Vision-related convergent gene losses reveal SERPINE3’s unknown role in the eye

Henrike Indrischek, Juliane Hammer, Anja Machate, Nikolai Hecker, Bogdan M. Kirilenko, Juliana G. Roscito, Stefan Hans, Caren Norden, Michael Brand, Michael Hiller

1 Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, 01307 Dresden, Germany
2 Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, 01187 Dresden, Germany
3 Center for Systems Biology Dresden, Pfotenhauerstr. 108, 01307 Dresden, Germany
4 LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt, Germany
5 Senckenberg Research Institute, Senckenberganlage 25, 60325 Frankfurt, Germany
6 Faculty of Biosciences, Goethe-University, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany
7 Center for Regenerative Therapies Dresden (CRTD), TU Dresden, Fetscherstraße 105, 01307 Dresden, Germany

# current affiliation: Center for Brain & Disease Research, VIB-KU Leuven, Herestraat 49, 3000 Leuven, Belgium; Department of Human Genetics, KU Leuven, Herestraat 49, 3000 Leuven, Belgium
§ current affiliation: DRESDEN-concept Genome Center, DFG NGS Competence Center, c/o Center for Molecular and Cellular Bioengineering (CMCB), TU Dresden, 01307 Dresden, Germany
$ current affiliation: Instituto Gulbenkian de Ciência, Rua da Quinta Grande 6, 2780-156 Oeiras, Portugal

*To whom correspondence should be addressed: Michael Brand <michael.brand@tu-dresden.de> & Michael Hiller <michael.hiller@senckenberg.de>

Abbreviations:

KO - knockout
WT - wild type
RPE - retinal pigment epithelium
MG - Müller glia
RT-qPCR - reverse transcription quantitative PCR
RNA-Seq - RNA-Sequencing
SNP - single nucleotide polymorphism
ISH - In situ hybridization
FDR - false discovery rate
Abstract

Despite decades of research, knowledge about the genes that are important for development and function of the mammalian eye and are involved in human eye disorders remains incomplete. During mammalian evolution, mammals that naturally exhibit poor vision or regressive eye phenotypes have independently lost many eye-related genes. This provides an opportunity to predict novel eye-related genes based on specific evolutionary gene loss signatures. Building on these observations, we performed a genome-wide screen across 49 mammals for functionally uncharacterized genes that are preferentially lost in species exhibiting lower visual acuity values. The screen uncovered several genes, including SERPINE3, a putative serine proteinase inhibitor. A detailed investigation of 381 additional mammals revealed that SERPINE3 is independently lost in 18 lineages that typically do not primarily rely on vision, predicting a vision-related function for this gene. To test this, we show that SERPINE3 has the highest expression in eyes of zebrafish and mouse. In the zebrafish retina, serpine3 is expressed in Müller glia cells, a cell type essential for survival and maintenance of the retina. A CRISPR-mediated knockout of serpine3 in zebrafish resulted in alterations in eye shape and defects in retinal layering. Furthermore, two human polymorphisms that are in linkage with SERPINE3 are associated with eye-related traits. Together, these results suggest that SERPINE3 has a role in vertebrate eyes. More generally, by integrating comparative genomics with experiments in model organisms, we show that screens for specific phenotype-associated gene signatures can predict functions of uncharacterized genes.

Significance statement

Genes with eye-related functions are maintained by selection in humans and other species where vision is crucial. In species that primarily use other senses such as smell and hearing, lack of selection can result in repeated losses of eye-related genes, yielding a specific evolutionary gene loss pattern. Since knowledge of genes important for eye function and human eye diseases is still incomplete, we utilized this evolutionary principle to uncover uncharacterized genes that may have eye-related functions. Our genome-wide screen revealed SERPINE3 as a top candidate. Experiments showed that SERPINE3 is predominantly expressed in eyes and that serpine3 knockout in zebrafish resulted in morphological eye defects. This shows how comparative genomics can reveal novel insights into the function of uncharacterized genes.
Introduction

Disorders affecting eyes range from subtle vision impairment to blindness and are among the most prevalent diseases in the human population (1). For example, an estimated 76 million people worldwide suffer from glaucoma (2), a disease involving optic nerve damage, and about 1.8 million people are blind due to age-related macular degeneration, a degenerative disease of the central retina affecting retinal pigment epithelium (RPE) and photoreceptor cells (3, 4).

Extensive research in the past decades identified many protein-coding genes with crucial roles in development and maintenance of different tissues and cell types in the eye as well as numerous genes that are associated with genetic eye disorders (1, 5). For example, the RetNet database (6) lists 271 genes associated with heritable retinal diseases. Even though the eye probably represents one of the best studied organs, our knowledge of the genes underlying eye diseases and disorders is still incomplete. For example, linkage analysis in patients with cataract, microcornea, microphthalmia, and iris coloboma identified new genomic loci linked to the diseases; however, the disease-causing genes have remained elusive (7-9). Similarly, the Cat-Map database (10) lists several additional cataract-associated loci where the underlying disease-causing gene has not been identified. Furthermore, there are still thousands of genes that have not been experimentally investigated in detail, leaving many genes where potential eye-related functions remain to be discovered. Indeed, systematic knockouts of 4364 genes in mouse detected ocular phenotypes for 347 genes, with 75% of them not been known as eye-related before (5). This indicates that vision-related genes as well as genes associated with genetic eye disorders remain to be identified and characterized.

Interestingly, many genes that are linked to human eye diseases are inactivated (lost) in non-human mammals that naturally exhibit poor vision (11, 12). For example, subterranean mammals, such as the blind mole rat, naked mole rat, star-nosed mole and cape golden mole, exhibit gene-inactivating mutations in genes implicated in cataract, retinitis pigmentosa, color or night blindness, or macular degeneration in human (e.g. ABCA4, BEST1, CRYBA1, EYS, GJA8, GNAT2, PDE6C, ROM1 and SLC24A1) (12-18). Symptoms that characterize these human eye diseases resemble traits found in these subterranean mammals, such as highly-degenerated retinae and lenses and sometimes blindness. Similarly, losses of the short wave sensitive opsin (OPN1SW), which is linked to color blindness in humans, occurred in several mammalian lineages such as cetaceans, bats, sloths and armadillos that are consequently inferred to have monochromatic vision (11, 19-23). In addition to losses of vision-related genes, subterranean mammals also exhibit widespread sequence and transcription factor binding site divergence in eye-related regulatory elements (24-26). Such mutations can cause mis-expression of target genes in ocular tissues such as the lens (27). Loss and divergence of vision-related genes and regulatory elements in these mammals is likely caused by the lack of natural selection on
maintaining functional eyes in a dark environment. Taken together, previous studies established a clear association between naturally occurring poor vision phenotypes and regressive evolution of the genetic machinery required for functional eyes (11, 17, 24, 25).

Here, we performed a genome-wide screen for genes preferentially lost in independent mammalian lineages with a low visual acuity with the goal of revealing currently uncharacterized genes, where vision-related gene loss patterns would predict vision-related functions. In addition to identifying losses of known vision-related genes, our screen revealed previously unknown losses of several functionally uncharacterized genes, among them SERPINE3, which is lost in 18 mammalian lineages that often do not use vision as the primary sense. We show that SERPINE3 is specifically expressed in eyes of zebrafish and mouse. Furthermore, by knockout of serpine3 in zebrafish, we show that the gene is required for maintenance of a proper retinal lamination and overall eye shape, which confirms the predicted vision-related function. Collectively, our results confirm that SERPINE3 has functions in vertebrate eyes and our discovery-driven study demonstrates how specific evolutionary divergence patterns can reveal novel insights into gene function (28).

Results

A genome-wide screen retrieves genes preferentially lost in mammals with low visual acuity

To uncover potentially unknown vision-related genes, we used the Forward Genomics framework (29) to search for associations between convergent gene losses and poor vision phenotypes in mammalian lineages, where poor vision has independently evolved. Vision is a complex multifaceted trait that may not be easily captured with a single variable. In our study, we decided to select mammals with poor vision based on visual acuity values for two reasons. First, visual acuity describes the ability of an animal to resolve static spatial details, which generally reflects how much an animal relies on vision in comparison to other senses (30). Second, visual acuity data is available for 49 placental mammals with sequenced genomes, enabling a comprehensive genomic screen (Figure 1 - source data 1). Using visual acuity values, we defined two groups. Low-acuity species have low visual acuity values \(< 1 \ (\log_{10}(va)<0)\), which comprises ten species (three echolocating bats, three rodents and four subterranean mammals) representing seven lineages (Figure 1A). All other species have visual acuity values \(> 1\) and comprise the group with higher visual acuity values.

Using this classification, we performed a genome-wide screen for genes that exhibit inactivating mutations (frame-shifting insertions or deletions, premature stop codons, splice site mutations, exon or gene deletions) preferentially in low-acuity species, similar to previous screens (31-33). Using a false discovery rate (FDR) cutoff of 0.05 and filtering for
genes lost in at least three low-acuity lineages, we obtained 29 genes that are each convergently lost in at least three lineages of low-acuity mammals (Figure 1B, Figure 1 - source data 2). These genes include several known vision-related genes, such as components of the photoreceptor signal transduction cascade (GUCA1B, ARR3), a factor required for retinal organization (CRB1), lens crystallins (CRYBA1, CRYBB3, CRYGS) and the cornea specific keratin 12 (KRT12). The set of 29 genes is enriched in vision-related functions such as the Gene Ontology term "visual perception" (p=1.8e-7), expression in the mouse lens (p=5.5e-3), and human eye diseases (cataract: p=8.5e-3) (Figure 1C, Figure 1 - source data 3). To confirm the specificity of these results, we performed a control screen for genes that are preferentially lost in high-acuity sister species of the low-acuity mammals. This control screen retrieved only two genes, none of which have known functions in the eye (Figure 1 - source data 4). Together, this shows that our genome-wide screen for genes preferentially lost in low-acuity species successfully retrieved known vision-related genes.

Interestingly, while 15 of the 29 top-ranked hits have no studied role in the eye (Figure 1B), many of these genes are expressed in human eyes. RSKR, LACTBL1, LBHD1 and ZNF529 even cluster in their expression pattern with other retina phototransduction and visual perception genes (34) (Figure 1 - source data 2). The preferential loss of these genes in species that exhibit lower visual acuity values and that have convergently lost other known vision-related genes predicts an uncharacterized vision-related function for some of these 15 genes (Figure 1A).

**SERPINE3 is convergently lost in low-acuity mammals**

We sought to experimentally test this prediction for a gene that is ranked highly in our screen. Since the first-ranked gene, TSACC (TSSK6 activating cochaperone), encodes a chaperone that is specifically expressed in testis (34, 35) and is therefore unlikely to have an eye-related function, we focused on the second-ranked candidate, SERPINE3 (serpin family E member 3). SERPINE3 is largely uncharacterized and classified as lost in 7 of the 10 low-acuity mammals (Figure 1, Figure 1 - source data 2). SERPINE3 independently accumulated inactivating mutations in all four subterranean species (cape golden mole, star-nosed mole, naked mole rat, blind mole rat). Within bats, SERPINE3 is inactivated in the two Myotis bats and the big brown bat (Yangochiroptera). These three species use echolocation instead of vision as the primary sense for hunting. In contrast, non-echolocating flying foxes (Pteropodidae) that rely more on vision, possess an intact SERPINE3 gene. To show that SERPINE3 loss is robustly associated with poor vision, we varied the visual acuity thresholds used to classify species as having low-acuity vision and explored an alternative approach for defining poor vision species based on a molecular loss signature of known vision-related genes. All three modified screens consistently retrieved SERPINE3 as one of the top-ranked hits (Figure 1 - source data 5-7), showing that this association is robust to the selected thresholds and phenotype definition.

**SERPINE3 became dispensable in many mammals that do not primarily rely on vision**
To explore the evolution of SERPINE3 in additional mammalian genomes, we made use of an orthology data set generated by the TOGA (Tool to infer Orthologs from Genome Alignments) method (36) that includes 381 additional placental mammalian species that were not part of the genomic screen. Interestingly, this substantially extended data set revealed a number of additional losses of SERPINE3, typically in species that do not rely on vision as their primary sense (Figure 2A). We only considered SERPINE3 losses for which at least one inactivating mutation could be validated (see Methods, Figure 2 - source data 1-2).

Further supporting SERPINE3 losses, analyses of publicly-available RNA-seq data indicates that remnants of inactivated SERPINE3 genes do not show relevant expression (Figure 2 - figure supplement 1, Figure 2 - source data 3). A detailed analysis of inactivating mutations indicates that SERPINE3 is inactivated in 70 of the 430 analyzed placental mammals and that the gene is convergently lost at least 18 times in placental mammal evolution (Figure 2 - figure supplements 2-8, source data 1).

For example, SERPINE3 is lost in several mammalian clades with nocturnal representatives that have a partially burrowing lifestyle such as pangolins (Manidae), aardvark and armadillos (Cingulata), which are characterized by proportionally small eyes (37, 38) (Figure 2A, Figure 2 - figure supplements 2-3). All representatives of these clades use smell and hearing as primary senses for perception of environmental clues and have a poor sense of vision (39). SERPINE3 has also been lost in the stem lineage of Pilosa (sloths and anteaters). Whereas extant sloths have rather high visual acuity values (10.2 for southern two-toed sloth) (40), the Pilosa ancestor was likely a digging/burrowing species (41), indicating that SERPINE3 loss originally occurred in a burrowing species. SERPINE3 is also lost in solenodon, European shrew and tenrecs (Tenrecidae), species with small eyes that use echolocation for close-range spatial orientation (42-44) (Figure 2B, Figure 2 - figure supplements 2, 4). For bats, our extended data set shows that SERPINE3 is convergently lost in seven echolocating lineages, but remains intact in all 13 analyzed Pteropodid bats, revealing a clear pattern of convergent losses of this gene restricted to bat lineages relying on laryngeal echolocation (Figure 2 - figure supplements 5-7). The sac-winged bat is the only laryngeal echolocating bat with an intact SERPINE3; however, selection rate analysis indicates that the gene evolves under relaxed selection (Figure 2 - source data 4). The available genomes support shared gene-inactivating mutations among many bat species, which likely represent ancestral events (Figure 2A right). Expanding upon the two convergent losses of SERPINE3 in subterranean rodents detected in the initial screen, the gene is also convergently lost in the fossorial Damara mole rat and evolves under relaxed selection in the Transcaucasan mole vole (Figure 2 - source data 4, Figure 2 - figure supplement 8), while being intact in all other 63 rodents. In the expanded data set, we further uncovered the loss of SERPINE3 in another clade of fossorial moles (Talpidae) within the order of Eulipotyphla (Figure 2 - figure supplement 4). A splice site mutation shared between dugong and manatee as well as patterns of relaxed selection indicate an ancestral loss of SERPINE3 in the ancestor of sirenians, which mostly use their tactile sense for navigation through murky water (45) (Figure 2 - figure supplement 2, Figure 2 - source data 1).
Consistent with our definition of poor vision based on visual acuity, the Florida manatee has a reduced ability to resolve spatial detail (va = 1.6 (46)) compared to its closest relatives, the African elephant (va = 13.16 (47)), which has an intact SERPINE3 gene. Finally, SERPINE3 is also lost in the cape elephant shrew, which is a mostly diurnal species with relatively large eyes (48); however, loss of SERPINE3 already occurred in the ancestral Afroinsectiphilia lineage (Figure 2 - figure supplement 2), which was presumably nocturnal (49).

Together, many independent gene losses in species that do not rely on vision as their primary sense predicts a vision-related function for SERPINE3 that became dispensable in these mammals, leading to convergent SERPINE3 losses due to relaxed selection.

SERPINE3 encodes a putative secreted proteinase inhibitor

Based on sequence homology, SERPINE3 is classified as a member of the serine proteinase inhibitor (SERPIN) family. Many members of this family are secreted into the extracellular space and inhibit their substrates by covalent binding (50, 51). We performed a SERPINE3 sequence analysis, which revealed that key sequence features of inhibitory serpins are well conserved among placental mammals, suggesting that SERPINE3 also functions as a secreted serine proteinase inhibitor (SI Text, Figure 2 - figure supplements 9-10).

Serpins have roles in coagulation, angiogenesis, neuroprotection and inflammation, and several serpins have been implicated in human diseases (51, 52). However, the functional role of SERPINE3 is largely unknown as the gene has never been studied in an animal or cellular model. A thorough literature search revealed that SERPINE3 is listed (often in the supplement) among many other genes, as differentially expressed in large-scale expression analyses. For example, Serpine3 is upregulated in the mouse retina in response to overexpressing neuroprotective factors (53). Serpine3 was also upregulated in complement component 3 (C3) knockout mice, which represent a model of the aged retina (54), and downregulated in the eye of knockout mice for Pcare, a causal gene for retinitis pigmentosa (55). Together, while this gene was never studied in greater detail, literature clues and the striking convergent gene loss pattern in mammals with poor vision suggest that SERPINE3 may have a functional role in the eye.

Serpine3 is specifically expressed in Müller glia in the adult zebrafish retina

To test the prediction that SERPINE3 has an eye-related function, we first analyzed its expression pattern in adult zebrafish (Figure 3). Zebrafish has proven as valuable model species for the study of eye genetics as it has a cone-dominated retina (~60% cones, ~40% rods) (56), similar to the central human retina. Furthermore, zebrafish have a single serpine3 ortholog that is located in a context of conserved gene order (Figure 4A), which makes zebrafish a suitable model for investigation of serpine3 expression and function. Reverse transcription quantitative PCR (RT-qPCR) analysis with two normalization genes (actb and rpl13a with n=8 and n=3 biological replicates, respectively) consistently revealed that the
highest *serpine3* expression is in the eye, followed by expression in the brain (Figure 3A, B, Figure 3 - source data 1-2). For *actb*, *serpine3* expression is significantly higher in eye than in other tissues (Figure 3A).

To better characterize *serpine3* expression in the eye, we next performed RT-qPCR on samples of RPE, retina and eye without retina and RPE (Figure 3C, Figure 3 - source data 3). *Serpine3* is expressed significantly higher in the retina compared to RPE only (two-sided unequal variances t-test, p=0.01, Figure 3C). Expression is barely detectable in whole eye without retina and RPE (two-sided unequal variances t-test, eye vs. retina, p=0.009, Figure 3C), suggesting that the expression signal obtained from whole eye RT-qPCR mostly originates from expression in retina.

To specify in which cell types *serpine3* is expressed, we finally performed *in situ* hybridization (ISH) on adult zebrafish retinae. *Serpine3* expression is detected in the inner nuclear layer throughout the retina with stronger signals in the ventral region close to the optic nerve (Figure 3E, F). Co-staining with different cell type specific markers reveals expression of *serpine3* in a fraction of glial fibrillary acidic protein (*gfap*)-positive Müller glia (MG) cells (Figure 3G, H, Figure 3 - figure supplement 1), whereas expression of *serpine3* mRNA was not detected in bipolar or amacrine cells (Figure 3 - figure supplement 2).

To also explore *SERPINE3* expression in mammals, we first used RT-qPCR to analyze *Serpine3* expression in mouse. Similar to zebrafish, mouse *Serpine3* expression is highest in whole eye, whereas expression was not detectable in colon, cortex, heart, liver, spleen and testis (Figure 3D, Figure 3 - source data 4). Next, we analyzed *SERPINE3* expression in other mammals and vertebrates using publicly available expression data sets. Despite sparser data, we found evidence for expression of *SERPINE3* in eyes of human, rat, cat, cow and chicken, which suggests a conserved expression pattern and a role in vertebrate vision (Figure 3 - source data 5).

In summary, our experiments in zebrafish and mouse together with available data sets of different vertebrates show that *SERPINE3* is expressed in the vertebrate eye, specifically in zebrafish in MG cells.

### Knockout of serpine3 in zebrafish leads to morphological defects in the eye including the retina

Next, we tested whether *serpine3* inactivation results in an eye phenotype. To this end, we deleted the transcription start site of *serpine3* with CRISPR-Cas9, denoted as the *serpine3* cbg17 allele (Figure 4A, Figure 4 - figure supplements 1-2). Adult homozygous *serpine3* cbg17 fish were viable and fertile. Using RT-qPCR and ISH, we confirmed that the deletion of the transcription start site abolished *serpine3* expression in retinae of adult homozygotes for *serpine3* cbg17 compared to wild type (WT) siblings (two-sided unequal variances t-test, p=0.049, Figure 4B, C, Figure 4 - source data 1).
We found that adult *serpine3*<sup>cbg17</sup> individuals frequently showed notches in the iris of one of the eyes (4 of 5 individuals, white arrows in Figure 4D, Figure 4 - source data 2, 3), which affected the eye’s overall shape. To confirm that these notch-caused shape deviations are caused by inactivation of *serpine3*, we generated an independent line, where we introduced an early frameshift in *serpine3*'s coding exon 1 with CRISPR-Cas9, denoted as the *serpine3*<sup>cbg18</sup> allele (Figure 4A, Figure 4 - figure supplements 1-2). *Serpine3*<sup>cbg18</sup> fish also showed a high frequency of notches in 13 of 20 individuals (Figure 4 - source data 2, 3). To quantify this shape deviation, we calculated iris solidity, which compares the area ratio of the eye’s outer shape (white line) and its concave shape (red dotted line in Figure 4D). Eyes of both *serpine3*<sup>cbg17</sup> and *serpine3*<sup>cbg18</sup> individuals had a significantly reduced solidity in comparison to their WT siblings (Wilcoxon rank sum test, p=0.009 and p=6.35e-9, respectively, Figure 4E). Iris circularity, another descriptor of the eye shape deviation, is also significantly reduced in homozygous individuals of both lines (Figure 4 - figure supplement 3). Together this indicates that mutations in *serpine3* cause eye shape deviations in zebrafish.

We next investigated the ocular morphology of adult fish of both lines and their WT siblings using hematoxylin/eosin sections (Figure 4F; Figure 4 - figure supplement 4-5). While ocular structure and optic nerves of all inspected eyes are largely normal (Figure 4 - figure supplement 4), the distance between retina and lens is reduced in homozygous fish of both alleles (Figure 4F, Figure 4 - figure supplement 5). A detailed inspection of the retinal organization revealed that although all retinal layers were present, the mutant retinae are generally less organized and structured compared to WT siblings. Retinal cells appear reduced in number and less densely packed. Most prominently, we noticed that rod outer segments and the pigmented RPE cells, which were aligned in WT fish, were not clearly separated in homozygous fish of both alleles (Figure 4F c-e, Figure 4 - figure supplement 5). Furthermore, we observed large clusters of pigmented cells in the photoreceptor layer (empty arrows, Figure 4F c-e) as well as single displaced pigmented cells in all retinal layers (yellow arrows, Figure 4F c-e). Similar alterations in retinal structure were detected in homozygous *serpine3*<sup>cbg18</sup> fish (Figure 4 - figure supplement 5). This shows that KO of *serpine3* in zebrafish results in morphological defects in the eye, characterized by differences in eye shape and retinal organization.

**Polymorphisms near human SERPINE3 are associated with human eye phenotypes**

We next analyzed recently published Genome Wide Association Study data for human single nucleotide polymorphisms (SNPs) associated with eye-related traits. This analysis revealed two such SNPs that are in linkage disequilibrium with SERPINE3 (Figure 4A, Figure 4 - source data 4). rs1028727 is located ~10 kb upstream of the SERPINE3 transcription start site and is associated with a decreased area of the optic nerve head (57). rs7327381 is located ~97 kb downstream of the SERPINE3 transcription start site and is associated with an increase in corneal curvature (58). This suggests that human SNPs
linked to SERPINE3 are associated with eye phenotypes, supporting a putative ocular function of human SERPINE3.
Discussion

Our study combines comparative genomics to predict genes having vision-related functions and experiments in zebrafish to confirm this prediction for a top-ranked candidate, the uncharacterized SERPINE3 gene. By conducting a genome-wide screen for genes that are preferentially lost in mammals with low visual acuity values, we uncovered both known vision-related genes as well as several genes that have no known eye-related function. One of the top-ranked candidates is SERPINE3, which we found to be independently lost at least 18 times in mammalian evolution, preferentially in species that do not use vision as the primary sense. For mouse and zebrafish, we show that the highest SERPINE3 expression is in the eye, which is corroborated by available expression data of other vertebrates. By generating the zebrafish serpine3\(^{cbg17}\) and serpine3\(^{cbg18}\) knockout (KO) lines, we show that inactivation of this gene results in abnormal eye shape and retinal lamination, revealing an eye-related function for serpine3. This is further supported by SERPINE3-linked polymorphisms that are associated with eye-related traits in human.

In zebrafish, serpine3 is expressed in a fraction of Müller glia (MG) cells. MG are the major retinal macroglia and perform numerous functions. By removing waste products and secreting (neuro)trophic substances and signaling molecules, they maintain the blood-retinal barrier and regulate vascularization (59). Most importantly, MGs are essential for the long-term viability of photoreceptors and other neuronal cell types (60). Our serpine3 KO fish exhibit a disorganization of RPE cells, a phenotype that resembles those previously observed in experiments that perturb MG function. For example, selective ablation of MG in adult mice results in eye defects including aggregation of RPE cells and displacement of pigment granules in the ganglion cell layer (61). Interestingly, we also observed that serpine3 KO affects eye shape, which cannot be readily explained by a direct MG-mediated effect. However, MG span the entire retina, connecting the extracellular space of retinal neurons, the vitreous and the capillaries at the apical retina (59, 62). It is therefore possible that a secreted protein, as predicted for serpine3 based on its conserved signal peptide, may also affect other eye subtissues. Finally, in zebrafish, MG are able to regenerate retinal neurons upon injury. However, serpine3 does not seem to be involved in this process as during stress-induced regeneration, it is upregulated in resting MG that do not proliferate (63).

In mammals, the SERPINE3 gene loss pattern and expression profile in several species also support an eye-related function; however, the cell type expression pattern may differ between mammals and zebrafish. While our experiments in zebrafish show serpine3 expression in MG, which is in agreement with (63), in mouse and human, SERPINE3 seems to be expressed in RPE (63-65) indicating that cell type specificity may differ between mammals and fish. As SERPINE3 is likely a secreted extracellular protein, it is possible that it has a similar function in mammals and zebrafish, despite secretion by different cell types. Whether this is the case or whether SERPINE3 function and protein expression pattern in the mammalian eye differs remains to be explored in future studies. At the molecular level,
this question may be addressed by investigating whether the molecular targets of 
SERPINE3 are conserved among vertebrates.

Of particular interest is elucidating the functional role of SERPINE3 in human. Several 
pieces of evidence indicate a potential role of this gene in anti-inflammatory processes and 
retinal survival. SERPINE3 is upregulated in human patients with age-related macular 
degeneration (66), a progressive eye disease that is linked to chronic inflammation and 
wound healing. Consistent with a role in retinal survival, mouse Serpine3 is upregulated after 
experimental overexpression of neurotrophin-4, a neuroprotective factor that promotes 
retinal survival (53). Furthermore, SERPINE3 is a hallmark gene of differentiated, healthy 
human RPE cells (67-69) that also have neurotrophic functions in the retina. It is thus 
conceivable that perturbation of proper SERPINE3 expression or function may influence 
age-related diseases or human eye phenotypes, as indicated by human polymorphisms that 
are linked to SERPINE3.

In addition to SERPINE3, our screen also revealed other candidate genes for which 
unknown eye-related functions are plausible. The LACTBL1 gene encodes a putative serine 
proteinase (70) and has an expression pattern similar to retina phototransduction genes 
(34). Furthermore, 20 kb upstream of LACTBL1 is a linked SNP (rs10158878) that is 
associated with refractive error in human (GWAS catalog, (71)). Another uncharacterized 
candidate gene with a strong visual acuity associated loss pattern is SAPCD1. A missense 
mutation (rs6905572) within SAPCD1 is associated with macular degeneration (dbSNP, 
(72)). The LRIT2 gene is linked to SNPs (rs12217769 and rs745480) that are associated 
with macular thickness and refractive error in human. Whereas this gene was not 
functionally characterized at the time our screen was conducted, a recent study showed that 
lrit2 knockdown in zebrafish led to a reduction in eye size (73). Finally, a high-throughput 
screen in mouse found that Pkd1l2 KO leads to abnormalities in lens morphology (74), but 
no follow-up experiments exist. A potential role of this gene in human disease is also 
indicated by an intronic SNP (rs873693), which is associated with age onset of myopia 
(GWAS catalog, (71, 75)). Thus, in addition to SERPINE3, for which we performed an initial 
functional characterization in zebrafish, our screen uncovered other promising, 
uncharacterized candidates that may have an eye-related function. Overall, this highlights 
the potential of comparative genomics to shed light on the functional roles of less 
characterized genes and to help to further identify human disease-causing genes (76).
Materials and Methods

Visual acuity values
We used publicly available visual acuity measurements (40, 77) (Figure 1 - source data 1 lists all primary references) to classify placental mammals into visual high-acuity and low-acuity species. Visual acuity can be measured by either behavioral experiments or calculated from the eye axial diameter, the peak ganglion cell density and a correction factor for diel activity (40). For species for which both measures were available, we used the behavioral visual acuity as this measure is more accurate (40). For species that lack visual acuity data, we used available visual acuity measurements of closely related species of the same genus or family (Figure 1 - source data 1). Three subterranean mammals lack available visual acuity measurements (cape golden mole, star-nosed mole, blind mole rat) but exhibit highly degenerated eyes. We therefore assumed a visual acuity of zero. In total, visual acuity values were obtained for 49 placental mammals that were included in a previously-generated whole genome alignment (78). Using a visual acuity threshold of one, we considered ten mammals (cape golden mole, naked mole rat, star-nosed mole, blind mole rat, big brown bat, little brown bat, David’s myotis bat, mouse, prairie vole, deer mouse) representing seven lineages, as low-acuity vision species. All other 39 species with visual acuity greater than one were considered as high-acuity vision species.

Forward genomics screen
To screen for genes that are preferentially lost in low-acuity placental mammals, we used a previously-generated data set of inactivated genes (33). Briefly, gene losses were detected with a pipeline that searches for gene-inactivating mutations and performs a number of filtering steps to distinguish between real mutations and artifacts related to genome assembly or alignment issues and exon-intron structure changes. A genome alignment with human (hg38 assembly) as the reference (78), human genes annotated by Ensembl (version 87) (79) and principal isoforms from the APPRIS database (80) were used as input. Based on the relative positions of inactivating mutations, a value measuring the maximum percent of the reading frame that remains intact (%intact) was computed for each gene and species (Figure 1 - source data 8). A gene was classified as lost if %intact was <60%. A gene was classified as intact if %intact was ≥ 90%.

Using the human Ensembl version 87 gene annotation, we excluded protein-coding genes that have missing data due to assembly incompleteness for more than 50% of low- or high-acuity species and genes that had identical %intact values across species, since such genes cannot show associations with lower visual acuity. This resulted in 13,172 genes. Figure 1 - figure supplement 1 provides a flowchart.

To search for genes that tend to have lower %intact values in the low-acuity group, we applied the Forward Genomics approach (16, 33) with a phenotype definition based on visual acuity to these 13,172 genes. We used phylogenetic generalized least squares (81) to
account for phylogenetic relatedness and ranked genes by the Benjamini-Hochberg corrected p-value (FDR<0.05). The list of p-values for all genes is provided in Figure 1 - source data 2. The resulting set of 68 genes is enriched in the GO-terms "visual perception (GO:0007601)" and "sensory perception of light stimulus (GO:0050953)" (Figure 1 - source data 9A). Finally, to obtain candidates that are convergently lost, we required that a gene was lost in at least three of the seven independent low-acuity lineages. This filtering step removed 37 genes and these genes are not enriched in any GO-term (Figure 1 - source data 9B). Of the remaining 31 genes, two genes are no longer contained in the latest Ensembl annotation (version 106) and were therefore dropped. The other 29 genes are presented in Figure 1 and considered in the enrichment analysis (below).

To test whether the identification of SERPINE3, which we experimentally investigated, is robust to the selected visual acuity threshold, we re-ran the screen after increasing or decreasing the threshold while keeping other parameters constant. Using a more inclusive definition of low-acuity species by increasing the visual acuity threshold to 1.5, which additionally considered manatee, the two flying foxes and rats as low-acuity species, identified a total of seven genes at an FDR of 0.05, with SERPINE3 at the first rank (Figure 1 - source data 5). Using a more restrictive definition of low-acuity species by decreasing the visual acuity threshold to 0.5 identified 31 genes at an FDR of 0.05 with SERPINE3 at rank 8 (Figure 1 - source data 6).

In addition to using phenotype definitions based on visual acuity values, we also tested a phenotype definition based on a molecular signature of losses of known vision-related genes. To this end, we ranked all 49 species based on their number of lost genes annotated with the GO-term “visual perception” (total of 93 genes, Figure 1 - source data 1). The set of ten species that have lost five or more visual perception genes is similar to the set of low-acuity species used in the original screen. The differences are that mouse and deer mouse are substituted by manatee and goat. Re-running the screen based on this molecular loss pattern of vision genes, retrieved SERPINE3 at rank 5 (Figure 1 - source data 7).

As a control to ensure that a Forward Genomics screen does not always retrieve vision-related genes, we ran a new screen, searching for genes preferentially lost in high-acuity sister species (elephant, rhinoceros, horse, two flying foxes, guinea pig, degu, squirrel) of the low-acuity mammals that we used in the original screen. All other species including the other high-acuity mammals were then treated as background (Figure 1 - source data 4).

**Enrichment analysis**

Enrichment analysis was performed using the Enrichr web service (82), which uses a two-sided Fisher’s exact test and corrects for multiple testing with the Benjamini-Hochberg method.

Investigating and validating SERPINE3 losses in additional genomes
To explore conservation and loss of SERPINE3 in additional mammalian genomes that became available since our initial screen, we used the TOGA method (Tool to infer Orthologs from Genome Alignments) (36). TOGA uses pairwise alignments between a reference (here human hg38) and a query genome, infers orthologous loci of a gene with a machine learning approach, and uses CESAR 2.0 (83) to align the exons of the reference gene to the orthologous locus in the query. TOGA then classifies each transcript by determining whether the central 80% of the transcript’s coding sequence encodes an intact reading frame (classified as intact) or exhibits at least one gene-inactivating mutation (classified as potentially lost). If less than 50% of the coding sequence is present in the assembly, the transcript is classified as missing.

Focusing on the evolutionarily conserved human SERPINE3 Ensembl (version 104) transcript ENST00000524365 (84), we analyzed the TOGA transcript classification for 418 assemblies that were not used in the initial screen (Figure 2 - source data 1). These assemblies represent 381 new placental mammal species. For species, where TOGA classified the SERPINE3 transcript as missing, we inspected the orthologous alignment chain to distinguish intact and lost orthologs from truly missing orthologs due to assembly gaps. Since assembly base errors can mimic false gene losses (32, 85), we validated inactivating mutations via two approaches. We considered a mutation as real if it is either shared with an independent assembly of the same or a closely-related species, or if alignments of raw DNA sequencing reads support the mutation. Support by raw DNA reads was assessed by aligning the genomic sequence around the mutation against the NCBI short read archive (SRA queried via NCBI megablast) (86) (Figure 2 - source data 2). With the exception of okapi, steenbok and fox, where SERPINE3 does not evolve under relaxed selection (Figure 2 - source data 4), we considered a gene loss as real if there is at least one validated inactivating mutation. We did not validate SERPINE3 mutations in the genomes of cyclops roundleaf bat and Stoliczka's trident bat, as both species have each eight mutations across several exons and their PacBio HiFi read-based assemblies have very high base accuracies (QV>60, indicating less than one base error per 1 Mb). In order to map SERPINE3 loss events on the phylogenetic tree, we searched for gene-inactivating mutations that are shared among phylogenetically related species, where parsimony indicates that these mutations and thus gene loss likely occurred in their common ancestor. Those mutations are shown in boxes in Figure 2 - figure supplements 2-8, all other mutations are the output of CESAR.

To explore if SERPINE3 remnants are still expressed, we mined expression for eleven SERPINE3-loss species by blasting the TOGA transcript annotation against publicly-available RNA-Seq data of eye or brain tissue. We counted the number of reads that have at least 95% identity (Figure 2 - source data 3).

Mammalian SERPINE3 have features of inhibitory, secreted SERPINs
Mammalian SERPINE3 carries an N-terminal signal peptide (Figure 2 – figure supplement 9) that is predicted to lead to their secretion into the extracellular space in 96% of all analyzed intact and complete SERPINE3. SignalP 5.0 did not predict the presence of a signal peptide for three species (probability<50%): fat dormouse, black flying fox and puma.

Sequence analysis of intact and complete SERPINE3s revealed that two key features of inhibitory serpins are well conserved among placental mammals (sequence alignment is provided as Supplementary File 1). First, the substrate determining residues P4-P4’ are conserved (positions 366-372 in human SERPINE3), whereby P1 denotes the substrate binding scissile bond, position 369 (Figure 2 – figure supplement 10) (87). This position is occupied by an arginine in SERPINE3 as is the case in the inhibitory SERPINE1 and SERPINE2 proteins. Second, the close-by hinge region (positions 355-361) is mostly occupied by small amino acids without prolines. This may allow the insertion of the hinge region into the A beta sheet, a key feature of serpins’ inhibitory mechanism (88).

Furthermore, AlphaFold2 (89) predicted a three-dimensional structure of the human SERPINE3 (Supplementary File 2) that is very similar to other native serpins (mean RMSD to native structures 1.6 A, Supplementary File 3) and adopts the native fold of serpins with an exposed, disordered reactive core loop for substrate binding that does not seem to adopt an alpha-helical conformation as in the non-inhibitory ovalbumin (90). Taken together, this suggests that SERPINE3 functions as a secreted serine protease inhibitor.

**Selection analysis**

For species with an unclear SERPINE3 loss status (mole vole, steenbok, okapi, fox, Steller’s sea cow) and species or clades that have an intact SERPINE3 but many close relatives have lost the gene (greater sac-winged bat, European hedgehog, elephants), we tested whether SERPINE3 evolves under relaxed selection. To this end, we used RELAX from the HyPhy suite (91) to test whether selection pressure was relaxed (selection intensity parameter K<1) or intensified (K>1) in this species or clade, which we labeled as foreground (Figure 2 - source data 4). We restricted the analysis to 327 SERPINE3 that are intact and complete (middle 80% of the coding sequence present) and treated those as background. Codon sequences were obtained from TOGA and aligned with MACSE v2.0 (92). This procedure was repeated using only one foreground species/clade at the time. The species tree used for the analysis is provided as Figure 2 - source data 4.

**Protein sequence analysis and structure prediction**

Signal peptides and the cellular location were predicted with the SignalP 5.0 webserver (93) and DeepLoc 1 (94), respectively, for all intact and complete SERPINE3 protein sequences as defined above. Protein sequences were aligned with muscle (95) and visualized with Jalview (96) (Supplementary File 1).
The three-dimensional structure of human SERPINE3 was retrieved from the AlphaFold2 web server (89) (Supplementary File 2). We calculated the root mean square distance (RMSD) to all homologous chains of existing crystal structures of close serpin relatives in native state (Supplementary File 3) after structural alignment in PyMOL (97). For each crystal structure, we averaged the RMSD for all chains.

**Mining gene function, expression and genetic variation sources**

Information on the function of genes discovered in our screen was obtained from GeneCards database (98), UniProt (99), Ensembl (84), Proteomics DB (100), the Human protein atlas (101), the Expression atlas and Single cell expression atlas (102). Expression in human eye for each candidate gene were obtained by averaging expression over all healthy, primary RNA-Seq data sets per tissue (cornea, RPE, retina) provided by the eyIntegration database (103). Cell lines were not included in the average. A gene was considered to be expressed if the Transcripts Per Million (TPM) value was >100 in cornea, RPE, or retina. Expression of SERPINE3 was further assessed by retrieving primary data sets from FantomCat (104), GEO profiles (105), and Bgee (106). Figure 3 - source data 5 provides the list of all data sets.

Phenotype associations of SNPs located in loci of interest were investigated based on the GWAS catalog (107), dbSNP (108) and PheGenI (109). Linkage of SNPs with a candidate gene in 30 human populations was investigated based on the GWAS catalog. To evaluate possible functional consequences of the respective SNPs, we overlapped their (projected) coordinates with regulatory elements from ENCODE for human (hg38) and mouse (mm10) via the web-based server SCREEN v. 2020-10 (110). Additionally, we investigated eye- and retina-associated regulatory elements in the Ensembl and UCSC genome browsers (84, 111).

**Animal husbandry**

Adult zebrafish (Danio rerio, AB line) were maintained at 26.5°C with a 10/14 h dark/light cycle (112). Embryos and larvae were raised at 28.5°C in the dark until six days old. For phenotyping, we used adult fish of both serpine3 KO lines (serpine3<sup>cbg17</sup>:19 months, serpine3<sup>cbg18</sup>, 11 months) generated in this study as well as their WT siblings. WT mice (Mus musculus, C57BL/6JolaHsd line) were maintained in a barrier system at 20-24°C with a 12/12 h dark/light cycle.

**Expression analysis by RT-qPCR**

Adult zebrafish >12 months were sacrificed by rapid cooling after anesthesia with MESAB. Adult mice (2 months, male) were sacrificed by cranial dislocation after carbon dioxide anesthesia. Tissues were dissected in ice-cold phosphate buffer (PBS), frozen in liquid nitrogen and stored at -80°C until further use. RNA was extracted from lysed, homogenized tissue with RNeasy mini or midi kits (Qiagen) according to manufacturer's instructions and reprecipitated if necessary. The RNA Integrity Number (RIN) was >7 for all tissues except...
spleen (RIN>6). Intact total RNA was reverse transcribed into cDNA using the ProtoScript® II First Strand cDNA Synthesis Kit (NEB) according to manufacturer’s instructions with random primers. RT-qPCR was performed after addition of SybrGreen (Roche). Expression relative to a normalization gene was calculated from Ct values according to the efficiency and delta delta Ct method. Specifically, relative ratios of Serpine3/serpine3 expression (Zf_serp_F: GAGACCCAAAACCTGCCCTT, Zf_serp_R: AGCCGAAATGACCAGTATTGA, Mm_serp_F: TGGAGCTTTTCAGAGGAGGT, Mm_serp_R: GATACTGAAGACAACCCCTGTGC) were obtained by using rpl13a (Zf_rpl13a_F: TCTGGAGAACTGTAAGAGGTATGC, Zf_rpl13a_R: AGACGCACAATCTTTGAGAGCGA) or actb (Zf_actb_F: CGAGCAGGAGATGGGAACC, Zf_actb_R: CAACCGAAACGCTCATTGC) as the reference gene for zebrafish and Rpl27 as the reference gene for mouse (Mm_Rpl27_F: TTGAGGAGCGATACAAGACAGG, Mm_Rpl27_R: CCCAGTCTCTCCACACAA). At least three biological replicates per sample group were analyzed, which were each represented by the average normalized relative ratio of three to six technical replicates. For zebrafish eye with actb as the normalizer, one biological replicate was an expression outlier for eye and liver, defined as exceeding the third quartile by more than three times the interquartile distance (Q3 + 3*interquartile distance). These two data points were excluded from Figure 3A but are contained in Figure 3 - source data 1.

ISH and FISH

For ISH and immunostainings, fish were sacrificed, eyes dissected and fixed in 4% paraformaldehyde/ 0.1 M PBS after removal of the lens. Eyes were embedded in gelatin/sucrose and sectioned (14 µm) with a cryostat. The serpine3 in situ probe spans the coding exons 3-8 (transcript: ENSDART00000132915.2). Using primers with restriction enzyme cut sites (forward, Not1: TAAGCGACCGCATGCCAG, reverse, BamHI: TAAGCGACCGCATGCCAG, ACAACTGCACCTAAAAACAGCAAC), serpine3 cDNA was amplified from total cDNA and cloned into the pCRII-topo vector (Invitrogen). The antisense probes were transcribed with SP6 polymerase, using a DIG-labeled NTP mix (Roche diagnostics). The (fluorescent) ISH were conducted as previously described with minor modifications (113): Hybridization and washing steps were performed at 60°C. For chromogenic in situ, sections were incubated with anti-digoxigenin-AP (Roche), diluted 1:4000 in DIG-blocking reagent (Roche) at 4°C overnight and subsequently developed with NBT/BCIP (Roche). For FISH, sections were washed in PBS immediately after quenching. Sections were blocked for 1 h with 2% blocking reagent in MABT (Perkin-Elmer) and then incubated with anti-digoxigenin-POD (Roche), diluted 1:500. The signal was detected with the TSA Plus Cy3/Cy5 kit (Perkin-Elmer).

Immunohistochemistry

Immunostainings for glial fibrillary acidic protein (ZRF1 from DSHB, 1:200), choline O-acetyltransferase (AB144P from Millipore, 1:500) and protein kinase C alpha (SC-208 from Santa-Cruz, 1:500) were performed after completion of the FISH protocol according to (113).
For chat, antigens were retrieved in preheated 10 mM sodium citrate buffer for 6’ at 85°C. Sections were washed in PBS and 0.3% PBSTx prior to primary antibody incubation. Following the protocol of (113), we washed the sections three times in PBSTx, and incubated in anti-goat or anti-rabbit IgG (H+L) Alexa 488-conjugated secondary antibodies (Invitrogen, 1:750).

**Generation of KO lines and genotyping**

To generate *serpine3* KO lines, deletions were introduced using the CRISPR-Cas9 system. Guides were chosen considering efficiency predictions of the IDT DNA CRISPR-Cas9 guide checker and ChopChopV2 (114) on the zebrafish assembly danRer7 and ordered as Alt-R CRISPR-Cas crRNA from IDT DNA. For each line, three guides were simultaneously injected into one-cell stage zebrafish embryos as ribonucleoprotein delivery using the Alt-R CRISPR-Cas system following the manufacturer’s instructions (0.5 fmol crRNA per embryo per guide, 0.68 ng Cas9 protein per embryo). The expected deletions were confirmed by Sanger sequencing in several founder individuals, one of which was chosen as the founder of each line (Figure 4 - figure supplement 1). Heterozygous cbg17 and cbg18 zebrafish were further outcrossed to WT fish for several generations and then bred to homozygosity.

More specifically, for *serpine3* cbg17, we abolished *serpine3* transcription by deleting the single transcription start site, which is supported by activating histone marks in zebrafish and is also well conserved in human and mouse (Figure 4 - figure supplement 1A, using the following guides: GGTATTTGTACTCTAATGAA (guide1), TGTACTCTAATGAAGGAAC (guide2), CTCACACAGGACAATCCGGCAGG (guide3). For genotyping, we used primers (For1: 5-GAAATCGCATGTCACGCAGAAAT-3, Rev2: 5-ATATCGGAACTGACATACTGAACG-3, Rev2.2: 5- GTGAGCTTCGTGTTTGTGGT-3) to amplify a region around the transcription start site (Figure 4 - figure supplement 2A, B, Figure 4, source data 5-6).

*Serpine3* cbg18 was generated by introducing a frame-shifting deletion in coding exon 1, which presumably results in three early stop codons when the transcript is translated. The following guides were used: TCTTCTGCAACTCGGGGCCA (guide4), TCTCTGTGAGCTTTGGTAG (guide5), AACACTCTGGTTCAGCTCGA (guide6) (Figure 4 - figure supplement 1B). We genotyped fish by amplifying a region around coding exon1 with the following primers: For3: 5-GGCATTGTTGAGATTCAGTAGTCA-3, Rev4: 5-CAGTTTACTCCTACCATTGACATC-3 (Figure 4 - figure supplement 2C, Figure 4, source data 7).

**Histology**

For hematoxylin/eosin stainings, fish were sacrificed and heads were fixed overnight at 4°C in 4% paraformaldehyde/ 0.1 M PBS and decalcified in 0.5 M EDTA in 0.1 M PBS for 3-4 days. Next, they were processed in a Paraffin-Infiltration-Processor (STP 420, Zeiss) according to the following program: ddH2O: 1×1′; 50% ethanol (EtOH) 1×5′; 70% EtOH 1×10′; 96% EtOH 1×25′; 96% EtOH 2×20′; 100% EtOH 2×20′; xylene 2×20′; paraffin 3×40%/60°C;
paraffin 1×60′/60°C. The heads were embedded in paraffin using the Embedding Center EG1160 (Leica). Semi-thin sections (2 µm) were cut on an Ultracut microtome (Mikrom) and counterstained using hematoxylin/eosin (HE, Sigma).

Microscopy, image processing and analysis

Imaging was performed using the ZEISS Axio Imager.Z1 provided by the CMCB Light Microscopy Facility. The images were processed in Fiji/ImageJ version 2.1.0 (115) (macroscopic eye images) and Adobe Illustrator. For macroscopic phenotyping, eyes were imaged with a Leica stereo microscope M165C. To parameterize the eye shape, the eye outline was first approximated by an oval and then manually corrected if necessary. Particle parameters of the final eye object were measured automatically in Fiji. The statistical analysis and visualization were conducted in R version 4.1.0 (2021-05-18) using the packages ggplot2 (116), tseries (117) and effsize. Outliers were not excluded. Comparing WT and KO individuals, we tested whether both genotypes have the same iris shape (estimated by iris solidity and circularity) using the Wilcoxon rank sum test after rejecting normality of the variables with a Jarque Bera Test. Both eyes of the same individual were treated as individual biological replicates, since we observed shape deviations often only in one eye of the same individual (Figure 4 - source data 2).

Animal licenses

All experiments in mouse and zebrafish were performed in accordance with the German animal welfare legislation. Protocols were approved by the Institutional Animal Welfare Officer (Tierschutzbeauftragter), and licensed by the regional Ethical Commission for Animal Experimentation (Landesdirektion Sachsen, Germany; license no. DD24-5131/354/11, DD24.1-5131/451/8, DD24-5131/346/11, DD24-5131/346/12).

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and the Supplementary Materials. The raw microscopy images of fish eyes (Figure 4 – source data 3) are available at https://genome.senckenberg.de/download/SERPINE3/. TOGA annotations of SERPINE3 and a visualization of gene-inactivating mutations are available at our UCSC genome browser mirror https://genome.senckenberg.de.

Competing interests

The authors have no competing interests.

Acknowledgment

We thank the genomics community for sequencing and assembling the genomes and the UCSC genome browser group for providing software and genome annotations. Experimental work would not have been possible (or as pleasant) without supporting hosting labs, especially Nadine Vastenhouw, Elisabeth Knust and Wieland Huttner. We also thank Nadine
Vastenhouw and her whole lab, Michael Heide and Mauricio Rocha as well as current and former members of the Hiller lab for helpful scientific discussion and comments on the manuscript. We thank the following facilities of MPI-CBG: Biomedical Services (especially fish unit), Cell technologies (Julia Jarrells), Sequencing and genotyping (Sylke Winkler), Light Microscopy, Scientific Computing, Computer Service Facilities as well as the Computer Service Facilities of MPI-PKS and the CMCB Histology (Susanne Weiche) and CMCB Light Microscopy Facility for their support. Work of HI and MH was supported by an exploration grant from the Boehringer Ingelheim Stiftung, the Max Planck Society and the LOEWE Centre for Translational Biodiversity Genomics (TBG) funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK). Work by JH, AM, SH and MB was supported by project grants of the German Research Foundation (Deutsche Forschungsgemeinschaft, project numbers BR 1746/3 and BR 1746/6) and an ERC advanced grant (Zf-BrainReg) to MB. This study was furthermore supported with a PhD scholarship from Studienstiftung des deutschen Volkes to JH.

References


36. B. M. Kirilenko *et al.*, TOGA integrates gene annotation with orthology inference at scale.


60. A. Bringmann et al., Cellular signaling and factors involved in Muller cell gliosis: neuroprotective and detrimental effects. Prog Retin Eye Res 28, 423-451 (2009).


**Figures**

**Figure 1:** A comparative genomics screen uncovered known and novel vision-related genes.

(A) Phylogeny of the species included in our screen (left). Visual acuity values on a log$_{10}$ scale are shown on the right as a bar chart; white boxes indicate the three subterranean mammals that lack acuity measurements but are functionally blind. Low-acuity species (red font) are defined here as species with visual acuity $<$1 ($\log_{10}(va)<0$). At a false discovery rate (FDR) threshold of 0.05, our screen retrieved 29 genes, which are preferentially lost in low-acuity species. Gene losses in individual species are denoted by red dots. The FDR value for the gene loss–phenotype association is shown at the top as a bar chart.

(B) List of 29 genes, together with known vision-related functions. Asterisk marks genes that have no known vision-related function but were mentioned in large-scale gene expression data sets of ocular tissues. Genes in bold are expressed in human eyes according to the eyelIntegration database (103) (see Methods). KO - knockout.

(C) Functional enrichments of the 29 genes reveals vision-related functions (Gene Ontology, GO of biological processes), associations with human eye disorders (OMIM), expression in ocular tissues in the mouse gene atlas (82) and eye phenotypes in mouse gene KOs (MGI Mammalian Phenotype level 4) among the top five most significant terms. Vision-related terms are shown as blue bars. The dashed line indicates statistical significance in a one-sided Fisher’s exact test after correcting for multiple testing using the Benjamini-Hochberg procedure (FDR 0.05).

**Figure 2:** SERPINE3 gene loss pattern across 430 mammalian species.

(A) Left: Phylogeny of mammalian species investigated for the loss of SERPINE3 with mapped gene loss events indicated as red crosses. Branches of major clades are colored (Rodentia – brown, Chiroptera – orange, Eulipotyphla – blue, Afroinsectiphila - red). The number of species investigated per clade is specified in parenthesis. For all loss lineages (red font), visual capability (classified as echolocating, fossorial/subterranean, vision as non-primary and primary sense) is displayed as pictograms at the right. Asterisk marks indicate species, where SERPINE3 evolved under relaxed selection but did not accumulate not more than one inactivating mutation. Right: The Serpin protein domain (Pfam) spans all eight protein-coding exons (boxes) of the intact human SERPINE3 gene (top). The secretion signal and the reactive core region are conserved in species with an intact SERPINE3 (Figure 2 - figure supplement 9-10). Gene-inactivating mutations are illustrated for three clades, with stop codon mutations shown in black, frame-shifting insertions and deletions shown in red and mutated splice site dinucleotides shown between exons in red. Deleted exons are shown as red boxes. Insets show codon alignments with inactivating mutations in red font. RB - roundleaf bat, HB - horseshoe bat.

(B) An ancestral deletion removed large parts of SERPINE3 in the tenrec lineage. UCSC genome browser (111) view of the human hg38 assembly showing the SERPINE3 locus and whole genome alignments between human and two tenrec species, visualized as a chain of co-linear local alignments. In these chains, blocks represent aligning sequence and double
lines represent sequence between the aligning blocks that do not align between human and tenrec. A large deletion removed the first five protein-coding exons of SERPINE3 in both species. Shared breakpoints (gray box) indicate that the deletion likely represents an ancestral event in Tenrecidae.

**Figure 3:** Serpine3 is expressed in zebrafish and mouse eyes.

(A,B) Expression of zebrafish serpine3 mRNA in relation to the normalization genes actb (A) and rpl13a (B) measured with RT-qPCR. Expression levels are consistently the highest in eye. A two-sided unequal variances t-test was used. Boxplots display first quartile, median and third quartile with whiskers extending to the maximum and minimum value. Measurements of technical replicates (n=3-6) are encoded in the same color, different colors represent different biological replicates (n=3-8). Relative rates are displayed as zero, if expression was below the detection limit.

(C) Serpine3 mRNA expression in different tissues of the zebrafish eye in relation to the reference gene actb measured with RT-qPCR for three biological replicates. Serpine3 is specifically expressed in the retina but not in other tissues of the eye. The expression level is significantly higher in retina without RPE compared to RPE only (two-sided unequal variances t-test). Although expression in RPE only is low, it is significantly higher than expression in eye tissue after removing retina and RPE (two-sided unequal variances t-test).

(D) Serpine3 mRNA expression in mouse in relation to the reference gene Rpl27 measured with RT-qPCR for three biological replicates. Serpine3 is specifically expressed in the eye but not in colon, cortex, heart, liver, spleen and testis.

(E-H) Serpine3 mRNA expression pattern in zebrafish retina. (E-F) Chromogenic *in situ* hybridization (ISH) shows localized expression of serpine3 (purple) in the retina, specifically in the inner nuclear layer (inlet). (G-H) Fluorescence *in situ* hybridization shows that serpine3 mRNA expression (yellow) is localized to cell bodies of Müller glia cells. Filaments of Müller glia cells are marked by the glial fibrillary acidic protein (ZRF1 antibody, green). Specificity of the serpine3 ISH probe is shown by absence of the signal in homozygous *serpine3*<sup>cbg17</sup> knockout fish (Figure 3 - figure supplement 1). Serpine3 mRNA is not expressed in bipolar or amacrine cells (Figure 3 - figure supplement 2). Nuclei are stained with DAPI (white). Scale bar is 200 µm in (E) and (G) and 20 µm in (F) and (H). INL – inner nuclear layer, ONL – outer nuclear layer, GCL – ganglion cell layer.

**Figure 4:** Serpine3 knockout in zebrafish causes defects in eye shape and retinal layering.

(A) UCSC genome browser visualization of the SERPINE3 genomic locus in human (hg38 assembly, top box) and zebrafish (danRer11 assembly, bottom box) shows that both species have a 1:1 ortholog with the same number of coding exons in a conserved gene order context. In the human locus, two single nucleotide polymorphisms (SNPs) are in linkage with SERPINE3 and associated with eye phenotypes. In zebrafish, we used CRISPR-Cas9 to generate two independent knockout (KO) lines. The position of guide RNAs is indicated as scissors. In the *serpine3*<sup>cbg17</sup> line, we deleted the promoter and first exon. In the *serpine3*<sup>cbg18</sup>
line, we introduced a 92 bp frame-shifting deletion in exon 2 (coding exon 1) that results in three early stop codons in the original reading frame.

(B) Relative expression of serpine3 mRNA in wild type (WT) zebrafish and serpine3\(^{cbg17}\) individuals quantified by RT-qPCR relative to the expression of rpl13a. Serpine3 mRNA expression is close to zero in serpine3\(^{cbg17}\) fish and significantly reduced in comparison to wild type fish (P = 0.049, two-sided unequal variances t-test). Technical replicates of the qPCR are shown as individual data points; different colors represent different biological replicates. Boxplots display first quartile, median and third quartile with whiskers extending to the maximum and minimum of the three biological replicates.

(C) In situ hybridization showing that serpine3 is expressed in the inner nuclear layer (INL) of WT zebrafish but not in the homozygous serpine3\(^{cbg17}\). Scale bar = 25 µm.

(D) Serpine3 knockout leads to changes in eye shape in adult, homozygous knockout (KO) fish of serpine3\(^{cbg17}\) and serpine3\(^{cbg18}\) lines in comparison to their WT siblings (18 and 11 months, respectively). In WT, the eye shape almost perfectly corresponds to the concave shape of the iris (overlay of white and red dotted lines). In contrast, many KO individuals have alterations in eye shape, evident by notches (arrow heads) in the white line that follows the iris. Scale bar = 1 mm.

(E) Iris solidity (ratio of eye shape/ concave eye shape) significantly differs between WT and KO siblings for both the serpine3\(^{cbg17}\) (16 vs. 10 eyes) and the serpine3\(^{cbg18}\) (40 vs. 40 eyes) line. A Wilcoxon Rank sum test was used. Boxplots display first quartile, median and third quartile with whiskers extending to the maximum and minimum within 1.5 times interquartile range. Outliers are shown in black. Iris circularity, another quantification measure for the phenotype, is shown in Figure 4 - figure supplement 3.

(F) Hematoxylin/eosin histology staining of the eye of serpine3\(^{cbg17}\) fish (22 months) reveals histological differences in comparison to their WT siblings (dorsal top, ventral bottom). In comparison to WT, distance between lens and retina of serpine3\(^{cbg17}\) fish is reduced (distance bars). The WT retina (top) has a distinct lamination with clear separation of the single retinal layers (a, b) as shown in the schematic (RPE – retinal pigment epithelium layer, POS – photoreceptor outer segment, ONL – outer nuclear layer, OPL – outer plexiform layer, INL – inner nuclear layer, IPL – inner plexiform layer, GCL – ganglion cell layer, RNFL – retinal nerve fiber layer). Although all retinal layers are present in serpine3\(^{cbg17}\) fish, the layering appears distorted and the density of cells is reduced (c-f). Specifically, the RPE cells display an altered distribution and even local clusters (empty arrows), and displaced pigmented cells emerge in all retinal layers (yellow arrows). This was confirmed also for the cbg18 allele (Figure 4 - figure supplement 5). Scale bar in the overviews = 200 µm, scale bar in the magnifications = 20 µm.
Legends of supplementary materials:

**Figure 1 – figure supplement 1.** Flow diagram of filtering steps applied in the Forward genomics approach.

The filtering procedure results in a list of 31 candidate genes. Two of the 31 genes are no longer contained in the latest Ensembl annotation version (v106) and were therefore dropped, resulting in 29 genes potentially associated with an eye-related function.

**Figure 2 - figure supplement 1.** The remnants of SERPINE3 are not expressed in the common vampire bat.

(A) UCSC genome browser screenshot of the common vampire bat (*Desmodus rotundus*) genome (118) showing the four remaining exons of SERPINE3 (red: TOGA annotation, the other exons are deleted in this bat) and three coding exons of the neighboring *INTS6* gene.

Eye RNA-seq from *D. rotundus* (22), shown as a wiggle track and read alignments, show that three SERPINE3 exons (blue highlight) are not expressed. The fourth SERPINE3 exon overlaps the 3’ UTR of the neighboring *INTS6* gene, which is located on the other strand.

This explains why reads cover this SERPINE3 exon and its flanks, but there is not a single read that supports splicing of this exon to other SERPINE3 exons. In contrast, *INTS6* exons are robustly expressed and show spliced alignments.

(B) For comparison, the conserved GLUL gene, which is a key enzyme in glutamate detoxification, is highly expressed in eyes of the common vampire bat.

**Figure 2 - figure supplement 2.** Gene-inactivating mutations support four independent losses of SERPINE3 in Afrotheria and Xenarthra (red dots).

Those shared mutations are marked by boxes, which indicate the loss of SERPINE3 in the common ancestor of related species according to parsimony. For tenrecs (*Microgale talazaci*, *Echinops telfairi*) and elephant shrew (*Elephantulus edwardii*), the putative region of exon 4 was likely deleted after the shared frame-shifting mutations. SERPINE3 in sirenia (branches marked in red) evolve under relaxed selection. Coloring and legend as in Figure 2. Gray exons denote missing information.

**Figure 2 - figure supplement 3.** Gene-inactivating mutations support a single loss of SERPINE3 in Pholidota.

A partial deletion of coding exon 6 and three frame-shifting deletions in exon 7 are shared among all four species, indicating a single loss of SERPINE3 in the Pholidota lineage. Coloring and legend as in Figure 2.

**Figure 2 - figure supplement 4.** Gene-inactivating mutations support three independent losses of SERPINE3 in Eulipotyphla (red dots).

The start codon is not intact in the four moles (a). Star-nosed mole (*Condylura cristata*) and Hispaniolan solenodon (*Solenodon paradoxus*) share two inactivating mutations in exon 3, although they are phylogenetically distant (b, c). Coloring and legend as in Figure 2.
Gene-inactivating mutations support one loss of SERPINE3 in Yangochiroptera.

The ancestral start codon is mutated in most Yangochiroptera (box). The most parsimonious explanation is a shared start codon mutation (red dot) in the ancestral lineage after split from the sac-winged bat (*Saccopteryx bilineata*). A back mutation to the regular start codon ATG likely occurred in *Molossus molossus* and *Artibeus jamaicensis*. Please see Figure 2 – figure supplement 6 for gene-inactivating mutations of 15 more Yangochiroptera assemblies. The sac-winged bat has an intact copy of the gene that evolves under relaxed selection (red branch) in addition to a second copy of SERPINE3, which was independently inactivated. Coloring and legend as in Figure 2. Gray exons denote missing information.

Figure 2 – figure supplement 6. Gene-inactivating mutations in additional Yangochiroptera.

This is an extension of Figure 2 – figure supplement 5. The start codon mutation shown is shared with most other Yangochiroptera (Figure 2 – figure supplement 5). Coloring and legend as in Figure 2. Gray exons denote missing information.

Figure 2 – figure supplement 7. Gene-inactivating mutations support six independent losses of SERPINE3 in Yinpterochiroptera.

Six independent inactivation events (red dots) occurred in SERINE3's coding sequence in Yinpterochiroptera. All Rhinolophidae share frame-shifting mutations in exon 4 (boxed), while Hipposideridea share a frame-shifting insertion in exon 1 and a splice site mutation at exon 2 (boxed). Coloring and legend as in Figure 2.

Figure 2 – figure supplement 8. Independent gene-inactivating mutations in other mammals.

A test for shifts in selective pressure revealed that SERPINE3 evolves under relaxed selection in the subterranean mole vole (*Ellobius lutescens*, marked in red), but not in the other cases of gene-inactivating mutations with unclear consequences (steenbok, okapi). A species-specific duplication with inactivation of one SERPINE3 copy occurred in white-footed mouse (*Peromyscus leucopus*). Coloring and legend as in Figure 2.

Figure 2 – figure supplement 9. Conservation of the signal peptide in mammalian SERPINE3.

The putative N-terminal signal peptide contains many hydrophobic residues in intact mammalian SERPINE3 and is predicted to guide secretion into extracellular space. We show the first 30 alignment columns. The sequence logo was generated with WebLogo (119).

Figure 2 – figure supplement 10. The hinge region and reactive core loop are conserved in intact mammalian SERPINE3.
The numbers refer to human SERPINE3, where R369 likely is the scissile bond (P1) within the reactive core loop (positions 366-372, P4-P4’). The hinge region is located at positions 355-361. The logo was generated with WebLogo (119).

**Figure 3 - figure supplement 1.** Serpine3 expression co-localizes with glial fibrillary acidic protein in the retina of wild type fish.

The fluorescence *in situ* signal for zebrafish serpine3 is specific for wild type (WT, yellow), where it is in proximity to staining of the ZRF1-antibody (green). This indicates expression of serpine3 by Mueller glia cells. Serpine3 signal is not present in homozygous serpine3<sup>cbg17</sup> siblings. In the overlay, nuclei (DAPI) are shown in white. ONL – outer nuclear layer, INL – inner nuclear layer, GCL – ganglion cell layer. Scale bar = 20 µm.

**Figure 3 - figure supplement 2.** Serpine3 expression does not co-localize with markers for bipolar or amacrine cells in the zebrafish retina.

Fluorescence *in situ* hybridization of serpine3 mRNA (yellow) does not show co-localization with anti-protein kinase C alpha (PKCa encoded by prkca, green) or choline O-acetyltransferase a (CHAT, encoded by chata, green) proteins. In the overlay, nuclei (DAPI) are shown in white. ONL – outer nuclear layer, INL – inner nuclear layer, GCL – ganglion cell layer. Scale bar overview = 20 µm, details = 20 µm.

**Figure 4 - figure supplement 1.** Knockout of serpine3<sup>cbg17</sup> and serpine3<sup>cbg18</sup> alleles in zebrafish by CRISPR-Cas9 is confirmed by sequencing results.

Sequenced reads are mapped against the danRer11 zebrafish genome assembly with Blat and are visualized in the UCSC genome browser (black) together with the serpine3 RefSeq annotation (blue).

(A) PCRs with primers For1 and Rev2 or For1 and Rev2.2 (green) amplify the expected regions around the transcription start site on chromosome 9 in wild type (WT_Rev2, WT_Rev2.2). Injection of CRISPR guides 1, 2 and 3 (orange) results in a deletion of about 400 bp for founder individuals 7 and 12 in serpine3<sup>cbg17</sup>. Offspring of founder 12 (394 bp deletion) was raised and further crossed. The location of a single serpine3 transcription start site is supported by annotation and activating histone marks H3K4me3 within the respective region (lower gray bar with darkness of color correlating with signal intensity).

(B) PCR with primers For3 and Rev4 (green) amplifies a region around coding exon1 of serpine3. Injection of CRISPR guides 4, 5 and 6 (orange) results in a deletion of about 100 bp for founders 1, 9, 18 and 19 in serpine3<sup>cbg18</sup>. For founder 9, this 92 bp deletion induces a frame-shift in the reading frame (bold) with three early stop codons (two shown as *, third stop located in coding exon 2) and offspring was grown. The deletion is equivalent to deletion of 31 amino acids (red) and +1 nt insertion. The amino acid encoded by a split codon is shown in blue.
Figure 4 - figure supplement 2. PCRs confirm expected CRISPR-Cas9 induced deletions in serpine3<sup cbg17</sup> and serpine3<sup cbg18</sup> fish. (A) Genotyping of the serpine3<sup cbg17</sup> with two primers (F0 generation). Injection of CRISPR guides into the one-cell embryo leads to mosaic embryos (several pooled at 72 hpf) with a strong wild type (WT) band at about 557 bp and a weak mutant band at about 161 bp, while uninjected embryos just have a single WT band (557 bp). The raw image is provided as Figure 4 – source data 5. (B) Genotyping of serpine3<sup cbg17</sup> with a mix of three primers (F2 generation). Heterozygous animals (H) have two bands, a WT (expected height: 286 bp) and mutant band (expected height: 161 bp), while homozygous mutants (M) and homozygous WT animals show a single band. NC – negative control (water). The raw image is provided as Figure 4 – source data 6. (C) Genotyping of serpine3<sup cbg18</sup> with two primers. Heterozygous animals (H) have two bands, a WT (expected height: 342 bp) and mutant band (expected height: 250 bp), while homozygous mutants (M) and homozygous WT animals show a single band. NC – negative control (water). The raw image is provided as Figure 4 – source data 7.

Figure 4 - figure supplement 3. Iris circularity differs between serpine3<sup cbg17</sup> and serpine3<sup cbg18</sup> fish and their respective wild type siblings (WT). Iris circularity is another descriptor for the deviation of eye shape and is defined as $4\pi \times \frac{\text{[Area]}}{\text{[Perimeter]}^2}$. A value of 1 indicates a perfect circle. Boxplots show that iris circularity significantly differs between WT (n=14) and serpine3<sup cbg17</sup> eyes (n=10). The same holds for the comparison of WT (n=40) and serpine3<sup cbg18</sup> eyes (n=40). A Wilcoxon Rank sum test was used. Boxplots display first quartile, median and third quartile with whiskers extending to the maximum and minimum within 1.5 times inter-quartile range. Outliers are shown in black.

Figure 4 - figure supplement 4. The optic nerve is intact in serpine3<sup cbg17</sup> and serpine3<sup cbg18</sup> fish as shown with a hematoxylin/eosin histology staining. No obvious difference in optic nerve morphology is observed between serpine3<sup cbg17</sup> (A) and serpine3<sup cbg18</sup> fish (B) and their respective wild type (WT) siblings. Scale bar = 200 µm.

Figure 4 - figure supplement 5. Hematoxylin/eosin histology staining of serpine3<sup cbg18</sup> eyes (14 months) reveals histological differences in comparison to their wild type (WT) siblings (dorsal top, ventral bottom). We show details of representative overview images for one eye of each genotype on the right. In comparison to WT (A), distance between lens and retina of serpine3<sup cbg18</sup> fish is reduced (distance bars). In the serpine3<sup cbg18</sup> eye (B), all retinal layers are present and distinguishable, although they are not as tightly packed and clearly separated as in the WT (c-f). Moreover, we observed displaced pigmented cells that are located mainly in the photoreceptor outer segment and the outer nuclear layer (yellow arrows, c-e). Furthermore,
the photoreceptor outer segment and RPE layer are not clearly separated in the 
serpine3<sup>cri18</sup> retina. Scale bar in the overviews represents 200 µm and in the magnifications 
20 µm.

Figure 1 – source data 1: Genome assemblies used in this study and visual acuity values of 
mammalian species.

Figure 1 – source data 2: List of candidate genes of the forward genomics screen based on 
visual acuity < 1 including a short description.

Figure 1 – source data 3: Enrichment of candidate genes in different processes/functions.

Figure 1 – source data 4: A control forward genomics screen does not retrieve vision-related 
genes.

Figure 1 – source data 5: List of candidate genes of the forward genomics screen based on 
a visual acuity threshold < 1.5.

Figure 1 – source data 6: List of candidate genes of the forward genomics screen based on 
visual acuity threshold < 0.5.

Figure 1 – source data 7: List of candidate genes of the forward genomics screen based on 
the molecular loss signature of known vision genes.

Figure 1 – source data 8: Reading frame %intact values for all investigated genes and 
species.

Figure 1 – source data 9: Enrichment of candidate genes in different stages of the filtering 
process.

Figure 2 – source data 1: SERPINE3 intactness in 447 mammalian assemblies.

Figure 2 – source data 2: Verification of inactivating mutations in SERPINE3 with raw DNA 
sequencing reads.

Figure 2 – source data 3: Expression of SERPINE3 remnants in public RNA-seq datasets.

Figure 2 – source data 4: Analysis of relaxation or intensification of selection in SERPINE3 
branches of interest.

Figure 3 – source data 1: qPCR analysis of a zebrafish organ series using the normalizer 
gene actb.
Figure 3 – source data 2: qPCR analysis of a zebrafish organ series using the normalizer gene *rpl13a*.

Figure 3 – source data 3: qPCR analysis of zebrafish eye subtissues.

Figure 3 – source data 4: qPCR analysis of mouse organs.

Figure 3 – source data 5: Data sets supporting expression of *SERPINE3* in tissues of the eye in vertebrates other than mouse and zebrafish.

Figure 4 – source data 1: qPCR analysis of zebrafish *serpine3*^cbg17^ homozygotes (KO) vs. WT.

Figure 4 – source data 2: Macroscopic eye phenotype in *serpine3* knockout lines.

Figure 4 – source data 3: Raw microscopy images of fish eyes.

Figure 4 – source data 4: Single nucleotide polymorphism near *SERPINE3* (13: 51,341,032-51,362,101) are associated with human eye phenotypes.

Supplementary File 1: Annotated protein alignment of intact and complete mammalian SERPINE3. .jvp format is to be opened in Jalview.

Supplementary File 2: Predicted structure of human SERPINE3 (AlphaFold 2).

Supplementary File 3: Comparison of 3D structures of crystallized SERPINs to the predicted human SERPINE3.
## Key Resources Table

<table>
<thead>
<tr>
<th>Reagent type (species or resource)</th>
<th>Designation</th>
<th>Source or reference</th>
<th>Identifiers</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>gene (Danio rerio)</td>
<td>serpine3</td>
<td>Ensembl</td>
<td>ENSDART000001 32915.2</td>
<td></td>
</tr>
<tr>
<td>gene (Homo sapiens)</td>
<td>SERPINE3</td>
<td>Ensembl</td>
<td>ENST0000052436 5</td>
<td></td>
</tr>
<tr>
<td>strain, strain background (Danio rerio)</td>
<td>WT (AB)</td>
<td>Zebrafish International Resource Center</td>
<td></td>
<td></td>
</tr>
<tr>
<td>strain, strain background (Mus musculus)</td>
<td>WT (C57BL/6JOl aHsd)</td>
<td>Envigio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>genetic reagent (Danio rerio)</td>
<td>cbg17</td>
<td>This paper</td>
<td></td>
<td>See Materials and Methods “Generation of KO lines and genotyping”</td>
</tr>
<tr>
<td>genetic reagent (Danio rerio)</td>
<td>cbg18</td>
<td>This paper</td>
<td></td>
<td>See Materials and Methods “Generation of KO lines and genotyping”</td>
</tr>
<tr>
<td>Antibody</td>
<td>Anti-digoxigenin-AP (polyclonal sheep)</td>
<td>Sigma-Aldrich</td>
<td>Cat. #: 11093274910, RRID: AB_2734716</td>
<td>IF (1:4000)</td>
</tr>
<tr>
<td>Antibody</td>
<td>Anti-digoxigenin-POD conjugated (polyclonal sheep)</td>
<td>Roche</td>
<td>Cat. #: 11207733910, RRID: AB_514500</td>
<td>IF (1:500)</td>
</tr>
<tr>
<td>Antibody</td>
<td>Cat No.</td>
<td>Supplier</td>
<td>RRID</td>
<td>Dilution</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>--------------------------</td>
<td>-----------------------------------------</td>
<td>---------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Anti-gfap ZRF-1 (monoclonal mouse)</td>
<td>Cat. #: zrf-1, RRID: AB_10013806</td>
<td>Zebrafish International Resource Center</td>
<td>IF (1:200)</td>
<td></td>
</tr>
<tr>
<td>Anti-choline acetyltransferase (polyclonal goat)</td>
<td>Millipore Cat. #: AB144P, RRID: AB_2079751</td>
<td>Sigma-Aldrich</td>
<td>IF (1:500)</td>
<td></td>
</tr>
<tr>
<td>Anti-protein kinase C alpha (polyclonal rabbit)</td>
<td>Cat. #: sc-208, RRID: AB_2168668</td>
<td>Santa Cruz Biotechnology</td>
<td>IF (1:500)</td>
<td></td>
</tr>
<tr>
<td>Anti-goat IgG (H+L) Alexa 488 (polyclonal chicken)</td>
<td>Cat. #: A21467</td>
<td>Invitrogen</td>
<td>IF (1:750)</td>
<td></td>
</tr>
<tr>
<td>Anti-rabbit IgG (H+L) Alexa 488 (polyclonal goat)</td>
<td>Cat. #: A11034</td>
<td>Invitrogen</td>
<td>IF (1:750)</td>
<td></td>
</tr>
<tr>
<td>recombinant DNA reagent</td>
<td></td>
<td></td>
<td></td>
<td>See Materials and Methods “Generation of KO lines and genotyping”</td>
</tr>
<tr>
<td>sequence-based reagent</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zf_serp_F</td>
<td>This paper</td>
<td>qPCR primer</td>
<td>GAGACCCAAAACCTGCCCTT</td>
<td></td>
</tr>
<tr>
<td>Zf_serp_R</td>
<td>This paper</td>
<td>qPCR primer</td>
<td>AGCCGGAATGACCGATATGTA</td>
<td></td>
</tr>
<tr>
<td>Mm_serp_F</td>
<td>This paper</td>
<td>qPCR primer</td>
<td>TGGAGCTTCAGAGGAGGGTA</td>
<td></td>
</tr>
<tr>
<td>sequence-based reagent</td>
<td>Forward Primer</td>
<td>Reverse Primer</td>
<td>Reference</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-----------</td>
<td>-------------</td>
</tr>
<tr>
<td>Mm_serp_R</td>
<td>GATACTGAAGACAAACCT</td>
<td>GTGC</td>
<td>This paper</td>
<td>qPCR primer</td>
</tr>
<tr>
<td>Zf_rpl13a_F</td>
<td>TCTGGAGGAACTGTAAGAG</td>
<td>GTATGC</td>
<td>10.1111/j.1745-7270.2007.00283.x</td>
<td>qPCR primer</td>
</tr>
<tr>
<td>Zf_rpl13a_R</td>
<td>AGACGCACAATCTTGGAGAG</td>
<td>CGA</td>
<td>10.1111/j.1745-7270.2007.00283.x</td>
<td>qPCR primer</td>
</tr>
<tr>
<td>Zf_actb_F</td>
<td>CGAGCAGGAGATGGGAACC</td>
<td></td>
<td>10.1007/s00441-020-03318-2</td>
<td>qPCR primer</td>
</tr>
<tr>
<td>Zf_actb_R</td>
<td>CAACGGAACCGCTATTGC</td>
<td></td>
<td>10.1007/s00441-020-03318-2</td>
<td>qPCR primer</td>
</tr>
<tr>
<td>Mm_Rpl27_F</td>
<td>TTGGAGAGCGATACAAGAC</td>
<td>AGG</td>
<td>This paper</td>
<td>qPCR primer</td>
</tr>
<tr>
<td>Mm_Rpl27_R</td>
<td>CCCAGTCTCTTCCACACAA</td>
<td>CAGG</td>
<td>This paper</td>
<td>qPCR primer</td>
</tr>
<tr>
<td>Guide1</td>
<td>GGTATTTGTACTCTAATGAA</td>
<td></td>
<td></td>
<td>Crispr guide</td>
</tr>
<tr>
<td>Guide2</td>
<td>TGTACTCTAATGAAAGGAAC</td>
<td></td>
<td></td>
<td>Crispr guide</td>
</tr>
<tr>
<td>Guide3</td>
<td>CTCACACAGGACAATCCGG</td>
<td>CAGG</td>
<td></td>
<td>Crispr guide</td>
</tr>
<tr>
<td>sequence-based reagent</td>
<td>This paper</td>
<td>Crispr guide</td>
<td>For1</td>
<td>Genotyping primer</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------</td>
<td>--------------</td>
<td>------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Guide4</td>
<td>This paper</td>
<td>Crispr guide</td>
<td>TCTTCTGCAACTCGGGGCCA</td>
<td></td>
</tr>
<tr>
<td>Guide5</td>
<td>This paper</td>
<td>Crispr guide</td>
<td>TCTCTGTGAGGCCTGGTAG</td>
<td></td>
</tr>
<tr>
<td>Guide6</td>
<td>This paper</td>
<td>Crispr guide</td>
<td>AACACTCTGGTCAGCTCAGA</td>
<td></td>
</tr>
<tr>
<td>For1</td>
<td>This paper</td>
<td>Genotyping primer</td>
<td>GAAATCGCATGTCACGCAGAAAT</td>
<td></td>
</tr>
<tr>
<td>Rev2</td>
<td>This paper</td>
<td>Genotyping primer</td>
<td>ATATCGGAACTGACATACTGAACG</td>
<td></td>
</tr>
<tr>
<td>Rev2.2</td>
<td>This paper</td>
<td>Genotyping primer</td>
<td>GTGAGCTTCTGTTTGTGGT</td>
<td></td>
</tr>
<tr>
<td>For3</td>
<td>This paper</td>
<td>Genotyping primer</td>
<td>GCCATTGTTGAGATTCAGTA GTCA</td>
<td></td>
</tr>
<tr>
<td>Rev4</td>
<td>This paper</td>
<td>Genotyping primer</td>
<td>CATTTACTCCTACCATTGACATC</td>
<td></td>
</tr>
<tr>
<td>Alt-R CRISPR-Cas 9 Nuclease</td>
<td>IDT</td>
<td>Cat. #: 10000735</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNeasy mini kit</td>
<td>Qiagen</td>
<td>Cat. #: 74106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial assay or kit</td>
<td>RNeasy midi kit</td>
<td>Qiagen</td>
<td>Cat. #: 75144</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------</td>
<td>--------</td>
<td>--------------</td>
<td></td>
</tr>
<tr>
<td>Commercial assay or kit</td>
<td>ProtoScript II First Strand Synthesis kit</td>
<td>NEB</td>
<td>Cat. #: E6560S</td>
<td></td>
</tr>
<tr>
<td>Commercial assay or kit</td>
<td>TSA Plus Cy3/Cy5 kit</td>
<td>Perkin-Elmer</td>
<td>Cat. #: NEL744001KT</td>
<td></td>
</tr>
<tr>
<td>Commercial assay or kit</td>
<td>Alt-R CRISPR-Cas9 tracr RNA</td>
<td>IDT</td>
<td>Cat. #: 1072532</td>
<td></td>
</tr>
<tr>
<td>Software, algorithm</td>
<td>Enrichr</td>
<td>10.1093/nar/gkw377</td>
<td>RRID:SCR_001575</td>
<td></td>
</tr>
<tr>
<td>Software, algorithm</td>
<td>TOGA</td>
<td>Kirilenko <em>et al.</em> TOGA integrates gene annotation with orthology inference at scale. under Review. <a href="https://github.com/hillerlab/TOGA">https://github.com/hillerlab/TOGA</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software, algorithm</td>
<td>CESAR 2.0</td>
<td>10.1093/bioinformatics/btx527</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software, algorithm</td>
<td>RELAX</td>
<td>10.1093/molbev/msu400</td>
<td>RRID:SCR_016162</td>
<td></td>
</tr>
<tr>
<td>Software, algorithm</td>
<td>MACSE v2.0</td>
<td>10.1093/molbev/msy159</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software, algorithm</td>
<td>Jalview</td>
<td>10.1093/bioinformatics/btp033</td>
<td>RRID:SCR_006459</td>
<td></td>
</tr>
<tr>
<td>Software, Human protein atlas</td>
<td>10.1038/nbt1210-1248</td>
<td>RRID:SCR_006710</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algorithm</td>
<td>Software</td>
<td>Version</td>
<td>DOI or Other Identifier</td>
<td>RRID:SCR_00798 9</td>
</tr>
<tr>
<td>-----------</td>
<td>----------</td>
<td>---------</td>
<td>--------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Gene Expression Atlas</td>
<td>PyMOL</td>
<td>Schrödinger</td>
<td>10.1093/nar/gkz947</td>
<td>RRID:SCR_00030 5</td>
</tr>
<tr>
<td>Fiji</td>
<td>10.1038/nmeth.2019</td>
<td>10.10228 5</td>
<td><a href="https://cran.r-project.org/src/base/R-3/">https://cran.r-project.org/src/base/R-3/</a></td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>Other</td>
<td>10.10190 5</td>
<td><a href="http://docs.ggplot2.org/current/">http://docs.ggplot2.org/current/</a></td>
<td></td>
</tr>
<tr>
<td>ggplot2</td>
<td>Other</td>
<td>10.10460 1</td>
<td><a href="https://CRAN.R-project.org/package=tseries">https://CRAN.R-project.org/package=tseries</a></td>
<td></td>
</tr>
<tr>
<td>tseries</td>
<td>Other</td>
<td>10.1038/s41587-019-0036-z</td>
<td><a href="https://cran.r-project.org/web/packages/effsize/index.html">https://cran.r-project.org/web/packages/effsize/index.html</a></td>
<td></td>
</tr>
<tr>
<td>effsize</td>
<td>Other</td>
<td>10.1038/s41587-019-0036-z</td>
<td>10.1093/bioinformatics/btx431</td>
<td></td>
</tr>
<tr>
<td>SignalP5</td>
<td>10.1093/nar/gkw398</td>
<td>RRID:SCR_01572 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeepLoc</td>
<td>10.1093/nar/gkw398</td>
<td>RRID:SCR_01572 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ChopChop v2</td>
<td>10.1093/nar/gkw398</td>
<td>RRID:SCR_01572 3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A phenotype association -\log_{10}(\text{FDR})

Cape golden mole
Manatee
Elephant
Star-nosed mole
Little brown bat
David's myotis (bat)
Big brown bat
Large flying fox
Black flying fox
Weddell seal
Pacific walrus
Ferret
Dog
Cat
White rhinoceros
Horse
Goat
Bison
Cow
Tibetan antelope
Minke whale
Killer whale
Dolphin
Bactrian camel
Alpaca
Pika
Rabbit
Naked mole rat
Degu
Guinea pig
Blind mole rat
Rat
Mouse
Deer mouse
Prairie vole
Squirrel
Bush baby
Tarsier
Squirrel monkey
Marmoset
Snub-nosed monkey
Green monkey
Baboon
Crab-eating monkey
Rhesus monkey
Orangutan
Bonobo
Chimpanzee
Human

0 1 2 3 4 5 6 7 8

B Gene

**Known vision-related function**

TSACC
SERPINE3
GUCA1B
CRYBA4
CRYBA1
RSKR
LACTBL1
PKD1L2
SAPCD1
OPTC
CRB1
GJA10
MAGEB1
ZNF529
GZMK
BCL2A1
KRT12
STX10
HSD11B1L
BIRC7
ZNF214
LBHD1

Expression in mouse gene atlas

**lens**
**retina**
**retinal pigment epithelium**
**dendritic cells**
**lymphoid**
**CD8a+**
**bladder**

C MGI Mammalian phenotypes

**total cataract**
**abnormal lens fiber morphology**
**abnormal lens morphology**
**increased mitochondrial number**
**increased corneal epithelium thickness**

**Expression in mouse gene atlas**

lens
retina
retinal pigment epithelium
dendritic cells
lymphoid
CD8a+
bladder

**OMIM Disease**
cataract
leber amaurosis
corneal dystrophy
microphthalmia
retinitis pigmentosa

**GO Biological process**
visual perception
sensory perception of light stimulus
G protein-coupled receptor internalization
regulation of lymphocyte apoptotic process
receptor guanylyl cyclase signaling pathway
Human genome Ensembl gene annotation v87
18,946 genes

17,946 genes

13,172 genes

68 genes

31 genes

Genome assemblies cover gene in >50% of high-acuity and >50% of low-acuity species

%intact values vary across species

Phylogenetic generalized least square corrected p-value < 0.05

Gene is independently lost in >2 low-acuity lineages

P-value correction
A  Common vampire bat genome (HLdesRot8A)  manual_scaffold_11:88,446,025-88,466,426

B  Common vampire bat genome (HLdesRot8A)  manual_scaffold_14:90,725,500-90,732,264
deletion of part of Exon 6; Exon 7, frame-shift -1, -2, -2

Manis tricuspis

-1

Phataginus tricuspis

-1

Manis pentadactyla

Manis javanica
details on 15 more Yangochiroptera are in a separate figure
<table>
<thead>
<tr>
<th>Species</th>
<th>Amino Acid</th>
<th>Codon 1</th>
<th>Codon 2</th>
<th>Codon 3</th>
<th>Codon 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Macrotus californicus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Micronycteris hirsuta</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Desmodus rotundus</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Phyllostomus discolor</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Tonatia saurophila</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Carollia perspicillata</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Sturnira hondurensis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Artibeus jamaicensis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Anoura caudifer</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Glossophaga soricina</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Leptonycteris yerbabuenae</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Pteronotus parnellii</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mormoops blainvillei</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Noctilio leporinus1</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Noctilio leporinus2</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SERPINE3 losses in rodents

Nannospalax galili

TAA

Heterocephalus glaber

TGA

Fukomys damarensis

Peromyscus leucopus

Peromyscus leucopus

Unclear SERPINE3 losses

Ellobius lutescens

Raphicerus campestris

Okapia johnstoni

SERPINE3 losses

TAA

Heterocephalus glaber

TGA

Fukomys damarensis

Peromyscus leucopus

Peromyscus leucopus

Unclear SERPINE3 losses

Ellobius lutescens

Raphicerus campestris

Okapia johnstoni
Iris circularity

WT

serpine3^{cbg17}

p=3.71E-02

WT

serpine3^{cbg18}

p=1.64E-05