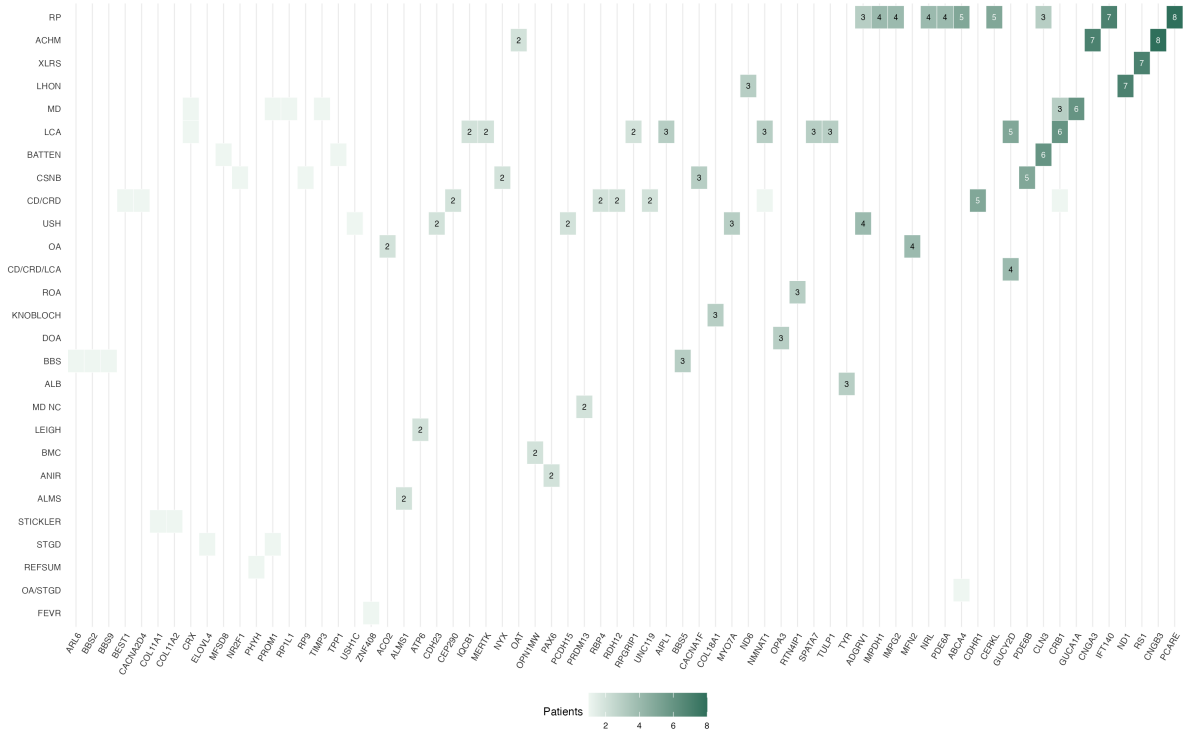
Supplementary Figure S7. Number of distinct disease-associated genes identified for each inherited retinal disease phenotype in the SARA cohort. Across 29 clinical phenotypes after excluding non-primary annotations, retinitis pigmentosa showed the broadest genetic heterogeneity (52 genes), followed by Leber congenital amaurosis (14 genes), cone/cone–rod dystrophy (9 genes), macular dystrophy and Usher syndrome (6 genes each), and congenital stationary night blindness (5 genes). These results highlight the asymmetric architecture of IRDs, in which a limited number of phenotypes account for a large fraction of the observed gene-level heterogeneity.

3.3 Phenotype–gene architecture of the SARA cohort (Core: at least 1%)



Supplementary Figure S8. Matrix representation of phenotype–gene associations accounting for at least 1% of genetically resolved cases. The recurrent core included 22 phenotype–gene pairs across 8 phenotypes and 21 genes. The largest associations corresponded to well-established relationships such as STGD–ABCA4, RP–RPGR, LHON–ND4, and RP–USH2A, illustrating the concentrated high-frequency architecture underlying a substantial fraction of the cohort.

3.4 Phenotype–gene architecture of the SARA cohort (Long-tail: less than 1%)



Supplementary Figure S9. Matrix representation of low-frequency phenotype–gene associations contributing to less than 1% of genetically resolved cases. The long-tail component included 107 phenotype–gene pairs involving 27 phenotypes and 94 genes. Most associations were observed in only one or two patients, illustrating the extensive molecular and phenotypic diversity underlying inherited retinal diseases. For readability, only the 10 most frequent low-frequency associations per phenotype are shown.

4 Supplementary Data

4.1 Supplementary Data File S1 and reproducibility script

The supplementary material is accompanied by Supplementary Data File S1, which contains the complete publication-ready SARA variant dataset used for the analyses reported in this manuscript.

A reproducibility script is also provided with the supplementary material. This script uses Supplementary Data File S1 as input and reproduces the main figure panels derived from the shared dataset.