Post-transcriptional control of gene expression by CELF1 in lens development and pathology

Matthieu Duot

COLLEGE SCIENCES
DOCTORAL DE LA VIE
BRETAGNE ET DE LA SANTE

Université de Rennes

THÈSE DE DOCTORAT DE

L'UNIVERSITE DE RENNES

ECOLE DOCTORALE N° 637
Sciences de la Vie et de la Santé
Spécialité : Biologie Cellulaire, Biologie du développement

Par
Matthieu DUOT

Contrôle post-transcriptionnel de l'expression génétique par CELF1 dans le développement et les pathologies du cristallin

Thèse présentée et soutenue à Rennes, le 24 octobre 2023
Unité de recherche : Institut de Génétique et Développement de Rennes (IDGR,UMR 6290)

Rapporteurs avant soutenance :

Isabelle Behm-Ansmant	CRCN CNRS, IMoPA Vandoeuvre-Les-Nancy
Julia Morales	DR2 CNRS, LBI2M Roscoff

Composition du Jury :

Présidente : Muriel Perron	DR1 CNRS, NeuroPSI Saclay
Examinateurs : Isabelle Behm-Ansmant	CRCN CNRS, IMoPA Vandoeuvre-Les-Nancy
Julia Morales	DR2 CNRS, LBI2M Roscoff
Dir. de thèse : Luc Paillard	PU Université de Rennes, IGDR Rennes
Co-Encadrant : Carole Gautier-Courteille	PU Université de Rennes, IGDR Rennes
Co-Encadrant : Salil A. Lachke	Associate Chair, University of Delaware Newark (USA)

Content Table

Acknowledgments

To begin with, I would like to thanks Drs Isabelle Behm-Ansmant, Julia Morales and Muriel Perron for participating to my thesis jury.

I would also like to express my gratitude to Drs Laurent Corcos and Odile Bronchain for their support throughout my PhD, during the CSI sessions.

Additionally, I extend my thanks to the University of Rennes for sponsoring my PhD for three years.

I want to express my thanks to Rennes Métropole for offering me an international mobility grant.

Obviously, I am truly grateful to Luc Paillard, my PI, and Carole Gautier-Courteille, my Co-supervisor, for placing their trust in me and welcoming me into the EGD team at the IGDR. You have been instrumental in guiding me through these years with numerous discussions on this project, helping me understand the complexity of post-transcriptional regulation, splicing, and the wonders of the RNA world. Your trainings has been invaluable.

Certainly, I thank all the members of the EGD team for their amazing help throughout these three years. Special thanks to David Reboutier for teaching me a lot about cell culture, organoid generation immunostaining imaging and so more, your supervision was irreplaceable for my PhD.

Thanks also to Yann Audic for helping me in debugging a large part of my code, Justine Viet for the generation the RNA-seq library, and to Agnès Mereau for your expertise in creating the iCLIP library, which was a cornerstone of my work. I'd also like to express my gratitude to Catherine Gaillard-Le-Goff, Franck Chesnel, and Xavier Le Goff for your valuable advice and support.

During my PhD, I had the wonderful opportunity to spend a year working with my Co-supervisor, Salil A. Lachke, at the University of Delaware. Indeed, Luc and Carole have had a longstanding collaboration with Salil's lab, resulting in three joint publications even before the beginning of my PhD. My doctoral research provided an opportunity to bring these two labs closer together, and a PhD student from Salil’s lab will have a similar opportunity to work for a year in a foreign laboratory (in Luc and Carole's teams). Therefore, I would like to thank the Doctoral School of Brittany for supporting this international collaboration.

I cannot thank Salil and the members of his lab enough for their warm welcome when I arrived in Newark in February 2022.

I would like to extend special thanks to Salil A. Lachke, my Co-supervisor, and my mentor at the University of Delaware. Thank you for the lengthy

discussions about the lens, the American science community, and many other subjects. Thanks for the immeasurable help with the preparation of the different publications that we wrote.

I also want to express my thanks to all the PhD students and post-docs from Lachke's lab, especially Sarah Coomson and Sanjaya Shrestha, but also Deepti Anand, Shaili Patel, Megha Oza, Sangeetha Kharidehal, and Jiniya Chakradhar. Thank you for the good mood that you brought to the lab and during all the conferences we were able to attend.

I would also like to express my gratitude to the members of the H2P2 platform for training me in the use of the Laser micro-dissector and assisting me with numerous staining procedures during my project. Special thanks to Roselyne Viel, Alain Fautrel, and Nicolas Mouchet.

Finally, I am immensely thankful for my family for their unconditional support throughout all these years. Thank you, Mom, Dad, and my little brother Adrien.

And thanks to all my friends who helped me stay motivated and forget about science after work. Since I'm afraid to forget anyone, I will just thank all the friends that I met in Rennes, Laval, Nantes, and Newark. I'm sure you will recognize yourselves.

Résumé Français

Introduction

Le cristallin, en anglais "lens", est un organe avasculaire caractérisé par des propriétés optiques spécifiques indispensables à la vision. Il réside entre la cornée et la rétine au sein de l'œil. Ce positionnement lui confère la capacité de focaliser la lumière issue d’objets externes sur la rétine.

Le cristallin est attaché à l’œil par le biais des fibres zonulaires. Ces fibres peuvent déformer le cristallin, modifiant ainsi sa forme et ses propriétés optiques. Cette déformation permet d'adapter la vision tant pour les objets éloignés que proches. De ce fait, il est impératif de préserver l'élasticité du cristallin tout au long de la vie afin d'assurer le bon fonctionnement de l'œil.

Le cristallin est constitué de deux types cellulaires: une monocouche de cellules épithéliales du cristallin (LEC) présentes sur la surface antérieure du cristallin, et des fibres cristalliniennes (FC). De plus, le cristallin est enveloppé par une membrane basale, la capsule.

Les LEC sont présentes en trois régions distinctes. La région pré-germinative se trouve dans la région la plus antérieure du cristallin, où les cellules ne se divisent pas. Elle est suivie par la zone germinative, la seule région où les cellules se divisent. À l'équateur du cristallin, les cellules entrent dans la zone de transition où elles cessent de se diviser et commencent leur différenciation en FC.

Les FC constituent la majeure partie du cristallin. Durant la différenciation des LEC en FC, ces cellules subissent des changements drastiques dans l'organisation de leur cytosquelette, formant ainsi des fibres allongées dans l'axe antéro-postérieur du cristallin. Les FC sont densément agencées pour faciliter le passage de la lumière. Ces cellules dégradent également leur noyau et leurs organites tout en produisant de grandes quantités de protéines particulières appelées cristallines. Cette transformation favorise la transmission de la lumière et protège les cellules contre les stress environnementaux.

Contrairement à la majorité des cellules, les FC ne disposent pas d’un mécanisme intrinsèque de mort cellulaire pour renouveler la population cellulaire du cristallin. En conséquence, les cellules différenciées persistent tout au long de la vie. L’accumulation d'agrégats protéiques en condition de stress au sein des FC conduit donc inévitablement à l'opacification du cristallin, engendrant ainsi la formation d'une cataracte liée à l'âge. L’apparition de la cataracte liée à l'âge peut varier en précocité en fonction de facteurs génétiques, environnementaux et/ou de la coexistence d'autres pathologies.

La cataracte demeure la principale cause de cécité à l'échelle mondiale. Elle peut être aisément traitée par chirurgie dans les pays industrialisés. Toutefois, aucun médicament préventif n'a encore été identifié en raison de la compréhension encore limitée de son étiologie.

Bien que la majorité des cas de cataracte soit liés à l'âge, des cataractes congénitales peuvent également se manifester (environ 5 naissances sur 10 000). Elles peuvent être déclenchées par des infections virales contractées *in utero*. Cependant, dans la majorité des cas, ces cataractes congénitales ont une origine génétique. Ainsi, plus de 400 gènes ont été identifiés comme étant associés à la cataracte congénitale, et ces mêmes gènes sont souvent associés à la précocité de l'apparition de la cataracte liée à l'âge. Parmi ces gènes, on trouve non seulement des gènes codant pour des cristallines, mais également des facteurs de transcription indispensables au maintien des LEC ou à leur différenciation en FC, par exemple *PAX6, PROX1, SOX2, L-MAF, MAF, MAFB*, etc.

Bien que les régulations transcriptomiques soient vitales pour la formation et la stabilité du cristallin, les régulations post-transcriptionnelles jouent un rôle particulièrement crucial dans cet organe. En effet, au cours de la différenciation des FC, ces cellules dégradent leur noyau, empêchant ainsi la transcription de nouveaux ARN messagers. Par conséquent, différents mécanismes de régulation post-transcriptionnelle sont nécessaires au niveau de la zone de transition. Ils ont pour rôle de réprimer la traduction de gènes associées aux LEC, tout en favorisant la traduction de gènes nécessaires aux processus de différenciation des FC (comme l'élongation des fibres, la perte de noyaux et d'organites, etc.). En outre, ces mécanismes permettent la stabilisation de certains ARN messagers, contribuant ainsi au maintien de la concentration importante de cristallines dans les FC tout au long de la vie.

Parmi les régulateurs post-transcriptionnels impliqués dans le développement normal du cristallin, de nombreux micro-ARN ont été identifiés, ainsi que des ARN longs non codants et des ARN circulaires. Cependant, les régulateurs post-transcriptionnels les plus étudiés dans le contexte du cristallin sont les protéines de liaison à l'ARN (RBP). Ces protéines sont capables d'interagir avec des ARN spécifiques, modifiant ainsi leur traductibilité et/ou leur stabilité. De plus, certaines de ces RBP peuvent également être localisées dans le noyau et influencer l'épissage alternatif des ARN prémessagers ciblés.

TDRD7 a été la première RBP associée au développement du cristallin. Elle a été identifiée en raison d'une mutation génétique causant une cataracte infantile. Des modèles animaux ont été utilisés pour confirmer l'importance de ce gène dans le développement du cristallin et le maintien de sa transparence. Suite à la découverte de TDRD7, le rôle de nombreuses autres RBP a commencé à être étudié dans le contexte du cristallin (ex :CAPRIN2, RBM24,...).

Parmi ces RBP impliquées dans la formation du cristallin, CELF1 occupe une place significative. Pendant ma thèse, j'ai exploré le rôle de CELF1 dans le développement du cristallin. Avant mon arrivée, mon équipe de recherche avait généré des souris KO pour *Celf1*. Ces souris présentent plusieurs pathologies, dont l'apparition de cataractes congénitales. L'absence de CELF1 impacte le développement du cristallin, aboutissant à une cataracte congénitale. Cette cataracte est caractérisée par l'absence de dénucléation des FC ainsi que de perturbations dans la structure du réseau d'actine et dans la morphologie des FC.

Jusqu'à présent, seuls quelques gènes directement régulés par CELF1 ont été identifiés dans le cristallin de souris. Parmi ceux-ci, *Dnase2b, p27Kip1* et *p21Cip1* jouent un rôle crucial dans la dégradation des noyaux des FC. De plus, deux facteurs de transcription majeurs dans le cristallin, PAX6 et PROX1, sont contrôlés par CELF1. De plus, de nombreux autres gènes sont probablement dérégulés, en absence de CELF1, sans que l'on sache à ce stade de savoir lesquels de ces gènes sont contrôlés directement ou indirectement par CELF1. De plus amples études sont donc nécessaires pour identifier l'ensemble des gènes directement régulés par CELF1 dans le cristallin, afin d'améliorer notre compréhension de l'étiologie de la cataracte.

L'un des objectifs majeur de ma thèse a donc été d'identifier de nouveaux ARN régulés par CELF1, contribuant ainsi à enrichir notre connaissance de la biologie du cristallin et de la formation de cataracte.

Celf1 fait partie de la famille des gènes *Celf*, qui comprend six membres numérotés de *Celf1* à *Celf6*. Les membres de cette famille codent des protéines caractérisées par la présence de trois motifs de reconnaissance de l'ARN (RRM) permettant leur liaison à leurs cibles ARN, ainsi qu'une région divergente entre leur deuxième et troisième RRM. Parmi les membres de la famille CELF, seul CELF1 est présente dans le cristallin.

Celf1 a été initialement identifié dans le contexte de la dystrophie myotonique de type 1 (DM1), une maladie rare à transmission autosomique dominante. Cette pathologie provoque chez les patients une faiblesse musculaire, des problèmes cardiaques, des troubles cognitifs et des cataractes infantiles. La DM1 est causée par une importante expansion d'une séquence répétée CTG dans la région 3'UTR du gène *Dmpk*.

À l'origine, il était supposé que CELF1 était piégée dans le noyau par l'ARN portant cette répétition de poly CUG. Cependant, il a été démontré que ce n'était pas CELF1 qui était retenue par l'ARN muté, mais une autre RBP appelée Muscleblind protein (MBNL1). Néanmoins, CELF1 demeure impliquée dans cette pathologie. Des modèles animaux ont montré que la surexpression de *Celf1* dans les muscles ou le cœur reproduit les modifications post-transcriptionnelles observées chez les patients atteints de DM1. L'hypothèse

actuelle suggère que CELF1 et MBNL1 ont un rôle antagoniste, et que la séquestration de MBNL1 par l'ARN muté entraîne une suractivité de CELF1.

CELF1 a été associée à d'autres pathologies telles que des maladies cardiaques, la maladie d'Alzheimer et divers types de cancers. Cependant, à ce jour, aucune étude n'a établi un lien entre CELF1 et l'apparition de cataractes chez l'Homme.

CELF1 peut réguler à un niveau post-transcriptionnel l’expression génétique d'autres gènes par le biais de différents mécanismes. Lorsque CELF1 se trouve dans le cytoplasme, elle peut se lier aux ARN messagers et altérer leur stabilité et/ou leur traductibilité, entraînant ainsi des modifications du niveau de protéines de ces gènes. Par ailleurs, lorsque CELF1 est localisée dans le noyau, il peut se lier aux ARN pré-messagers et moduler leur épissage alternatif. Cette modulation entraîne la création d'ARN messagers différents, ce qui peut générer des isoformes protéiques alternatives possédant des fonctions et/ou des localisations cellulaires distinctes. De plus, les modifications apportées aux ARN messagers peuvent influencer leurs interactions avec d'autres régulateurs, tels que les microARN ou d'autres RBP.

Pour identifier les gènes contrôlés par CELF1, des méthodes telles que le Cross-Linking ImmunoPrecipitation (CLIP) ou le RNA ImmunoPrecipitation (RIP) ont été utilisées pour mettre en évidence les ARN qui interagissent directement avec CELF1. Cependant, ces interactions peuvent différer en fonction du contexte (organe, âge, espèce). À notre connaissance, aucune étude n'a encore examiné des données de CLIP ou RIP dans le cristallin pour CELF1. Afin d'identifier les ARN interagissant avec CELF1 dans cet organe j’ai, durant ma thèse, analysé les données d'un iCLIP-seq réalisé sur des cristallins de souris.

La plupart des modèles utilisés pour étudier l'œil et le cristallin sont des modèles animaux, tels que les embryons de poulet qui ont été largement utilisés dans les premières études sur l'embryologie du cristallin. Des modèles aquatiques comme le poisson zebrafish ou l'amphibien xénope ont également été utilisés pour évaluer l’impact de l'inactivation de gènes soupçonnés d'être liés à la cataracte. Cependant, la souris reste le modèle le plus couramment utilisé pour la recherche sur le cristallin. Ce modèle a permis de confirmer le rôle de nombreux gènes dans le développement et le maintien de la transparence du cristallin.

La recherche *in cellulo* sur le cristallin est plus complexe, car le cristallin est principalement constitué de FC qui n'ont pas de noyaux et ne peuvent pas donc être cultivées. Certaines études ont été menées sur les LEC, mais il est difficile d'induire leur différenciation complète en FC en culture en 2D. Par conséquent, différentes équipes ont exploré la possibilité de créer des modèles 3D de cristallin pouvant imiter partiellement les processus survenant durant le développement.

Ces équipes ont développé différents modèles, chacun présentant des caractéristiques intéressantes pour la recherche sur le cristallin. Cependant, ces modèles nécessitent l'utilisation de cellules souches embryonnaires ou de cellules pluripotentes induites, ainsi que différents facteurs de croissance. Cela rend l'utilisation de ces modèles plus complexe et coûteuse. Au cours de ma thèse, j'ai travaillé sur la caractérisation d'un nouveau modèle d'organoïde du cristallin, facile a générer en grande quantité.

Objectif de la thèse

L'objectif principal de ma thèse était d'identifier de nouveaux gènes directement régulés par CELF1 dans le cristallin et de comprendre l'importance de ces régulations post-transcriptionnelles dans le développement du cristallin, et les causes de l'apparition de cataracte en absence de CELF1. J'ai d'abord réalisé une analyse transcriptomique du cristallin de souris nouveau-né avec une déficience en *Celf1* pour comprendre l'impact global de l'absence de régulation par CELF1, résultant en une publication en 2023 (chapitre I). Ensuite, j'ai identifié des gènes dont l'épissage alternatif est directement contrôlé par CELF1 en intégrant les résultats de plusieurs techniques omiques, dont deux RNA-seq et un iCLIP-seq réalisés sur des cristallins de souris (chapitre II). Enfin, j'ai contribué au développement d'un nouveau modèle d'organoïde pour le cristallin. Grâce à des analyses transcriptomiques approfondies de ce modèle, j'ai démontré sa capacité à partiellement reproduire le développement du cristallin ainsi que la différenciation des FC. Ces résultats ont été présentés dans un manuscrit actuellement en cours de soumission (chapitre III).

Résultats

Les études précédentes réalisées par mon équipe d'accueil ont confirmé le rôle crucial de CELF1 dans le contrôle de l'expression génique post-transcriptionnelle au sein du cristallin. L'inactivation conditionnelle (cKO) de *Celf1*, réalisée par une Cre recombinase sous le contrôle du promoteur *Pax6* spécifique de l'œil, entraîne le développement de cataractes congénitales. Dans le chapitre I, je présente mon travail visant à décrypter les changements transcriptomiques survenant dans le cristallin en l'absence de CELF1. Pour ce faire, j'ai réalisé une analyse d'expression différentielle sur des données d’un RNA-seq obtenues à partir de cristallin de souris témoins et cKO postnatals.

Par cette approche, j'ai identifié 987 gènes exprimés différentiellement en absence de CELF1. Une majorité de ces gènes sont surexprimés en l'absence de CELF1. Étant donné le rôle connu de CELF1 pour promouvoir la dégradation des ARNm, une proportion significative de ces gènes surexprimés pourrait être sous la régulation directe de CELF1.

L'intégration des données CLIP-seq obtenues à partir de cellules HeLa et des scores d'enrichissement dans le crystallin obtenus par la base de donnée iSyTE a révélé un schéma intéressant. Notamment, les gènes présentant des niveaux d'expression accrus sont peu spécifiques du cristallin. Les ARNm codés par une partie significative d'entre eux interagissent avec CELF1. Il pourrait donc s'agir de gènes responsables de fonctions générales dans la cellule, dont la surexpression de ces gènes pourrait avoir des effets préjudiciables sur le développement du cristallin. La répression par la CELF1 de ces gènes serait nécessaire pour la différenciation du cristallin.

Inversement, les gènes dont l'expression est réduite en absence de CELF1 sont généralement fortement exprimés dans le cristallin. Ainsi, ils pourraient être cruciaux pour un développement normal du cristallin. Leur expression réduite en l'absence de CELF1 pourrait contribuer à la formation des cataractes congénitales observées chez la souris cKO.

De plus, en combinant ces données avec d'autres ensembles de données transcriptomiques obtenues à partir d'analyses de microarray de souris cKO à différents stades de développement, nous avons pu affiner et prioritiser un sous-ensemble de gènes les plus susceptibles d'être impliqués dans la formation de la cataracte observée chez les souris cKO. Parmi ces candidats, on retrouve le facteur de transcription *Prdm16* ainsi que le facteur d'élongation *Ell2*. Cette sélection de gènes ouvre de nouvelles perspectives pour des investigations sur les mécanismes aboutissant à la formation d’une cataracte.

Dans le chapitre II, j’ai voulu me concentrer sur la régulation nucléaire par CELF1, le contrôle de l’épissage alternatif des ARN prémessager ciblés.

Dans cette étude, un iCLIP-seq sur la protéine CELF1 réalisé sur des cristallins de souris adultes a permis d'identifier 1 728 gènes dont les ARN interagissent directement avec CELF1. Étant donné que cet iCLIP-seq a été effectué sur des cristallins entiers, les ARN ciblés par CELF1 correspondent aux deux types cellulaires du cristallin, LEC et FC.

Ainsi, ces données peuvent être extrêmement précieuses pour étudier à la fois le rôle cytoplasmique et nucléaire de CELF1 dans le cristallin. 5 % des sites de liaison identifiés sont présents dans les régions introniques. Cela confirme l'activité nucléaire de CELF1 dans le cristallin, car CELF1 peut interagir avec ces régions pré-ARNm uniquement dans le noyau.

Deux RNA-seq ont été utilisées pour identifier des gènes présentant des épissages anormaux dans le cristallin en absence de CELF1 chez les souris adultes et des nouveaux-nés. Les données RNA-seq sur les cristallins de nouveaux-nés sont celles obtenues à partir de souris cKO et déjà été étudiées dans le contexte du niveau d’expression des gènes (cf chapitre I). Les données RNA-seq sur les cristallins adultes correspondent à des souris inactivées pour

Celf1 de manière constitutive. Ces données n'avaient pas été traitées auparavant. J'ai ainsi identifié des jonctions d'épissage utilisées de façon différentes chez le témoin et en absence de CELF1 chez 438 et 1 061 gènes, respectivement, chez les nouveaux-nés et les adultes.

En intégrant les données de iCLIP-seq avec les deux ensembles de données RNA-seq et en effectuant un processus de sélection manuelle, j'ai sélectionné les 22 gènes les plus prometteurs. Parmi ceux-ci, 10 codent de protéines du cytosquelette. Nous nous sommes concentrés sur ces 10 gènes, puisque les crystallins déficients en *Celf1* présentent des défauts du cytosquelette. Nous avons ensuite validé par RT-PCR l'épissage différentiel de 7 de ces gènes : *Ablim1, Ctnna2, Clta, Septin8, Sptbn1, Ywhae* et *Ank2*.

Avec le nouvel outil de prédiction de structure protéique AlphaFold2 (AF2), j'ai prédit les différences structurelles des isoformes de protéines de ABLIM1, CTNNA2, CLTA, SEPTIN8, SPTBN1, YWHAE et ANK2.

Ainsi AF2 a prédit des résultats intéressants pour SPTBN1. CELF1 réprime la formation des isoformes dépourvus du domaine de Pleckstrin Homology (PH). La fonction exacte du domaine PH reste inconnue dans le contexte du cristallin. Toutefois, dans d'autres organes, il est impliqué dans la localisation des protéines à la membrane. Par conséquent, une probable mauvaise localisation de l'isoforme sans PH pourrait potentiellement contribuer aux anomalies du cytosquelette observées dans le cristallin des souris déficientes en *Celf1*.

Les prédictions d'AF2 suggèrent également un changement conformationnel dans l'hélice alpha principale de CLTA. Cette altération est impliquée dans la modulation de la taille de vésicules dans les neurones. Dans le contexte du cristallin, ce changement conformationnel pourrait potentiellement perturber la formation d'une structure spécifique des FC ("ball and sockets") qui permet une forte cohésion des FC. Cette perturbation structurelle pourrait avoir des implications sur l’agencement des FC, ce qui pourrait contribuer à la formation d'une cataracte.

Pour les autres gènes candidats validés (*Ablim1, Ctnna2, Septin8, Ywhae* et *Ank2*), AF2 n'a pas été en mesure de prédire la conformation du domaine gagné ou perdu en l’absence de CELF1. Ces domaines sont des régions intrinsèquement désordonnées (IDR). Les IDR sont des segments de protéines qui nécessitent des interactions avec d’autres protéines pour adopter leur conformation appropriée. Ces IDR jouent un rôle crucial dans les interactions protéine-protéine. Ainsi, nous supposons que les IDR contrôlés par CELF1 dans ABLIM1, CTNNA2, SEPTIN8, YWHAE et ANK2 dans le cristallin pourraient aboutir à différents réseaux d'interaction protéine-protéine. L'identification des intéractomes spécifiques de chaque isoformes pour chacun de ces gènes aidera à comprendre les mécanismes moléculaires sous-jacents à la formation de la cataracte dans les crystallins déficients en *Celf1*. Cette compréhension

pourrait potentiellement révéler de nouveaux gènes essentiels pour le développement du cristallin.

Enfin dans le chapitre III, je présente la caractérisation d'un nouveau modèle d'organoïde de crystallin.

Notre protocole peut générer des organoïdes présentant des propriétés optiques intéressantes en seulement 10 jours, sans aucun facteur de croissance. Ces organoïdes sont transparents et, de manière similaire au cristallin, sont capables de focaliser la lumière.

Nous avons effectué une analyse RNA-seq pour discerner les modifications transcriptomiques lors de la transition de la culture 2D à la culture 3D. De manière intéressante, l'expression des gènes davantage exprimés dans les organoïdes 3D sont augmente lors du développement du cristallin chez les souris. Parmi ces gènes figurent des facteurs de transcription spécifiques du cristallin tels que *Eya1, Meis1, Prox1, Hsf4* et *Maf*. De plus, des gènes codant pour des protéines de liaison aux ARN liées au cristallin bien connues telles que *Tdrd7* et *Caprin2*, ainsi que des membres de voies de signalisation cruciales dans le développement du cristallin (*Jag1, Notch3, Tgfb2, Tgfb3, Fgfrl1, Ephb6, Epha7, Efna1)* ont été identifiés. Ces changements soutiennent l'idée que notre modèle subit, au moins partiellement, des processus similaires à celles observés dans le cristallin en développement.

Tout comme l'organisation structurale du cristallin, notre modèle d'organoïde présente des régions histologiques distinctes. Nous en avons identifié trois : (i) une région externe avec des noyaux ronds, (ii) une région intermédiaire avec des noyaux allongés, et (iii) une région interne avec des noyaux compacts. De manière similaire au cristallin, seules les cellules de la région externe peuvent se diviser, tandis que celles des autres régions cessent de se diviser pour initier leur différenciation, impliquant des modifications de morphologie et de transcriptome. Il convient de noter que, contrairement au cristallin, les cellules de la région interne de ces organoïdes ne subissent pas de dégradation nucléaire complète. Cependant, malgré cette différence, le modèle conserve sa transparence.

Pour caractériser les changements transcriptomiques entre les différentes régions, nous avons utilisé la microdissection par capture laser pour séparer les régions interne et externe avant de séquencer leurs transcriptomes respectifs. Comme prévu, les gènes liés au cycle cellulaire sont surexprimés dans la région externe. En effet, ces cellules sont les seules capables de se diviser. Certains gènes associés à la capsule ont également été identifiés (par exemple, *Lmnb1*). En revanche, dans la région interne sont sur-exprimés des gènes spécifiques des FC, notamment des facteurs de transcription tels que *Prox1* et *Maf*, des protéines du cytosquelette telles qu'*Ank2* et des protéines cristallines telles que *Cryab*. Le marquage par immunofluorescence a validé

cette régionalisation: la laminine, une protéine de la capsule, est principalement localisée dans la région externe, tandis que les niveaux de HSF4 et CRYAB sont plus élevés dans la région interne. De plus, le marquage de la protéine de la membrane externe mitochondriale TOMM20 est réduit dans la région interne. Cela suggère que ces cellules amorcent la dégradation des mitochondries dans le cadre de leur processus de différenciation.

Il est important de noter que ce nouveau modèle d'organoïde offre des perspectives prometteuses pour l'étude de la formation de la cataracte. Lorsqu'ils sont exposés à des molécules connues pour induire des cataractes, ces organoïdes perdent leur transparence caractéristique et leur capacité à focaliser la lumière, de manière similaire aux cristallins présentant une cataracte. De plus, en inactivant le gène *Celf1*, les organoïdes deviennent opaques et perdent également leur capacité à focaliser la lumière, comme dans le cristallin des souris déficientes en *Celf1*. Cela fait de ce modèle un outil précieux pour étudier les changements moléculaires qui surviennent dans le développement de la cataracte suite à un stress environnemental ou à une mutation dans un gène associé à la cataracte. Le coût abordable de ce modèle devrait permettre le criblage de composés pour la prévention ou le traitement de la cataracte.

Conclusion

En conclusion, durant mon projet de thèse j'ai mis en évidence les altérations transcriptomiques qui ont lieu dans le cristallin en absence de CELF1. De plus, j'ai réussi à identifier des gènes susceptibles d'être directement régulés par CELF1, que ce soit par la modulation de la stabilité des ARNm ou l'épissage alternatif des pré-ARNm. Pour valider l'importance de ces régulations dans la formation de la cataracte, j'ai contribué à développer un nouveau modèle d'organoïde qui sera utile pour étudier différents types de cataracte, y compris ceux résultant de la déficience en CELF1. Ce modèle novateur offre non seulement des informations sur les mécanismes sous-jacents à la cataracte liée à la déficience en *Celf1*, mais constitue également un outil pour la communauté de recherche sur le cristallin afin d'étudier d'autres types de cataracte.

Introduction

The ocular lens exhibits unique characteristics that enable it to focus light onto the retina, ensuring clear vision for the organism. In this introduction, I will highlight the two main cell types of the lens, namely the lens epithelial cells (LEC) and the fiber cells (FC), as well as the process of lens formation. Opacification of the lens leads to cataracts, a major pathology associated with the lens.

Congenital cataracts can result from genetic mutations that hinder proper LEC maintenance or the differentiation of LEC into FC, underscoring the critical role of gene expression controls in preventing cataract formation. During FC differentiation, cells lose their nuclei and, consequently, their transcriptomic activity, emphasizing the significance of post-transcriptional gene expression regulation in lens biology. This introduction will delve into the role of post-transcriptional regulations in lens formation, focusing on key players such as micro-RNAs, non-coding RNAs, and RNA binding proteins.

Our team has previously demonstrated the critical role of the RNA-binding protein CELF1 in ensuring proper lens development. I will describe CELF1's molecular functions in regulating the expression levels and splice profiles of its targeted RNAs. While CELF1 was initially identified in myotonic dystrophy type 1 disease, it plays a crucial role in the development and/or activity of several organs.

Lens pathology studies often employ various animal models. However, recent advancements have led to the development of numerous *in vitro* 3D organoids, specifically for lens research. I will review the advantages and limitations of these different models, aimed at providing a comprehensive understanding of lens pathologies with the potential for future therapeutic interventions.

A. Ocular lens organization and development

A.1. Anatomy and functional regionalization of the lens

A.1.a: 0verview of lens organization

The ocular lens is a unique organ that is essential for the visual system. Its unique properties are necessary for a clear vision. The lens is localized in front of the eye, just behind the iris and the cornea. It is attached to the eye throught zonule fibers and the ciliary muscles (Figure 1, left). The lens is surrounded by a basement membrane, the lens capsule. It is composed of two distinct cell types, a monolayer of lens epithelial cells (LEC) and fiber cells (FC). The LEC are outward facing, at the anterior surface of the lens. They can proliferate and differentiate into FC, the main cell type that forms the bulk of the lens (Figure 1, right). FC are elongated cells, unable to divide. They pack tightly and leave no inter-cellular space. During their differentiation, the FC lose their nuclei and organelles and they synthesize massive amounts of specific proteins named crystallins. This create the OFZ (organelle free zone) in the center of the lens. Those changes support lens transparency.

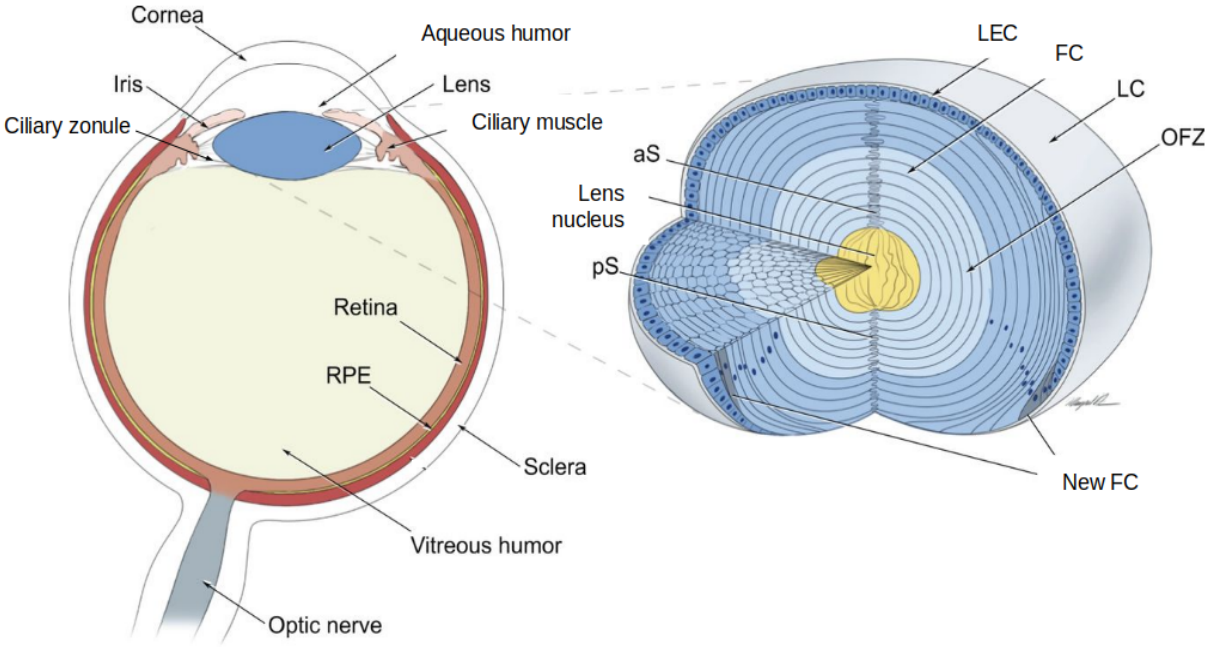


Figure 1: Eye and lens structure. *The lens is positioned between the aqueous and vitreous humors of the eye. Its purpose is to focus the light onto the retina. Toward this goal, the bulk of the lens is composed of transparent fiber cells that are tightly packed. The primary fiber cells form the lens nucleus (in yellow). New secondary fiber cells differentiate from the epithelial monolayer at the equator of the lens. RPE: Retinal Pigment Epithelium; LEC: Lens Epithelial Cell; FC: Fiber Cell; LC: Lens Capsule; OFZ: Organelle Free Zone; aS: anterior Suture; pS posterior Suture. Adapted from Cvekl et al. 2014* [1]

Lens transparency, combined with its spherical shape, allows light to be focused onto the retina. The lens can also modify its optical properties, allowing the eye to change the focus point depending on the distance to the observed object. This adaptability is called visual accommodation. In primates, including human, the accommodation is managed through lens deformation. When the eye looks at a distant object, the unaccommodated lens is flattened by the tension of the zonules fiber. To change the focus point to look at a closer object, the ciliary muscles contract, which releases the tension of the zonule fibers on the lens. With its natural elasticity, the lens maintains a spherical shape. This modifies the optical properties of the lens and allows the eye to have a closer focus point[2,3]. It is worth noting that some nocturnal non-primate species like rats and mice have a reduced need to accommodate their vision. Their ciliary muscles are smaller, consistent with a reduced accommodation capacity[4,5].

A.1.b. The epithelium and the capsule

The lens epithelial cells (LEC) are present in the anterior peripheral part of the lens. In the mouse lens, labeling proliferating epithelial cells allowed the identification of three distinct regions (Figure 2), from the posterior to the anterior region: (i), the transition zone where the epithelial cells initiate their differentiation into fiber cells, and where no cell proliferate anymore; (ii), the germinative zone, the only site where cells divide; (iii), the pre-germinative zone, the most anterior region where the proliferating activity is almost null[6].

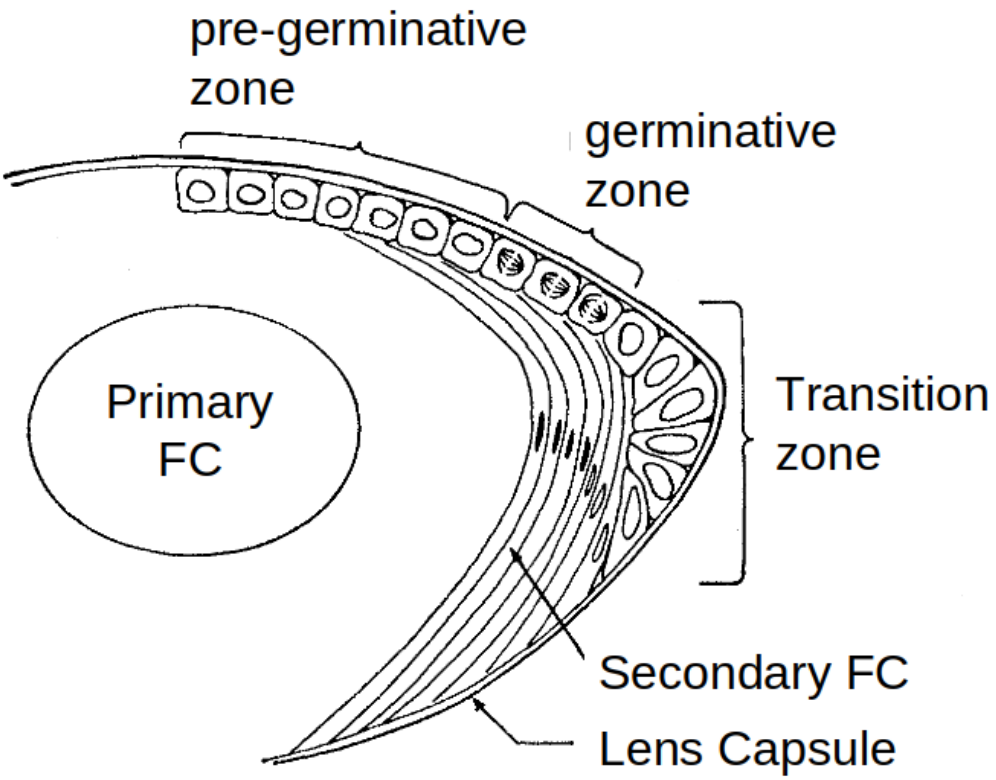


Figure 2: The three distinct regions of the lens epithelium: (i), the transition zone at the lens equatorial region where the cells initiate their internalization and their differentiation into secondary fiber cell; (ii) the germinative region where the cells proliferate; (iii) the pre-germinative zone at the anterior pole of the lens with almost no proliferation. FC: Fiber Cell. Adapted from Papaconstantinou et al. 1967 [7]

Pioneering work by Coulombre's team in the 60s revealed that the properties of the LEC are modulated by different regulatory factors originating from the vitreous humor. They showed that, in E5 chicken embryos, the proper polarization of the lens is dependent on the orientation of the lens in the eye. This suggested a communication between the lens and the rest of the eye, probably through the vitreous humor[8,9]. The vitreous humor is the gel-like medium that is present between the retina and the lens. Multiple regulators are

present in the vitreous humor to allow proper lens polarization and homeostasis. The main factor is FGF (fibroblast growth factor). FGF can cause LEC proliferation, migration or differentiation depending on its concentration. FGF is secreted by the retina. Consequently, FGF concentration in the vitreous humor follows a gradient, with a higher concentration in the posterior than in the anterior region. Therefore, the LEC at the transition zone are submitted to a higher FGF concentration than LEC in more anterior regions. High FGF drives LEC differentiation into FC. The LEC in the germinative zone are exposed to a smaller amount of FGF, which induces their proliferation and migration to the transition zone. Finally, the LEC in the pre-germinative zone have a low proliferation rate due to a low FGF concentration[10].

In accordance with the responsiveness of LEC to FGF signaling, RNA-seq of dissected LEC and FC revealed a high expression of FGF receptors in both cell types. However, these data did not reveal significant difference between LEC and FC in terms of FGF receptors expression. Conversely, other signaling pathways are apparently more active in the LEC than in the FC. For example, multiple Wnt, Notch and TGFβ receptors have been detected in the LEC[11,12]. Other regulators than FGF are present in the vitreous humor, potentially to regulate LEC proliferation, migration and differentiation.

The lens is enclosed by a basal membrane, the lens capsule. The capsule is required for lens development. It consists of an extracellular matrix containing laminin, heparan sulfate proteoglycans and a network of collagen IV. The capsule is used as an anchor point for the lens cells. LEC can bind to Collagen IV and laminin through membranes integrins[13]. Furthermore, the lens capsule can also serve as a sponge for growth factors. As discussed above, multiple growth factors like FGF come from the vitreous humor to control LEC fate. The FGF co-localizes with the heparan sulfate proteoglycans in the lens capsule[14]. This interaction is necessary for presentation of the FGF to the lens cell receptor[15]. Thus, the lens capsule contributes to FGF gradient formation and allows the activation of the FGF pathway which is critical for lens development (see A.3.b).

A.1.c. The fiber cells

The bulk of the lens is composed of fiber cells (FC), which are elongated organelle-free cells. There are two types of FC in the lens, the primary and the secondary FC. The primary FC are present in the core of the lens (Figure 2). They correspond to the differentiated posterior LEC originating from the lens vesicle during the first steps of lens embryogenesis. As they are still present in adults, they are among the oldest cells within adult organisms. They form the lens nucleus present at the center of the lens. The secondary fiber cells are more peripheral (Figure 2). They correspond to the LEC differentiated at the

transition zone during later life stages. The production of secondary FC occurs throughout the entire life span. They correspond to the lens cortex that surrounds the lens nucleus.

FC differentiation is a critical process for lens development. The cells undergo multiple transformations. To allow the light to pass through the lens without obstacle, the FC degrade their nuclei and organelles. This creates an organelle free zone in the center of the lens (OFZ, see Figure 1). Troubles during the differentiation of the FC can result in the accumulation of DNA and/or organelle debris in the lens. This can lead to the opacification of the lens. Even small amounts of material can result in lens pathology. In mice, the accumulation of DNA and organelle debris in the lens is correlated with age related cataract (ARC)[16]. In mice inactivated for *Dnase2b*, the gene encoding the deoxyribonuclease in charge of degrading DNA after karyolysis during FC maturation, defective nuclei degradation causes a congenital cataract[17]. DNASE2B co-localizes with lysosomal proteins such as LAMP-1, suggesting that lysosomes are involved in nuclei degradation[18].

Autophagy is a process that occurs within most cell types, to degrade defective mitochondria and other organelles. During this process, a double membrane structure named autophagosome encapsulates the mitochondria and organelles in the cytoplasm. The autophagosome fuses with a lysosome, leading to the degradation of the mitochondria and organelles[19]. In human lens, autophagy-related genes are enriched in the FC[20]. One such autophagy-related gene is *FYCO1,* which can cause human congenital cataract when mutated[21]. The impact of *FYCO1* deficiency in autophagy and cataract was validated in a mouse model, and experiments in cultured cells showed that a mutation on *Fyco1* reduces the autophagic flux of the cell[22]. Another interesting autophagy-related gene is *Bnip3l.* It is associated with autophagosome-mediated mitochondria degradation in erythrocytes[23]. It is also highly expressed in FC, and mouse FC disrupted for *Bnip3l* retain their mitochondria and organelles (Endoplasmic reticulum and Golgi apparatus). Interestingly, in the mouse *Bnip3l KO* model, nuclei are still degraded. This suggests that nuclei and organelle degradation in the lens are regulated by different processes[24]. In accordance with these genetic data, electron microscopy in human and chicken lens confirmed the presence of mitochondria enclosed in autophagosomes in cortical FC[25]. In addition, Morishita and colleagues have described an autophagy-independent process of organelle degradation. They showed the involvement of PLAAT family member genes, which encode phospholipases, for organelle degradation. Zebrafish Plaat1 and mouse PLAAT3 are recruited by organelles presenting membrane damages. The interaction of PLAAT proteins with the organelle leads to their membranes rupture, and thus to the degradation of the organelle. The deficiency of PLAAT family member genes in model animals causes lens opacification, indicating that this autophagosome-

independent pathway of organelle degradation is required during FC differentiation [26]. Those different studies show the involvement of multiple pathways for nuclei and organelles degradation. All those pathway are necessary for the complete transparency of the lens.

During FC differentiation, the cells elongate in the antero-posterior direction. Following this elongation, cells are aligned in the direction of the major light pathway, implying that the light passes through a limited number of cell membranes (Figure 3). The processes regulating this elongation are still unknown. Multiple hypotheses have been proposed to explain the molecular mechanisms that support FC elongation[27]. Microtubule polymerization at the apical tip of the FC was believed to play a role in the elongation of those cells[28] but the inhibition of microtubule assembly does not prevent the elongation of FC in cell cultures [29]. This suggests that, even if microtubule formation in the lens plays a role in lens development, it is not the main factor for FC elongation. The FC are rich in F-actin, a cytoskeleton protein[30,31]. They seem to play a critical role for FC elongation. Indeed, in cell cultures, the pharmacological inhibition of F-actin formation prevents cultured cells elongation[32]. The F-actin is not the only cytoskeleton member responsible for FC elongation. The F-actin network is known to interact in the FC with other cytoskeleton proteins. Some of them are present in most cell types including FC (ANK2, SPTBN1) whereas other are virtually FC-specific (BFSP1, BFSP2)[31,33]. Indeed immunofluorescence experiments on lenses have revealed that specific intermediate filament proteins on the FC cytoskeleton change their cellular localization during lens development. An accumulation of vimentin, phakosin and filensin is observed during development, suggesting that they could play a role in FC elongation. When the adult stage is reached, all those proteins end up in the plasma membrane, leaving the center of the fiber cell clear[34]. This shows the key role of the cytoskeleton for proper FC elongation. Other types of proteins are also necessary for FC elongation. CDH2 (N-cadherin) is a protein involved in cell-to-cell junction, between LEC and cortical FC at the place where FC initiate their differentiation. The genetic inactivation of *Cdh2* leads to abnormal lens development with aberrant FC elongation[35]. More studies should investigate the cytoskeleton modifications involved in FC elongation to understand the interactions between the different cytoskeleton proteins.

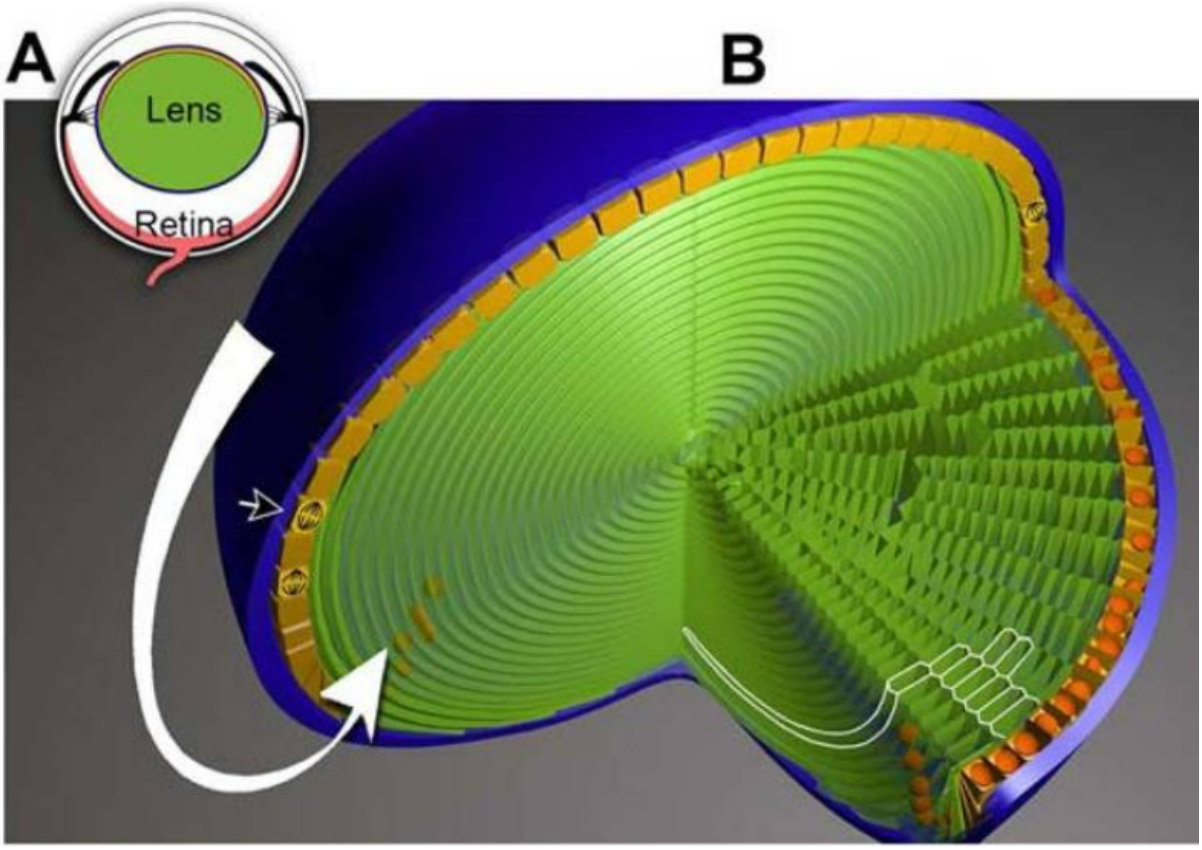


Figure 3: Fiber cell organization in the lens. *A, Orientation of the lens in the eye with the epithelial monolayer (in orange) at the anterior of the lens. B, Cellular organization in the lens with in orange the lens epithelial cells that proliferate at the germinative zone (black arrow) and then migrate to the transition zone (white arrow) where they differentiate into fiber cells (in green). During the differentiation they pack themselves with their hexagonal shape to create the best possible light path. Adapted from Bassnett and Sikic 2017* [36]

Above, we have reviewed two mechanisms that support lens transparency: the degradation of nuclei and organelles that prevent light scattering by autophagy-dependent and independent mechanisms, and the orientation of the FC along the light pathway. In addition, the gaps between adjacent cells are minimized as much as possible, as they would otherwise be filled with extracellular matrix likely to reduce lens transparency. Toward this goal, the cells form an hexagonal lattice (Figure 3B). Proteomic assays on mouse FC membranes were run to reveal the proteome changes that occur during FC differentiation[31]. The major FC membrane protein is AQP0 (Aquaporine-0, also named major intrinsic protein or MIP). AQP0 acts as a water pore and is greatly involved in cell to cell junctions in the lens[37]. The other main membrane proteins are connexins (GJA8, GJA3 and GJA1). The connexins allow inter-cellular communications and stabilize cell to cell junctions. Especially, GJA8 is essential for the formation of ball-and-socket structures[38], which are inter-digitation structures at the FC membrane necessary for proper FC packing[38–40]. In human, numerous mutations on *AQP0*, *GJA8*, *GJA3* and *GJA1*

have been associated to lens development issues resulting in congenital cataract[37,38,41–45].

Simultaneously to those changes, the FC express a high quantity of crystallins. Crystallins represent the vast majority of transcripts and proteins in the FC [30,33]. The majority of the crystallins belong to the families alpha or beta-gamma. The crystallins are water-soluble proteins that allow the lens to acquire a correct refractive index[46]. More specifically, members of the crystallin family alpha (CRYAA and CRYAB) play a key role to protect the lens against stress throughout life. When a protein in the lens is denatured, e.g. following an oxydative stress or an exposure to UV light, it can form insoluble aggregates. Crystallin alpha have a chaperone activity that highly reduces the formation of those aggregates. As the aggregates would otherwise scatter light, crystallin failure leads to early-onset cataract [47,48]. The crystallins beta-gamma correspond to the same super-family, but crystallins beta form oligomeres complex whereas crystallins gamma are only present in monomers. In total the crystalline beta/gamma family is composed of 14 genes in human. The role of those proteins is still unclear. It has been observed that they can bind Ca2+[49], which can impact the formation of cataract[50,51]. Nevertheless, those proteins have a key role for lens transparency, since multiple mutations on those genes causing cataract have been described[41,52].

A.2. Lens embryology

In mice, key steps of lens development occur between embryonic day 9 (E9) and embryonic day 19 (E19) (Figure 4).

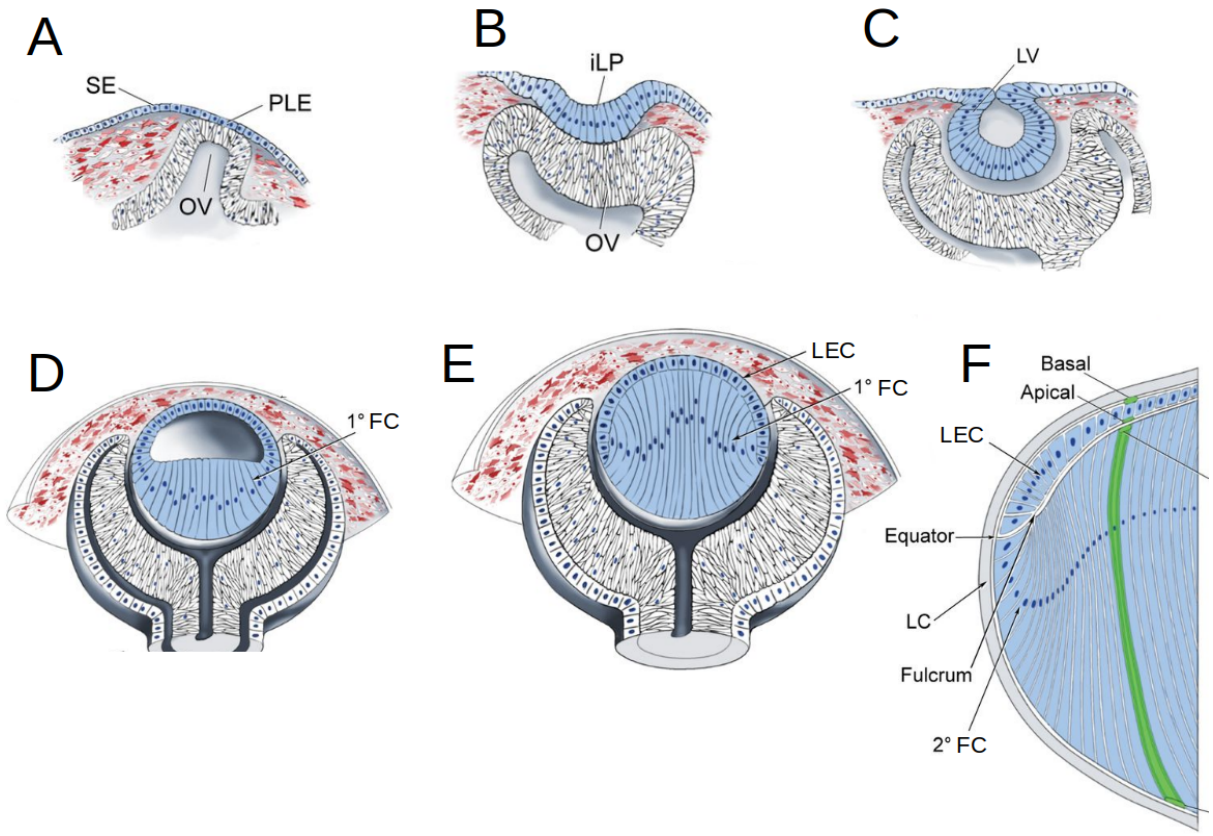


Figure 4: Mouse lens embryology. *Schematic showing the lens development at E9 (A), E10 (B), E11 (C), E13 (D), E15 (E) and post-natal (F). SE: Surface Ectoderm; PLE: Prospective Lens Ectoderm; OV: Optic Vesicle; iLP: invaginating Lens Placode; LV: Lens Vesicle; 1° FC: Primary Fiber Cell; LEC: Lens Epithelial Cell; 2°FC: secondary Fiber Cell; LC:Lens capsule. Adapted from Cvekl et al. 2014* [1]

At E9, the optic vesicle (the presumptive retina) approaches the surface ectoderm (Figure 4A). At E10, the optic vesicle and the ectoderm start their invagination forming the lens placode (Figure 4B). At E11, the lens placode closes on itself, incorporating epithelial cells from the ectoderm. The newly formed spherical structure is called the lens vesicle. It is now detached from the ectoderm (which will form the cornea) (Figure 4C). At E13, the posterior cells of the lens vesicle start their differentiation into primary fiber cells. They elongate to fill the lens vesicle (Figure 4D). At E15, the secondary fiber cells start their differentiation from proliferative epithelial cells in the germinative zone (Figure 4E). At E19, the karyolysis in primary fiber cells is almost complete. The secondary fiber cells continue to be formed [53]. At birth (E21), the secondary fiber cells continue their maturation to leave the center of the lens

that becomes fully transparent (Figure 4F). Comparing the weight of the lens (dry or wet) across multiple species showed that the lens continues to form secondary fiber cells all along life. This is accompanied by an increase in cell compaction and an increase in the dry weight/ size of the lens[54]. In human, the lens continuously grows during the whole life: rapidly during the prenatal life, then much more slowly after birth[55].

A.3. Controls of lens development

A.3.a Transcription factors

The formation of the eye and more specifically the lens, is driven by multiple transcription factors (TF). TF control the precise genetic expression necessary for lens formation and the maintenance and differentiation of the various cell types in the lens. A single TF can regulate the expression of other TFs, leading to the formation of a complex regulatory network. A schematic representation of this regulatory network for the major TF in the lens is given in Figure 5.

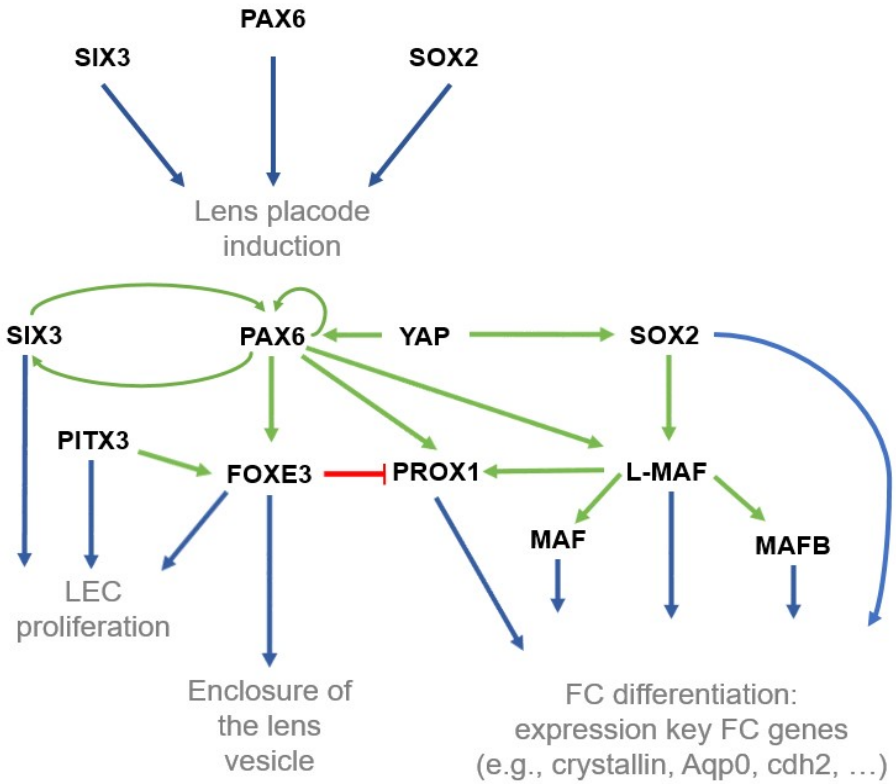


Figure. 5: Regulation network between the transcription factors during lens development and its homeostasis. *See references in the text*

The induction of eye formation is controlled by the TF PAX6. Pioneering studies in Drosophila showed that the ectopic expression of *Pax-6* is sufficient

to induce the development of ectopic eyes[56]. Conversely, in mice and humans, loss of PAX6 lead to the absence of eyes[57,58]. In addition to its key role in eye formation, *PAX6* is highly expressed in the nuclei of LEC during embryonic development[59,60]. To investigate the functions of PAX6 in lens development, a mouse model was developed where *Pax6* is conditionally inactivated at E10, after PAX6-mediated eye induction. This showed that *Pax6* expression in the lens placode is required for lens development. In absence of PAX6, the FC have differentiation failures[61]. As a TF, PAX6 regulates the transcription of other TF such as *Prox1*, *L-maf, Six3* and also *Pax6* itselff[62–64]. This probably puts PAX6 at the top of the gene regulatory network that controls lens formation.

Among the targets of PAX6, SIX3 is another major TF in lens development. In association with PAX6, it controls the expression of multiples crystallins, the major components of lens fiber cells[65,66]. The inactivation of *Six3* in mouse models prevents the formation of the eye and thus also of the lens[67,68]. Chromatin immunoprecipitations and luciferase reporter assays showed that, in mice, SIX3 controls the expression of *Pax6* in a feedback loop, and also of *Sox2*[67,69]. Like PAX6, the ectopic induction of *Six3* promotes the formation of an ectopic eye in medaka embryonic fish[70]. In the lens, the main role of SIX3 seems to be the activation of *PAX6* and *SOX2* to initiate eye formation. However, in non-lens organs such as the brain, multiple other genes are controlled by SIX3, for example a member of the Wnt pathway[67]. This suggests that *SIX3* may directly regulate other key genes for lens development.

Similarly to *PAX6* and *SIX3,* the deficiency of *SOX2* during eye formation leads to developmental issues. In mice, the conditional KO of *Sox2* causes the absence of the lens in the eye[71]. In human, *SOX2* mutations can cause anophtalmia[72] or coloboma, a rare eye developmental pathology impacting different tissues in the eye including the lens[73]. In addition a genome-wide association study in human associated *SOX2* to age-related cataract[74]. The genes directly regulated by SOX2 are still not fully known. In chick embryos, SOX2 induces the expression of crystallins[75]. Multiple studies have shown that PAX6 and SOX2 interact to initiate the transcription of their target genes[75–77]. So we can hypothesize that SOX2 shares a number of targets with PAX6. Interestingly, it has been recently observed that SOX2 can bind to RNA in vivo and in mouse embryonic stem cells. This opens the hypothesis that SOX2 can exert post-transcriptional controls on a subset of mRNA[78].

While the transcription factors PAX6, SIX3 and SOX2 are globally instrumental for eye development as a whole, the MAF family of TF play more lens-specific roles in the eye. The MAF family of bZip trancription factors include three members in the lens, *L-MAF*, *MAFB* and *MAF* (also named *C-MAF*). The deletion of either of these genes, or mutations, cause cataract[41,79]. In chick, the expression of *L-MAF* is controlled by PAX6 and SOX2 after the formation of

the lens placode[80]. Then, L-MAF induces the expression of *MAFB* and *MAF*[81]. All the MAF transcription factors control the expression of key lens genes such as the crystallins[79,82,83], the TF *Prox1*[80] or specific membrane proteins like connexins or AQP0[81,84].

FOXE3 is a TF abundantly present during the initial stages of mouse lens embryonic development[30]. FOXE3 remains present in the developed lens but is only localized in the LEC[11,12,18]. Mutations of *FOXE3* cause anophtalmia[85]. The expression of *FOXE3* is controlled by the TF PITX3[86]. FOXE3 seems to mostly play a role for the proliferation and maintenance of the LEC, as it has been show in *in cellulo* studies[87]. During the first steps of eye formation, FOXE3 participates to the correct enclosure of the lens vesicle by promoting the proliferation of the LEC[88–90]. Indeed, a mouse model was generated where *Foxe3* is KO. These mice present major lens defects with microphatalmia and congenital cataract. At E14.5 a connection between the cornea and the lens epithelium persists, revealing a defect enclosure of the lens vesicle. Furthermore, the *Foxe3* KO lens also presents FC differentiation defects associated with a defective regulation of FC genes like *Prox1* or gamma-crystallin*[90].* The regulation of these genes was also investigated in an other mouse model expressing *Foxe3* ectopically under the control of an alpha crystallin promoter. Those mice have cataracts associated with the mis-regulation of *Prox1* and other FC key genes [91]. Hence, a high expression of *Foxe3* in the FC is detrimental for lens development. Accordingly, *Foxe3* expression is specific to the LEC[11,12,18]. There, it prevents the differentiation into FC of the LEC that are not at the transition zone through, at least partially, the inhibition of the FC specific TF *Prox1*. RNA-seq on zebrafish lens mutant for *Foxe3* revealed thousands of mis-regulated genes[92]. This shows that *Foxe3* mis-regulation causes massive transcriptomic changes. Even if most of those genes are probably indirectly controlled by FOXE3, FOXE3 may be involved in other lens development processes.

PITX3, a TF, can cause microphthalmia if mutated in mouse lens. The absence of PITX3 transcriptional regulation prevents FC differentiation, and causes mis-expression of the key lens genes *Prox1, Foxe3, Aqp0* and multiple crystallins[93]. *Aqp0* [94] and *Foxe3* [86] have been identified as direct targets of PITX3. In human, numerous mutations on *PITX3* have been linked to congenital cataract[41,95,96]. ISH showed that *Pitx3* expression in the mouse lens starts at the lens placode stage[97]. Then, IF showed the localization of PITX3 in the LEC[98].

PROX1 is a TF localized in the cytoplasm in the lens placode, and then in the LEC. It is relocalized in the nuclei of differentiating FC, before karyolysis[59,99].

Prox1 inactivation in mouse lens causes FC differentiation and elongation defects, associated with the mis-expression of FC specific genes like *Cdh2*, *Aqp0*, and some crystallins and cell-cycle inhibitors[100,101]. RNA-seq of *Prox1*-deficient lenses highlighted numerous mis-expressed genes, revealing that PROX1 is a central TF for FC differentiation[101,102].

Finally, YAP encodes a downstream transcription factor of the Hippo signaling pathway. In mice, *Yap* deficiency causes early-onset cataract[103–105]. *Yap* KO lenses present several LEC defects, including reduced proliferation, increased senescence [104,105], abnormal cell shape, and premature FC differentiation[106]. Furthermore, RNA-seq analysis of *Yap*-deficient lenses identified numerous mis-regulated genes including those encoding the TF PAX6 and SOX2, and diverse crystallins[105]. Cell cultures showed that *YAP* expression is induced by FGF signaling. However, at high FGF concentrations, YAP is localized in the cytoplasm, making it inactive as a transcription factor. In the lens, FGF is distributed following a decreasing posterio-anterior gradient arising from the vitreous humor and the retina. Hence, YAP is nuclear at the germinative zone, but cytoplasmic at the transition zone, which is located in a more posterior position and where FGF concentration is higher[103,104]. In the germinative zone where YAP is active, it induces the transcription of genes necessary for LEC proliferation and prevents the cells to initiate their differentiation into FC prematurely. In the transition zone where YAP is exported outside the nucleus, the inhibition of YAP activity allows the cells to initiate their differentiation into FC. How FGF and Hippo signaling pathways interact to control YAP subcellular location in not known [103].

A.3.b Signaling pathways

Alongside the TFs, the lens development is also influenced by multiple signaling pathways that play crucial roles in various processes. Table 1 provides a summary of the functions of these diverse signaling pathways in lens development.

Table 1: List of the signaling pathways involved in lens development and homeostasis

Signaling pathway	Functions in the lens
FGF (Fibroblast Growth Factor)	Lens placode formation, lens polarization, LEC proliferation, primary and secondary FC differentiation
Notch/JAG1	LEC proliferation ,Secondary FC differentiation
Transforming Growth Factor β (TGFβ)	Represses lens placode formation, LEC proliferation, FC differentiation?
BMP (Bone morphogenetic protein)	Lens placode formation, Primary FC differentiation
Eph/Ephrin	FC differentiation (regulation of the actin network)
Hippo	LEC proliferation

The vitreous humor influences lens development[9]. As seen above, one of the factors present in the vitreous humor that could explain this influence is FGF. FGF is secreted by the retina and diffuses in the vitreous humor. This creates a FGF gradient through the lens. The FGF gradient regulates LEC differentiation into FC, or their proliferation[10]. The importance of FGF signaling pathway for fiber cells differentiation is demonstrated by the phenotype of mice inactivated for the FGF receptors (FGFR1, FGFR2, and FGFR3). They have severe microphthalmia, and the primary fiber cells are unable to start their differentiation[107]. In LEC cultures, FGF induces the phosphorylation of ERK1/2 in a dose-dependent manner. ERK1/2 are MAP kinases that in other cell types control different processes like cell proliferation or cell differentiation[108]. In LEC, FGF induction causes cell proliferation or cell differentiation depending on the FGF dose, but the effect of FGF on proliferation and differentiation is abolished in the presence of an ERK1/2 phosphorylation inhibitor. This indicates that FGF controls LEC proliferation and differentiation through ERK and a MAP kinase cascade. However, the expression of beta-crystallins is induced by FGF even at doses of ERK1/2 phosphorylation inhibitors that abolish FGF-induced morphological changes. This indicates that FGF modulates FC differentiation via MAP kinases-independent pathways[109]. Notably, FGF induces the expression and cellular localization of the transcription factor YAP, and YAP contributes to regulating the cellular fate of the LEC. YAP expression is regulated by ERK1/2[103].

Whether or not the FGF-dependent, ERK1/2-independent, pathway that induces beta-crystallin expression is mediated by YAP is still unclear.

FGF also plays a role during placode formation. The deletion of two FGF receptors that are expressed during placode formation (FGFR1/2) causes lens morphological defects, correlated with a reduced expression of lens key genes like *Sox3*, *Foxe3* and alphaA crystallins. It does not change PAX6 levels. This shows that it is not the only factor controlling lens development[110]. For example, in LEC cultures, FGF induces LEC proliferation, morphological changes and the expression of specific markers (CDH2, c-MAF, PROX1 and alpha-crystallin), but the TGFβ/Smad has to be inhibited to maintain those LEC characteristics [111]. FGF also promotes the expression of Wnt and Notch pathway genes[112,113].

The Notch signaling pathway also plays a key role for lens formation. Mice conditionally inactivated for *Notch2*, the Notch receptor expressed in the lens, show microphthalmia and cataract. DNA microarray hybridization showed the global mis-expression of key lens genes in *Notch2* cKO lenses[114]. Additionally, the inactivation of *Rbp-j*, a critical component of the Notch pathway, in mice leads to a similar lens defect[115]. JAG1 is the major Notch membrane ligand in lens. In ISH and IF, it is present in the mouse lens during early development. Its conditional inactivation leads to the absence of lens once the mouse reaches adult stages. *Jag1* deficiency affects secondary fiber cells development[116]. *Jag1* is expressed in the LEC in the fulcrum, the region between the FC and the LEC at the transition zone and in the germinative region. This difference is regulated by the FGF gradient on the lens. Only LEC at the transition zone reach the FGF stimulus threshold necessary to express *Jag1*[113]. The bidirectional signaling of the Notch/JAG1 pathway is instrumental in regulating the dramatic transcriptomic changes that occur when initially proliferative cells internalize and differentiate into FC (Figure 6)[113]. However, the absence of JAG1 in the LEC within the germinative region caused a unidirectional signaling from the FC to the LEC (Figure 6). This unidirectional Notch/Jag1 pathway promoting the proliferation of these cells and thereby sustain the LEC population in the lens[115]. Among other, the Notch pathway regulates the expression of *Cdh2*, *Foxe3* and cell-cycle proteins[113,117].

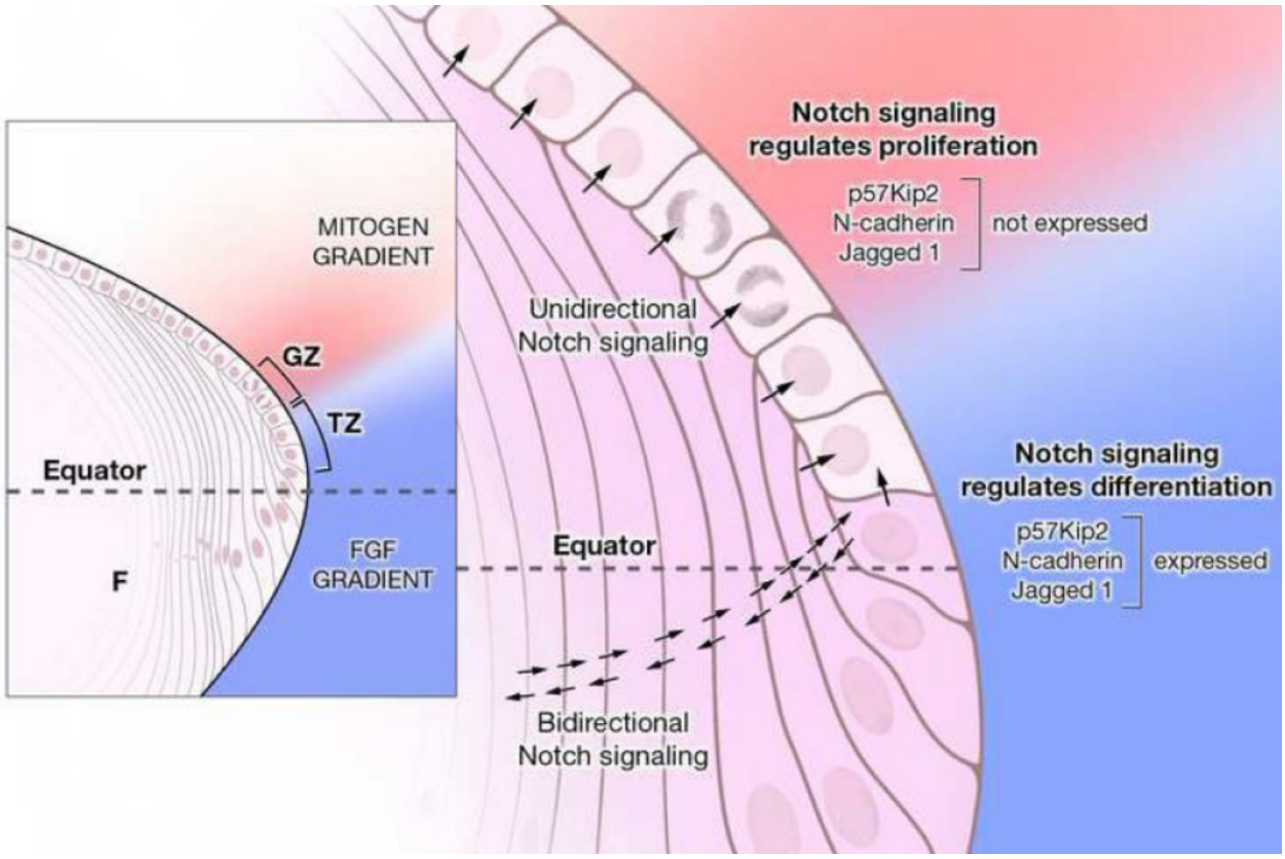


Figure 6: The multiple roles of the Notch pathway. *In the germinative zone (GZ) of the lens, the lens epithelial cells (LEC) are exposed to a low concentration of FGF, which prevent the expression of Jag1. The Notch pathway maintain these cells in an epithelial state and promote their proliferation. Whereas, the LEC at equator of the lens enter in the transition zone (TZ), where FGF levels are higher and can promote Jag1 expression. In this region, the Notch/Jag1 pathway promote the differentiation of these LEC into fiber cells, while simultaneously suppressing their proliferation through the expression of P57kip2, a cell cycle inhibitor. Figure from Saravanamuthu et al 2009* [113]

The Transforming Growth Factor β (TGFβ) superfamily plays a crucial role in regulating lens development. Mutations in certain members of this superfamily, such as *Tgfb1*[118] and *Tgfb3*[119] have been linked to human cataract. Similarly, the absence of *Tgfb2* has been found to lead to significant ocular defects in a mouse model[120]. Interestingly, during the induction of lens placode, the TGFβ/SMAD3 pathway needs to be deactivated to ensure the proper expression of *Pax6* and thus the normal formation of the lens placode. Furthermore, it's noteworthy that TGFβ2 can induce the proliferation, migration and trigger the epithelial-to-mesenchymal transition (EMT) of the cultured LEC *in cellulo*[121–123]. This suggests that TGFβ might have a role in maintaining the LEC population and potentially influencing the differentiation of FC.

Within the TGFβ superfamily, the Bone Morphogenetic Proteins (BMP) also play a significant role in lens development. The BMP signaling pathway is required for the proper formation of the lens placode. Specifically BMP4 is

necessary for the invagination of the lens placode which will form the lens. BMP4 mouse mutants have no eye even if an optic cup is still present. BMP4 deficiency does not affect PAX6 and SOX3 levels in the lens placode, but a reduced expression of *Sox2* is observed [124]. Other members of the BMP pathway are involved in lens formation. For example, *Bmp7* KO mice present a similar defect of the lens placode preventing lens formation[125]. Similarly, the absence of two BMP receptors (*Bmpr1a* and *Acvr1)* prevents lens formation. In more details, the BMP pathway regulates cell proliferation and the survival of the placode epithelial cells needed for the invagination of the lens placode. The BMP pathway controls the levels of the key TFs SOX2 and FOXE3, but also of the alpha A crystallin[126]. Beyond these early roles in eye formation, the BMP pathway also regulates lens development and specifically FC differentiation. After the closure of the lens vesicle, BMP4 is present in the retina where it diffuses to the lens through the vitreous humor. The use of a BMP ligand inhibitor NOGGIN cause a reduced the development of the primary FC which lead to a smaller lens. The mutation of a BMP receptor (*Bmpr1b*) prevents a normal primary FC differentiation[127]. It is worth noting that in human the polymorphism BMP4-V152A was linked to congenital cataract[128].

The Eph/Ephrin are membrane receptors that can be associated with other receptors like the FGF receptor[129]. Multiple human mutations on *EPHA2* have been associated to age related or congenital cataract[130–132]. This suggests an important role of the Eph/Ephrin pathway for lens development. The Eph/Ephrin pathway controls FC shape and cell-cell interactions. The mutation of Ephrin-A5 in mouse lens causes cataract. The FC present a defective packing associated with mis-localized CDH2[133,134]. Mouse models deficient in *Epha2* show fulcrum defects with mis-localized CDH2 and F-actin network. FC packing is also disorganized[135].

The Hippo pathway promotes the reduction of cell proliferation and is used to control organ size. In mammals, the Hippo pathway works as a kinase cascade, starting by MST1/2 and ending by the phosphorylation of the transcription factor YAP, which inhibits its transcriptional activity[136]. In lens, the Hippo pathway is activated by the high FGF concentration in the transition zone[103]. The Hippo pathway inhibits YAP activity through its phosphorylation. The phosphorylated YAP cannot enter the nucleus, preventing the expression of LEC specific gene and allows FC differentiation[103,106].

A.3.c Genes identified as lens-enriched

Lachke's team has developed a computational tool, iSyTE (integrated Systems Tool for Eye gene discovery; http://bioinformatics.udel.edu/Research/iSyTE), to facilitate the identification of key genes for lens development. They used transcriptomic microarray data from mouse embryonic lens (E10.5, E11.5 and E12.5) and the rest of the whole body of embryonic mouse. With the use of an in silico substraction method (gene expression in the lens / gene expression in the whole body), they calculated a lens enrichment score[137]. Their hypothesis was that if a gene is more expressed in the lens than in the rest of the body (enriched), this gene is more likely to be critical for lens development and/or homeostasis. Thus, this gene could be a potential target for new studies. Obviously, some genes can be vital not only for the lens, but also for other organs. Such genes may present a low enrichment score, even if they could be interesting for lens studies. With this limitation in mind, this tool is meant to be used as a prioritization tool. iSyTE was then improved to iSyTE 2.0 with transcriptomic data from 9 time points (E10.5 to P56)[30]. The most enriched genes encode without surprise the different crystallins and some membrane proteins like AQP0, some connexins (GJA3 and GJA8) and different spectrins (SPTB, SPTBN2). But this tool also successfully identified genes previously unsuspected to be involved in lens development, like those encoding the RNA-binding proteins TDRD7[138] ,CAPRIN2[139], and CELF1[140].

Symbol	Rank	Dev E10.5 affy430	Dev E11.5 affy430	Dev E12.5 affy430	Dev E16.5 affy430	Dev E17.5 affy430	Dev E19.5 affy430	Dev P0 affy430	Dev P2 affy430	Dev P56 affy430
Foxe3	-	19.85	33.31	20.32	7.11	7.78	2.49	6.27	7.77	3.99
Maf	-	13.32	12.9	17.05	18.83	11.1	15.74	11.84	9.59	9.71
Mafb	-	3.44	5.98	5.51	4.31	3.89	4.44	5.17	3.2	1.3
Pax6	-	24.56	19.7	14.42	6.32	8.71	15.24	8.11	13.13	10.19
Pitx3	-	14.71	18.56	16.07	11.75	10.7	5.91	11.1	7.01	6.03
Prox1	-	18.25	37.09	37.84	29.26	28.9	62.29	36.63	18.27	27.08
Six3	-	20.27	15.34	13.33	14.11	10.09	16.13	9.82	5.8	6.74
Sox2	-	1.67	-1.88	-3.85	-2.35	-4.08	-2.91	-2.82	-2.33	1.21
Yap1	-	1.13	-1.04	-1.18	-1.72	-1.47	-1.47	-1.72	-1.86	-2.36

Figure 7: Enrichment scores during mouse lens development of genes encoding some transcription factors. *The enrichment score corresponds to the log2-ratio of the hybridization signal in the lens at a precise developmental stage to the hybridization signal in the whole embryonic body. Data from iSyTE2* [30]

As examples, Figure 7 shows the enrichment scores of a number of genes encoding transcription factors and involved in lens development. All those TF are enriched in the lens compared to the whole body at E10.5, during lens placode invagination. PROX1, known to regulate FC differentiation[100], is mostly enriched at E19.5 when the primary FC finish their differentiation and the secondary FC start to differentiate. The expression of *Maf* and *Mafb* is stable throughout lens development (no probe has been designed for L-maf). The other TF are mostly enriched at an earlier development stage, before primary FC differentiation. This is in accordance with the studies highlighting the role of some of the TF for lens induction and the first lens developmental steps.

A.4. Lens diseases.

A.4.a. Genetic cataracts

Lens opacification, also called cataract, is mostly age-related and is essentially due to the UV and oxydative stress all along life. But more rarely some cases of congenital cataract can occur, with a low prevalence of 4.24 per 10,000 people[141]. Congenital cataract often has genetic causes, although it can also have non-genetic causes, like the exposure of the pregnant mother to some viruses. Age-related cataract (ARC) essentially has non-genetic causes, but genetic variations modulate the susceptibility of the elderly to have cataract. The Cat-Map database lists the known mutations in human and animal models known to cause cataract[41]. More than 400 genes have been associated with congenital and/or ARC. Interestingly, some genes present mutations associated with both ARC and congenital cataract, for example *GJA8*[142,143]. This suggests that most "cataract genes" could lead to either congenital cataract or ARC, depending on the type of mutation. Most of the listed genes are known for their critical roles in lens development and homeostasis, like crystallin genes, connexins (*GJA3*, *GJA8*), and cytoskeleton genes (*VIM*, *BFSP1* and *BFSP2*). Numerous mutations in the autophagy related gene *FYCO1* result in the formation of cataract. Some collagen genes necessary for lens capsule formation and maintenance are also present such as *COL1A1*, *COL4A1*, *COL4A5*, among others. Listed RNA-binding proteins include *TDRD7*, *CELF1*, *CAPRIN2*, and listed transcription factors include *SOX2*, *PAX6*, *PROX1*, *PITX3*, *c-MAF*, *FOXE3*, *YAP.*

A.4.b. Non-genetic congenital cataracts

Non-genetic factors can cause congenital cataract. *In utero*, viral infections by varicella[144] or rubella[145,146] can cause multiple developmental disorders including microphthalmia and/or cataract. Due to vaccination those cases have become rare in the last decades. More recently, Zika virus infection, *in utero*, has been shown to cause multiple major developmental issues, including congenital cataract[147,148].

Different factors can also favor the appearance of a cataract. For example diabetic patients present a higher prevalence of cataract[149,150]. The etiology of these cataracts is still unclear, but the main hypothesis stands that diabetic patients present a higher level of glucose in their vitreous humor. This would impact the proper regulation of the LECs at the germinative and transition zones[151,152]. Other types of stress on the lens can increase the appearance of a cataract. Exposure to UV radiations is associated with increased risks of cataract formation[153]. UV light increases the oxidative stress of the lens with the denaturation of the FC proteins. Accumulation of denatured proteins creates aggregates and thus leads to opacification[154].

A.4.c. Other lens pathologies

Beyond cataract, several lens disorders exist, but they are easily covered by corrective glasses. Presbyopia is a highly common vision pathology, developing after the age of 40. Presbyopia is described as the loss of accommodation capacity. This may be caused by the loss of lens and/or capsule elasticity or the loss of ciliary body contractility[155,156].

Myopia is the most common vision pathology. Myopia has genetic and environmental causes. Patients with myopia cannot clearly see distant objects. This is caused by an elongation of the eyes. This implies that the lens cannot focus the light properly on the retina when the lens is flattened to observe distant objects [157].

At the opposite hyperopia correspond to an incapacity to clearly see close objects. Similarly to the myopia, the lens cannot focus on the retina. This can be due to a short eye size or to an optical defect on the cornea and/or the lens[158].

B. Post-transcriptional controls in lens development

We have seen above that transcription factors are at play to control lens development and homeostasis. Beyond transcriptional controls, another layer of gene regulation was more recently discovered to also play a key role in lens biology. Post-transcriptional regulations are the controls of gene expression that are exerted at the RNA level. They include the regulations of alternative splicing, as well as the regulations of RNA stability and translation. They play a crucial role during fiber cell differentiation, especially considering that nuclei are degraded during this differentiation. In the absence of fiber cell nuclei, these cells lose the ability to actively transcribe new mRNA molecules. Therefore, the regulation of gene expression at the post-transcriptional level becomes vital for controlling the differentiation process and to preserve the homeostasis of the lens over the entire lifespan.

B.1. MicroRNA

MicroRNAs (miRs) are small (17-25 nt) non-coding RNAs that regulate gene expression post-transcriptionally. miR can bind to mRNA with a fully or partially complementary sequence. Thereby, they repress their expression by inhibiting translation and/or inducing their degradation. A single miR can have multiple target mRNAs, and a single mRNA can be targeted by multiple miRs. Hence, the miRs create a complex interconnected network of post-transcriptional regulations.

DICER is a ribonuclease essential for miR maturation. It is present in the mouse lens and retina, with notably higher expression levels observed in the embryonic lens compared to the adult lens. Deficiency of *Dicer* in the lens leads to microphthalmia and aphakia, the absence of lens in the eye[159]. Those result demonstrate the key role of post-transcriptional controls globally exerted by the miRs in the lens.

More specifically, multiple miRs necessary for the development and homeostasis of the lens have been identified. Here, we will review the functions of some of them and their involvement in cataract. These data are summarized in Table 2. Comparing, in infants (1-4 years), the expression levels of lens-related miRNAs between lenses from postmortem donors free of ocular diseases and lens from congenital cataract patients, revealed a positive correlation between the overexpression of *miR-182* and lens opacification[160]. The putative causal relationships between the overabundance of *miR-182* and the occurrence of congenital cataracts remain unknown. However, this

suggests that the increased abundance of this miR and the repression of its mRNAs targets can perturb proper lens development and contribute to the appearance of congenital cataract. Conversely, the expression of *miR-182* seams to reduce age-related cataract by protecting the lens epithelial cells against oxidative stress. This effect could be mediated by the repression of *NOX4,* a gene coding for a NADPH oxidase able to generate reactive oxygen species (ROS)[161,162]. The accumulation of oxidative stress over life is know to be an important factor for the development of the age related cataract (ARC)[163].

miR-204 under-expression has been correlated with human congenital cataract[160]. Its necessary role for lens development was confirmed in the medaka fish, where its repression by a morpholino antisense oligonucleotide blocking its maturation leads to major eye defects. *miR-204* targets *Meis2*, which encodes a transcription factor involved in controlling the expression of *Pax6*, a major transcription factor for eye development[164]. *miR-204* also targets *Ankrd13A,* whose repression allows proper cell motility and elongation for lens development[165]. In addition to congenital cataract, *miR-204* down-regulation is also observed in age-related cataract. *miR-204* may be involved in the regulation of the oxidative stress[166]. In diabetic patients, the down-regulation of *miR-204* is correlated with the development of diabetic cataracts[167,168]. The mRNA encoding the main TGFβ receptor (*TGFBR1*) is degraded by a *miR-204*-dependent pathway. TGFβ signalling is a major trigger of the epithelial-mesenchymal transition (EMT). Hence, the down-regulation of *miR-204* probably contributes to increased EMT of lens epithelial cells, resulting in cataract[167]. The under-expression of *miR-204* is also associated with posterior capsule opacification through the regulation of genes including *Meis2*, *Zeb1* and *Smad4* [169–171].

Similarly to *miR-204*, *miR-124* has been found to be under-expressed in a cohort of human congenital cataract lenses[160]. A mouse model was used to show that *Vim*, which encodes the lens cytoskeleton protein Vimentin, is one of the targets regulated by *miR-124* in the lens. Its over-expression when *miR-124* is under-expressed may participate in the formation of congenital cataract[172]. Furthermore, the under-expression of *miR-124* can participate in age-related cataract owing to the over-expression of *Dapk1* and *Bcl2l11* which regulate LEC apoptosis and oxidative stress response[173,174]. However, another article reports that, at the opposite, the over-expression of *miR-124* is associated to age-related cataract. *In cellulo* studies identified *Spry2* and *Mmp-2* as direct targets of *miR-124.* The mis-regulation of those genes perturbs cell viability and promotes apoptosis[175]. These conflicting reports show the importance of a precise level of *miR-124* for the homeostasis of the lens, and so probably also for its development.

miR-184 is the most abundant miR in mouse lens [176,177]. It is also highly expressed in human lens and is found in human exosomes in the aqueous humor[178]. *MiR-184* mutation can cause ocular diseases with multiple symptoms including congenital or pediatric (1-2 year old) cataract. Mutations on miR-184 prevent its correct binding to *INPPL1* (inositol polyphosphate-like 1) and *ITGB4* (integrin, beta4) mRNA. It competes with *miR-205*, a more effective repressor, on these transcripts. Hence, the reduced capacity of mutant *miR-184* to bind to these transcripts results in their accelerated decay[179–183]. Recent studies have also suggested that mutations in *miR-184* may cause a congenital cataract phenotype by perturbing the metabolic pathway of the lens epithelial cells via the mis-regulation of *Aldh5A1* and *Gabra3*[182]. A *miR-184* KO zebrafish model has lens defects such as microphthalmia and cataract. This confirms the essential and evolutionary conserved role of *miR-184* for lens development[184].

To our knowledge, only the four above microRNAs have been associated with congenital cataract: *miR182*, *miR204*, *miR124* and *miR184*. However, other microRNAs are associated with other types of cataract. Age-related cataract (ARC), the most common type of cataract, has been extensively studied, resulting in the identification of a high number of miRs in this context. For instance, the down-regulation of *miR-29a* can lead to the up-regulation of its target *TDG* (thymine DNA glycosidase), which causes oxidative damage repair issues leading to ARC[185]. In the same miRNA family, the down-regulation of miR-29b results in *SMAC* over-expression, which causes ARC by promoting LEC apoptosis[186].

Globally, multiple mis-regulated miRs have been associated to human ARC by regulating LEC proliferation, apoptosis and/or migration such as miR-15a-3p[187], *miR-34a*[188], *miR-211*[189], *miR-196a-5p* [190], *miR-378a-5p*[191], *miR-630*[191], miR-23a-3p[192,193], and others.

Diabetic patients show a high prevalence of cataract (diabetic cataract, DC). The complete etiology of the DC is still unknown. The main hypothesis is that the high glucose amount in the vitreous humor perturbs the normal regulation of the lens epithelial cells causing abnormal proliferation, apoptosis and/or EMT. DC have been associated to numerous mis-regulated miRs. For example, the down-regulation of *miR-144-3p* that targets *NRF2* induce the EMT[194]. The down-regulation of *miR-30a* during DC increases the protein levels of its targets *SNAI1* and *BECN1* resulting in EMT and autophagy [152,195]. Likewise, several other miRs have been identified in DC lens. A non-exhautive list of microRNAs involved in cataracts is given in Table 2.

After cataract surgery, the remaining lens epithelial cells can undergo proliferation and EMT leading to a posterior capsule opacification (PCO) also known as a secondary cataract. *miR-204* has been associated with PCO[169], and it is also the case for numerous other miR. For example, the under-expression of *miR-34a* is associated with PCO. The overexpression of its targets *c-Met*, *Snail1* and *Notch1* causes cell proliferation and EMT[171,196,197]. Interestingly this miR is also implicated in ARC[188] and in congenital cataract (see above). This indicates that a single miR can be involved in different types of lens disorders.

Table 2: Non-exhaustive table of miRs involved in lens defects. CC: Congenital Cataract; ARC: Age related cataract; DC:Diabetic cataract; PCO; Posterior Capsule Opacification

miR	species	Identify targeted gene(s)	Lens defect	reference
miR-182	Human	*Nox4*	CC/ARC	160–162
miR-204	Human/ medaka fish	*Meis2/ Ankrd13A*	CC/ DC/ microphthalmia/ lens development defect	160,164,167,168, 198
miR-124	Human/Mice	*Vim*	CC	160,172,199
miR-184	Human/ zebrafish	*Inppl1/Itgb4/Aldh5A1/ Gabra3/Atp1a3a/ Nck2a*	CC/ microphthalmia	179–184
miR-29a	Human	*TDG*	ARC	185
miR-29b	Human	*SMAC*	ARC	186
miR-15a-3p	Human	*BCL2 / MCL1*	ARC	187
miR-34a	Human	*NOTCH2/SMAD2/ NOTCH1/ c-MET/SNAIL1*	ARC/DC/PCO	171,188,196,197, 200
miR-211	Human	*p53/BAX*	ARC	189
miR-196a-5p	Human	*?*	ARC	190
miR-378a-5p	Human	*E2F3*	ARC	191
miR-630	Human	*E2F3*	ARC	191
miR-23a-3p	Human	*BCL-2*	ARC	192
miR-30a-5p	Human	*BECN1/SNAL1*	DC	152,195
miR-205-3p	Human	*MMP16*	DC	151
miR-144-3p	Human	*NRF2*	DC	194
miR-29b	Human	*CACNA1c*	DC	201
miR-211-5p	Human/Mice	*E2f3/Sirt1*	DC	202,203
miR-214-3p	Human	*MMP2*	DC	204
miR-199a-5p	Human	*SP1*	DC	205
miR-26a-5p	Human	*ITGAV*	DC	174
miR-31	Human	*FGF7*	PCO	206
miR-22-3p	Human	*HDAC6*	PCO	207
miR-92b-3p	Human	*COL1A2*	PCO	122
miR-377-3p	Human	*CTGF/ COL1A2*	PCO	121,208
miR-181a	Human	*C-MET/SLUG/COX-2*	PCO	209
miR-497-5p	Human	*CCNE1/FGF7*	PCO	123
miR-3666	Human	*IGF-1*	PCO	210
miR-26a	Human	*SMAD4/FANCE*	PCO	211,212
mIR-204	Human	*ZEB1*	PCO	171

Table 2 shows that, at that time, only 4 miRs (*miR-182*, *miR-204*, *miR-124* and *miR-184*) have been related to congenital cataract. The majority of the

miRs identified correspond to more common types of lens opacification such as ARC, DC or PCO that are more studied than congenital cataract. But as we saw with *miR-204* or *miR-34a*, the same miR can be involved in different types of cataract. Additionally, miR are known to regulate numerous genes simultaneously. Consequently, the RNA targets described in the table only represent a small fraction of all the genes regulated by each miR. Even so, some genes are regulated by different miRs like *E2f3* or *c-Met.* This suggests that different miRs in the lens may regulate similar pathways. The miR in the lens seem to form an interconnected network of post-transcriptional regulators. We can suppose that the set of genes regulated by these miR may lead to different types of cataract depending on the levels of mis-regulation and the biological context. So it is reasonable to assume that among all the miR involved in ARC, DC and/or PCO, some of them may play a crucial role for the proper post-transcriptional regulation of genes involved in lens development. Mutation or mis-regulation of those miRs could potentially lead to the development of congenital cataract. The functions of those miRs should be more studied to gain valuable insights into the molecular mechanisms underlying lens development and the pathogenesis of congenital cataracts. Globally, it would be worth studying the numerous miRs identified in specific types of cataract in the context of other cataracts and in normal lens development.

B.2. Long non coding RNA

Beyond microRNAs, other major post-transcriptional regulators are the long non coding RNA (lncRNA). These are a class of RNA longer than 200 nucleotides that do not code for a protein. They can modulate the activity of other post-transcriptional regulators by acting as an RNA sponge by capturing miRNA or RNA-binding proteins.

While lncRNA have been largely studied in several biological and pathological contexts like cancers, their roles in lens development have been hardly investigated. A mutation in the lncRNA *RP1-140A9.1* is associated with the Volkmann type congenital cataract. The mechanisms leading to the formation of the cataract are still unknown but the mutation may lead to the retention of the intronic sequence of *RP1-140A9.1* lncRNA. This could lead to the capture of the RNA-binding protein eIF4AIII and/or the miRNA *miR-1207*, and thus perturb the regulation of their respective targets leading to the defect lens development[213]. Another lncRNA, *Malat1*, is one of the most abundant RNA during lens development in mice[30]. Its over-expression and thus the reduced activity of the miR that it sequesters have been associated with ARC, DC and

PCO[194,214–216]. It is worth noting that one of the miR identified as interacting with *Malat1* is miR-204, which is associated with congenital cataract (see above) [164]. Hence, the mis-expression of *Malat1* at a more early time point of lens development could also lead to congenital cataract. Other lncRNA are described in Table 3.

B.3. Circular RNA

Circular RNA (circRNA) are a recently described type of non-coding RNA. They are produced during RNA splicing, when a donor splice site is linked to an acceptor splice site at a 5' position. Together with the classical sense exon-exon junctions, this back-splice junction produces a circular RNA. Back-splicing is a rare event. However, the resulting circular RNAs are extremely stable as they are resistant to any exoribonuclease and can only be degraded by endonuclease-dependent pathways. Hence, the concentrations of circRNAs in living cells is far from negligible. There, they act like the lncRNA as RNA sponges for miRNA and/or RNA-binding proteins.

At this time, no circRNA has been correlated to congenital cataract, but some have been linked to DC[168,202] or PCO[122,206,208] (Table 3). Notably, the Circ *KMT2E* is supposed to bind to *miR-204*[168], which is involved in congenital cataract[164]. Thus, like for the lncRNA, a mutation or mis-regulation of the circRNA *KMT2E* could perturb proper lens development and participate to the apparearance of a congenital cataract. Extended studies are necessary to investigate the role of the CircRNA during the lens development.

*Table 3: Non-exhaustive table of non-coding RNA correlated with lens defect. LncRNA: Long non-coding RNA; CircRNA: Circular RNA; CC: Congenital Cataract; ARC: Age related cataract; DC:Diabetic Cataract; PCO; P*osterior Capsule Opacification

Type of non-coding RNA	RNA name	species	Targeted miR/RBP	Lens defect	ref
LncRNA	RP1–140A9.1	human	*miR-1207*/eIF4AIII	CC	213
LncRNA	KCNQ1OT1	human	*miR-223/miR-26a-5p/ miR-29c-3p/ miR-377-3p/ miR-124-3p*	ARC/DC	121,174,211 ,217,218
LncRNA	NEAT1	human	*miR-124-3p/ miR-26a-5p/ miR-34a/ miR-204/ miR-205-3p*	ARC/DC/PCO	151,171,173 ,212
LncRNA	MALAT1	human	*miR-144-3p/miR-204/ miR-26a*	ARC/DC/PCO	194,214–216
LncRNA	PVT1	human	*miR-214-3p*	DC	204
LncRNA	PLCD3-OT1	human	*miR-224-5p*	ARC	219
LncRNA	TUG1	human	*miR-196a-5p/ miR-29b*	ARC	186,190
LncRNA	GAS5	human	*miR-204-3p*	DC	167
LncRNA	lncRNA H19	human	*miR-29a*	ARC	185
LncRNA	XIST	human	*miR-34a*	DC	200
CircRNA	Circ-KMT2E	human	*miR-204*	DC	168
CircRNA	Circ-PAG1	human	*miR-211-5p*	DC	202
CircRNA	Circ-CARD6	human	*miR-31*	PCO	206
CircRNA	Circ-PRDM5	human	*miR-92b-3p*	PCO	122
CircRNA	Circ-MKLN1	human	*miR-377-3p*	PCO	208

B.4. RNA-binding proteins

RNA-binding proteins (RBP) are a large group of different proteins interacting with RNA (pre-mRNA or mRNA) to modulate their activity through diverse mechanisms. Here we describe the RBP known to be involved in lens development or homeostasis. A summary of these data is given in Table 4.

B.4.a. TDRD7

TDRD7 has been the first RBP associated with lens development, owing to the discovery of a homozygous mutation on this gene causing human pediatric cataract[138,220,221]. The functions of *Tdrd7* during lens development were investigated in animal models. *Tdrd7* is expressed in mouse and chicken embryo eyes, specifically in the fiber cells. The inactivation of T*drd7* in mouse and chick embryos leads to eye abnormalities including opacification of the lens[138]. In a Chinese population, a SNP (Single Nucleotide Polymorphism) within *TDRD7* locus is associated with the risk for age-related cataract onset. This is another example that supports the notion that variations on the same genes can be linked to both congenital and age-related cataracts[222]. TDRD7 is localized in the lens in RNA Granules (RG). RG are important for mRNA regulation (stabilization, degradation, etc). In cell cultures, the inactivation of *TDRD7* reduces the number of RG, specifically stress granules and processing bodies[138]. Beyond its probable role in RG formation, the role of TDRD7 in the lens is still partially unknown. As an RBP, TDRD7 could regulate the expression of multiple genes and thus regulate different pathways. By microarray hybridization, it was found that, in the absence of TDRD7, a large amount of mRNA and miRNA are mis-regulated[138,176]. In mice, TDRD7 regulates proper autophagosome maturation by targeting *Tbc1d20* mRNA. The autophagosome plays a key role for lens development by degrading the nuclei and organelles of the fiber cells. Hence the mis-regulation of *Tbc1d20* in the absence of TDRD7 may explain at least partially the cataract phenotype observed in the *Tdrd7* deficient lens[223]. TDRD7 also binds to *Hspb1* transcript, which leads to a decrease of HSPB1 protein level, a chaperone protein that can sequestrate mutant crystallin[224]. *Hspb1* KD in Xenopus leads to lens or eye defects. This confirm the necessary role of *Hspb1* for the proper development of the lens[225]. However, the micro-array and RIP assays have suggested numerous other mRNA targets for TDRD7, for example *Crybb3*[138] *or Act2*[225]. Their incorrect regulation in the absence of TDRD7 can be involved in cataract formation.

B.4.b. CAPRIN2

Lachke's team developed iSyTE (integrated Systems Tool for Eye gene discovery),a database for the mRNA expression of thousand of genes in mice lens from E10.5 to E12.5. iSyTE identified *Caprin2* as highly enriched in the embryonic mouse lens[137]. Then, iSyTE2.0 (the improved version of iSyTE, see above) confirmed the enrichment of *Caprin2* in the lens, mostly between E16.5 and P2[30]. This suggested a possible role for the RNA-binding protein CAPRIN2 in lens development. Accordingly, *Caprin2* KO or cKO eyes have defects such as smaller lens nucleus (which is formed by the primary fiber cells) and a high proportion of Peter anomaly, a stalk between the lens and the cornea[139]. In human lens cell lines, a ribopuromycilation assay (which relies on the detection of puromycin incorporation into nascent polypeptide chains) revealed that over-expression of *CAPRIN2* reduces global translation. IP-MS and Co-IP identified protein-protein interactions between CAPRIN2 and EIF3B, a translation initiation factor. The interaction between CAPRIN2 and EIF3B in lens could be required to reach a proper translation level that supports normal lens development[226]. Two different GWAS associated the *CAPRIN2* locus with the susceptibility to age-related cataracts[227,228] This condition also relates to the small lens nucleus observed in *Caprin2* deficient mice, as ARC lens present a smaller lens nucleus size than transparent lens at the same age[229]. This correlation suggests that CAPRIN2 deficiency could lead to the formation of ARC. Hence, as for TDRD7, CAPRIN2 is involved in both congenital and age-related cataracts

B.4.c. RBM24

Rbm24 codes for an RBP, mostly studied in cardiac and muscle differentiation. It has been recently demonstrated that *Rbm24* is also expressed in the lens of vertebrates such as Xenopus, chicken or mice during their embryonic development[230]. Similarly to TDRD7, RBM24 is recruited into stress granules, suggesting a post-transcriptional role[231]. Mice disrupted for *Rbm24* present embryonic lethality and eye defects such as microphthalmia and anophtalmia[232]. Similarly, morpholino antisense oligonucleotide-mediated knock-down of *Rbm24* in zebrafish embryos causes microphthalmia and/or cataract depending on the rate of inactivation[233]. RNA-seq on whole zebrafish larvae identified a large number of mis-regulated genes upon *Rbm24* knock-down, including key lens genes such as crystallins. Most of the mis-regulated lens genes are repressed. RIP followed by RT-qPCR demonstrated the interaction between RBM24 and *Lhx2, Sox2, Cryaa* and *Hsp70* mRNA, which are expressed in the lens. Over-expression of *Rbm24* increases the half-life of these

transcripts. A luciferase assay showed that RBM24 increases the stability of the mRNA of *Sox2* by interacting with its 3'UTR[232–234]. Together, these data show that RBM24 post-transcriptionnally up-regulates the expression of several key genes in lens development.

B.4.d. A quick survey on other RNA-binding proteins in lens

The Polypyrimidine Tract Binding Proteins (PTBPs) are a family of 3 RBP (PTBP1-3). In brain, the PTBP proteins modulate the alternative splicing of *DLG4* (which encodes the membrane-associated guanylate kinase PSD-95) and *CAMK2A* (which encodes a calcium/calmodulin-dependent protein kinase CaMKIIα). These two genes are also expressed in the lens and their transcripts are alternatively spliced like in the brain. In IF, *PTBP2* is expressed in rat and mouse adult lens[235]. This suggests that PTBP proteins, and notably PTBP2, regulates the alternative splicing of *Dlg4* and *Camk2a* and probably other genes in rodent lenses. Bitel's teams has characterized more precisely the expression of *PTBPs* in adult mouse lenses. *PTBP1* is expressed in lens epithelial cells whereas PTBP2 is found in the cytoplasm of the fiber cells[199]. However, in iSyTE 2.0, only *PTBP1* is lens-enriched at an early development stage. *PTBP2* and *PTBP3* show a lower general expression, and are not more expressed in the lens than in the whole body[30]. Further studies are required to reveal the roles of PTBP family members in the regulation of alternative splicing during lens development.

STAU2 (Staufen homolog 2) is an RBP targeting double stranded RNA. ISH and IF showed its expression in the retina and in the lens of chicken embryos. Overexpression of *Stau2* leads to big eye phenotype. Inversely, knock-down of *Stau2* leads to microphthalmia. Reduced *Stau2* expression is also correlated with a reduced expression of *Hes1* and *Sox2*. These two genes are known for their involvement in the early stages of eye development[236]. But it is still unknown if they are directly regulated by STAU2. Beyond its general role in eye development, no specific role for STAU2 is known in lens.

Ferritin is an intracellular protein complex involved in iron storage, hence the response to iron-mediated oxidative stress. It is composed of light and heavy chains, the latter being encoded by the *TFH1* gene. A mutation in the 5'UTR of *TFH1* causes a human autosomal dominant genetic pathology, the hereditary hyperferritinemia cataract syndrome (HHCS). In low iron concentrations, the RBPs iron regulatory proteins 1 and 2 (IRP1 and IRP2) interact with the 5'UTR of *TFH1*. This blocks *TFH1* translation. At high iron concentrations, IRP1 and IRP2 are no longer able to interact with *TFH1* 5'UTR,

allowing efficient *TFH1* translation and a decrease of free iron concentration. In HHCS patients, the mutation prevents the binding of IRP1 and IRP2 regardless of iron concentration. Together, these data show that the regulation of *TFH1* by IRP1 and 2 is necessary to prevent the juvenile cataract phenotype observed in HHCS patients[237]. We do not know if these proteins regulate other genes during lens development.

RNPC3 is a component of the minor spliceosome, which is responsible for the splicing of a special type of intron, the U12-type. Those introns represent less than 1% of all introns[238]. Biallelic mutations on *RNPC3* cause a pathology affecting multiple organs including the eye with a congenital cataract[239,240]. The impact of those mutations on RNPC3 is still unknown, but it changes an amino acid that is conserved among vertebrates. Since the genetic pathology is recessive, we can suppose that the mutation prevents RNPC3 proper activity in the minor spliceosome complex and thus modifies the splicing of U12 type introns. Further studies will be needed to identify the genes possessing U12 type intron in the lens that are mis-spliced when *RNPC3* is mutated.

QKI (QUAKING) is an RBP that regulates alternative splicing and modulates mRNA degradation and/or translation[241]. Eye-specific inactivation of *Qki* in mice causes a congenital cataract with 100% penetrancy. This cataract is associated with a mis-regulation of genes involved in cholesterol biosynthesis. Shin et al. have demonstrated that an isoform of *Qki* (*Qki*-5) can interact with and control the translation of *Srebp2*. This mRNA encodes a transcriptional regulator that controls the transcription of cholesterol biosynthesis genes[242]. Deficiencies in the biosynthesis of cholesterol-related metabolites cause cataract [243]. Hence, the cataract caused by the inactivation of *Qki* can result from an abnormal metabolism of cholesterol due to a loss of QKI-mediated translational control of *Srebp2*. However, we can suppose that QKI post-transcriptionally controls other genes to allow correct lens development. Interestingly, a GWAS associated the *QKI* gene with age-related cataract[227].

Finally, CELF1 is an RBP that I have worked on during my thesis. It will be described in more detail elsewhere (see below). Briefly, *Celf1* is the only member of the *Celf* family to be expressed in the lens[30]. Its key role for lens development was demonstrated in multiple animal models where its deficiency causes lens defects (microphthalmia, congenital cataract)[140]. The absence of CELF1 in the lens leads to large transcriptomic perturbations. Some CELF1's RNA targets in the lens have already been identified. For example, *Dnase2b* mRNA may be stabilized by CELF1, which increases its protein levels, whereas the expression of $p27^{Kip1}$ and $p21^{Cip1}$ are repressed by CELF1. These regulations participate in the proper nuclei degradation of the fiber cells[140]. CELF1 also

regulates the expression of two main transcription factors in the lens, PAX6 and PROX1. CELF1 fine-tunes the expression levels of these two genes, with PAX6 being only found in the LEC and PROX1 in the fiber cells[59]. CELF1 also seems to supervise the cytoskeletal organization during fiber cell differentiation. In the absence of CELF1, a major disorganization of the F-actin network occurs, with a dramatic change in fiber morphology. Among the mis-expressed genes, we can find *Actn2*, a cytoskeleton gene highly enriched in the lens, and *Sptb,* a spectrin gene. These two RNAs are mis-spliced in *Celf1* cKO lenses[140]. Further studies are needed to confirm the direct or indirect regulation of those gene by CELF1, and to identify new RNA target regulated by CELF1, and I contributed to these studies during my PhD.

Table 4: Table of RNA binding proteins (RBP) involved in lens defect. CC: Congenital Cataract; ARC: Age related cataract

RBP	species	Targeted miR/RBP	Lens defect	ref
TDRD7	Human/ Mice/ Xenopus	*Tbc1d20/ Hspb1/ Crybb3/ Act2*	CC	138,176,221–223,225
CAPRIN2	Human/Mice	unknow	Peter anomaly/ ARC	139,226–228
RBM24	Mice/ Zebrafish	*Lhx2/ Sox2/ Cryaa/ Hsp70*	microphthalmia/ anophtalmia/ CC	230,233,234
PTBPs	Mice	*Dlg4 / Camk2a*	?	235
STAU2	Chicken	*Hes1/ Sox2*	microphthalmia	236
IRP1 & IRP2	Human	*TFH1*	CC	237
RNPC3	Human	unknow	CC	239,240
QKI	Human/ Mice	*Srebp2*	CC/ ARC	227,241,242
CELF1	Mice/ Xenopus/ Zebrafish	*Dnase2b/ $p27^{Kip1}$/ $p21^{Cip1}$/ Pax6/ Prox1*	microphthalmia/ CC	59,140

C. The RNA-binding protein CELF1

In the previous chapter, we have reviewed the post-transcriptional controls of gene expression in lens. Essentially two types of regulatory molecules are at play, regulatory RNA (miRNA, lncRNA, circRNA) and RNA-binding proteins (RBP). Here, we will focus on one specific RBP, CELF1, as a part of my PhD thesis is devoted to the role of CELF1 in lens. CELF1 is a member of the CELF family of RNA-binding proteins, and we will firstly review this family.

C.1. The CELF family of RNA-binding proteins

The gene *CELF1* (CUGBP Elav-Like Family Member 1) encodes an RNA Binding Protein. CELF1 is also know as EDEN-BP, ETR1, Bruno2 or CUG-BP1 in different organisms. *CELF1* belongs to the *CELF* family of RNA-binding proteins, which includes 6 genes (*CELF1-6*) in mammals. All CELF proteins are composed of 3 RRMs (RNA Recognition Motif), the second and the third being separated by a divergent domain (Figure 8)[244]. The RRMs allow CELF1 to recognize and to bind to specific RNA sequences, either pre-mRNA in the nucleus or mRNA in the cytoplasm. This interaction between CELF1 and its ligand RNAs supports CELF1-mediated regulations.

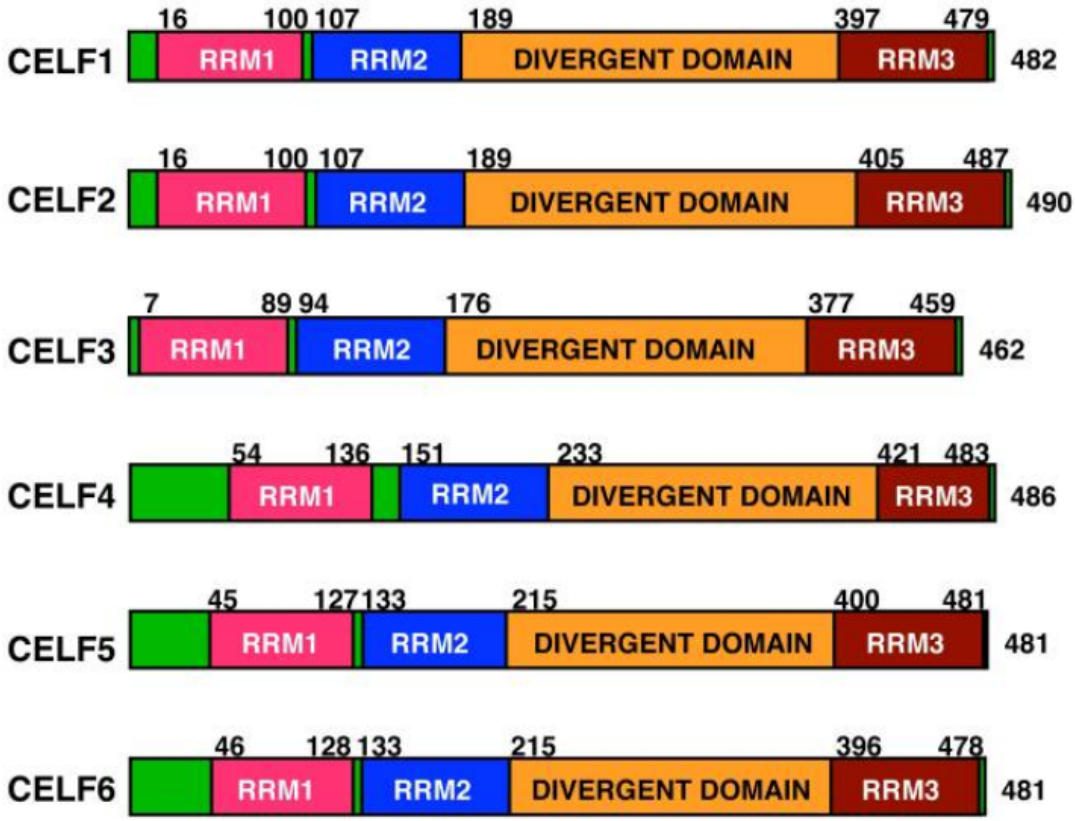


Figure 8: Structure of the 6 CELF human proteins, with three highly conserved RNA Recognition Motifs (RRM) and a divergent domain between RRM2 and RRM3. From Dasgupta et al. 2012 [244]

Based on the human protein atlas, the *CELF1-6* genes have different expression patterns (Figure 9). All *CELF* genes are expressed in a part or all of brain regions. While CELF4 and CELF5 are only present in brains, *CELF2* is also expressed in the lung, lymphoid tissues, the gastrointestinal tract and the female reproductive organs. CELF6 is found in the gastrointestinal tract and in testis. *CELF3* is enriched in the brain, endocrine tissue, gastrointestinal track and the placenta. *CELF1* and *CELF3* are expressed ubiquitously or almost ubiquitously, but *CELF3* is apparently expressed at a lower level (Figure 9)[245].

Noteworthingly, those data have been obtained from human samples collected on adult donors. They may not represent correctly the expression of those genes during the early stages of development. Indeed, in mice, multiple studies have shown by western blot a reduction of the CELF1 protein level between embryos and adult stages, in liver, stomach, lung or heart[246,247].

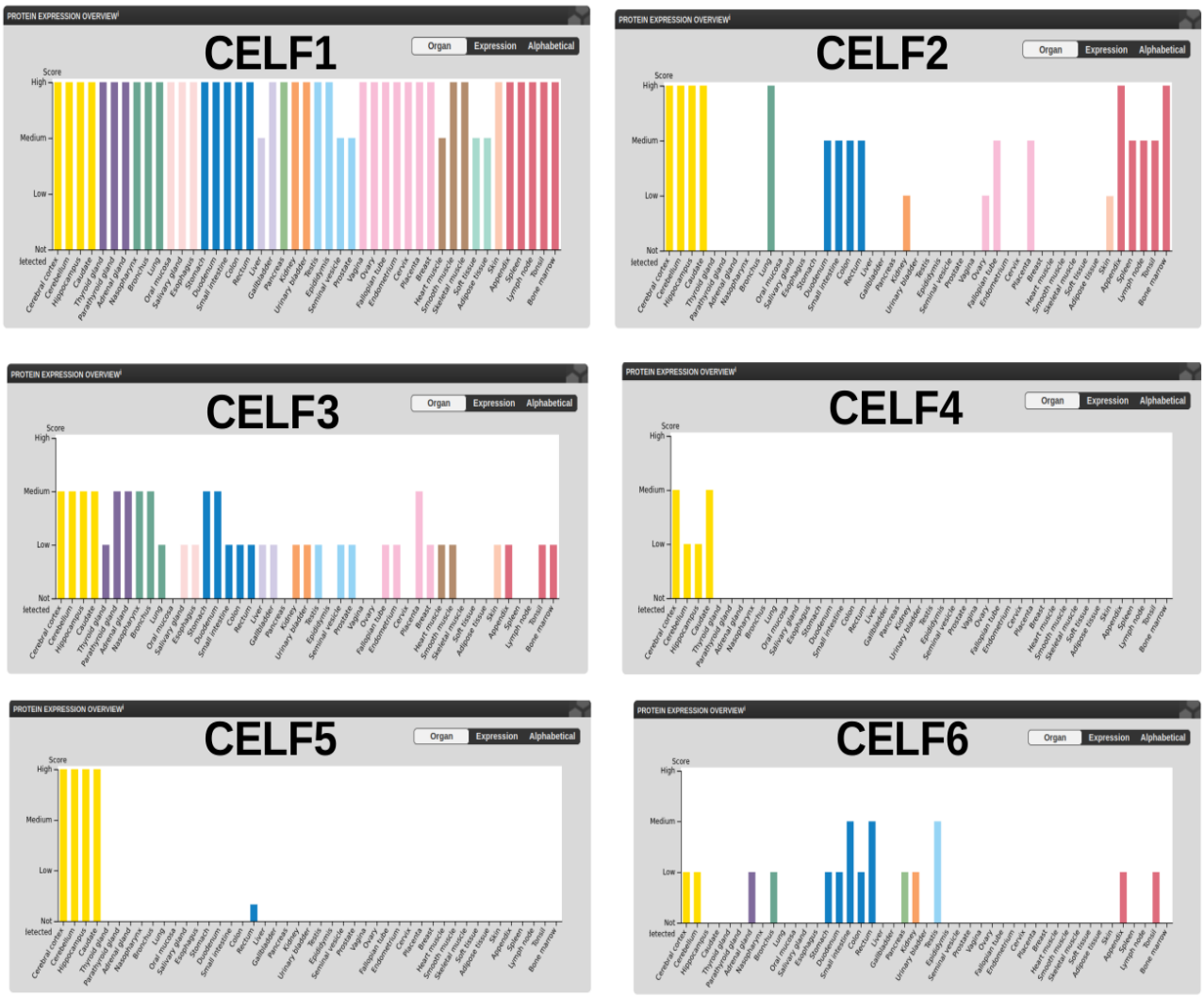


Figure 9: Expression patterns of the 6 human *CELF* genes, evaluated at the protein level (reported by the human proteome atlas [245]). All *CELF* genes are expressed in nervous tissues (yellow).

How are *CELF* genes expressed in lens? In chick (embryonic stage 35), an ISH (In Situ Hybridation) revealed a strong *Celf1* signal in the lens, in addition to the neural tube, brain, heart, lung, liver, and digestive organ[248]. That *Celf1* is highly expressed in lens was also observed in zebrafish, xenopus and mice[140,249]. As explained above, Lachke's team has developed iSyTE 2.0 (integrated Systems Tool for Eye gene discovery version 2.0), a database for the mRNA expression of thousand of genes in mice lens from E10 to P56[30]. This database returns gene expression in lens, but also gene enrichment defined as the ratio of gene expression in lens to gene expression in enucleated whole body (see above A.3.c). In iSyTE 2.0, *Celf1* is strongly expressed and enriched in the lens since the first day of eye development (Figure 10). *Celf3-5* are not expressed in the lens (Figure 10). Even if *Celf2* is apparently weakly expressed, ISH on chicken lenses did not confirm *Celf2* expression in the lens[248]. Hence, *Celf1* is the only *Celf* gene to be strongly expressed in the lens. Its high

enrichment during the first steps of lens development suggests that CELF1 may play important roles that cannot be compensated for by other CELF RBPs.

A

Symbol	Rank	Dev E10.5 affy430	Dev E11.5 affy430	Dev E12.5 affy430	Dev E16.5 affy430	Dev E17.5 affy430	Dev E19.5 affy430	Dev P0 affy430	Dev P2 affy430	Dev P56 affy430
Celf1	-	5594.34	5836.93	6603.21	4608.22	5123.1	7478.49	5676.04	3522.36	4598.95
Celf2	-	565.02	410.88	259.39	315.96	310.68	458.12	248.07	541.58	772.11
Celf3	-	45.05	60.02	50.26	74.04	75.15	53.27	72.17	46.93	75.32
Celf4	-	45.06	43.29	47.48	77.71	70.89	73.64	53.88	214.91	67.96
Celf5	-	16.86	26.23	20.02	39.07	21.84	20.6	19.18	38.54	12.08

B

Symbol	Rank	Dev E10.5 affy430	Dev E11.5 affy430	Dev E12.5 affy430	Dev E16.5 affy430	Dev E17.5 affy430	Dev E19.5 affy430	Dev P0 affy430	Dev P2 affy430	Dev P56 affy430
Celf1	-	1.55	1.63	1.99	1.34	1.5	2.2	1.63	1.05	1.35
Celf2	-	-1.29	-1.7	-2.73	-2.29	-2.26	-1.55	-2.88	-1.34	1.09
Celf3	-	-1.84	-1.42	-1.56	-1.16	-1.26	-1.65	-1.22	-1.86	-1.17
Celf4	-	-5.57	-4.46	-4.9	-3.42	-3.61	-3.32	-4.31	-1.11	-3.61
Celf5	-	-6.9	-4.96	-5.89	-3.72	-5.41	-5.71	-6.47	-3.03	-9.87

Figure 10: ***Lens specific expression of the members of the Celf family, deduced from micro-array hybridization.*** *(A) Normalized hybridization signals. (B) Enrichment scores, corresponding to the ratio of the hybridization signal at a precise developmental stage to the hybridization signal in the whole embryonic body. Data from iSyTE 2.0* [30]

C.2. Diseases and phenotypes associated with defective expression of *CELF1*

C.2.a. CELF1 and Mytonic Dystrophy, type 1

In human, *CELF1* was initially identified by groups investigating the etiology of Myotonic Dystrophy, type 1 (DM1). This autosomal dominant disease (world prevalence 1 in 2,000 to 5,000 births) leads to multiple symptoms such as muscle weakness, cardiac condition, cognitive dysfunctions and cataract[250]. It is caused by a CTG repeat extension in the 3'UTR of the gene *DMPK* (Dystrophy Myotonic Protein Kinase). In DM1 patients, who are heterozygous for *DMPK*, the *DMPK* RNA arising from the transcription of the morbide allele is retained in nuclear foci, resulting in a ~50% reduction of DMPK protein level[251,252]. Reduced DMPK may contribute to DM1. However, the phenotype of mice inactivated for *Dmpk* is not related to human DM1[253]. This suggests that the symptoms of DM1 essentially have other causes than reduced DMPK. It was found that the CTG repeat extension leads to the transcription of a "toxic" *DMPK* RNA with large CUG repeats in the 3'UTR, and that the toxicity of this RNA is the major trigger of DM1. Accordingly, mice that express genes with large CUG repeats but with a coding frame unrelated to *DMPK* (e.g. GFP) largely recapitulate DM1[254,255]. It has been hypothesized that the toxicity of the *DMPK* RNA with expanded CUG is due to the poly(CUG) titrating away specific RNA-binding proteins, thereby modifying the metabolism of their normal targets. Identifying the RBPs able to interact with the poly(CUG) then became a major objective in DM1 research. CELF1 (previously named CUGBP1, CUG-binding protein 1) was initially identified based on its reported ability to bind to the poly(CUG) [251]. However, it became rapidly clear that CELF1 is not the protein titrated away by the poly(CUG) expansion in DM1. Among other arguments, the capacity of CELF1 to interact with a poly(CUG) sequence has not been confirmed in other experiments including SELEX or CLIPseq [256–259]. Rather, the Muscleblind proteins (MBNL1 and MBNL2) are now considered as major CUG-interacting proteins, whose titration in DM1 patients strongly contributes to the pathological state[260–262].

Quite surprisingly nevertheless, *CELF1* remains involved in DM1, but unlike initially proposed. It is overexpressed in DM1 patients[263]. In mice, overexpression of *Celf1* in heart or muscle can cause lethality or muscle deficiency depending on the level of overexpression. The list of misregulated genes following *Celf1* over expression and in DM1 pathology are similar, with an elevation of myogenesis factor (P21, MEF2A)[264] and similar mis splicing events for multiple genes.(e.g. *Tnnt2, Mtmr1* and *Clcn1*)[265]. Furthermore, in the

mouse model expressing the toxic RNA (DMPK 3'UTR with the large extension) specifically in the heart, the toxic RNA leads to an over-expression of *Celf1* and to the mis splicing of *Tnnt2* transcript.

The current model of DM1 pathology states that MBNL1 is the RBP that directly binds to the toxic RNA. MBNL1 and CELF1 are antagonist regulators for the splicing of pre-mRNA [247,259,266,267] MBNL1 sequestration by the toxic RNA accompanied by an increased activity of CELF1 by poorly elucidated mechanisms cause the mis-regulation of the genes regulated in opposite directions by MBNL1 and CELF1[262].

C.2.b. Roles of CELF1 in other human diseases

in addition to DM1, *CELF1* has been associated with other human pathologies. In type 1 diabetes patient, *CELF1* is overexpressed in heart, and this is associated with a large amount of mis-spliced genes that can lead to cardiac complication[268]. A large proportion of those mis-spliced pre-mRNA in the heart of diabetic patients are directly regulated by CELF1[269]. The pathological impact of *CELF1* over-expression on the heart has also been reported for cardiac hypertrophy[270]. A large scale genome-wide association study identified *CELF1* as a new human susceptibility locus for Alzheimer's disease [271]. Finally, *CELF1* has been reported to play a role in the proliferation and severity of different types of cancer such as gastric cancer[272], glioma [273,274], breast cancer[275], melanoma[276], hepatic stellate cell cancer[277], lung cancer[278], oral cancer[279] and colorectal cancer[280]. Except for DM1 patients who suffer from cataract among other symptoms, there is no reported role for CELF1 in human cataract to our knowledge.

C.2.c. Phenotypes associated with *Celf1* inactivation in animal and cellular models

CELF1-associated diseases in human are most often correlated with *CELF1* overexpression. To get additional insights into the functions of CELF1, gene inactivation or knock-down (KD) have been made in different animal and cellular models. In worms (*Caenorhabditis elegans*), the ortholog of *CELF1* (*etr-1*) is muscle specific. RNAi-mediated KD leads to embryonic lethality[281], highlighting the importance of *Celf1* during early development. In cell cultures, the deficiency of *Celf1* in chicken heart cells causes myofibrillar structure disorganization[282]. In *Xenopus,* morpholino antisense oligonucleotide-mediated KD revealed that *celf1* (previously nammed EDEN-BP in this species) is

necessary for somite segmentation and eye development. Furthermore Celf1 regulates the mRNA encoding XSu(H), a protein that plays a central role in Notch signaling which is implicated in somite segmentation[283]. Embryo staining showed that the KD of *celf1* alters the expression of other Notch pathway genes (*esr5*, *esr9*, and *x-delta2*), with a likely impact on cell fate determination during development[140,282,284]. In Zebrafish, *celf1* KD similarly causes defects in somite segmentation, but also in the development of endoderm-derived organs (gut tube, liver, pancreas)[285]. In more recent studies, lens development defects such as microphthalmia (small eyes) have been observed in zebrafish deficient in *Celf1* [140,285,286].

In mice, a constitutive KO model has been generated. *Celf1* null mice present multiple pathologies including a high rate of embryonic lethality, a male sterility caused by defective spermiogenesis due to the overexpression of the enzyme aromatase, which converts testosterone into estradiol, and a fully penetrant congenital cataract[140,284,287]. To circumvent the problems raised by using constitutively disrupted mice, another mouse model was generated. Here, a constitutive exon of *Celf1* was floxed. Following a cross with a Cre-expressing mouse strain, *Celf1* is only inactivated in tissues and organs that express the Cre recombinase. This model was initially used with a Cre recombinase controlled by the *Pax6* promoter to inactivate *Celf1* in presumptive eyes, including lens [140]. I used this conditional inactivation model during my thesis.

As an alternative to constitutively or conditionally inactivated KO mice, mice that express a modified *Celf4* gene were engineered. CELF4 was deleted of the 2 N-terminal RRMs. This truncated protein presents a dominant negative activity by disrupting proteins-protein interaction of potentially all endogenous CELF proteins (CELF1 and other CELF-family members). This perturbs the endogenous CELF-regulated alternative splicing events[288]. Using this model, Guangju's team has shown the role of CELF1 and possibly other CELF proteins in heart regeneration after an apical resection of the heart in neonatal mice[289]. All those models confirm the key role of CELF1 during the embryonic development of multiple organs including the lens. It can be noted that in human, no pathogenic mutation of *CELF1* has been reported[290]. As in *C Elegans* and mice[140,281] the absence of CELF1 leads to embryonic lethality, we can suppose that human germinal mutations within *CELF1* gene would lead to an early embryonic death preventing any observation of those mutations.

C.3. Molecular functions of CELF1

CELF1 can regulate post-transcriptionally gene expression through different mechanisms depending of its cellular localization. In the nucleus (Figure 11A), CELF1 controls the alternative splicing of pre-mRNAs (e.g. inclusion or skipping of alternative exons). This modifies the ratio of synthesized mRNA isoforms. The protein isoforms produced from different mRNA isoforms can have different molecular functions. In addition, the different mRNA isoforms can have different stabilities or translational efficiencies, so that alternative splicing can result in different amounts of produced proteins. In the cytoplasm (Figure 11B), CELF1 can modulate the stability and the translation of its target mRNA, at least in part by stimulating their deadenylation (shortening of the poly(A) tail).

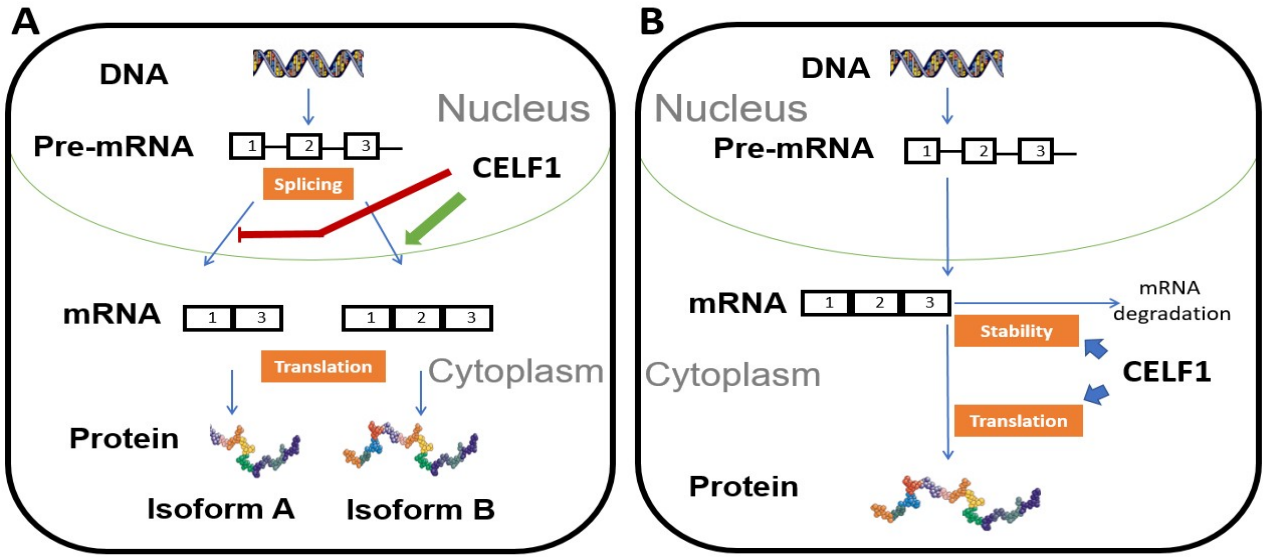


Figure 11: CELF1-mediated post-transcriptional regulation depends on its localization. *(A), When CELF1 is nuclear, it can modulate the alternative splicing of it target pre-mRNA. (B), When CELF1 is cytoplasmic it can control the stability and/or translation of its mRNA targets.*

Paillard's team first described one of the cytoplasmic roles of CELF1 in Xenopus embryos. CELF1 binds to the EDEN (embryo deadenylation element) sequence, a G/U rich element, within *Eg5* and *c-mos* 3'UTR. After fertilization, this leads to the deadenylation (removal of the poly(A) tail) of those maternal mRNA[291]. Further studies have shown that Xenopus *Celf1* can also bind and act as a deadenylation factor for microinjected human *c-Jun* mRNA [292], and that human CELF1 can replace Xenopus *Celf1* in Xenopus egg extracts to target the rapid deadenylation of EDEN-containing RNAs. Hence, human CELF1 also

behaves as a deadenylation-targetting factor at least in a heterologous system[293]. In Xenopus embryos, mRNA deadenylation is temporally uncoupled from mRNA decay, whereas in most models the removal of the poly(A) tail leads to nearly immediate mRNA degradation. Consequently, the impact of CELF1 on mRNA stability rather than directly on mRNA deadenylation was tested in other organisms. In murine myoblast cells (C2C12), the stability of the mRNA *Tnf* has been show to be directly regulated by CELF1 [294]. Still in C2C12 cells, a global analysis using a RIP-chip (RNA Immuno Precipitation) correlated the stability of specific mRNA with their binding score with CELF1. The half-life of those mRNA increased when *Celf1* expression was reduced by siRNA [295]. The impact of CELF1 on the stability of its target mRNA was validated in a human cell line (HeLa) for *Tnf*, *c-Jun, JunB* and other mRNA [296,297]. In HeLa cells, CELF1 binds to the GU rich element in the 3'UTR of mRNA. In vitro assays validated that the CELF1 dependent decay is mediated by the 3'UTR GU-rich element[298]. As CELF1 interacts with PARN, a poly(A)-specific exoribonuclease, it was suggested that the capacity of CELF1 to direct the rapid decay of bound mRNAs is due to an enhancement of deadenylation [296].

Another less described cytoplasmic role of CELF1 is the modulation of the translation of its target mRNAs. In cell cultures, CELF1 has been demonstrated to be necessary for the recruitment of the 40S ribosomal subunit onto the IRES (Internal Ribosome Entry Site) localized in the 5'UTR of the *Shmt1* transcript. Thereby, it stimulates the translation of this transcript that is initiated by a non canonical, cap-independent, mechanism [299]. In a cell-free translation system, CELF1 has been shown to increase *MEF2A* translation[264]. In a crustacean species (*Daphnia magna)*, in vivo silencing and over expression experiments demonstrated the role of CELF1 in the repression of the expression of *Dsx1*. CELF1 regulation reduces protein level without modifying RNA level, suggesting an effect on mRNA translation rather than mRNA decay. It regulates this gene through a GU rich element in the 5' UTR[56.] In mouse intestinal tissues, CELF1 binding to *Myc* 3'UTR does not affect mRNA levels but the protein level is reduced[301]. Similarly, in mouse neonatal cardiomyocytes, CELF1 binds to *Pebp1* transcript, which results in a reduction of the protein levels without altering the transcript levels[270]. Finally, in a more global approach, it was shown that TGFbeta-induced epithelial-to-mesenchymal-transition (EMT) of human MCF10A cells requires CELF1-mediated translational stimulation of several mRNAs. This up-regulation requires CELF1 binding to G/U rich elements in the 3'UTR [302]. Together, these data suggest that CELF1 may modulate mRNA translation by at least two different mechanisms: it enhances translation by binding to an IRES in cap-independent translation, and it modulates translation in either direction by binding to the 3'UTR. Whether translational modulation is mediated by changes in poly(A) tail length has not been tested.

CELF1 possesses also a nuclear role. It can influence pre-mRNA alternative splicing, thus generating alternative mRNA isoforms. This function was first described in the context of DM1 pathology were *cTNT* pre-mRNA presents a mis-splicing event (an inclusion of exon 5) associated with the increased activity of CELF1 in the nuclei of DM1 patients cells[303]. Further studies have highlighted similar mis-splicing profiles in the muscle or the heart of DM1 patients and in mouse models that overexpress *Celf1* [265,266,304]. The mis-splicing of several mRNA in the heart of mouse models for type 1 diabetes is controlled by CELF1[269]. Also in the heart, the absence of CELF1 (KO) leads to mis-splicing of mRNA directly interacting with CELF1[305]. In HeLa cell, CELF1 is localized in the nucleus and mediates the alternative splicing event of the pre-mRNAs encoding the hyaluronic acid receptor CD44 and the Insulin Receptor[275,306].

In lens, few studies have identified alternative splicing events potentially regulated by CELF1. In mouse lenses conditionally inactivated for *Celf1*, *Sptb* is differentially spliced[140]. *Sptb* codes for a Spectrin, a cytoskeleton protein interacting with the cell membrane and the actin network to control the shape of the fiber cells[307]. In the absence of CELF1, a SPTB isoform with a shortened C-terminal domain is produced in mouse lens. This could perturb fiber cell cytoskeleton network, which may contribute to the cataract phenotype observed in the *Celf1* KO mouse model. However, the direct regulation of *Sptb* by CELF1 has not been validated, and at that time no global studies has been made on alternative splicing regulation mediated by CELF1 in the lens. In another report, *CELF1* was overexpressed in the human lens epithelial cell (LEC) line SRA01/04 and differentially spliced mRNAs were identified by transcriptomic approaches[308]. The authors used a RIP-seq approach to identify, among the differentially spliced mRNAs upon *CELF1* overexpression, those that directly interact with CELF1 and are therefore presumably directly controlled by CELF1. However, the RIP data were obtained in HeLa cells. Even if HeLa and SRA01/04 cells are both of human origin, they present a radically different context for the regulation by CELF1. This is due to the presence of different transcripts and different other RBP competing or interacting with CELF1. Consequently, CELF1 targets specific of the lens and more likely to cause lens defect when mis-regulated are still unknown. In addition, since the SRA01/04 cell line is derived from LEC, the nuclear localization of CELF1 in the SRA01/04 cell line has to be questioned. Indeed, Siddam et al. have shown that in the lens CELF1 is only present in the cytoplasm of the LEC and not in the nucleus[140]. It is only in the transition zone that CELF1 is imported into the nucleus. Consequently, lens epithelial cells may not be a relevant model to study CELF1-mediated control of alternative splicing in the lens. Rather, splicing defects in SRA01/04 cells overexpressing *CELF1* may arise from the defective regulation at the cytoplasmic level of other RBPs.

C.4. The repertoire of CELF1 RNA targets

In the above data, transcriptomic approaches (RNAseq) were used to identify mRNAs with a modified abundance and/or splicing pattern following *CELF1* depletion or overexpression. However, finding that an mRNA is differentially regulated in cells with modified levels of CELF1 proteins does not prove that CELF1 directly controls this mRNA. To help identify the genes directly controlled by CELF1, different methods were used to establish a repertoire of RNA directly interacting with CELF1. Cross-Linking ImmunoPrecipitation (CLIP) analysis was used to identify the first non biased set of RNA ligands of CELF1 in the mouse hindbrain[309], and later in chicken embryonic heart[248]. CLIP is derived from ChIP (Chromatin ImmunoPrecipitation). UV light is used to create a covalent bond (a cross-link) between the RBP (e.g. CELF1) and its bound RNAs. The RNAs are then clived into small fragment (50-200 nt) by limited RNase digestion. The RNA-RBP complex are then immunopurified with an antibody targeting the RBP. After electrophoresis and transfer to a pure nitrocellulose membrane to remove any free RNA, the RNAs are recovered by proteinase K digestion. Then, they are retro-transcribed into cDNA and sequenced[310]. In the initial articles, the RNAs associated with CELF1 were identified by classical (Sanger) sequencing, allowing the identification of a limited subset of RNA ligands. Later, the CLIP technique was adapted for deep sequencing analysis (CLIP-seq)[311]. Coupling CLIP with deep sequencing does not only dramatically increase the number of identified RBP ligands, but also allows the identification of the binding sites between the RNA and the RBP. CLIPseq of CELF1 were published in different organs or cell line in mice [259,312] or human[257].

RNA ImmunoPrecipitation (RIP) is an older and easier alternative to CLIP. It differs from CLIP by the absence of cross-linking prior to the immunoprecipitation, and by the lack of limited RNase digestion. It cannot bring any information on the localization of the binding site between the RNA and the RBP. For CELF1, it was used, combined with the identification of co-immunoprecipitated RNAs by microarray hybridization (RIP-chip[295,298,313]) or deep sequencing (RIP-seq[276,314]), in mouse[295],Xenopus[313] or human[276,298,314]

A summary of all published CLIP and RIP experiments coupled with various technologies to identify RNAs is given in Table 5. Those studies are very useful to know the direct interactions between CELF1 and its target RNA, hence the repertoire of RNAs controlled by CELF1, in the cell type or tissue where the experiment has been made. However, in a different context such as a different developmental stage, a different organ or a different species, the landscape of RNA bound by a given RBP can strongly differ. For instance, different transcripts can be present, as well as different RBP that can potentially interfere (competition or synergy) with CELF1 binding. From now, no publication has described a CLIP or RIP assay on the lens in any species, and identifying the

repertoire of RNA bound by CELF1 in mouse lens by CLIP-seq constitutes a part of my PhD thesis.

Table 5: Overview of CELF1 CLIP and RIP experiments. *The strong differences in the number of detected ligands results from the technology used to identify ligand RNAs (Sanger sequencing, deep sequencing, microarray hybridization), but also from different thresholds for significance.*

Number of identified targets	Method	Organ/ cell line	Organism	Reference
153	RIP-chip	Xenopus egg	Xenopus	[313]
206	CLIP	P8 hindbrains	Mouse	[309]
613 (cytoplasmic RNA)	RIP-chip	Hela (cervical cancer cell line)	human	[298]
881 (cytoplasmic RNA)	RIP-chip	C2C12 (myoblast cell line)	Mouse	[295]
2123 (3'UTR)	CLIP-seq	C2C12 (myoblast cell line)	Mouse	[267]
1137 (3'UTR)	CLIP-seq	C2C12 (myoblast cell line)	Mouse	[259]
4170 (3'UTR)	CLIP-seq	Heart 16 weeks	Mouse	[259]
3636 (3'UTR)	CLIP-seq	Muscle 16 weeks	Mouse	[259]
159	CLIP	embryonic heart	Chicken	[248]
1560	CLIP-seq	Hela (cervical cancer cell line)	Human	[257]
5933	RIP-seq	Hela (cervical cancer cell line)	Human	[314]
3386	RIP-seq	SK-Mel-103 (Melanoma cell line)	Human	[276]
2561	RIP-seq	UACC-62 (Melanoma cell line)	Human	[276]

D. Lens models

In the above chapters, I have reviewed the controls that govern lens development and diseases, with a focus on post-transcriptional regulations and especially the RNA-binding protein CELF1. The wealth of data come from a large number of experiments carried out using specific models to examine various hypotheses. Here, I will describe the animal and *in vitro* models used in lens and cataract research.

D.1. Animal models

D.1.a. Chicken

Obtaining chicken embryos is relatively easy and does not require sacrificing the mother, unlike in mammalian models. In addition, the morphology of the chicken lens is similar to that of humans, with an anterior layer of epithelial cells that differentiate into FC at the equatorial region. Thus the chicken was historically used by Coulombre's laboratory to describe lens development. They have used this model to identify the communication between the lens and the rest of the eye[315], the importance of the orientation of the lens in the eye for the polarization of the lens and more specifically the influence of the vitreous humor on differentiation of the primary FC[8,9].

D.1.b. Fish and amphibian models

Aquatic models are excellent models for studying developmental processes. Their embryos are translucent and they undergo a fast embryonic development. They entirely develop in vitro, making their observation much easier than other models. They are generally cheap models, at least compared to mammals. Gene inactivations are easy to perform, either by injecting morpholino antisense oligonucleotides, or more, recently by CRISPR/Cas9-mediated genome engineering. However, the optical properties of aquatic models lenses differ from those of most mammalian species, as they are suited to accomodating light coming from a liquid medium. With this limit in mind, essentially two species are used as aquatic models in eye research, zebrafish and Xenopus.

The impact of gene inactivation on the lens can be relatively easily tested in zebrafish [92,316]. Furthermore, since it is possible to generate numerous zebrafish embryos quickly, this model can potentially be used to screen pro or anti- cataract compounds[317]. It is worth noting that zebrafish lens development differs from that of humans. During the formation of the primary FC, the cells do not elongate linearly as in mammals and birds. Instead, they elongate in a circular fashion, similar to the secondary FC[318]. This indicates that while lens development in zebrafish is largely similar to that in humans, this model may have a partially different regulatory network, which may limit the study of certain genes.

As tetrapods, the amphibians Xenopus are evolutionary closer to human than zebrafish. They have been used to test the impact of the deficiency of

lens genes [140,319–322]. Xenopus lens formation is mostly similar to human. However, contrary to the mammalian lens, the LEC are also found in the posterior region of the lens[322]. This model can be used like the zebrafish to screen potential pro or anti cataract drugs.

D.1.c. Mammalian models

The most frequently used mammalian model is the mouse. Mice exhibit a lens development pattern that is similar to that of humans. Moreover, they are amenable to genetic modifications, to test the effect of altering the expression (inactivation, over-expression) of candidate genes. Site-directed mutations can be introduced, to mimic mutations found in human patients and test their effect on cataract etiology. The alpha crystallins have been along the first potential cataract genes that were tested in mice. The interest for this family of genes came from their very high expression in the lens[323], and from the identification of human mutations in alpha crystallin genes associated with congenital cataracts[324]. Brady et al. generated a mouse model inactivated for the alpha A crystallin (*Cryaa*) gene. As these mice had cataract, this model confirmed that the deficiency of *CRYAA* leads to the formation of congenital cataracts in human. Additionally, this model has provided insights into the molecular function of this protein [325]. Other mouse models have been utilized to examine the impact of *Cryaa* point mutations previously identified in human[326]. Multiple other genes have also been studied using mouse models, including other crystallins[327–333], connexins[38,43,44,334], membrane protein[107,335,336], cytoskeleton proteins[335,337–339], transcriptional factors[67,68,71,79,82,91,93,100,101,103–105,340], RNA-binding proteins[59,139,140,225,232,242] and other genes[18,341–344] (Table 6).

Table 6: Non exhaustive list of lens genes associated with cataract identified using the mouse model (see references in the text)

Gene family	Genes
Crystallins	*Cryaa/ Cryab/ Cryab1/ Cryab2/ Crybb2/ Cryga/ Cryab/ Cryagc/ Cryagep1*
Connexins	*Gja1/ Gja3/ Gja8*
Cytoskeleton proteins	*Vim/ Bfsp1/ Bfsp2/ Ank2*
Other membrane proteins	*Aqp0/ Epha2/ Itgb1/ Fgfr1/ Fgfr2/ Fgfr3*
Transcription factors	*Six3/ Sox2/ Yap/ Foxe3/ Pitx3/ Prox1/ Pax6/ Maf*
RNA-binding proteins	*Celf1/ Tdrd7/ Caprin2/ Rbm24/ Qki*
Other genes	*Fyco1/ DnaseIIB/ Coc4a1/ Hsf4/ Aldh1a1*

The mouse model also enables the generation of conditional KO model (cKO), where the targeted gene is only inactivated in the targeted organ. To achieve that, the mouse's genome is modified to include two Cre recognition sites that flank one or multiple exons of the targeted gene. The gene is then "floxed" (flanked by LoxP). The floxed mouse is then crossed with another genetically modified mouse that expresses the Cre DNA recombinase in a specific cell type or tissue. The Cre DNA recombinase recognizes the two Cre recognition sites and removes the floxed DNA region lying between them. The choice of which exon(s) should be floxed is based on the desired outcome: their deletion can either remove RNA sequences, including the start codon, resulting in no protein expression, or cause a shift in the Open Reading Frame, leading to the appearance of a premature stop codon and activation of the NMD pathway (nonsense-mediated mRNA decay), which causes rapid degradation of the mRNA[345].

The Cre-dependent conditional KO approach allows to analysis the impact of lens- or at least eye-specific gene deletion. For instance, a constitutive KO of *Celf1* in the mouse leads to an increased embryonic lethality and a male sterility[284]. However, by employing eye-specific inactivation of *Celf1,* it became easier to generate mouse models specifically for studying the role of *Celf1* in the lens[140]. This eye-specific inactivation confirms that the observed phenotype of these models (e.g. cataract) arises from gene inactivation in the eye rather than in other organs. The reason why conducting these experiments is crucial is that the inactivation of a gene in the whole body may indirectly affect factors such as growth factors and/or glucose levels in the blood and/or vitreous humor. This would result in lens defects, even if the gene itself is not directly involved in lens development.

In most cases for eye studies, a *Pax6*-driven Cre is used. The Cre recombinase is expressed under the control of a *Pax6* promoter, starting at E9.5 in the lens placode. However, other Cre lines may be utilized to investigate the role of genes during later stages of development. For example,

the MRL39-Cre is only expressed in the lens fiber cells at E12.5, during the differentiation of the primary FC[346]. Therefore, the availability of different Cre recombinase lines specifically expressed at various stages in the lens allows for the examination of the diverse roles of genes of interest throughout various stages of development.

Similar to mice, rats have been used as models to study cataract formation, e.g., to validate the pathological impact of mutations on the lens[347,348]. The rat model is particularly useful to study diabetic cataracts since diabetes can be easily induced in this model[349,350].

Among mammals, the dog occupies a specific place in medical research. It is essentially not amenable to experimentation. However, due to specific breeding practices, the many dog breeds are a valuable resource to discover genes involved in genetic diseases. Inbreeding can result in breed-specific genetic diseases, including the formation of cataracts in the German Pinscher or the Jack Russell Terriers [351,352]. Veterinary studies on companion dogs have led to the identification of cataract-associated mutations on *HSF4*[353] and *FYCO1* [354]. These genes were already known to be involved in the cataract in mice and humans. However, due to the significant number of companion dogs, veterinary studies on these animals could potentially lead to the identification of new genes linked to cataracts.

D.2. *In vitro* models

D.2.a. Cell cultures from lens explants

Ex vivo models are more cost-effective and easier to use compared to animal models. Numerous three-dimensional models have been developed to mimic lens formation and FC differentiation. These models will be presented below and their main characteristic are described in Table 7.

Initially, lens epithelial cells (LEC) were obtained from rat lens epithelial explants, which correspond to the anterior epithelium of the lens. Experiments using rat lens explants have been conducted to study the impact of various factors (such as growth factors or proteins in the capsule) on LEC. These studies have shown that lens explants can mimic some aspects of cataract formation, such as the deposition of an extracellular matrix or the expression of EMT markers such as the alpha-smooth muscle actin under TGFβ induction or culture on vitronectine substract. Vitronectine is a glycoprotein present in the lens epithelium[355–357]. Interestingly, in similar conditions with a TGFβ induction, these LEC can also mimic some aspects of FC differentiation. The cells change

their morphology with a rapid elongation and express beta-crystallin, which are characteristic features of FC[355]. This suggests that under appropriate conditions, these cells can be used to study FC differentiation.

In an attempt to reconstitute the 3-dimensional organization of early lenses from lens explants, O'Connor et al. dissected postnatal (P21) rat anterior lens explants, taking epithelial cells and the capsule, but removing all fiber cells. They associated explants by pairs to expose the capsules to the extracellular medium while the apical faces of the epithelial cells faced each other within the explant pair. Explant pairs aim at mimicking early lens vesicles where only LEC are present without any fiber cell. After 30 days of culture in a medium containing bovine vitreous humor, the explant pair gained an ovoid shape and the ability to focus light. Furthermore, in IF, those explant pairs started to express FC genes, for instance the gamma crystallins[358]. This study provides evidence that the paired explant model can replicate certain aspects of lens development. However, the production of this model necessitates to sacrifice animal and requires a significant culture time.

It is also possible to obtain human LEC from eye surgical procedures. Human LEC can be cultured for an extended period of time (450 days). A basement membrane can then be observed. This basement membrane exhibits a high degree of similarity with the lens capsule. It is composed of collagen IV and laminin. When cultured at a sufficient density, the cells form lentoid bodies. The lentoid bodies are multi cellular complexes that originate from the cultured monolayer of LEC. They form round structures that end up by detaching from the plate[359]. Furthermore, adding specific growth factors such as fibroblast growth factor (FGF) on LEC cultured in 2D leads to the formation of lentoid bodies expressing gamma-crystallin[111] (Figure 12). Since these crystallins are more specific for the FC than for the LEC[11,12,18], this suggests a proteomic change mimicking the differentiation. The lentoid bodies spontaneously arising from explants could be a good model to study lens formation, but their production rate is low and their size can differ. This lack of reproducibility makes this model difficult to use.

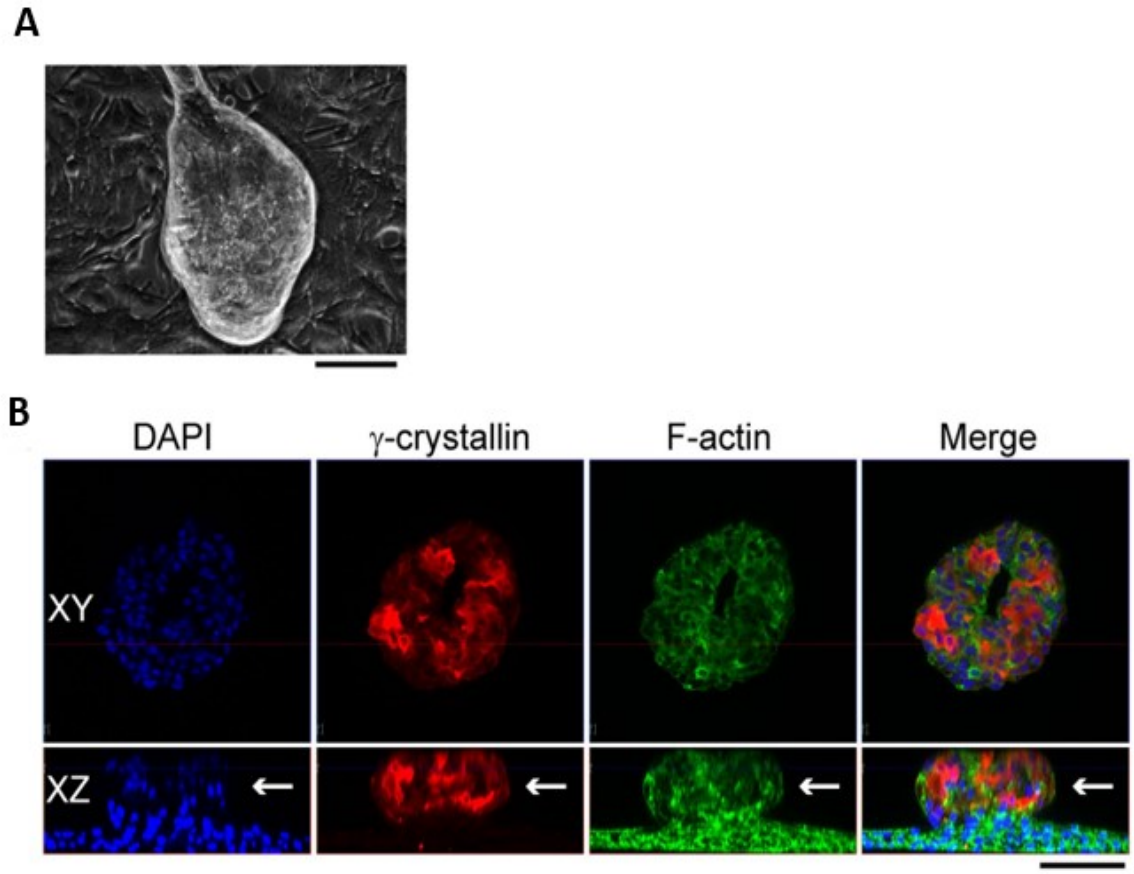


Figure 12: Lentoid body spontaneously emerging from mouse LECs. *(A) Phase contrast image of a lentoid body emerging from the 2D cultured mouse LEC. (B) Immuno-Fluorecence imaging of the lentoid by antibodies against γ-crystallin. Nuclei are stained by DAPI and F-actin is stained by phalloidin. Scale bar is 100µm. From Wang et al. 2017* [111]

D.2.b. 3D in vitro models of lens from pluripotent cells

The LEC obtained from lens explants correspond to aged anterior LEC, which may differ from earlier LEC, e.g. at the lens vesicle stage. While previous studies have shown their utility to a certain point, they may not be the best models to mimic lens development. The use of embryonic stem cells (ES) or induced pluripotent stem cells (iPS) differentiated into LEC may be a better model to study lens formation. This is of uppermost importance for studies on human, since obtaining early LECs is almost always impossible.

A first study demonstrated the feasibility to differentiate mouse ES cells into LEC. When cultured with specific growth factors (basic fibroblast growth factor, dexamethasone, cholera toxin), these cells formed eye-like structures after 11 days. The eye-like structures expressed eye-specific genes, although the lens-type cells were mixed with other cells in the structure[360].

Subsequently, Cvekl's teams developed a protocol to differentiate human ES cells into LEC-like cells, in 35 days, using growth factor at specific time-point

(Table 7). Those LEC-like cells express lens specific genes (eg: crystallins, *BFSP1/2, AQPO*) and form spontaneously numerous lentoid bodies, more than in previous LEC cultures[361]. The protocol was next modified to include a step ofmagnetic-activated cell sorting (MACS) to select only ROR1+ cells. ROR1 (receptor tyrosine kinase-like orphan receptor 1) is a membrane protein specific to the LEC. Hence, ROR1 is a good antigen to select LEC via MACS. This protocol allows the production of thousands of similar “micro-lenses”. Those micro-lenses present lens-specific characteristics, such as the ability to focus light and the presence of a lens-capsule-like membrane rich in collagen IV and laminin. Those micro-lenses can mimic a cataract phenotype (a global opacification) in contact with cataract-causing drugs[362]. Proteomic assays revealed a drastic change between the initial ROR1+ cell selected by the MACS and the micro-lenses arising from these cells after several days of culture. The protein levels of numerous crystallins are increased during the formation of the micro-lens. Other specific lens proteins such as GJA8 and VIM are found in those micro-lenses. The cells inside the micro-lenses seem to initiate a FC differentiation, as they have fewer organelles, although they still have nuclei[363].

Certain amphibian species like the newt have the capacity to regenerate their lens in case of injuries. Their iris cells differentiate into LEC that form the new lens. Based on this observation, Imaizumi’s teams derived human iPS from iris (H-iris iPS)[364]. Next, they differentiated those cells into LEC likely to form lentoid bodies useful for lens studies using Cvekl’s differentiation protocol (corresponding to the model #1 in their paper, see Table 7). Their rationale was that, if the reprogrammation into iPS is not complete, starting from iris cells is more advantageous to obtain LEC than starting from other cells more frequently used in iPS approaches.

To generate new models, they devised different variants of the initial protocol (#1) by using a rotational culture device, alternating between inclined and horizontal positions at different time points. In modified protocol #2, cells form aggregates and cell elongation occurs on one side of the aggregate. This resembles the polarization observed during the differentiation of primary FC in the lens. In modified protocol #3, cells form heterogeneous aggregates with different characteristics. Some aggregates feature two-cell layer vacuoles expressing lens-capsule genes (Collagen IV). In modified protocol #4, they generated spheroids with two distinct cell populations: a bulk of cells and an external layer of epithelial cells, resembling the organization of the lens with FC forming the bulk and a monolayer of LEC at the anterior part of the cells. Finally, in modified protocol #5, the cell aggregates express *CRYBB2,* which is specific to FC[11,12,18]. All these cell aggregates hold potentials as useful tools for studying specific characteristics of the lens. However, it's worth noting that they are opaque, which limits their suitability for cataract studies[364].

Yoa's team developed the "fried-egg" method, a protocol for deriving human induced pluripotent stem cells (iPS) into lens epithelial cells (LEC) that can form lentoid bodies (Table 7). This method involves two rounds of cell cluster selection during a 25-day differentiation process. The differentiation of those cells is validated by the expression of lens-specific genes (eg: crystallins, *AQP0*, ...), and the appearance of a lens capsule identified through CDH1 and collagen IV protein localization. The lentoid bodies exhibit magnification properties similar to lens but cannot focus light. The disadvantage of this approach comes from the requirement of multiple growth factors and the selection of specific sub-populations of cells[365]. Comparative proteomic analysis was run on the lentoid bodies produced from hIPS with the “fried-egg” protocol. This analysis confirmed the presence of lens-specific proteins (LEC and Fiber cells proteins) in the lentoid body[366].

The “fried-egg” differentiation protocol has been used to study cataract formation. The lentoid bodies generated from the differentiated LEC are transparent, but their transparency decreases over-time. This may mimic age-related cataract (ARC). The time-dependent opacification is accelerated by the presence of H_2O_2, a known chemical inducer of cataract. The opacification of the lentoid body is associated with a reduced expression of the autophagy gene *LC3BII*. A reduced expression of *LC3BII* is also found in patients with ARC. Therefore, this model could be used to study the autophagy issues taking place in ARC[367]. The fried-egg protocol was used in a high-impact article claiming that cataract caused by mutant crystallin aggregation could be reversed by a cholesterol derivative. Lanosterol was described to reduce the severity of these cataracts both in animal models[243] and in lentoid bodies[368]. However, the capacity of lanosterol to treat cataract was questioned in subsequent studies [369]. Nevertheless, lentoids bodies seem to be a good model to screen drugs potentially impacting ARC. This model could also be used to test congenital cataract mutation.

Indeed the development of human iPS studies opens the possibility to obtain iPS from patients with specific mutations. In this purpose, iPS from patients with mutations on *CRYBB2* or *CRYGD* have been differentiated into LEC thought the “fried-egg” differentiation protocol. The lentoid bodies obtained from the mutant LEC present an abnormal opacification. Remarkably, the lentoid bodies from the patient with the most intense phenotype (mutation on *CRYGD*) present a more intense opacification[370]. Thus this model could be used to test specific mutations suspected to cause congenital cataracts. It could also be used to understand the molecular mechanisms leading to the pathology.

Table 7: Characteristics of the different in cellulo model the lens studies

Differentiation protocol	Cells used	Culture condition	Time of culture	Specificity of the model
Explant pairs	Rat lens explant	Medium mixed with Vitreous humor[355]	30 days	Transparent and can focus the light
Cvekl's team protocol	Human ES cells	Specific growth factors added at different times (Yang et al.[361])	35 days	Express lens markers
Cvekl's team protocol; ROR1+ selected cells	Human ES cells	Specific growth factors added at different times (Yang et al.[361]) and MACS selection of ROR1+ cells[362]	27-48 days	Express lens markers, can focus light and can mimic cataract formation
Imaizumi's teams protocol #2	H-iris iPS	Specific growth factors added at different times (Yang et al.[361]) and different types of cell culture rotation[364]	35 days	Express lens markers, show spheroid polarization, but are not transparent
Imaizumi's teams protocol #3	H-iris iPS	Specific growth factors added at different times (Yang et al.[361]) and different types of cell culture rotation[364]	35 days	Show 2-cell layer vacuoles expressing lens capsule genes, but are not transparent
Imaizumi's teams protocol #4	H-iris iPS	Specific growth factors added at different times (Yang et al.[361]) and different types of cell culture rotation[364]	49 days	Express lens markers, spheroid surrounded by a thin layer of epithelial cell, but are not transparent
Imaizumi's teams protocol #5	H-iris iPS	Specific growth factors added at different times (Yang et al.[361]) and different types of cell culture rotation[364]	49 days	Express lens markers, including *CRYBB2* (a FC specific gene), but are not transparent
"fried-egg" protocol	hiPS	Specific growth factors added at different times and selection of sub-populations (Fu et al.[365])	25 days	Express lens markers, are transparent, can magnify light but cannot focus light. Loses its transparency over time

Table 7 summarizes the characteristics of these cellular models. It is noteworthy that all of these models exhibit interesting characteristics for lens studies but also have significant drawbacks. The explant pair model requires the sacrifice of rats, which limits the number of explant pairs that can be produced. The other models require growth factors, which are sensitive and expensive compounds. Additionally, the culture of embryonic stem cells (ES) and human-induced pluripotent stem cells (hiPS) can be challenging, particularly for extended time periods as required in these protocols (25-49 days). The inclusion of special steps in certain protocols, such as MACS cell selection, cell cluster selection, or culture using a rotational device, increases the difficulty of using these models. The development of a new model that would recapitulate most of the characteristics of the lens and can be generated in large quantities without requiring challenging culture conditions would greatly enhance the ability to easily screen for pro or anti-cataract compounds. A part of my PhD thesis was devoted to developing such a model, as will be described in the Results part.

E. PhD thesis objectives

The main focus of my PhD thesis was to identify the RNAs directly targeted by CELF1 in the lens and to investigate the consequences of their mis-regulation in the absence of CELF1. These mis-regulations are likely to contribute to the development of cataracts observed in *Celf1* KO mice. CELF1-mediated regulations can occur through cytoplasmic modulations, affecting mRNA stability and/or translation into protein, or nuclear controls, influencing the alternative splicing of pre-mRNAs and resulting in the production of different mRNA isoforms.

First, I conducted an in-depth analysis of transcriptomic data obtained from RNA-seq carried out on RNA extracted from either control or *Celf1* cKO newborn mouse lenses. This data was integrated with other data arising from microarray hybridization of control and *Celf1* cKO newborn and 6-day post-natal mouse lenses. By doing so, we identified global alterations in the lens transcriptome, revealing mis-regulated genes and pathways involved in the formation of congenital cataracts. This analysis enabled us to prioritize genes already known to be associated with cataracts and also identify novel candidate genes critical for proper lens development. These results were published in 2023, and the article is included in the "Results" part of my thesis manuscript (Chapter I).

Next, I focused on the identification of defects in splicing patterns in *Celf1* cKO lenses. With the objective to identify RNA directly regulated by CELF1, I integrated several omics analyses: two RNA-seq datasets and a CELF1 iCLIP-seq dataset made on mouse lens. I checked the altered splicing patterns of a few selected genes by RT-PCR. This integrative approach resulted in the identification of several genes exhibiting alternative splicing events presumably directly controlled by CELF1. Currently, no manuscript has been written, and these results are given in the "Results" part of my thesis manuscript (Chapter II).

Finally, to test the impact of mis-regulating the previously identified CELF1 RNA targets on cataract formation, I contributed to developing a novel and user-friendly lens organoid model. I will present the characterization of this new 3D model for lens pathology research. To demonstrate that this model successfully mimics some aspects of lens development and fiber cell differentiation, we employed 3' end RNA-seq on laser-captured specific regions of these lens organoids. This model could be easly produce *en-mass* for other study on the field of the lens development or for the screening of anti-cataract drugs. These results are presented in manuscript currently submitted for publication, which is in the "Results" part of my thesis manuscript (Chapter III).

Chapter I: Global transcriptomic disruption in CELF1 deficient lens

Abstract

The previous studies conducted by my host team have successfully confirmed the crucial role of CELF1 in post-transcriptional controls within the lens. The conditional knockout (cKO) of *Celf1* was achieved through a floxed allele of *Celf1* the eye-specific expression of the Cre recombinase under the control of the *Pax6* promoter. It results in the development of congenital cataracts. In this chapter, I present my work aimed at unveiling the transcriptomic disturbances occurring in the lens due to the absence of CELF1. To achieve this, I performed differential expression analysis on RNA-seq data obtained from postnatal control and cKO mouse lenses.

As a result, 987 genes displayed a differential expressionfollowing specific criteria: a log2(fold change) above 1.5 in absolute value, an expression (log2 counts per million) above 1, and a significance threshold (FDR, false discovery rate) below 0.05. Notably, most differentially expressed genes had elevated expression levels in the absence of CELF1. Given CELF1's known role in promoting mRNA degradation, a significant proportion of these upregulated genes might be under direct regulation by CELF1.

The integration of CLIP-seq data obtained from HeLa cells and lens-enrichment scores derived from the iSyTE database revealed an interesting pattern. Notably, genes with increased expression levels tend not to be lens-specific, and a substantial portion of them display direct RNA interaction with CELF1. This suggests that the elevated expression of these genes might have detrimental effects on lens development and differentiation, and that the regulatory role of CELF1 in repressing these genes is necessary for maintaining proper lens function.

Conversely, genes showing reduced expression levels tend to be lens-specific. Thus they could be critical for normal lens development. Their reduced expression in the absence of CELF1 could contribute to the formation of the observed congenital cataracts in the cKO mice.

Moreover, by comparing the data with other transcriptomic datasets obtained from microarray analyses of cKO mice at postnatal day 6 and newborn stages, we were able to refine and prioritize a subset of highly confident genes that are likely to be involved in the observed cataract formation. This selection of genes holds great promise as potential candidates for further investigation into the underlying mechanisms of cataract development.

Siddam et al. 2023, published in Cells

Siddam, Archana D., Matthieu Duot, Sarah Y. Coomson, Deepti Anand, Sandeep Aryal, Bailey A. T. Weatherbee, Yann Audic, Luc Paillard, and Salil A. Lachke. 2023. "High-Throughput Transcriptomics of *Celf1* Conditional Knockout Lens Identifies Downstream Networks Linked to Cataract Pathology" *Cells* 12, no. 7: 1070. https://doi.org/10.3390/cells12071070

High-throughput transcriptomics of *Celf1* conditional knockout lens identifies downstream networks linked to cataract pathology

Archana D. Siddam [1]#, Matthieu Duot [1,2]#, Sarah Y. Coomson [1]#, Deepti Anand [1], Sandeep Aryal [1], Bailey A. T. Weatherbee [1], Yann Audic [2], Luc Paillard [2]* and Salil A. Lachke [1,3]*

[1] Department of Biological Sciences, University of Delaware, Newark, DE 19716 USA
[2] Univ Rennes, CNRS, IGDR (Institut de génétique et développement de Rennes, UMR 6290, Rennes, F-35000 Rennes, France
[3] Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE 19716 USA
# Contributed equally to the work
* Correspondence: salil@udel.edu, luc.paillard@univ-rennes1.fr

Abstract: Defects in development of the ocular lens can cause congenital cataract. To understand the various etiologies of congenital cataract, it is important to characterize the genes linked to this developmental defect and to define their downstream pathways that are relevant to lens biology and pathology. Deficiency or alteration of several RNA-binding proteins, including the conserved RBP Celf1 (CUGBP Elav-like family member 1), has been described to cause lens defects and early onset cataract in animal models and/or humans. Celf1 is involved in various aspects of post-transcriptional gene expression control, including regulation of mRNA stability/decay, alternative splicing and translation. *Celf1* germline knockout mice and lens conditional knockout (*Celf1*^{cKO}) mice develop fully penetrant cataract in early postnatal stages. To define the genome-level changes in RNA transcripts that result from Celf1 deficiency, we performed high-throughput RNA-sequencing of *Celf1*^{cKO} mouse lenses at postnatal day (P) 0. *Celf1*^{cKO} lenses exhibit 987 differentially expressed genes (DEGs) at cut-offs of >1.0 log2 counts per million (CPM), ≥ ±0.58 log2 fold-change and < 0.05 false discovery rate (FDR). Of these, 327 RNAs were reduced while 660 were elevated in *Celf1*^{cKO} lenses. The DEGs were subjected to various downstream analyses including iSyTE lens enriched-expression, presence in Cat-map, and gene ontology (GO) and representation of regulatory pathways. Further, a comparative analyses was done with previously generated microarray datasets on *Celf1*^{cKO} lenses P0 and P6. Together, these analyses validated and prioritized several key genes mis-expressed in *Celf1*^{cKO} lenses that are relevant to lens biology, including known cataract-linked genes (*e.g.*, several crystallins, *Dnase2b*, *Bfsp1*, *Gja3*, *Pxdn*, *Sparc*, *Tdrd7*, *etc.*) as well as novel candidates (*e.g.*, *Ell2*, *Prdm16*). Together, these data have defined the alterations in lens transcriptome caused by Celf1 deficiency, in turn uncovering downstream genes and pathways associated with lens development and early-onset cataract.

Keywords: Lens; Eye; Cataract; RNA-binding protein; Cugbp1; Development; Post-transcriptional control; Transcriptome; RNA-sequencing; Microarrays

1. Introduction

Morphogenesis of the vertebrate ocular lens has been studied for over 100 years [1]. In addition to uncovering key principles in developmental biology, understanding the process of lens formation has helped identify genetic causes underlying human lens defects, such as congenital cataracts [2,3]. Indeed, thus far, several regulatory pathways involved in lens development have been identified [4]. While the majority of these studies were focused on signaling and transcriptional regulation [4,5], research over the past ~10 years has shown that RNA-binding protein (RBP)-based post-transcriptional control of gene expression plays key roles in lens development [6,7]. These findings have shown that the expressions of several RBPs, namely, Caprin2, Celf1 (Cugbp1), Rbm24 and Tdrd7, are

conserved in lens development across multiple vertebrate species [8,9,10,11,12,13,14,15,16]. Deficiency or mutation in these RBPs in animal models or humans are associated with eye and/or lens defects/cataracts [17,18,19,20,21,22,23]. Functional studies have indicated that these RBPs have a distinct role in spatiotemporal control over key factors in lens development. However, compared to our understanding of signaling and transcription, our knowledge on lens regulatory networks impacted by perturbation of these RBPs is limited.

CELF1 has three RNA-recognition motifs (RRMs) that allow it to bind to its target RNAs and is known to mediate RNA localization, decay/stability, alternative splicing and translation [24,25,26,27]. It has been shown that in the majority of the cases, binding of Celf1 protein to its target mRNA results in the destabilization of the latter [28]. Previously, we demonstrated that *Celf1* germline knockout (KO) or conditional KO (cKO) in the lens results in fully penetrant congenital cataracts in mice. Celf1-knockdown in fish and frogs also results in lens defects, suggesting that Celf1 plays an important role in vertebrate lens development [9]. We previously characterized specific aspects of *Celf1* deficiency-based lens defects in mice, demonstrating that Celf1-mediated negative control at the translational level over the cyclin-D kinase inhibitor $p27^{Kip1}$ was important for achieving optimal phosphorylation of nuclear lamin proteins, which in turn is critical for fiber cell nuclear envelope breakdown in normal lens development. This, in addition to *Celf1*'s positive control of mRNA expression levels of the nuclease *Dnase2b*, was found to be necessary for nuclear degradation in fiber cells [9]. Subsequently, we showed that Celf1 also played a role in achieving proper protein levels and spatiotemporal distribution of key transcription factors (TFs) in the lens. Indeed, Celf1 was found to be necessary for restriction of the expression of Prox1 protein to fiber cells and that of Pax6 to the anterior epithelium of the lens (AEL), as well as early fiber differentiating cells in normal lens development [10].

While these studies have uncovered specific aspects of Celf1 function in the lens, high-throughput RNA-sequencing (RNA-seq)-based transcriptome analyses of Celf1-deficient mouse lenses has not been described. Such an analysis will identify, on the genome-level, different mRNAs that are altered upon Celf1 deficiency, shedding further light on Celf1's role in lens development and offering new explanations regarding how alterations in its downstream pathways may contribute to lens pathology in $Celf1^{cKO}$ mice. In the present study, we address this critical knowledge gap by performing RNA-seq analysis on newborn lenses from $Celf1^{cKO}$ mice and identifying cohorts of differentially expressed genes (e.g., *Cryab*, *Cryba2*, *Cryba4*, *Crybb1*, *Crybb2*, *Cryga*, *Crygb*, *Crygc*, *Crygd*, *Cryge*, *Crygf*, *Dnase2b*, *Bfsp1*, *Gja3*, *Pxdn*, *Sparc*, *Tdrd7*, etc.) and pathways (e.g., structural constituents of eye lens, lens development in camera-type eye, lens fiber cell differentiation, etc.) associated with lens development and cataracts.

2. Materials and Methods

2.1 Animals

The University of Delaware Institutional Animal Care and Use Committee (IACUC) reviewed and approved the animal protocols described in this study. The Association for Research in Vision and Ophthalmology (ARVO) statement for the use of animals in ophthalmic and vision research was followed for animal experiments. The strategy for generating *Celf1* lens-specific conditional knockout mice is previously described [9]. Briefly, breeding was set up to generate mice (referred to as *Celf1*^{cKO}) carrying one *Celf1* germline knockout allele, (referred to as *Celf1*^{lacZKI}), one *Celf1* conditional knockout allele (exon five flanked by *loxP* sites, referred to as *Celf1*^{flox}) and the lens Cre deleter mouse line P0-3.9GFPCre (The Jackson Laboratory: 024578; henceforth referred to as *Pax6GFPCre*) that initiates Cre expression in the lens placode at embryonic day E9.5 [8,9,29]. GFPCre protein is detected to be highly and predominantly expressed in cells of the lens and pancreatic lineage in this deleter line [29,30], which has been used for generating lens-conditional knockout [8,9,10,29,31]. In the past, mice heterozygous for the *Pax6GFPCre* allele were not found to exhibit any lens defects and were used as a control [9]. *Celf1*^{flox} mice without the *Cre* allele were used as a control unless otherwise noted. Briefly, the breeding scheme was as follows. Mice containing *Celf1*^{lacZKI} allele were crossed with *Pax6GFPCre* transgenic mouse line to generate *Pax6GFPCre*:*Celf1*^{lacZKI}. These were in turn crossed with mice homozygous for the *Celf1* allele in *loxP* sites flank exon 5 (*Celf1*$^{flox/flox}$) to generate mice that carried one allele of Pax6GFPCre, one allele of *Celf1*^{lacZKI} and one allele of *Celf1*^{flox}. These mice were of mixed backgrounds with contributions from C57BL/6 and FVB strains. Plugs were checked and the day of birth was designated as postnatal day 0 (P0).

2.2 Lens RNA isolation

Lens tissue was micro-dissected from the control and *Celf1*^{cKO} mice, flash-frozen on dry ice and stored at −80 °C until further use. Two P0 lenses were pooled per biological replicate, and three biological replicates each were used for the control and *Celf1*^{cKO} mice for RNA isolation using the RNeasy mini kit (Qiagen, Germantown, TN, USA) for RNA-sequencing. For microarray analysis, total RNA was isolated using RNeasy mini kit (Qiagen) from P6 lenses (1 lens per biological replicate) from *Celf1*^{cKO} and control (*Celf1*$^{lacZKI/+}$) mice. RNA quality was evaluated by Bioanalyzer at the University of Delaware and RNA samples with an RNA quality number (RQN) above 8 were considered for microarrays or library preparation and RNA-sequencing.

2.3 RNA-sequencing and analysis

Total RNA from the control and *Celf1*^{cKO} P0 mouse lens tissue was used for RNA-sequencing (strand-specific, paired-end 150 bp-libraries) using the Illumina HiSeq 2500 sequencer at the University of Kansas Medical Center Genomics Core. FastQC was used to evaluate the quality of raw paired-end reads. The RNA-sequencing data reported here is submitted to the NCBI Gene Expression Omnibus (GEO) database under series

GSE227293. Raw sequences were trimmed to remove the adaptor sequence. Trimmed sequences were aligned on to the mouse genome (GRCm38.p6) with the STAR software (STAR(2.7.8a)) [32], and only uniquely mapped reads were retained for downstream analysis. Reads were associated to genes by featureCount (v2.0.0) [33], and only genes with >0.2 CPM (counts per million) were considered for differential gene expression analysis. The R package edgeR [34] was used to identify differentially expressed genes (DEGs), Fold Change (FC) > 1.5 (|logFC| > 0.58), False Discovery Rate (FDR) < 0.05 and an average expression in log2 CPM > 1.0.

2.4 Microarrays analysis

Total RNA from *Celf1*^{cKO} and control (*Celf1*$^{lacZKI/+}$) mouse lenses at stage P6 were isolated as stated above and microarray analysis was performed using the MouseWG-6 v2.0 BeadChip platform (Illumina, San Diego, CA, USA) following previously described protocols [9]. The previously unreported microarray data on stage P6 is submitted to NCBI Gene Expression Omnibus (GEO) database under series GSE225303. Previously generated microarray data on *Celf1*^{cKO} and control (wild-type) mouse lenses at stage P0 (deposited in GEO, GSE101393) was also used for comparative analysis in the present study. For these datasets, gene expression was estimated based on the fluorescence signal intensity of probes associated to specific genes. In cases where multiple probes were associated to the same gene, the expression of the gene was calculated as the logarithmic average of the signals from all probes assigned to the gene. Only the probes that had a fluorescence signal intensity significantly higher than background in at least two samples were retained for downstream differential gene expression analysis. The R package edgeR [34] was used to identify DEGs, with FC > 1.5 (|logFC| > 0.58), FDR <0.05 and an average expression signal > 2 5 (LogSignal > 4.6) cut-offs.

2.5 Prioritization of DEGs by Cat-Map, iSyTE, Expression in Fiber vs. Epi, and Pathways analysis

Celf1^{cKO} lens DEGs known to be associated with cataracts were identified by comparing individual gene names (mouse gene name) for DEGs to the 454 genes (human gene name) listed in the database CatMap (vOct 21) [35]. This identified a subset of genes that were differentially expressed in *Celf1*^{cKO} lenses and whose deficiency or alterations were also associated with human cataracts.

2.5.1 Cat-Map: Cataract associated genes

Celf1^{cKO} lens DEGs known to be associated with cataracts were identified by comparing individual gene names (mouse gene name) for DEGs to the 454 genes (human gene name) listed in the database CatMap (vOct 21) [35]. This identified a subset of genes that were differentially expressed in Celf1^{cKO} lenses and whose deficiency or alterations were also associated with human cataracts.

2.5.2 iSyTE: Gene expression enrichment in the lens

To determine the expression enrichment score of the *Celf1*^{cKO} lens DEGs in the normal lens—compared to the whole embryonic body (WB)—in different development stages, we used the database iSyTE [36]. iSyTE contains microarray data from normal mouse lenses at different development stages and the WB reference dataset. The lens-enriched expression scores of the *Celf1*^{cKO} lens DEGs were calculated as the maximum expression of the genes at either E10.5, E12.5, E14.5, E16.5, E17.5, E19.5 or P0, normalized by their expression WB. *Celf1*^{cKO} lens DEGs with an expression enrichment score >1.5 (|logFC| > 0.58) were considered to have enriched expression in normal lens development.

2.5.3 Celf1^{cKO} DEGs preferentially expressed in normal lens fiber cells or epithelial cells

To identify *Celf1*^{cKO} DEGs preferentially expressed in either normal lens fiber cells (FCs) or the anterior epithelial lens (AEL; also referred to as lens epithelial cells) previously generated data on these lens cell types was used [37]. This data is based on RNA-seq analysis on wild-type (WT) mice at stage P0, which identified 3516 and 3975 genes to be preferentially expressed in FCs and AEL, respectively, based on cut-offs of padj < 0.05 and FC > 1.5.

2.5.4 Celf1^{cKO} DEGs independently identified as RNA targets of CELF1 protein by CLIP-seq in a human cell line

To identify the subset of *Celf1*^{cKO} DEGs directly regulated by CELF1 protein, we examined previously generated crosslinked immunoprecipitation coupled with RNA-sequencing (CLIP-seq) data using CELF1 antibody on the human cell line Hela [38]. In this study, RNAs encoded by 3025 human genes that are bound in cellulo by the CELF1 protein have been identified by CLIP-seq in human Hela cells. We found 2825 (93.4%) mouse orthologs corresponding to these identified targets. Comparative analysis was done between these orthologs and the *Celf1*^{cKO} lens DEGs to identify genes that are recognized as RNA targets of CELF1 proteins.

2.5.5 Gene Ontology (GO) term and pathways analysis

The R package ClusterProfiler (v3.18.0) [39] was used to identify Gene Ontology (GO) terms enriched in *Celf1*^{cKO} DEGs by GO enrichment analysis as well as gene set enrichment analysis (GSEA), GO biological process (BP), GO cellular component (CC), and molecular functions (MF). GO analysis in turn led to insights into specific pathways that are altered due to Celf1 deficiency in the lens.

2.5.6. Immunostaining Analysis

Immunostaining was performed as previously described [11,40]. Briefly, mouse embryonic head tissue from stage E16.5 was fixed in 4% paraformaldehyde (PFA) (prepared in 1× phosphate buffer saline, PBS) for 30 min on ice, followed by transfer to 30% sucrose overnight at 4 °C. Once the tissue settled at the bottom, indicating that it was equilibrated, it was mounted in OCT (Tissue Tek, Torrance, CA, USA), frozen and stored at −80 °C until cryosectioning. Cryostat was used to obtain sections of 16 µm thickness. For

immunostaining, sections were blocked in a solution of 5% chicken serum, 1% BSA, 0.1% Tween (prepared in 1× PBS) for 1 h at room temperature (RT). The section was subjected to primary antibody (Cryg antibody, Santa Cruz Biotechnology #sc-22415, at 1:100 dil. in 5% chicken serum; E-cad antibody, Cell Signaling #4065, at 1:100 dil. in 5% chicken serum) by overnight incubation at 4 °C. On the following day, slides were washed three times with 1× PBS and incubated for 1 h at RT with the secondary antibody, chicken anti-goat IgG conjugated to Alexa Fluor 488 (1:200 dil.) or anti-rabbit IgG conjugated to Alexa Fluor 594 (1:200 dil.) (Life Technologies, Carlsbad, CA, USA) with the nuclear stain DRAQ5 (1:2000 dil.) (Biostatus Limited, Leicestershire, UK). Slides were washed three times in 1× PBS, mounted and imaged using Zeiss LSM 780 confocal configured with Argon/Krypton laser (488 nm and 561 nm excitation lines) and Helium Neon laser (633 nm excitation line) (Carl Zeiss Inc., White Plains, NY, USA). Adobe Photoshop CS6 (Version: 13.0.0) was used for adjustment of brightness/contrast applied consistently for all images.

3. Results

3.1. Generation of RNA-seq datasets from Celf1^{cKO} and control lenses

Lenses were micro-dissected, and RNA was isolated from stage P0 *Celf1*^{cKO} and control mice as described in detail in the Methods section. An experimental and computational pipeline was developed for RNA-seq analyses (Figure 1). Paired-end, 150 bp-long libraries were prepared, sequenced and analyzed using this strategy. For control and *Celf1*^{cKO} samples, an average of 55.1 million reads per replicate were obtained and aligned using STAR software (STAR(2.7.8a)) [32] (Table S1). On average, 77.3% of the reads were uniquely mapped to the *Mus musculus* reference genome (GRCm38.p6) (Table S1).

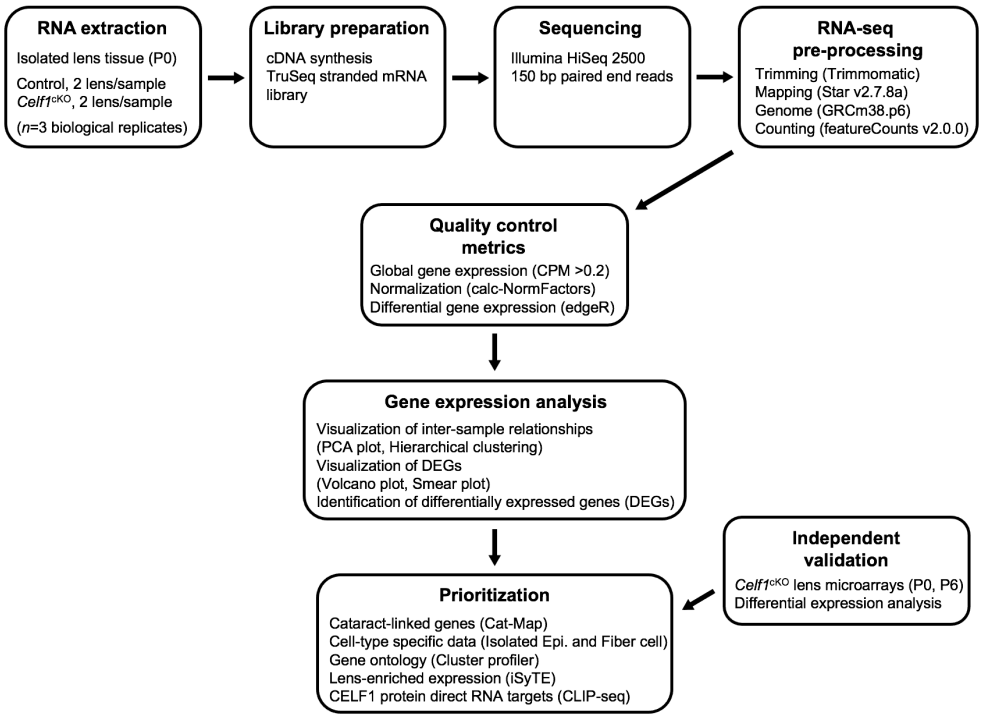


Figure 1. RNA-sequencing (RNA-seq) analysis flowchart. A flowchart outlining the experimental design and bioinformatics pipeline to determine differentially expressed genes between control and *Celf1*^{cKO} lens and their downstream analysis.

3.2. Quality control of RNA-seq datasets

We first examined the Celf1 transcript profiles in *Celf1*^{cKO} and control lenses by visualization of the RNA-seq mapped reads using the software IGV (2.8.10) (mm10) [41] and found Celf1 mRNA to be reduced in *Celf1*^{cKO} lenses (Figure 2A, Table S2). Since the conditional knockout strategy involves removal of exon 5 (Figure 2B), we quantified the inclusion of exon 5 in Celf1 mRNA in *Celf1*^{cKO} and control lenses. It is expected that exon 5 will be deleted by Cre recombinase driven by the *Pax6GFPCre* allele only in *Celf1*^{cKO} lenses, which in turn will result in a premature stop codon. This analysis shows that while control lenses had normal inclusion of exon 5, on average, *Celf1*^{cKO} lenses had 48.3% reduced *Celf1* transcripts that contained exon 5 (Figure 2C,D). Together, these data validate Cre-mediated deletion of *Celf1* in *Celf1*^{cKO} lenses. To assess the quality of the datasets on the global level, principal component analysis (PCA) was performed, which showed that control replicate samples clustered together and separately from *Celf1*^{cKO} replicate samples (Figure 3A). Additionally, hierarchical clustering between samples clearly separated control replicates from *Celf1*^{cKO} replicates (Figure 3B). Together, these analyses validate that Cre-mediated recombination of the *Celf1* conditional knockout allele driven by the *Pax6GFP* Cre-deleter line resulted in global transcriptome changes in the *Celf1*^{cKO} lens. This further confirmed that although Cre-deletion did not result in all *Celf1* transcripts being devoid of exon 5 in *Celf1*^{cKO} lenses, it was sufficient to generate transcriptome changes that result in lens defects.

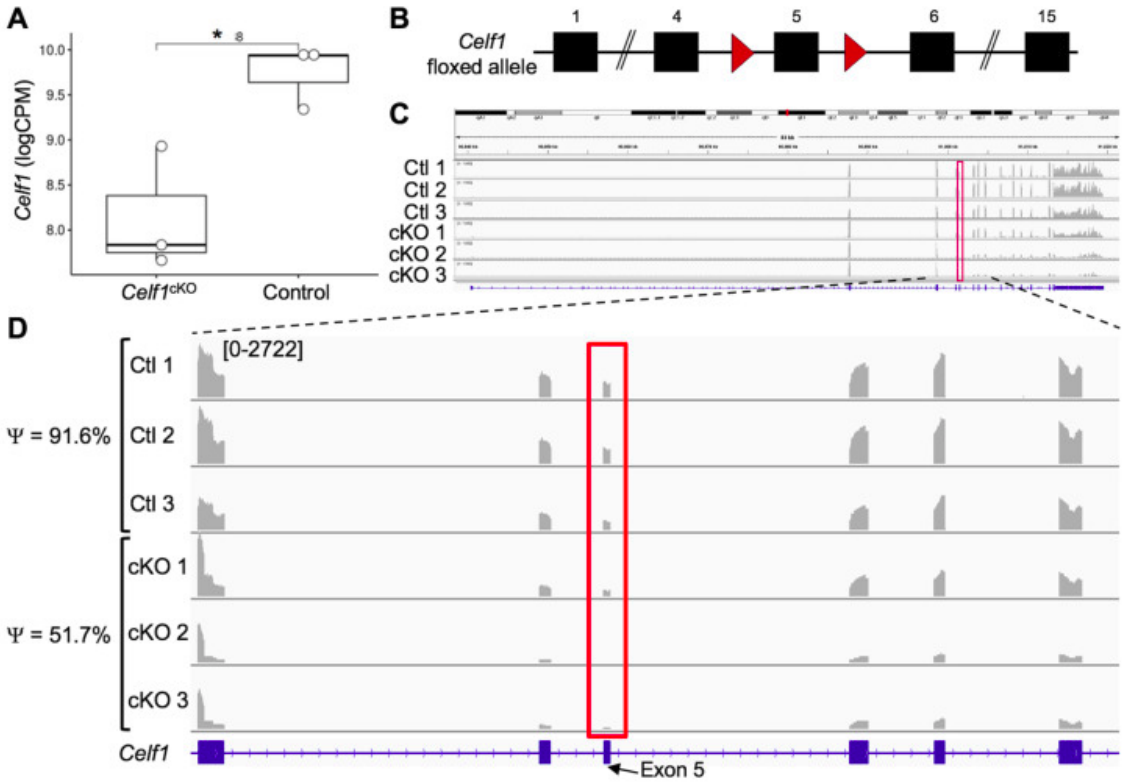


Figure 2. RNA-seq confirms reduction of *Celf1* mRNA in *Celf1*^{cKO} mouse lens. (A) *Celf1* mRNA levels in logCPM (counts per million) are significantly reduced in *Celf1*^{cKO} lenses compared to control (n=3) as estimated by student t test (asterisk indicates $P < 0.05$). (B) Schematic of Celf1 floxed allele showing exon 5 flanked by *loxP* sites (red arrowheads). (C) Visualization of the mapped reads on mouse *Celf1* locus, which at high magnification (D) shows that compared to control lens samples 1-3 that exhibit inclusion of *Celf1* exon 5 in an average of 91.6% of transcripts (represented by Ψ), *Celf1*^{cKO} lens samples 1-3 show an average of only 51.7% of transcripts to include *Celf1* exon 5 (represented by Ψ). The proportion of transcripts containing exon 5 is estimated by exon junction analysis.

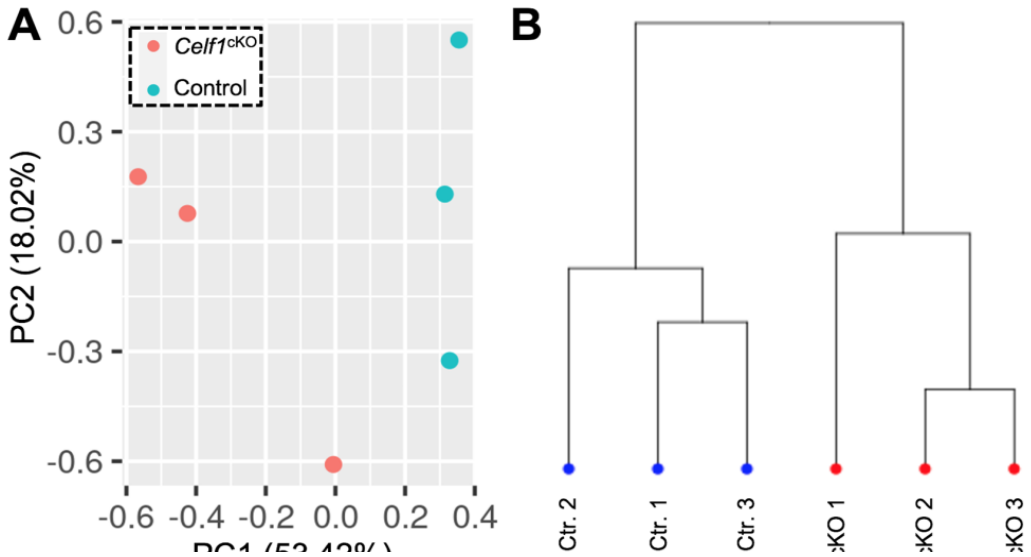


Figure 3. Validation of replicates for control and *Celf1*^{cKO} lens RNA-seq datasets. (A) Principal component analysis (PCA) of RNA-seq samples shows principal component 1 (PC1) segregates the control replicates from *Celf1*^{cKO} replicates. PC1 is responsible for 53.42% of the variance. (B) The control and *Celf1*^{cKO} replicates can be segregated as per hierarchical clustering analysis based on expression of all genes.

3.3. Identification of differentially expressed genes (DEGs) in Celf1^{cKO} lens

Based on cut-off criteria of normalized expression counts >1 log2 counts per million (CPM) averaged across all replicates, >0.58 log2 fold-change and false discovery rate (FDR) < 0.05, a total of 987 differentially expressed genes (DEGs) were identified in *Celf1*^{cKO} lenses, which is visualized by a volcano plot and a smear plot (Figure 4A,B). Of the 987 DEGs, 660 are found to be elevated while 327 are found to be reduced in *Celf1*^{cKO} lenses (Table S2). Further, RNA-seq analysis confirmed the reduction in *Dnase2b* mRNA and elevation of *p21* (*Cdkn1a*) mRNA in *Celf1*^{cKO} lenses (Table S2), as was expected based on our previous findings [9].

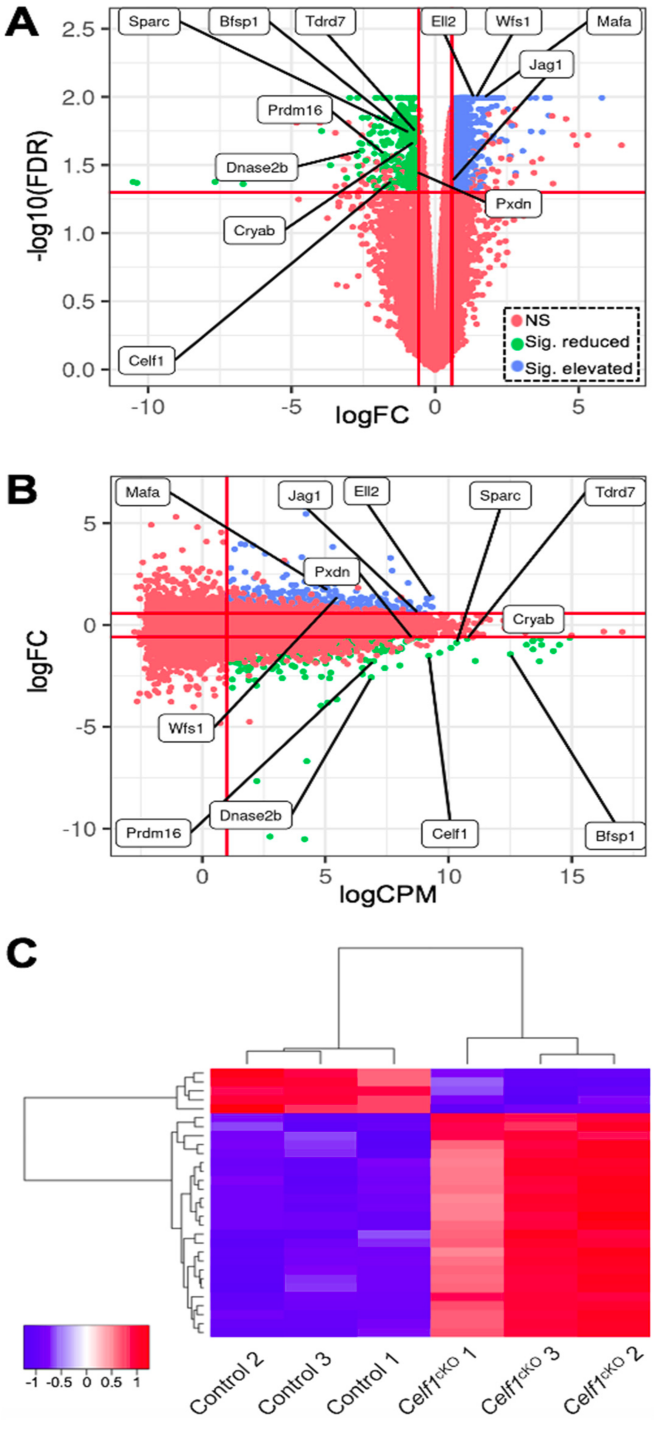


Figure 4. Identification of differentially expressed genes (DEGs) in *Celf1*^{cKO} lens. (**A**) Volcano plot and (**B**) smear plot show the different cut-offs used to identify DEGs between *Celf1*^{cKO} and control lens RNA-seq samples. With cut-off FDR < 0.05, |logFC| > 0.58, logCPM > 1.0, 660 genes and 327 genes were found to be significantly elevated and reduced, respectively, between *Celf1*^{cKO} and control lens samples. NS, not significant. Key cataract/lens-relevant DEGs are labelled in (**A**,**B**). (**C**) Heat map of all significant DEGs in *Celf1*^{cKO} lenses compared to control.

3.4. Relevance of Celf1^{cKO} *Lens DEGs to Lens Development and Cataracts*

Next, we sought to prioritize Celf1^{cKO} lens DEG candidates that are relevant to lens development and are involved in cataract pathology. Toward this goal, we performed

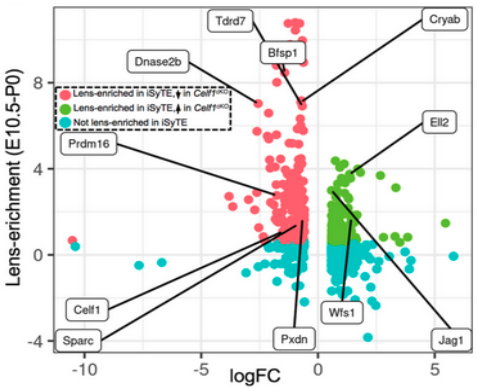
comparative analyses with publicly available datasets relevant to lens biology and pathology. For identifying DEGs linked to cataracts, we used the Cat-Map database [35]. For identifying DEGs exhibiting enriched expression in embryonic lens development, we used the iSyTE database [36]. For identifying DEGs that are preferentially expressed either in the epithelium or fiber cells, we used transcriptome datasets on isolated epithelial and fiber cells [37].

Table 1. Celf1cKO lens DEGs linked to cataract in the Cat-Map database.

Symbol	logFC	logCPM	FDR	Symbol	logFC	logCPM
Lgsn	-3.957	4.826	0.018	*Tdrd7*	-0.673	10.755
Gjb6	-3.08	1.224	0.039	*Cryba4*	-0.671	13.735
Dnase2b	-2.565	6.866	0.025	*Pxdn*	-0.667	8.535
Lctl	-1.985	8.025	0.014	*Flnb*	-0.664	5.836
Crybb2	-1.763	11.087	0.033	*Crybb1*	-0.623	14.92
Celf1	-1.511	9.204	0.041	*Jag1*	0.582	8.78
Bfsp1	-1.428	12.513	0.015	*Psmc3*	0.587	7.955
Adgrl2	-1.326	5.119	0.026	*Ube2a*	0.597	4.673
Crygb	-1.275	14.257	0.015	*Nploc4*	0.6	6.468
Cryba2	-1.195	13.429	0.021	*Sec23a*	0.681	7.201
Gja3	-1.12	10.1	0.022	*Klc1*	0.724	8.255
Dnmbp	-1.109	8.04	0.012	*Ercc6*	0.803	4.453
Ulk4	-1.07	2.037	0.02	*Atm*	0.811	4.526
Crygd	-0.973	14.5	0.013	*Rnf149*	0.857	3.842
Cryga	-0.968	13.166	0.012	*Agps*	0.907	5.422
Sparc	-0.874	10.326	0.018	*Ptn*	1.047	5.652
Sord	-0.815	4.669	0.018	*Polr3b*	1.08	4.807
Loxl1	-0.793	6.509	0.024	*Pqbp1*	1.247	5.673
Crygc	-0.788	13.425	0.012	*Pgrmc1*	1.258	6.345
Ace	-0.785	5.561	0.046	*Wfs1*	1.421	5.521
Fkrp	-0.745	4.324	0.022	*Mafa*	1.684	5.134
Cryab	-0.73	13.439	0.021			

3.4.1. Prioritization of Celf1^{cKO} lens DEGs using the Cat-Map database

Comparison of the 987 DEGs with Cat-Map identified 43 genes (including *Celf1*) that are linked to cataract in humans and/or animal models (Table 1). These genes include several crystallins (*e.g.*, *Cryab*, *Crybb2*, *Cryga*, *etc.*), membrane proteins (*e.g.*, *Gja3*), signaling pathway proteins (*e.g.*, *Jag1*) and other RBPs (*e.g.*, *Tdrd7*). Altered expression of these genes may together contribute to the lens defects observed in *Celf1*^{cKO} mice.



889* DEGs	Reduced in *Celf1*^{cKO} (*n*=282)	Elevated in *Celf1*^{cKO} (*n*=607)
Lens-enriched in iSyTE between E10.5-P0	203 (71.7%)	207 (34.1%)
Not Lens-enriched in iSyTE	79 (28.3%)	400 (65.9%)

*98 DEGs not found in iSyTE

C

889* DEGs	Reduced in *Celf1*^{cKO}	Elevated in *Celf1*^{cKO}
Lens-enriched in iSyTE between E10.5-P0 (*n*=410)	203 (49.5%)	207 (50.5%)
Not Lens-enriched in iSyTE (*n*=479)	79 (16.5%)	400 (83.5%)

Figure 5. Examination of iSyTE-based lens-enriched expression of Celf1^{cKO} lens DEGs. (A) Quadrant plot of Celf1^{cKO} lens DEGs (logFC on x-axis) and their respective lens-enrichment score between stages E10.5 and P0 as per iSyTE (y-axis). The threshold of lens-enrichment is >1.5 fold-change in the lens as per in silico subtraction-based comparison to WB reference dataset. (B) Quantification table of reduced and elevated Celf1^{cKO} lens DEGs with respect to their lens-enrichment score in normal lens development. (C) Quantification table, in which percent of lens-enriched DEGs or lens non-enriched DEGs are either reduced or elevated in Celf1^{cKO} lens DEGs.

3.4.2. Prioritization of Celf1^{cKO} lens DEGs using the iSyTE database

The 987 DEGs were examined for their potential lens-enriched expression in iSyTE at stages E10.5, E12.5, E14.5, E16.5, E17.5, E19.5 and P0. Out of 282 reduced genes in the *Celf1*^{cKO} lens, 71.7% (n = 203) were found to have lens-enriched expression in at least one of the stages examined (Figure 5A,B; Table S3). In contrast, out of 607 elevated genes in the *Celf1*^{cKO} lens, the majority of the genes 65.9% (n = 400) did not exhibit lens-enriched expression (Figure 5A,B; Table S3). Furthermore, when only the DEGs that do not exhibit lens-enriched expression are considered, a vast majority (83.5%) are found to be elevated in iSyTE (Figure 5C). In this analysis, 98 DEGs were not found in iSyTE lens microarray datasets. This may be due to differences between the two transcriptomics approaches (see Discussion). This analysis suggests that Celf1 may contribute to maintaining normal lens developmental transcriptome by negatively regulating genes not normally enriched in the lens as well as positively regulating genes, likely indirectly, which are normally enriched in the lens. This analysis also identified new potential regulators in lens development (e.g., *Ell2*) (Table S3).

3.4.3. Prioritization of Celf1^{cKO} lens DEGs using isolated epithelial and fiber cell transcriptome data

At early stages of embryonic development, Celf1 exhibits high expression in fiber cells and in later stages is also expressed in epithelial cells [9]. This suggests that it may play a role in transcriptome regulation of both cell types. To examine the impact of Celf1 deficiency in genes preferentially expressed in either epithelial or fiber cells, we compared the 987 DEGs with previously described transcriptome data from isolated epithelial and fiber cells from mouse P0 lenses. First, of the 327 genes reduced in *Celf1*^{cKO} lenses, the majority of the genes (72.5%, n = 237) were found to be preferentially expressed in either fiber or epithelial cells (Table 2; Table S4). Of these, the majority of the genes (55.6%, n =

182) were preferentially expressed in fiber cells (Table 2; Table S4). In contrast, 16.8% of reduced genes (n = 55) were preferentially expressed in epithelial cells. Next, of the 660 genes elevated in *Celf1*^{cKO} lenses, the majority of the genes (52.7%, n = 348) did not show preferential expression in either fiber cells or epithelial cells. Of these, the majority of the genes (32.0%, n = 211) were preferentially expressed in fiber cells compared to epithelial cells (15.3%, n = 101). These data indicate that a deficiency of Celf1 has a substantial impact on transcripts expressed in both fiber cells and epithelial cells. However, the extent of Celf1's impact is greater on fiber cells compared to epithelial cells. Indeed, independent validation by immunostaining shows that the fiber cell-enriched gamma crystallins are reduced in *Celf1*^{cKO} lenses, in agreement with RNA-seq analysis (Figure S1). Finally, similar to iSyTE data analysis, the upregulated genes in *Celf1*^{cKO} lenses appear to not be enriched in either epithelial or fiber cells in normal lens development. Further, the majority of the elevated genes that are not enriched in epithelial or fiber cells (52.7%, n = 348) are also not enriched in iSyTE, suggesting that in normal lens development, Celf1 is necessary, either directly or indirectly, to repress the expression of these genes.

Table 2. *Celf1*^{cKO} lens DEGs preferentially expressed in normal lens fiber cells (FCs) or anterior epithelium of the lens (AEL).

Total DEGs (n=987)	**Reduced in *Celf1*^{cKO} lens (n=327)**	**Elevated in *Celf1*^{cKO} lens (n=660)**
Preferentially exp. in FCs	182	211
Preferentially exp. in AEL	55	101
Not preferentially exp. in FCs or AEL	90	348

3.4.4. Prioritization of Celf1^{cKO} lens DEGs using CLIP-seq data identifying direct RNA targets of CELF1 protein in a human cell line

Celf1 encodes a protein containing three RRMs that enable it to bind to its target RNAs and mediate post-transcriptional regulation of gene expression. Previously, CLIP-seq analysis with a CELF1 antibody has been applied to identify the direct-bound RNA targets of CELF1 protein in the human cell line, Hela [38]. Comparative analysis showed that 32.2% (n = 318) of the 987 *Celf1*^{cKO} lens DEGs are also identified in this CLIP-seq dataset. Further, within these 318 DEGs that are directly bound by CELF1 protein, the majority (83.0%; n = 264) are found to be significantly elevated in *Celf1*^{cKO} lenses. While 21 of these 318 *Celf1*^{cKO} lens DEGs were not found in iSyTE, 32.4% (n = 103) of these DEGs are found to exhibit enriched expressed in normal lenses in iSyTE, while 61.0% (n = 194) are not lens-enriched (Table 3). Of the 318 DEGs, the majority are not preferentially expressed in either FCs or AEL and are elevated in *Celf1*^{cKO} lenses (n = 149) (Table 4). Among the DEGs preferentially expressed in either cell type, the majority are preferentially expressed in FCs (n = 104).

Table 3. *Celf1*^{cKO} lens DEGs selected by cross-linked immunoprecipitation (CLIP) that exhibit enriched expression in normal lens as per iSyTE.

DEGs selected by CLIP (*n*=297[1])	Reduced in *Celf1*^{cKO} lens (*n*=50)	Elevated in *Celf1*^{cKO} lens (*n*=247)
Lens enriched-exp. in iSyTE	26	77
Not lens enriched-exp. in iSyTE	24	170

[1]21 of the DEGs selected by CLIP were not found in iSyTE

Table 4. *Celf1*^{cKO} lens DEGs selected by cross-linked immunoprecipitation (CLIP) that are preferentially expressed in normal lens fiber cells (FCs) or anterior epithelium of the lens (AEL).

DEGs selected by CLIP (*n*=318)	Reduced in *Celf1*^{cKO} lens (*n*=54)	Elevated in *Celf1*^{cKO} lens (*n*=264)
Preferentially exp. in FCs	30	74
Preferentially exp. in AEL	9	41
Not preferentially exp. in FCs or AEL	15	149

3.5. Gene ontology and pathway analysis of Celf1^{cKO} lens DEGs

Next, we examined the different pathways that were represented in *Celf1*^{cKO} lens DEGs. Toward this goal, we performed pathway enrichment analysis by examining gene ontology (GO) enrichment separately on all DEGs, elevated DEGs and reduced DEGs, compared to all the genes expressed in the RNA-seq data. In parallel, GSEA (gene set enrichment analyses) were performed on all the genes in the RNA-seq data that had an expression of at least 1 logCPM and based on their logFC rank. Further, we performed the GI enrichment analysis on a subset of these DEGs that are found to have enriched expression in the lens by iSyTE described in Section 3.4.2. We also performed this analysis on a subset of these DEGs that are preferentially expressed in epithelial or fiber cells described in Section 3.4.3. GSEA analysis was not performed on this subset of DEGs because a large number of genes are required for optimal analysis. This analysis identified "structural constituent of eye lens" (GO:0005212), "lens development in camera-type eye" (GO:0002088), and "visual perception" (GO:0007601) among the top enriched GO terms in 327 reduced DEGs in *Celf1*^{cKO} lenses (Figure 6; Table S5). The same GO terms were identified among reduced DEGs that exhibit enriched expression in normal lens development as per iSyTE (Figure 7) as well as reduced DEGs that are preferentially expressed in normal fiber cells (Figure 8), and these GO terms were also identified by the GSEA analysis as reduced (Table S5). Further, among the reduced DEGs that are preferentially expressed in fiber cells, the GO term "lens fiber cell differentiation" (GO:0070306) was also found to be significantly enriched. These GO categories identified candidate genes with known functions in the lens and/or those associated with cataracts. Among these are several crystallins, *Bfsp1*, *Gja3*, *Tdrd7*, etc. (Table S5). Only two GO terms were found to be enriched among reduced DEGs that are preferentially expressed in epithelial cells. These are "extracellular matrix" (GO:0031012) and "cell projection membrane" (GO:0031253). GSEA analysis and GO enrichment analysis of all the 660 elevated DEGs, or 207 elevated DEGs with enriched expression in normal lens development as per iSyTE, or 211 elevated DEGs preferentially expressed in normal fiber

cells, commonly identified, among others, the GO terms, "proton transporting ATPase activity, rotational mechanism" (GO:0046961) and "cytoplasmic vesicle membrane" (GO:0030659) to be enriched. Additionally, in all the 660 elevated DEGs, the GO terms "calcium-dependent protein binding" (GO:0048306), "clathrin coat of coated pit" (GO:0030132), "organelle subcompartment" (GO:0031984) and "protein kinase inhibitor activity" (GO:0004860) were also found to be enriched (Table S5). The majority of the GO terms described above were also identified when all the DEGs or the DEGs with lens-enriched expression, or the DEGs preferentially expressed in fiber cells were considered. Further, of the 453 elevated DEGs that do not have an enriched expression in normal lenses, the GO terms "cytoplasmic vesicle membrane" (GO:0030659), "proton-transporting V-type ATPase complex" (GO:0016471) and "proteasome complex" (GO:0000502) were found to be enriched. Finally, GO term analysis of reduced DEGs that were also identified in CLIP showed enrichment of the terms related to "positive-regulation of brown fat cell differentiation" (GO:0090335) (Figure 9). Among elevated DEGs also identified in CLIP, the GO terms, "translational initiation" (GO:0045948), "clathrin-coated vesicle" (GO:0030136), "ribonucleoprotein complex binding" (GO:0043021) and "calcium-dependent protein binding" (GO:0048306) were identified (Table S5). A subset of these GO terms was also found to be enriched when all DEGs that are identified by CLIP were considered. Thus, GO term analysis identifies pathways whose perturbations contribute to the cataract pathology observed in *Celf1*^{cKO} lenses, which are further highlighted in the Discussion below.

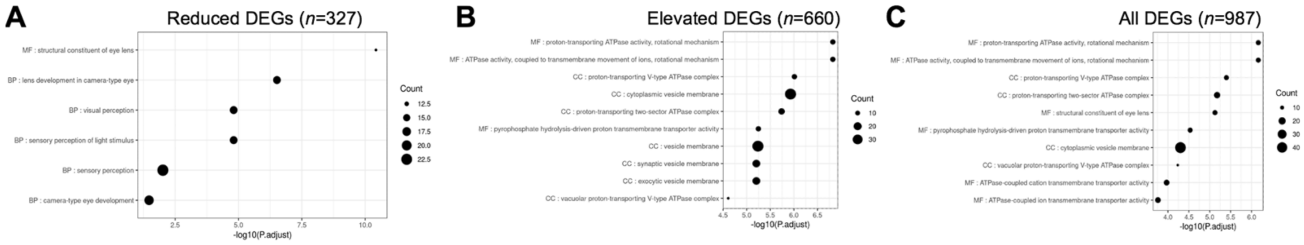


Figure 6. Gene ontology (GO) and pathway analysis on *Celf1*^{cKO} lens DEGs obtained from RNA-seq. The top significant GO terms enriched in (A) reduced DEGs, (B) elevated DEGs, and (C) all DEGs in *Celf1*^{cKO} lens.

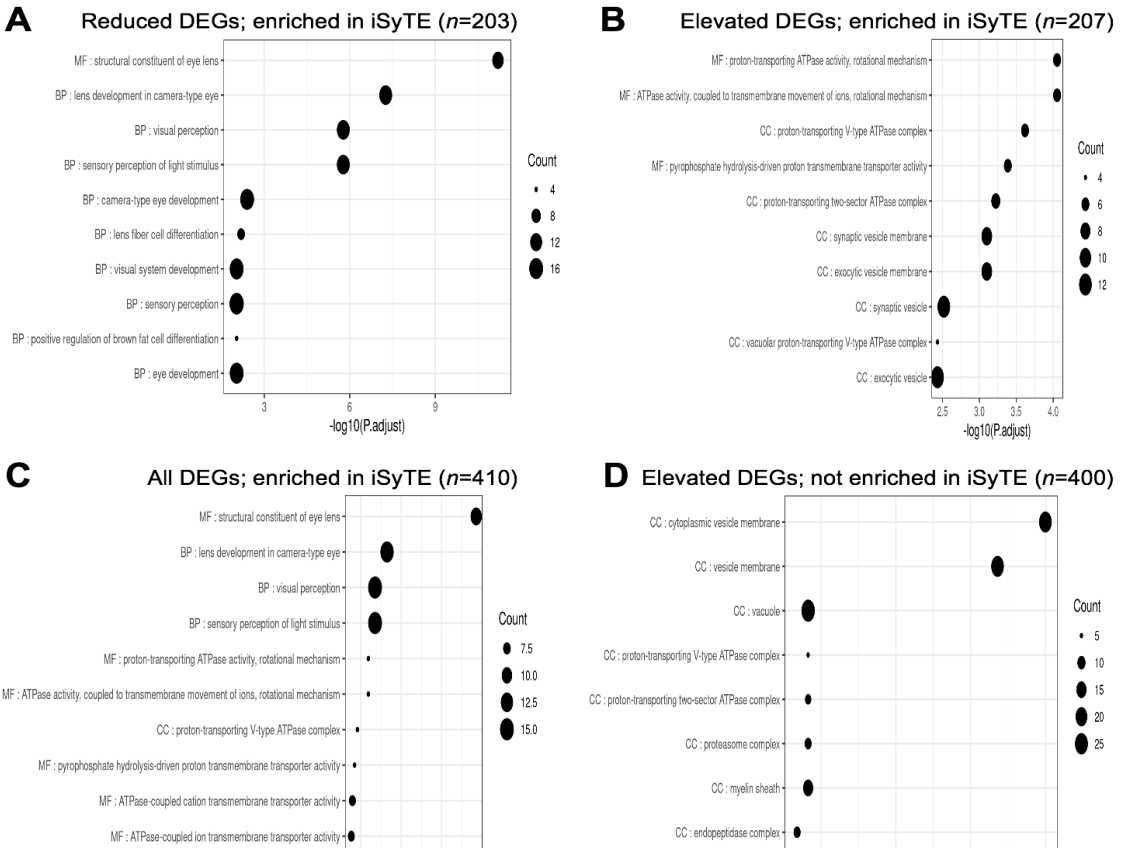


Figure 7. Gene ontology (GO) and pathway analysis on *Celf1*^{cKO} DEGs analyzed by iSyTE. The top significant GO terms enriched in (A) reduced DEGs, (B) elevated DEGs, and (C) all DEGs that exhibit enriched expression in normal lens as per iSyTE. (D) The top significant GO terms enriched in *Celf1*^{cKO} lens elevated DEGs that do not exhibit enriched expression in normal lens.

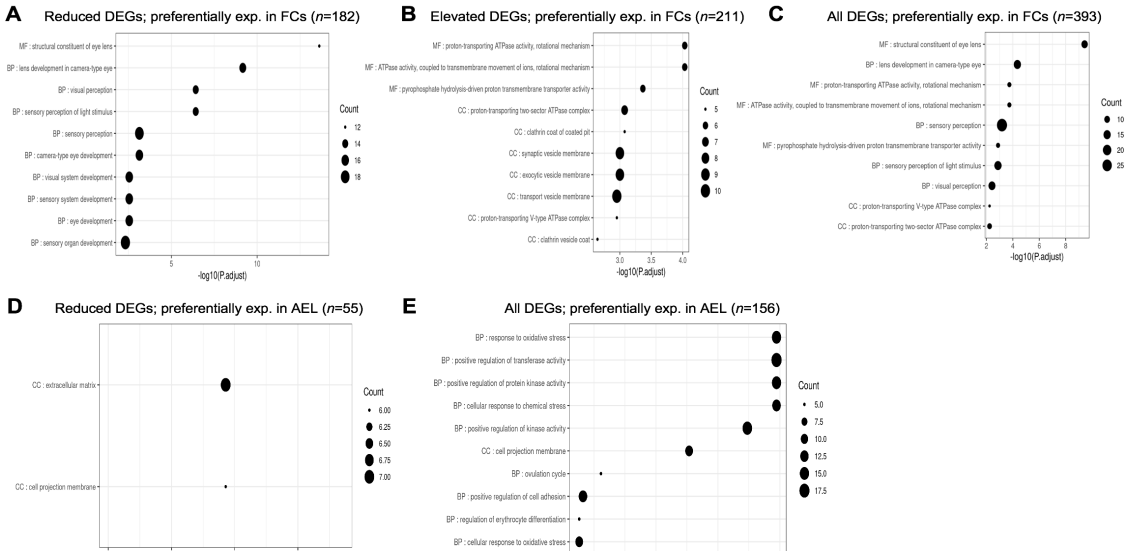


Figure 8. Gene ontology (GO) and pathway analysis on *Celf1*^{cKO} lens DEGs preferentially expressed in the anterior epithelium of the lens (AEL) and fiber cells (FCs). The top significant GO terms enriched in (A) reduced DEGs, (B) elevated DEGs, and (C) all DEGs that are preferentialy expressed in normal FCs. The top significant GO terms enriched in (D) reduced DEGs and (E) all DEGs that are preferentialy expressed in normal AEL. No significant GO terms were identified in *Celf1*^{cKO} elevated DEGs expressed in normal AEL.

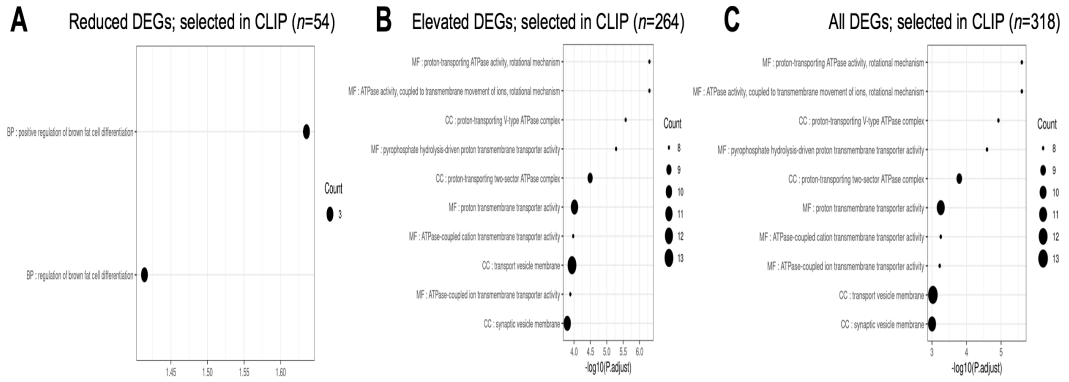


Figure 9. Gene ontology (GO) and pathway analysis on *Celf1*^{cKO} lens DEGs that are also identified in CLIP-seq dataset. The top significant GO terms enriched in (A) reduced DEGs, (B) elevated DEGs, and (C) all DEGs that are also identified in CELF1 cross-linked immunoprecipitation followed by RNA-sequencing (CLIP-seq) data on the human cell line (HeLa cells).

3.6. Comparative analysis of Celf1^{cKO} lens DEGs identified by RNA-seq and microarrays

Next, we sought to compare *Celf1*^{cKO} lens DEGs identified by RNA-seq with DEGs that are identified by expression microarrays so as to provide independent validation of the DEGs that can be used for prioritization of candidates. For *Celf1*^{cKO} lenses, published expression microarray data is available for stage P0. There is also unpublished expression microarray data on *Celf1*^{cKO} lenses for stage P6. We first performed differential expression analysis on *Celf1*^{cKO} lens microarray data for stage P0 and P6. This analysis identified 549 DEGs at P0 and 665 DEGs at P6 (Figure 10 and Figure S2; Table S6). Of these, 322 were elevated and 227 were reduced at P0, while 304 were elevated and 361 were reduced at P6. Comparative analysis identified 174 DEGs to be commonly elevated and 78 DEGs to be commonly reduced between RNA-seq and microarrays at P0. Comparative analysis identified 158 DEGs to be commonly elevated and 90 DEGs to be commonly reduced between RNA-seq and microarrays at P6. There is a higher number of DEGs found to be elevated or reduced by the RNA-seq approach at P0. This may be due to technical differences in the two approaches. While microarrays are limited by a predetermined number of genes represented on the array, RNA-seq has no such limitation. Further, while RNA-seq provides individual sequence reads, microarrays depend on probe binding kinetics which may impact their sensitivities. Further, there is a higher number of DEGs found to be mis-expressed by microarrays at P6 compared to P0, which is expected because of the progression of the lens defects. Together, this analysis provides independent validation of numerous DEGs that are mis-expressed upon Celf1 deficiency in the lens (Table S6). Furthermore, among the *Celf1*^{cKO} lens DEGs commonly identified by RNA-seq and microarrays, 84 found at the RNA-seq (P0)–microarray (P0) comparison and 54 found at the RNA-seq (P0)–microarray (P6) comparison were also found to be directly bound by Celf1 protein as per CLIP data (Table S7).

A

	RNA-seq (P0) (*n*=987)	Microarrays (P0) (*n*=549)	Microarrays (P6) (*n*=665)
Elevated *Celf1*^{cKO} DEGs	660 (66.9%)	322 (58.7%)	304 (45.7%)
Reduced *Celf1*^{cKO} DEGs	327 (33.1%)	227 (41.3%)	361 (54.3%)

B

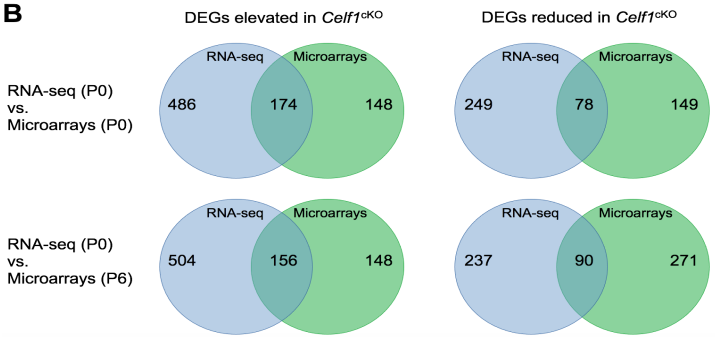


Figure 10. Comparative analysis of *Celf1*^{cKO} lens DEGs obtained from RNA-seq and microarrays. (A) List of elevated and Reduced differentially expressed genes (DEGs) in *Celf1*^{cKO} lens as identified by RNA-seq at stage P0, microarrays at stage P0 and P6. (B) Venn diagrams indicating elevated or reduced *Celf1*^{cKO} lens DEGs commonly- or exclusively-identified by RNA-seq and microarrays analysis at P0 and P6.

4. Discussion

Celf1 encodes an RNA-binding protein that has been associated with various tissue development/cell differentiation and developmental defects/diseases. Celf1 has a role in cells as different as sperm, muscle, and lens cells, among others, and its alterations are associated with various types of cancer and developmental defects including heart defects, myotonic dystrophy and cataracts [6,25,27,42,43,44,45,46,47,48,49]. As an RBP, Celf1 can mediate gene expression control by directly binding to target RNAs and impact their intracellular localization, splicing, stability/decay or translation [38].

Mouse models of *Celf1* deficiency exhibit cataracts and other pathologies [9,10,43]. In the past, different proteins/pathways that are altered due to Celf1-decificiency have been characterized (e.g., p27^{Kip1}, Dnase2b, Pax6 and Prox1). For example, previous work described how Celf1 post-transcriptionally controls the dosage of p27^{Kip1} protein by reducing it in fiber cell differentiation, while also being necessary for optimal levels of the nuclease Dnase2b in the lens. Further, Celf1 also functions to negatively control p21^{Cip1} in the lens. Together, these actions of Celf1 result in proper degradation of fiber cell nuclei thereby contributing to optimal refraction of light and lens transparency [9]. Celf1 is also necessary for proper spatiotemporal expression of Prox1 and Pax6 transcription factors in lens development; the disruption of which further contributes to the lens defects [10]. Additionally, an absence of Celf1 is expected to lead to changes to the lens transcriptome that result in lens defects and cataracts. To gain insights into such global perturbations, we performed high-throughput RNA-seq and examined the differentially expressed genes in the *Celf1*^{cKO} lens. While there have been reports on RNA-seq on Celf1 perturbations, in the

context of the lens, these are limited to cell lines and not the lens tissue [50,51]. The only transcriptome data available on *Celf1*-deficient lens tissue is on a microarray platform. While microarrays are informative, they have limitations as they depend on probe binding kinetics and can only report on a predefined set of genes. On the other hand, RNA-seq does not present such limitations and offers greater depth of global changes in transcripts.

While the *Celf1*^{cKO} lens exhibits 987 DEGs, interestingly the majority of genes were found to be elevated, suggesting that Celf1 protein—either directly or indirectly—has negative control over these transcripts in normal lenses. However, because 327 were found to be reduced in *Celf1*^{cKO} lenses, this suggests that Celf1 is also necessary for positive control over these genes that may be important for proper lens development. We aimed to identify both, "established" cataract-linked genes as well as potentially novel candidates that are differentially expressed upon Celf1 deficiency. To address the former, we performed analysis of *Celf1*^{cKO} RNA-seq data with respect to the known cataract-linked genes contained in the Cat-Map database. This helped identify the established cataract-linked genes that are significantly impacted because of Celf1 deficiency. On the other hand, the iSyTE database informs on both established cataract-linked genes as well as novel genes that are relevant to lens biology. Therefore, we also performed comparative analysis of *Celf1*^{cKO} RNA-seq data with respect to the genes recognized as lens-enriched in the iSyTE database. Indeed, in addition to identifying known cataract-linked genes, this analysis also identified novel genes with potential functions in the lens. We elaborate below on these findings.

Among the DEGs, 43 genes, including *Celf1*, have previously been linked to cataracts in humans or animal models as per Cat-Map. The majority of these (>60%) were found to be reduced in *Celf1*^{cKO} lenses. This includes *Dnase2b* which is significantly reduced in *Celf1*^{cKO} lenses. Because Dnase2b is necessary for proper nuclear degradation in lens fiber cell differentiation [52] and was also previously found to be a direct RNA target of Celf1 protein in the lens [9], this finding renders confidence in the RNA-seq data. Additionally, several other genes linked to human cataracts were found to be significantly reduced in *Celf1*^{cKO} lenses. These include the crystallins *Cryab*, *Cryba2*, *Cryba4*, *Crybb1*, *Crybb2*, *Cryga*, *Crygb*, *Crygc* and *Crygd*, the connexins *Gjb6* and *Gja3*, the membrane protein *Bfsp1*, the extracellular matrix associated peroxidase *Pxdn*, as well as other post-transcriptional regulatory proteins such as *Tdrd7* [35]. Interestingly, *Celf1*^{cKO} lenses also exhibit significant reduction of *Sparc*, whose deficiency is known to cause cataracts in mice [53]. Because the majority of these key cataract-linked genes are preferentially expressed in fiber cells, this suggests that Celf1 has a major function in controlling fiber cell transcriptome. This finding also suggests that significant reduction of these cohorts of cataract-linked genes may contribute to the cataract pathology in *Celf1*^{cKO} lenses. Furthermore, 16 DEGs that are associated with cataracts were found to be elevated in *Celf1*^{cKO} lenses. This suggests that Celf1 is necessary for optimal transcript levels of genes (neither too high, nor too low) that are critical for lens transparency.

While Cat-Map allows identification of a subset of DEGs that are known to be associated with cataracts, to gain further insights into the impact of *Celf1* deficiency on transcripts relevant to lens biology, comparative analysis was performed using iSyTE. This allowed identification of DEGs that exhibit enriched expression in normal lenses, which has previously been found to be predictive of functions in the lens [30,36,54,55,56]. Thus, mis-regulation of such candidates can potentially contribute to lens pathology. The majority of the genes that are reduced upon *Celf1*-deficiency are found to exhibit enriched expression in normal lens development. Thus, it can be hypothesized that the sub-optimal expression levels of these lens-enriched transcripts could contribute to the cataract pathology. Interestingly, the majority of the DEGs that are elevated in *Celf1*^{cKO} lenses are not found to have enriched expression in normal lens development. Thus, this suggests that elevated expression of these transcripts upon *Celf1*-deficiency may contribute to alterations in lens development.

While iSyTE lens-enrichment is helpful, iSyTE data is primarily based on whole lenses. Celf1 is known to be highly expressed in fiber cells, but later in development it is also known to be expressed in epithelial cells, suggesting that it may have a function in both cell types in the lens [9,10]. Therefore, *Celf1*^{cKO} lens DEGs were examined for their preferential expression in normal isolated lens epithelial or fiber cells. The majority of the DEGs were found to be not preferentially expressed in either cell type and were also found to be elevated in *Celf1*^{cKO} lenses. Thus, similar to the iSyTE lens-enrichment analysis, the cell-type (gene expression in either epithelial or fiber cells) specific analysis reinforces the hypothesis that this subset of non-enriched, elevated *Celf1*^{cKO} DEGs should not be expressed at such high transcript levels for proper lens development. Apart from the DEGs that do not exhibit cell type-preferred expression, the majority of the remaining DEGs are preferentially expressed in fiber cells. This suggests that Celf1's impact on the lens is primarily through its function in fiber cells. However, it should be noted that *Celf1*^{cKO} DEGs were identified from whole lens samples, which are expected to have higher levels of transcripts from fiber cells compared to epithelial cells. Thus, the sensitivity toward examination of epithelial DEGs is comparatively low. In future, this can be addressed by performing spatial transcriptomics, for example, by conducting RNA-seq on isolated epithelial and fiber cells from *Celf1*^{cKO} lenses. Alternately, this can also be addressed by performing single cell RNA-seq analysis on *Celf1*^{cKO} lenses.

The above analyses inform on the overall impact of Celf1 on normal lens development. To gain insights into the subset of DEGs that are potential direct RNA targets of Celf1 protein, comparative analysis was performed with CLIP-seq data on CELF1 protein in a human cell line. Although this data is from humans, and not mice, and from a non-lens cell line, this analysis identified 32% of *Celf1*^{cKO} lens DEG transcripts to be directly bound by the CELF1 protein in cellulo. The majority of these direct RNA targets of Celf1 protein are found to be elevated in the *Celf1*^{cKO} lens. This supports the hypothesis that Celf1 protein is necessary to directly bind and repress the expression of hundreds of RNAs that

are not highly enriched in the normal lens. Among the cell-type preferentially expressed DEGs that are also identified in CLIP-seq, the majority are preferentially expressed in fiber cells, suggesting that the direct impact of Celf1 is higher in the lens fiber cells compared to lens epithelial cells.

Of the *Celf1*^{cKO} lens total DEGs, 32% represent a high number of Celf1-direct RNA targets identified, especially considering that this cell line is not lens-derived and thus may not optimally represent lens gene expression, in addition to other caveats such as the suboptimal expression of Celf1 accessory protein/RNA. Thus, it can be hypothesized that the number of direct DEG RNA targets of the Celf1 protein in the lens may be even higher. This can be addressed in the future by performing CLIP-seq on lens cell lines or whole lens tissue. Further, among the direct RNA targets of Celf1 identified by CLIP-seq, is the transcription elongation factor for RNA polymerase II 2 (Ell2), which exhibits highly enriched expression in normal lens development as per iSyTE. Interestingly, Ell2 expression is significantly elevated in *Celf1*^{cKO} lenses, suggesting that Celf1 protein may function to achieve optimal transcript levels of this key regulatory protein, which is involved in transcription control. Thus, this analysis gives new insights into the specific Celf1 targets that are common regardless of the difference in these cell types (lens vs. Hela) and furthermore are also indicative of the similarities in Celf1 function across different species, namely, mouse and human.

To identify pathways that are altered upon Celf1-deficiency, GSEA analysis and GO term analysis was performed on *Celf1*^{cKO} lens total DEGs, as well as the subset of DEGs prioritized by different approaches. Broadly, DEGs reduced in *Celf1*^{cKO} lenses were found to be enriched in pathways that are relevant to lens development (e.g., "structural constituent of eye lens" (GO:0005212) and "lens development in camera-type eye" (GO:0002088)), while the elevated DEGs represented pathways not enriched in normal lenses (e.g., "proton transporting ATPase activity, rotational mechanism" (GO:0046961), "cytoplasmic vesicle membrane" (GO:0030659) and "calcium-dependent protein binding" (GO:0048306)). Additionally, the GO term "lens fiber cell differentiation" (GO:0070306) was enriched for DEGs that are reduced upon Celf1-deficiency and are also preferentially expressed in normal fiber cells, suggesting that key fiber cell expressed genes are under positive control of Celf1. Interestingly, the GO terms "extracellular matrix" (GO:0031012) and "cell projection membrane" (GO:0031253) were enriched for DEGs preferentially expressed in normal epithelial cells, suggesting that Celf1 may have a distinct role in positive regulation of these processes in epithelial cells. Finally, the GO term "positive-regulation of brown fat cell differentiation" (GO:0090335) was enriched in the subset of reduced DEGs that are direct RNA targets of Celf1 protein. This GO term contained the transcription factor Prdm16, which is independently found to exhibit high lens-enriched expression in iSyTE, especially at/beyond secondary fiber cell differentiation at E16.5. Thus, alteration of *Prdm16* expression in *Celf1*^{cKO} lenses, its identification as a direct RNA target of Celf1 protein, and its enriched expression in normal lenses together make

Prdm16 a high-priority candidate whose role in lens development and pathology can be examined in the future.

Together, these various analyses provide insights into lens pathology in *Celf1*^{cKO} mice and identified numerous promising candidates that may be critical for proper lens development. To further prioritize direct RNA targets of Celf1 protein that play a key role in the lens, we used previously reported as well as new microarray transcriptomic analysis on *Celf1*^{cKO} lenses at different postnatal stages (P0 and P6). This allows independent validation of hundreds of DEGs identified by RNA-seq in the *Celf1*^{cKO} lens. Along with the various prioritization approaches described in this report, especially the CLIP-seq analysis that identified direct targets of Celf1 (in addition to other parameters such as, lens-enriched expression in iSyTE, preferential expression in epithelial or fiber cells, etc.), the microarray data identifies high-confidence candidates in the lens for future studies. These analyses show that upon Celf1 deficiency, a cohort of cataract-linked genes are mis-expressed (e.g., the crystallins *Cryab*, *Cryba2*, *Cryba4*, *Crybb1*, *Crybb2*, *Cryga*, *Crygb*, *Crygc* and *Crygd*, the connexins *Gjb6* and *Gja3*, the membrane protein *Bfsp1*, the extracellular matrix associated peroxidase *Pxdn*, and the post-transcriptional regulator *Tdrd7*), in addition to alterations in distinct pathways, thus indicating that multiple factors likely contribute to the manifestation of the cataract defect. Importantly, the present analyses identify as yet unappreciated and novel high-priority candidates in the lens for defining new pathways involved in lens biology (e.g., Ell2 and Prdm16) that likely also contribute to the cataracts resulting from Celf1 deficiency. In particular, the following targets are promising. Ell2, a transcription elongation factor, that functions in a fundamental regulatory process—considered ubiquitously important—in transcription. This is because Ell2 facilitates the release of the RNA Polymerase II from its "pause" in early stages of transcription, which in turn allows the enzyme to proceed with transcription of its target genes. The present study shows that Celf1 functions in controlling the proper dosage of Ell2 in the lens, and thus opens up a new direction in lens research by encouraging the question: are factors like Ell2—that play a critical role in a ubiquitously important regulatory process—specifically recruited for controlling expression of key genes in the lens, a tissue that is known to produce extremely high levels of transcripts that in turn get translated into abundant levels of proteins (e.g., crystallins, which reach concentrations of 450 mg/mL in the lens). Further, this study, by prioritizing Prdm16 which is significantly reduced in Celf1-deficient lenses, has led to the identification of a new transcription factor in the lens, further investigation of which will advance the understanding of gene expression control in this tissue.

5. Conclusions

This study reports on the impact of Celf1-deficiency on the early postnatal lens transcriptome. Application of various analyses such as identification in Cat-Map, lens-enriched expression in iSyTE, preferential expression in epithelial or fiber cells, identification as direct RNA target in CLIP-seq data, and GO term enrichment provides

insights into key transcriptomic events that are under the control of Celf1 in normal lens development and whose alterations contribute to lens pathology, which includes several established cataract-linked genes such as crystallins, connexins, membrane proteins, etc. Finally, along with independent validation by microarrays, this study provides a new cohort of high-confidence genes (e.g., *Ell2*, *Prdm16*, etc.) for future investigations in lens development.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/cells12071070/s1, Figure S1: *Celf1*^{cKO} lens exhibits reduced gamma crystallin expression, Figure S2: Volcano plot and smear plot of *Celf1*^{cKO} lens microarray datasets at P0 and P6, Table S1: Summary of RNA-seq reads and mapping, Table S2: Differential gene expression analysis of *Celf1*^{cKO} lens by RNA-seq, Table S3: iSyTE lens-enriched expression of *Celf1*^{cKO} lens DEGs identified by RNA-seq, Table S4: Epithelial- and fiber cell-preferred expression of *Celf1*^{cKO} lens DEGs identified by RNA-seq, Table S5: GSEA analysis of *Celf1*^{cKO} lens RNA-seq data, Table S6: Gene ontology analysis of *Celf1*^{cKO} lens DEGs, Table S7: Differential gene expression analysis of *Celf1*^{cKO} lens by microarrays at P0 and P6, Table S8: *Celf1*^{cKO} lens DEGs commonly identified by RNA-seq and microarrays.

Author Contributions: Conceptualization, A.D.S. and S.A.L.; methodology, A.D.S., M.D., S.C., Y.A.; validation, A.D.S., S.C., S.A. and B.A.T.W.; formal analysis, M.D., S.C. and D.A.; writing—original draft preparation, A.D.S., M.D., S.C. and S.A.L.; writing—review and editing, A.D.S., M.D., S.C., D.A., S.A., B.A.T.W., Y.A., L.P. and S.A.L.; supervision, L.P., S.A.L.; project administration, S.A.L.; funding acquisition, S.A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Institutes of Health (NIH), grant number R01 EY021505 to S.A.L. D.A. was supported by Knights Templar Pediatric Ophthalmology Career Starter Grant Award. A.D.S. and S.A. were supported by a Fight For Sight Summer Fellowship. S.A. also received support from Sigma Xi through a fellowship.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Association for Research in Vision and Ophthalmology (ARVO) Statement for the use of animals in ophthalmic and vision research, and approved by the Institutional Review Board (Institutional Animal Care and Use Committee (IACUC)) of University of Delaware (protocol number 1226 approved October 1, 2022).

Informed Consent Statement: Not applicable.

Data Availability Statement: The RNA-sequencing data and the microarray data described here be will submitted to the Gene Expression Omnibus.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Harland, R.M.; Grainger, R.M. Xenopus Research: Metamorphosed by Genetics and Genomics. *Trends Genet.* **2011**, *27*, 507–515, doi:10.1016/j.tig.2011.08.003.
2. Graw, J. Mouse Models of Cataract. *J. Genet.* **2009**, *88*, 469–486.
3. Shiels, A.; Hejtmancik, J.F. Inherited Cataracts: Genetic Mechanisms and Pathways New and Old. *Exp Eye Res* **2021**, *209*, 108662, doi:10.1016/j.exer.2021.108662.
4. Cvekl, A.; Zhang, X. Signaling and Gene Regulatory Networks in Mammalian Lens Development. *Trends Genet.* **2017**, *33*, 677–702, doi:10.1016/j.tig.2017.08.001.
5. Lachke, S.A.; Maas, R.L. Building the Developmental Oculome: Systems Biology in Vertebrate Eye Development and Disease. *Wiley Interdiscip Rev Syst Biol Med* **2010**, *2*, 305–323, doi:10.1002/wsbm.59.
6. Lachke, S.A. RNA-Binding Proteins and Post-Transcriptional Regulation in Lens Biology and Cataract: Mediating Spatiotemporal Expression of Key Factors That Control the Cell Cycle, Transcription, Cytoskeleton and Transparency. *Exp Eye Res* **2022**, *214*, 108889, doi:10.1016/j.exer.2021.108889.
7. Dash, S.; Siddam, A.D.; Barnum, C.E.; Janga, S.C.; Lachke, S.A. RNA-Binding Proteins in Eye Development and Disease: Implication of Conserved RNA Granule Components. *Wiley Interdiscip Rev RNA* **2016**, *7*, 527–557, doi:10.1002/wrna.1355.
8. Dash, S.; Dang, C.A.; Beebe, D.C.; Lachke, S.A. Deficiency of the RNA Binding Protein Caprin2 Causes Lens Defects and Features of Peters Anomaly. *Dev. Dyn.* **2015**, *244*, 1313–1327, doi:10.1002/dvdy.24303.
9. Siddam, A.D.; Gautier-Courteille, C.; Perez-Campos, L.; Anand, D.; Kakrana, A.; Dang, C.A.; Legagneux, V.; Méreau, A.; Viet, J.; Gross, J.M.; et al. The RNA-Binding Protein Celf1 Post-Transcriptionally Regulates

P27Kip1 and Dnase2b to Control Fiber Cell Nuclear Degradation in Lens Development. *PLoS Genet.* **2018**, *14*, e1007278, doi:10.1371/journal.pgen.1007278.
10. Aryal, S.; Viet, J.; Weatherbee, B.A.T.; Siddam, A.D.; Hernandez, F.G.; Gautier-Courteille, C.; Paillard, L.; Lachke, S.A. The Cataract-Linked RNA-Binding Protein Celf1 Post-Transcriptionally Controls the Spatiotemporal Expression of the Key Homeodomain Transcription Factors Pax6 and Prox1 in Lens Development. *Hum Genet* **2020**, *139*, 1541–1554, doi:10.1007/s00439-020-02195-7.
11. Dash, S.; Brastrom, L.K.; Patel, S.D.; Scott, C.A.; Slusarski, D.C.; Lachke, S.A. The Master Transcription Factor SOX2, Mutated in Anophthalmia/Microphthalmia, Is Post-Transcriptionally Regulated by the Conserved RNA-Binding Protein RBM24 in Vertebrate Eye Development. *Hum. Mol. Genet.* **2020**, *29*, 591–604, doi:10.1093/hmg/ddz278.
12. Lachke, S.A.; Alkuraya, F.S.; Kneeland, S.C.; Ohn, T.; Aboukhalil, A.; Howell, G.R.; Saadi, I.; Cavallesco, R.; Yue, Y.; Tsai, A.C.-H.; et al. Mutations in the RNA Granule Component TDRD7 Cause Cataract and Glaucoma. *Science* **2011**, *331*, 1571–1576, doi:10.1126/science.1195970.
13. Barnum, C.E.; Al Saai, S.; Patel, S.D.; Cheng, C.; Anand, D.; Xu, X.; Dash, S.; Siddam, A.D.; Glazewski, L.; Paglione, E.; et al. The Tudor-Domain Protein TDRD7, Mutated in Congenital Cataract, Controls the Heat Shock Protein HSPB1 (HSP27) and Lens Fiber Cell Morphology. *Hum. Mol. Genet.* **2020**, *29*, 2076–2097, doi:10.1093/hmg/ddaa096.
14. Lorén, C.E.; Schrader, J.W.; Ahlgren, U.; Gunhaga, L. FGF Signals Induce Caprin2 Expression in the Vertebrate Lens. *Differentiation* **2009**, *77*, 386–394, doi:10.1016/j.diff.2008.11.003.
15. Nakazawa, K.; Shichino, Y.; Iwasaki, S.; Shiina, N. Implications of RNG140 (Caprin2)-Mediated Translational Regulation in Eye Lens Differentiation. *J Biol Chem* **2020**, *295*, 15029–15044, doi:10.1074/jbc.RA120.012715.
16. Shao, M.; Lu, T.; Zhang, C.; Zhang, Y.-Z.; Kong, S.-H.; Shi, D.-L. Rbm24 Controls Poly(A) Tail Length and Translation Efficiency of Crystallin MRNAs in the Lens via Cytoplasmic Polyadenylation. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 7245–7254, doi:10.1073/pnas.1917922117.
17. Chen, J.; Wang, Q.; Cabrera, P.E.; Zhong, Z.; Sun, W.; Jiao, X.; Chen, Y.; Govindarajan, G.; Naeem, M.A.; Khan, S.N.; et al. Molecular Genetic Analysis of Pakistani Families With Autosomal Recessive Congenital Cataracts by Homozygosity Screening. *Invest. Ophthalmol. Vis. Sci.* **2017**, *58*, 2207–2217, doi:10.1167/iovs.17-21469.
18. Fernández-Alcalde, C.; Nieves-Moreno, M.; Noval, S.; Peralta, J.M.; Montaño, V.E.F.; Del Pozo, Á.; Santos-Simarro, F.; Vallespín, E. Molecular and Genetic Mechanism of Non-Syndromic Congenital Cataracts. Mutation Screening in Spanish Families. *Genes (Basel)* **2021**, *12*, doi:10.3390/genes12040580.
19. Kandaswamy, D.K.; Prakash, M.V.S.; Graw, J.; Koller, S.; Magyar, I.; Tiwari, A.; Berger, W.; Santhiya, S.T. Application of WES Towards Molecular Investigation of Congenital Cataracts: Identification of Novel Alleles and Genes in a Hospital-Based Cohort of South India. *Int J Mol Sci* **2020**, *21*, doi:10.3390/ijms21249569.
20. Tan, Y.-Q.; Tu, C.; Meng, L.; Yuan, S.; Sjaarda, C.; Luo, A.; Du, J.; Li, W.; Gong, F.; Zhong, C.; et al. Loss-of-Function Mutations in TDRD7 Lead to a Rare Novel Syndrome Combining Congenital Cataract and Nonobstructive Azoospermia in Humans. *Genet. Med.* **2019**, *21*, 1209–1217, doi:10.1038/gim.2017.130.
21. Tanaka, T.; Hosokawa, M.; Vagin, V.V.; Reuter, M.; Hayashi, E.; Mochizuki, A.L.; Kitamura, K.; Yamanaka, H.; Kondoh, G.; Okawa, K.; et al. Tudor Domain Containing 7 (Tdrd7) Is Essential for Dynamic Ribonucleoprotein (RNP) Remodeling of Chromatoid Bodies during Spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10579–10584, doi:10.1073/pnas.1015447108.
22. Zheng, C.; Wu, M.; He, C.-Y.; An, X.-J.; Sun, M.; Chen, C.-L.; Ye, J. RNA Granule Component TDRD7 Gene Polymorphisms in a Han Chinese Population with Age-Related Cataract. *J. Int. Med. Res.* **2014**, *42*, 153–163, doi:10.1177/0300060513504702.
23. Choquet, H.; Melles, R.B.; Anand, D.; Yin, J.; Cuellar-Partida, G.; Wang, W.; 23andMe Research Team; Hoffmann, T.J.; Nair, K.S.; Hysi, P.G.; et al. A Large Multiethnic GWAS Meta-Analysis of Cataract Identifies New Risk Loci and Sex-Specific Effects. *Nat Commun* **2021**, *12*, 3595, doi:10.1038/s41467-021-23873-8.
24. Bauermeister, D.; Claußen, M.; Pieler, T. A Novel Role for Celf1 in Vegetal RNA Localization during Xenopus Oogenesis. *Dev Biol* **2015**, *405*, 214–224, doi:10.1016/j.ydbio.2015.07.005.
25. Vlasova-St Louis, I.; Dickson, A.M.; Bohjanen, P.R.; Wilusz, C.J. CELFish Ways to Modulate MRNA Decay. *Biochim. Biophys. Acta* **2013**, *1829*, 695–707, doi:10.1016/j.bbagrm.2013.01.001.
26. Zheng, Y.; Miskimins, W.K. CUG-Binding Protein Represses Translation of P27Kip1 MRNA through Its Internal Ribosomal Entry Site. *RNA Biol* **2011**, *8*, 365–371.

27. Ladd, A.N.; Charlet, N.; Cooper, T.A. The CELF Family of RNA Binding Proteins Is Implicated in Cell-Specific and Developmentally Regulated Alternative Splicing. *Mol. Cell. Biol.* **2001**, *21*, 1285–1296, doi:10.1128/MCB.21.4.1285-1296.2001.
28. Vlasova, I.A.; Tahoe, N.M.; Fan, D.; Larsson, O.; Rattenbacher, B.; Sternjohn, J.R.; Vasdewani, J.; Karypis, G.; Reilly, C.S.; Bitterman, P.B.; et al. Conserved GU-Rich Elements Mediate MRNA Decay by Binding to CUG-Binding Protein 1. *Mol. Cell* **2008**, *29*, 263–270, doi:10.1016/j.molcel.2007.11.024.
29. Rowan, S.; Conley, K.W.; Le, T.T.; Donner, A.L.; Maas, R.L.; Brown, N.L. Notch Signaling Regulates Growth and Differentiation in the Mammalian Lens. *Dev. Biol.* **2008**, *321*, 111–122, doi:10.1016/j.ydbio.2008.06.002.
30. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* **2013**, *29*, 15–21, doi:10.1093/bioinformatics/bts635.
31. Liao, Y.; Smyth, G.K.; Shi, W. FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30*, 923–930, doi:10.1093/bioinformatics/btt656.
32. Chen, Y.; Lun, A.T.L.; Smyth, G.K. From Reads to Genes to Pathways: Differential Expression Analysis of RNA-Seq Experiments Using Rsubread and the EdgeR Quasi-Likelihood Pipeline. *F1000Res* **2016**, *5*, 1438, doi:10.12688/f1000research.8987.2.
33. Shiels, A.; Bennett, T.M.; Hejtmancik, J.F. Cat-Map: Putting Cataract on the Map. *Mol. Vis.* **2010**, *16*, 2007–2015.
34. Kakrana, A.; Yang, A.; Anand, D.; Djordjevic, D.; Ramachandruni, D.; Singh, A.; Huang, H.; Ho, J.W.K.; Lachke, S.A. ISyTE 2.0: A Database for Expression-Based Gene Discovery in the Eye. *Nucleic Acids Res.* **2018**, *46*, D875–D885, doi:10.1093/nar/gkx837.
35. Zhao, Y.; Zheng, D.; Cvekl, A. A Comprehensive Spatial-Temporal Transcriptomic Analysis of Differentiating Nascent Mouse Lens Epithelial and Fiber Cells. *Exp. Eye Res.* **2018**, *175*, 56–72, doi:10.1016/j.exer.2018.06.004.
36. Le Tonquèze, O.; Gschloessl, B.; Legagneux, V.; Paillard, L.; Audic, Y. Identification of CELF1 RNA Targets by CLIP-Seq in Human HeLa Cells. *Genom Data* **2016**, *8*, 97–103, doi:10.1016/j.gdata.2016.04.009.
37. Yu, G.; Wang, L.-G.; Han, Y.; He, Q.-Y. ClusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters. *OMICS* **2012**, *16*, 284–287, doi:10.1089/omi.2011.0118.
38. Thorvaldsdóttir, H.; Robinson, J.T.; Mesirov, J.P. Integrative Genomics Viewer (IGV): High-Performance Genomics Data Visualization and Exploration. *Brief Bioinform* **2013**, *14*, 178–192, doi:10.1093/bib/bbs017.
39. Beisang, D.; R., P.; Vlasova-St. Louis, I.A. CELF1, a Multifunctional Regulator of Posttranscriptional Networks. In *Binding Protein*; Abdelmohsen, K., Ed.; InTech, 2012 ISBN 978-953-51-0758-3.
40. Kress, C.; Gautier-Courteille, C.; Osborne, H.B.; Babinet, C.; Paillard, L. Inactivation of CUG-BP1/CELF1 Causes Growth, Viability, and Spermatogenesis Defects in Mice. *Mol. Cell. Biol.* **2007**, *27*, 1146–1157, doi:10.1128/MCB.01009-06.
41. Cifdaloz, M.; Osterloh, L.; Graña, O.; Riveiro-Falkenbach, E.; Ximénez-Embún, P.; Muñoz, J.; Tejedo, C.; Calvo, T.G.; Karras, P.; Olmeda, D.; et al. Systems Analysis Identifies Melanoma-Enriched pro-Oncogenic Networks Controlled by the RNA Binding Protein CELF1. *Nat Commun* **2017**, *8*, 2249, doi:10.1038/s41467-017-02353-y.
42. Chaudhury, A.; Cheema, S.; Fachini, J.M.; Kongchan, N.; Lu, G.; Simon, L.M.; Wang, T.; Mao, S.; Rosen, D.G.; Ittmann, M.M.; et al. CELF1 Is a Central Node in Post-Transcriptional Regulatory Programmes Underlying EMT. *Nat Commun* **2016**, *7*, 13362, doi:10.1038/ncomms13362.
43. House, R.P.; Talwar, S.; Hazard, E.S.; Hill, E.G.; Palanisamy, V. RNA-Binding Protein CELF1 Promotes Tumor Growth and Alters Gene Expression in Oral Squamous Cell Carcinoma. *Oncotarget* **2015**, *6*, 43620–43634, doi:10.18632/oncotarget.6204.
44. Matsui, T.; Sasaki, A.; Akazawa, N.; Otani, H.; Bessho, Y. Celf1 Regulation of Dmrt2a Is Required for Somite Symmetry and Left-Right Patterning during Zebrafish Development. *Development* **2012**, *139*, 3553–3560, doi:10.1242/dev.077263.
45. Timchenko, N.A.; Cai, Z.J.; Welm, A.L.; Reddy, S.; Ashizawa, T.; Timchenko, L.T. RNA CUG Repeats Sequester CUGBP1 and Alter Protein Levels and Activity of CUGBP1. *J. Biol. Chem.* **2001**, *276*, 7820–7826, doi:10.1074/jbc.M005960200.
46. Philips, A.V.; Timchenko, L.T.; Cooper, T.A. Disruption of Splicing Regulated by a CUG-Binding Protein in Myotonic Dystrophy. *Science* **1998**, *280*, 737–741.

47. Xiao, J.; Jin, S.; Wang, X.; Huang, J.; Zou, H. CELF1 Selectively Regulates Alternative Splicing of DNA Repair Genes Associated With Cataract in Human Lens Cell Line. *Biochem Genet* **2022**, doi:10.1007/s10528-022-10324-2.
48. Xiao, J.; Tian, X.; Jin, S.; He, Y.; Song, M.; Zou, H. CELF1 Promotes Matrix Metalloproteinases Gene Expression at Transcriptional Level in Lens Epithelial Cells. *BMC Ophthalmol* **2022**, *22*, 122, doi:10.1186/s12886-022-02344-8.
49. Nishimoto, S.; Kawane, K.; Watanabe-Fukunaga, R.; Fukuyama, H.; Ohsawa, Y.; Uchiyama, Y.; Hashida, N.; Ohguro, N.; Tano, Y.; Morimoto, T.; et al. Nuclear Cataract Caused by a Lack of DNA Degradation in the Mouse Eye Lens. *Nature* **2003**, *424*, 1071–1074, doi:10.1038/nature01895.
50. Gilmour, D.T.; Lyon, G.J.; Carlton, M.B.; Sanes, J.R.; Cunningham, J.M.; Anderson, J.R.; Hogan, B.L.; Evans, M.J.; Colledge, W.H. Mice Deficient for the Secreted Glycoprotein SPARC/Osteonectin/BM40 Develop Normally but Show Severe Age-Onset Cataract Formation and Disruption of the Lens. *EMBO J* **1998**, *17*, 1860–1870, doi:10.1093/emboj/17.7.1860.
51. Anand, D.; Lachke, S.A. Systems Biology of Lens Development: A Paradigm for Disease Gene Discovery in the Eye. *Exp. Eye Res.* **2017**, *156*, 22–33, doi:10.1016/j.exer.2016.03.010.
52. Lachke, S.A.; Ho, J.W.K.; Kryukov, G.V.; O'Connell, D.J.; Aboukhalil, A.; Bulyk, M.L.; Park, P.J.; Maas, R.L. ISyTE: Integrated Systems Tool for Eye Gene Discovery. *Invest. Ophthalmol. Vis. Sci.* **2012**, *53*, 1617–1627, doi:10.1167/iovs.11-8839.
53. Anand, D.; Kakrana, A.; Siddam, A.D.; Huang, H.; Saadi, I.; Lachke, S.A. RNA Sequencing-Based Transcriptomic Profiles of Embryonic Lens Development for Cataract Gene Discovery. *Hum. Genet.* **2018**, *137*, 941–954, doi:10.1007/s00439-018-1958-0.
54. Aryal, S.; Anand, D.; Hernandez, F.G.; Weatherbee, B.A.T.; Huang, H.; Reddy, A.P.; Wilmarth, P.A.; David, L.L.; Lachke, S.A. MS/MS in Silico Subtraction-Based Proteomic Profiling as an Approach to Facilitate Disease Gene Discovery: Application to Lens Development and Cataract. *Hum. Genet.* **2020**, *139*, 151–184, doi:10.1007/s00439-019-02095-5.

Supplementary Figure

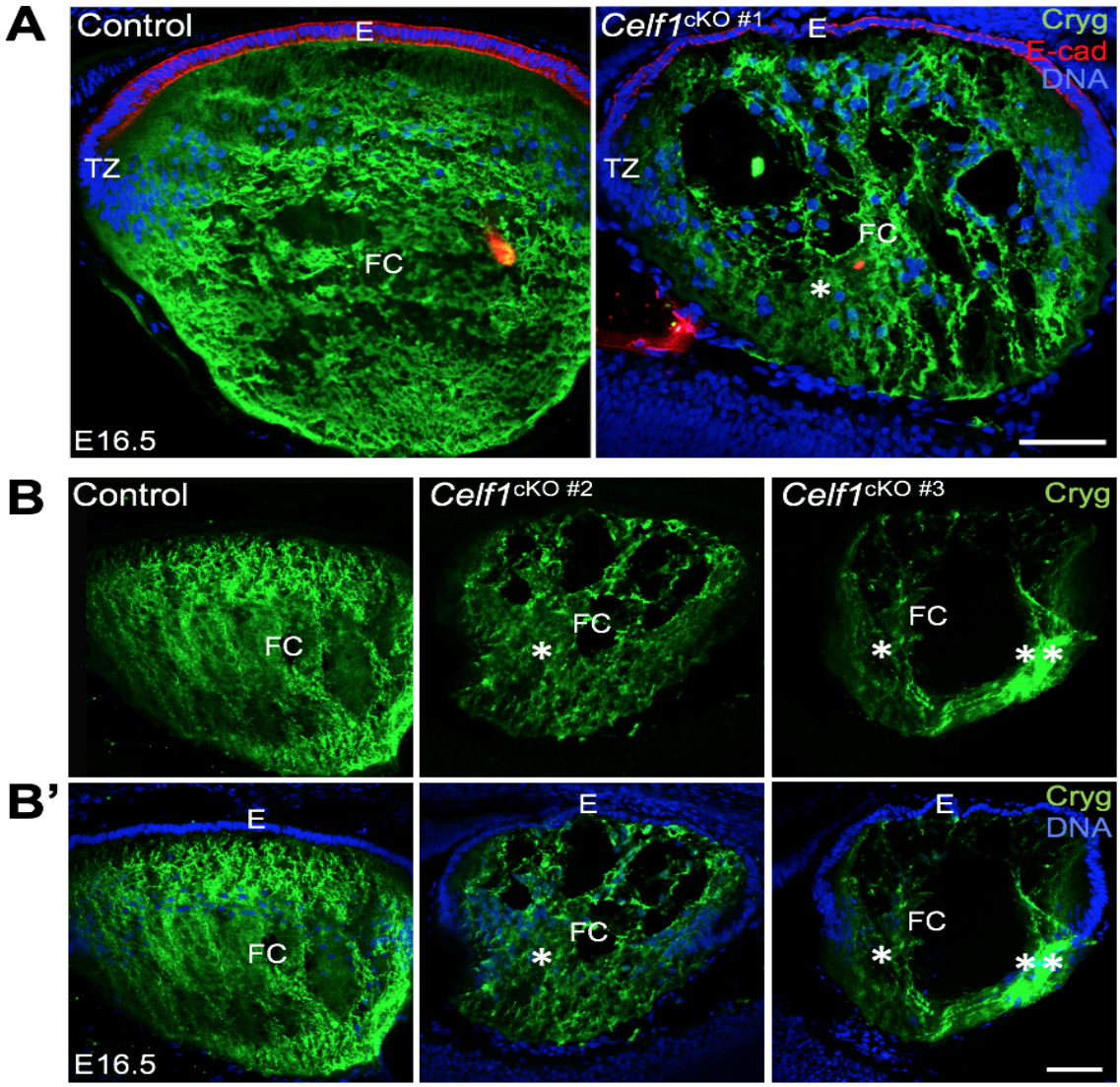


Fig S1. *Celf1*^{cKO} lenses exhibits reduced gamma crystallin expression. (**A**) Immunostaining shows gamma Crystallin (Cryg, green; asterisk) is reduced in fiber cells in E16.5 *Celf1*$^{cKO\ \#1}$ mouse lens. E-cadherin (E-cad, red) staining is used as a marker for the epithelium. DNA is stained by DRAQ5. (**B**) Two additional replicates of *Celf1*$^{cKO\ \#2,\ \#3}$ lenses stained with Cryg, and an additional replicate of control and (**B'**) merged with DNA. Reduced Cryg is indicated by asterisk. Please note: *Celf1*^{cKO} lenses exhibits abnormal retention of nuclei and large gaps in the tissue, which are not sectioning artifacts except for *Celf1*$^{cKO\ \#3}$ lens in a specific region where a folded tissue is indicated by two asterisks. These lens phenotypic defects are described in detail in Siddam *et al.* (2018) *PLOS Genetics* 14(3):e1007278.Abbr.: E, epithelium; TZ, transition zone; FC, fiber cells. Scale bar, 75 μm.

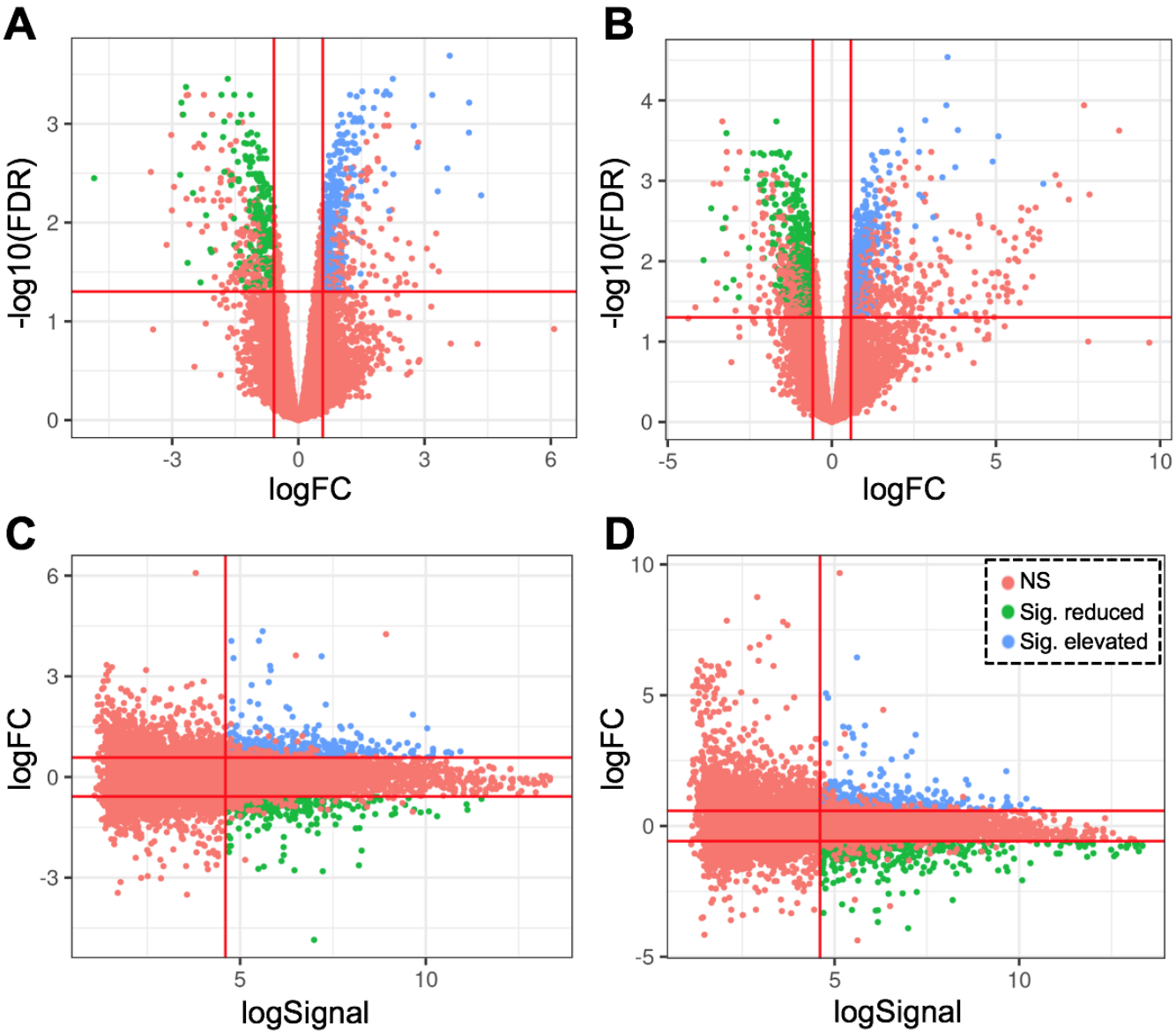


Fig S2. Volcano plot and smear plot of *Celf1*^{cKO} lens microarray datasets at P0 and P6. Volcano plots at postnatal stages (A) P0 and (B) P6, and Smear plots at stages (C) P0 and (D) P6. These plots show the differentially expressed genes (DEGs) between *Celf1*^{cKO} and control lens microarrays at P0 and P6 at different cut-offs. Cut-offs used were FDR <0.05, logFC >0.58, logSignal > 4.6. NS, not significant.

Chapter II: Alternative splicing in the lens: identification of CELF1-controlled RNA by a multi-omic analysis

A. Introduction

CELF1 plays a crucial role as a key regulator of gene expression during lens development. In a previous study, my host team demonstrated that the deficiency of *Celf1* in mouse lenses resulted in significant mis-regulation of the lens transcriptome, correlated with cataract formation. Notably,it has shown that CELF1 regulates *Dnase2b, Cdkn1b, Pax6* and *Prox1* in the lens[140,371]. Both PAX6 and PROX1 are major transcription factors involved in lens development and homeostasis. It is therefore highly plausible that numerous mis-regulated genes identified in the previous chapter [371] are indirectly controlled by CELF1 through its influence on *Pax6*, *Prox1*, or other mRNA encoding transcriptional or post-transcriptional regulators that are directly regulated by CELF1 itself.

To gain a better understanding of the role of CELF1 in the lens and the molecular mechanisms underlying lens opacification in its absence, it would be highly beneficial to identify the genes directly targeted by CELF1. Multiple technologies have been employed in previous studies, to identify the interactions of CELF1 with RNA (pre-mRNA or mRNA) in various cell types. However, these datasets may present some limitations due to their distinct contexts from the developing lens. The radically different transcriptomes and the presence of other post-transcriptional regulators could result in specific regulatory patterns observed only within the lens. Therefore, to identify the genes regulated by CELF1 in the lens, it is crucial to identify the specific ligands of CELF1 within the lens itself. This would provide valuable insights into the direct molecular interactions and regulatory pathways involving CELF1 in lens development and the pathogenesis of lens opacification.

In this study, we used a iCLIP-seq analysis on the lens of adult mice to identify the RNA molecules that interact with CELF1. iCLIP-seq analysis not only allows for the identification of RNA targets, but also enables the precise localization of CELF1 binding sites on the RNA. This information about localization is crucial in prioritizing CELF1 RNA targets. As described in the introduction, CELF1 exhibits different post-transcriptional activities depending on its cellular localization. When CELF1 is located in the cytoplasm, it primarily regulates mRNA stability and/or translation. In these cases, CELF1 often binds to the 3' untranslated region (UTR) of the mRNA. On the other hand, when CELF1 is nuclear, it regulates alternative splicing of pre-mRNA into mRNA. To do

so, CELF1 binds to the pre-mRNA near the alternative splice junction. Therefore, the precise localization of CELF1 binding sites on their ligand RNAs can be correlated with RNA-seq data from lenses of mice deficient in *Celf1* to identify nuclear pre-mRNA targets of CELF1. This approach enables the association of alternative splicing events with specific CELF1 binding sites, providing valuable insights into the regulatory role of CELF1 in the lens.

This work is focused on CELF1-mediated regulation of alternative splicing in the lens. I used two RNA-seq datasets to identify the repertoire of mRNA that are alternatively spliced in the absence of CELF1. One dataset is the one already used in our previous article and corresponds to control and *Celf1* cKO newborn mouse lenses[3]. The other one is from control and constitutive *Celf1* KO adult mouse lenses and has not been published yet.

By integrating these two datasets with the iCLIPseq dataset, I found that CELF1 controls the splicing of *Ablim1, Ctnna2, Clta, Septin8, Sptbn1, Ywhae* and *Ank2* in the lens.

B. Materials and Methods

B.1. iCLIP-seq analyses

B.1.a Generation of the iCLIP library

The iCLIP library was generated from adult mouse lenses in my host team before I joined it following the protocol from Huppertz et al[372]. The protocol to generate iCLIP libraries is summarized in Figure 13. Briefly, a covalent bond is created between CELF1 and the RNAs (pre-mRNAs and mRNAs) by an exposure of the freshly dissected lenses to UV-C light. After cell lysis, the RNA are clived into small RNA fragments and then immuno-purified using a CELF1-specific antibody. After denaturing electrophoresis, the proteins and protein-RNA complexes are transferred to a pure nitrocellulose membrane. As pure nitrocellulose interacts with proteins but not RNA, the only RNA remaining onto the membrane following the transfer are those covalently bound to an RNA-binding protein. The nitrocellulose membrane is then cut above the position of CELF1 (to take into account the linked RNA fragment that increases the apparent molecular weight of CELF1). The RNA previously linked to CELF1 are then eluted from the nitrocellulose membrane by partial digestion with proteinase K. This leaves a small polypeptide on the RNA binding site. An RNA adaptor is linked to the 3’ end of these RNA. This allows the reverse-transcription of the RNA into cDNA, with a DNA adaptor presenting a complementary sequence to the RNA adaptor, a restriction site, a second adaptator sequence and a specific barcode. The reverse-transcriptase is stopped by the remaining polypeptide close to the binding site. Ultimately, this will allow the precise localization of this binding sites as the 5' most nucleotides of the sequenced fragments. Then, the cDNA are circularized and linearized using the restriction site added in the DNA adapter. Thus the sequence that is complementary to the RNA will be flanked by known adaptor sequences. This enables the subsequent amplification and deep-sequencing of the library (Figure 13).

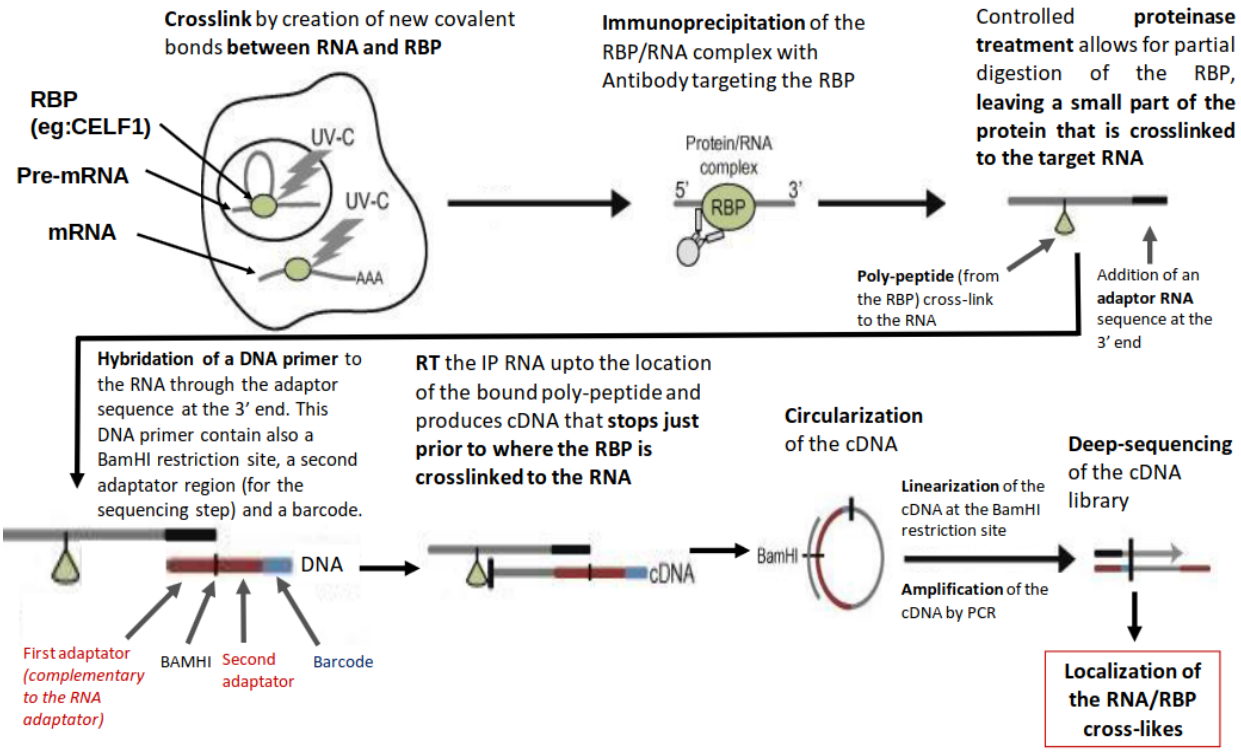


***Figure 13:* Summarized protocol to generate an iCLIP-seq library.** *Adapted from Huppertz et al* [372]

B.1.b Library Sequencing and analysis

While the iCLIP library had been prepared before I joined my PhD team, I was in charge of analysing the sequencing data. The RNA from the iCLIP library were sequenced by an Illumina sequencer (strand-specific, single-end 50 bp reads). The raw reads were analysed by the specific software for the analysis of iCLIP-seq data iCount (2.0.1)[373]. The raw reads were trimmed to remove the adaptor sequence and demultiplexed using the specific barcode. Only reads with at least 19 nucleotide were kept. The trimmed sequences were mapped onto the mouse genome (GRCm38) with the STAR[374] software following iCount instructions. Only the reads mapping a single genomic region were kept. Since the reads end at the position where the reverse-transcription was stopped by the polypeptide, the cross-link site between the RNA and CELF1 is localized.

The significance of the identified cross-link sites was calculated based on the method proposed by Gage's teams[375]. This allows to obtain a FDR (False Discovery Rate) score for each cross-link signal based on their score (the number of reads associated with this localization) and the score of the neighboring cross-link signals (within a half windows of 3 nt) that are compared to a random distribution of these scores across the RNA. The FDR score allows the selection of significant cross-link signals, particularly in highly expressed

genes that may present numerous cross-link signals due to an aspecific interaction with CELF1. The significant cross-link signals are named peaks.

The peaks were then clustered if they are closer than 20 nt. The score of the cluster corresponds to the sum of all the scores of the peaks and of non-significant cross-link signals within the cluster. Only the clusters with a size greater than 5 nt and a score higher than 10 were retained in this study.

B.2 RNA-seq from adult mouse lenses dataset

The constitutive *Celf1* KO mice are described in Kress et al.[284]. As for the iCLIP-seq, I was in charge of analysing the sequencing data obtained from biological samples that were prepared before I joined my PhD team. Three biological replicates for each control and each KO mice were obtained. Each biological replicate corresponds to a single mouse. The RNA were extracted by the RNAeasy kit (Quiagen) following the manufacturer protocol. The quality of the RNA was validated by a bioanalyser at the University of Rennes. Samples with a RIN (RNA Integrity number) score of at least 5 were used to generate the library. This library was sequenced by an Illumina (Miseq) sequencer (strand specific; paired-end; 150 bp).

The raw reads were trimmed to remove the adaptor sequences. The trimmed reads were mapped using the software STAR[374] (2.4.2a) on the mouse genome (GRCm38). Two successive mappings of the reads were made. This strategy allows a better quantification of non-annotated junctions. During the first mapping the reads that cannot be mapped according to the used annotation file are used by STAR to discover new exon-exon junctions. To obtain accurate quantification of these newly discovered junctions, a second mapping is made, adding the newly described junctions in the annotation file[376,377].

B.3 RNA-seq from newborn mouse lenses dataset

The generation of the new-born RNA-seq library and sequencing has been previously described[371]. To accurately quantify the junctions in these RNA-seq data, a similar strategy as that used for the adult dataset was implemented. The STAR software (version 2.7.8a) was used for mapping and analysis, following the two-step mapping approach as described for the adult dataset.

B.4 Identification of significant alternative spliced junctions

For both adult and newborn datasets, the R package DESXeq (1.36.0)[378] was used with a junction-centric approach. Only the junctions that were used at least six times in different replicates were retained for the analysis. DEXSeq

identifies the significant alternatively spliced junctions (sigAS), if they present a FDR<0.05 and an absolute log2 Fold Change (|log2FC|)>0.5.

B.5 Identification of candidate genes

The bash tool bedtools (2.27.1)[379] was used to integrate the different deep-sequencing analyses. It was used to identify genes with at least one CELF1 cluster inside or closer than 300bp to a significant alternatively spliced junction (sigAS) in either the adult or the newborn dataset.

B.6 RT-PCR validation

The RNA extracted from lenses of either control or *Celf1* KO mice were extracted by the RNAeasy kit (Quiagen) following the manufacturer protocol. For each condition, two biological replicates were used. These RNA were reverse-transcribed with the kit RT SuperScript™ II (Thermo Fisher), and amplified with specific primers using the kit GoTaq (Promega), following the manufacturer protocol. The primers used to quantify the different levels of each mRNA isoforms for the candidate genes are described in Table 8. Transcript levels were normalized to the housekeeping gene *B2m*. The PCR products were migrated on 8% acrylamide gels, and the band size was used to validate the specific amplification of the targeted transcript. The ImageJ software (1.53f51) was used to quantify the band signal on the gel. The transcript signal was calculated as the measured band signal/PCR product size (bp).

Table 8: Primers used for the RT-PCR validation of the alternative splicing events of the candidate genes. *FW: Forward; Rv: Reverse*

Primer target	Primer sequence
Ablim1 Fw exon 20	TAGCCGGCACAGTTACACTC
Ablim1 Rv exon 23	TGCTGTTTGTAGATGGGTGG
Nupr1 Fw exon 1	GCATGGTGTGCCATGATGAAGC
Nupr1 Fw exon 2	GAGGCGAGAGCTTTCCACG
Nupr1 Rv exon 3	AGGTTTAGAGGTTGCTGGGAAGG
Nupr1 Rv exon 4	GCTAGGGCGGTTGGTATTGG
Bfsp1 Fw exon 1	CTCGACGAGTTCCGCAGC
Bfsp1 Rv exon 2	TGTTAAGCCGTTCCAGCATTTC
Bfsp1 Rv exon 3	GCAGAAACTGTGCCTCCAACTG
Clta Fw exon 4	AGCTGGAAGAGTGGTATGCG
Clta Rv exon 7	TTGGCCTGTTTGCTGGACTT
Ivns1abp Fw exon 7	GCGTAGCATCTGGGAGAATG
Ivns1abp Rv exon 8 (extended)	TTCTCGGCTCTTCAGTCTTGG
Ivns1abp Rv exon 9	TGGTCATCATCACTGCCAAAC
Sptbn1 Fw exon 31	TTGGAGCTACTGGAAGTGCG
Sptbn1 Rv exon 32 (extended)	ATGAATGGTCACTGGCTGTCC
Sptbn1 Rv exon 33	GCTGGAAGTGCAGATTTGGC
Klc1 Fw exon 14	CCTTGGAGCACTTTACCGACG
Klc1 Rv exon 15	GCAGCACATGCCTCACTCCT
Klc1 Rv exon 16	TCCCTTCAGCTTCCTAACCA
Ctnna2 Fw exon 1	CGAGAAACTCCCACCGACC
Ctnna2 Rv exon 3	AGTCTTTCCACTGTGAGTGTCC
Ywhae Fw exon 5	TGACGCAATTGCAGAACTGG
Ywhae Rv exon 6	TGCAGCGCTTCTTTATTCTGC
Ank2 Fw exon 44	GAGTGACACCCAGCAGTCAG
Ank2 Rv exon 47	GAGGTCGTCTTGGTCCAGTG
B2m Fw	TGGTGCTTGTCTCACTGACC
B2m Rv	CCGTTCTTCAGCATTTGGAT

B.7 Structural prediction of the candidate protein isoforms

The prediction tool AlphaFold2[380] (2.0.0) is used to predict a structural difference between the protein isoforms encoded by the different mRNA isoforms of the validated candidate genes. The protein structural prediction are visualized with the software ChimeraX (1.2.5)[381].

B.8 Cat-Map: Cataract Associated Genes

The genes identified through the iCLIP analysis and known to be associated with cataracts were cross-referenced with the CatMap database (vOct 21)[41]. The CatMap database contains a collection of 454 genes associated with cataracts using human gene names as identifiers. To establish correspondence between the iCLIP-identified genes (which were annotated using mouse gene names) and the genes listed in CatMap, a comparison was performed to identify matching gene names between the two datasets.

An hyper-geometric test was made using the R package stats (4.0.3), with 14,609 genes considered as expressed in the lens. The 14,609 genes were identified in the adult RNA-seq as the genes presenting at least one used junction.

C. Results

C.1. iCLIP-seq analyses

The iCLIP-seq analyses allow the identification and localization of cross-link signals between CELF1 and RNA in the lens. We made statistical tests to determine if each retrieved cross-link signal (sum of the sequencing reads that end at a given genomic position) is significant compared to a random distribution of cross-link signals all along each transcript. This is particularly useful for genes that are highly expressed in the lens such as the crystallin genes. For these genes, there is a high probability to detect a CELF1 cross-link signal only because of the high abundance of the transcript rather than owing to a specific interaction between CELF1 and the RNA. Comparing the score of each experimental cross-link signal within one given RNA with the scores obtained after randomly distributing all along the same RNA the reads mapped on the same gene allows to discard these false positives. This way, I identified 106,220 significant cross-link signals (hereafter referred to as peaks). The peaks that are in the same region (less than 20 nt apart) were clustered together. Only the clusters with a minimum score of 11 (sum of the scores of all the sequenced reads within the cluster) and a minimal length of 6 nt were kept and considered as CELF1 binding sites (Figure 14 A).

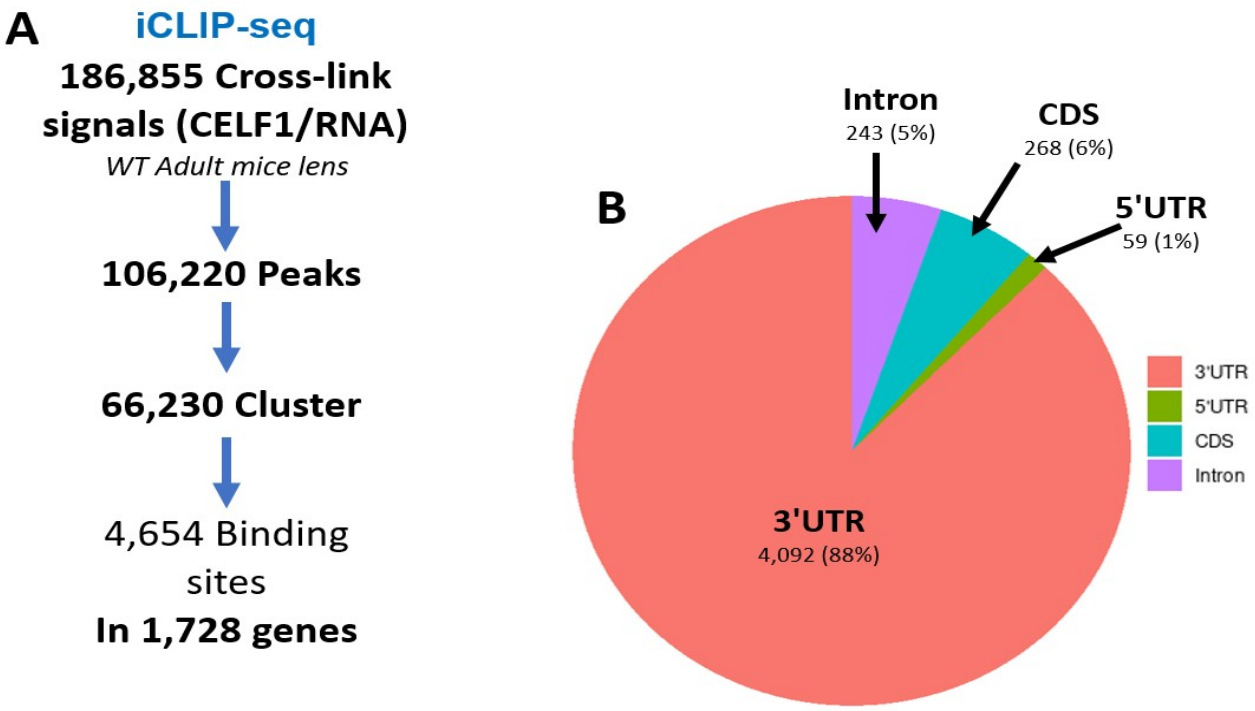


Figure 14: iCLIP-seq identifies CELF1 binding sites on RNA in adult mouse lenses. *(A) Schematic of the pipeline used to analyse the iCLIP-seq data. Cross-link signal, any position within the genome corresponding to the 5' end of a read (several reads can end on the same genome position, and the number of reads correspond to the signal score); Peak, significant signal, where significance arises from the comparison with a random distribution of the reads along the gene; Cluster, group of peaks in the same region; binding site, cluster above a threshold in terms of score and length. (B), Distribution of the CELF1 binding sites on the RNA of coding genes. UTR :Untranslated region; CDS: Coding sequence.*

With this approach, I identified 4,654 CELF1 binding sites in 1,728 genes expressed in the lens. In previous studies[257], CELF1 was shown to preferentially bind to 3' untranslated region (3'UTR). This is similar to what we show here (4,092/4,654 sites, 87.9%, are in the 3'UTR, Figure 14B). However, 253 (5%) binding sites are also present in intronic regions, confirming that in the lens CELF1 binds pre-mRNA suggesting a nuclear activity.

The 1,728 genes whose RNA interact with CELF1 in the lens are highly enriched in genes associated with cataract according to the database CatMap. CatMap contains 456 genes, and 96 of them have CELF1 binding sites (p-value = 8.56e-09, hypergeometric test, considering that 14,609 genes are expressed in the adult lens with at least one exon-exon junctional read). Among them we find *Cdkn1b*, *Pax6* and *Prox1* that have already been described as regulated by CELF1 in the lens[59,140]. Furthermore, other transcripts coding for proteins known for their role during lens development have CELF1 binding sites, such as transcription factors (e.g. *Sox2, Six3, Yap1)*, other RBPs, cytoskeleton and extracellular compound genes, and crystallins.

C.2 Identification of differentially spliced RNAs

To identify the RNA that are differentially spliced in the absence of CELF1, I used two RNA-seq datasets. The first dataset corresponds to new-born mice conditionally inactivated for *Celf1* through an eye-specific Cre promoter (*Pax6*), compared to control lenses of the same age. It has already been described in our previous paper but only for the differential expression and not for alternative splicing[371]. The second dataset corresponds to lens samples from adult mice with a constitutive disruption of *Celf1*, achieved through the insertion of an nLACZ sequence in the first exon of the *Celf1* gene. Here the control lenses are from adult mice. This dataset will be referred to as the "Adult" dataset for the remainder of the study.

To identify splice events that are significantly different in *Celf1*-inactivated and control lenses, a junction-centric approach was employed. This approach allows the identification of non-annotated splicing events, taking into consideration the possibility that certain alternative splicing events specific to the lens might not be present in the available mouse genomic annotation. Thus, a two-step mapping allows the identification of new junctions not previously annotated and their correct quantification (see Materials and Methods)[377]. Among all the junctions that were identified in the lens, the junctions are considered as significantly different between control and *Celf1* cKO or KO lenses at the following thresholds: FDR < 0.05 and |log2FC| > 0.5 (Figure 15 A', B').

In the new-born dataset, a total of 717 significant alternatively spliced junctions (sigAS) were identified (Figure 15A), while in the adult dataset, 1,969 sigAS were detected (Figure 15B). These sigAS correspond to 438 genes in the new-born dataset and 1,061 genes in the adult dataset.

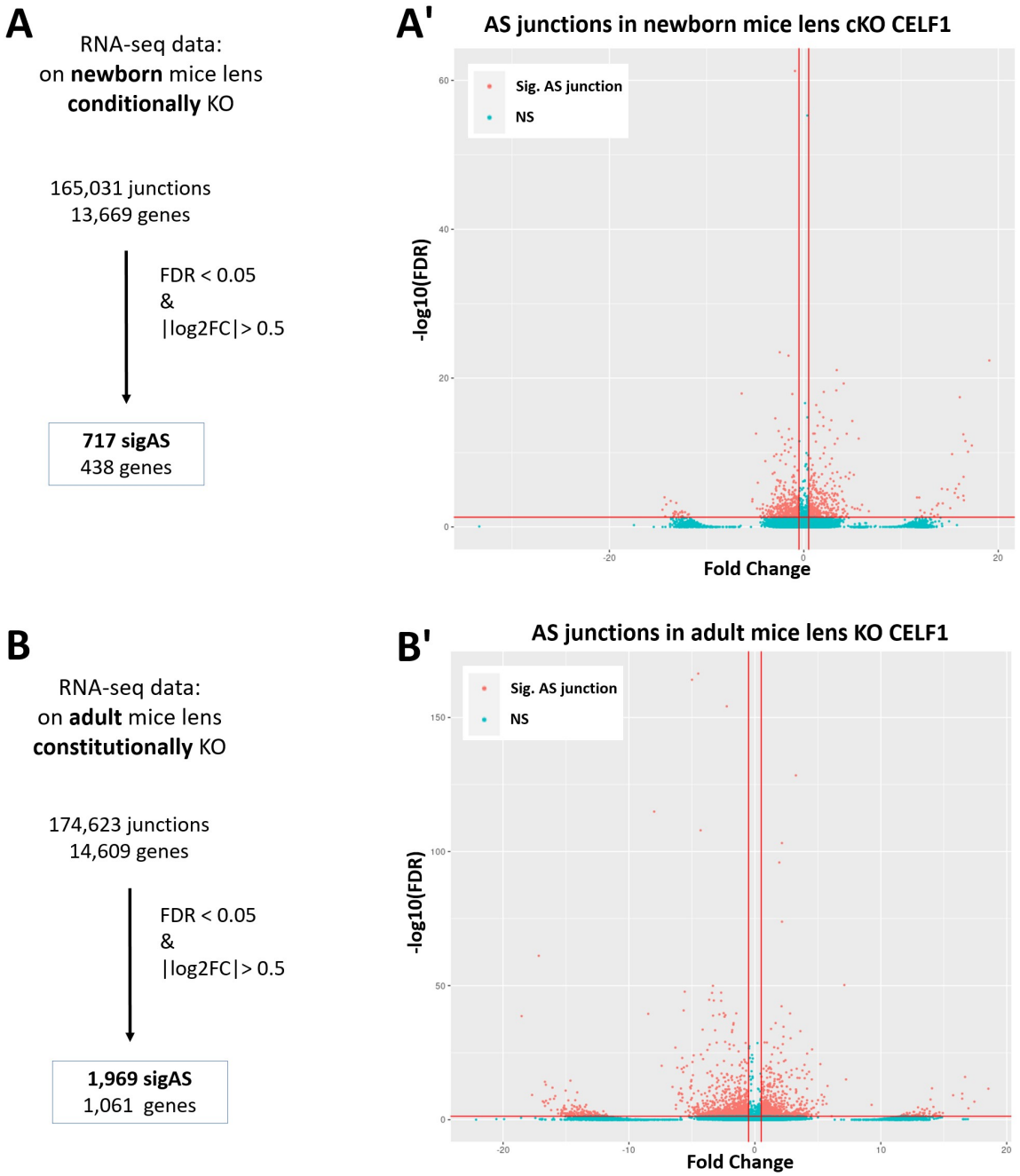


Figure 15: Identification of significant alternatively used junctions in Celf1 cKO and KO mouse lenses. *Result of the junction centric analysis of an RNA-seq on newborn (A) and adult (B) lenses. In the Volcano plots(A', B'), each dot corresponds to a junction. Junctions presenting a positive FC are more used in the control lens, while inversely, junctions with a negative are more used in the KO lens. FDR: False Discovery Rate; FC: Fold Change, log2 scale; sigAS: significant Alternatively Spliced junctions.*

C.3 Toward identifying CELF1-regulated splicing: integrating binding sites and alternative junctions

Our next goal was to identify genes whose splicing is directly regulated by CELF1 in lens. Towards this goal, we reasoned that RNA-binding proteins generally control splice sites (acceptor site, donor site or branch point) at a reasonable distance from their binding site. Therefore, we focused on the genes that have a CELF1 binding sites inside a significant junction (between the start and the end of the junction), or less than 300 nt apart from the start or the end of the junction. The many genes with CELF1 binding sites and significant alternative splice junctions but too far from each other were not selected.

With this methodology, I identified 28 and 55 genes with a CELF1 binding site associated with a significative alternative junction in newborn and adult lenses, respectively. Sixteen genes are common in both datasets. I looked at the splicing of all 67 genes in control and *Celf1* disrupted lenses from the RNAseq data in the IGV viewer. I selected high confidence candidate genes based on:

(i) the specificity of the localization of CELF1 binding sites,

(ii) the existence of the junction in previous annotations,

(iii) the presence of the spliced exon in the RNA-seq sequence at a reasonable expression level.

For example, the gene *Ablim1* has 3 significant alternative junctions, from exon 20 to exon 22, from exon 22 to exon 23, and from exon 20 to exon 23 (exon 21 is not used in lens). In addition, (i) it has a clear CELF1 binding site within intron 22, (ii) it has annotated isoforms containing or devoid of exon 22 (exon 20 is joined to exon 23), and (iii) exon 22 is significantly included in at least one condition (in control lenses, see Figure 16A). Exon 22 is by far more included in control than in KO or cKO lenses (Figure 16A). Thus we can propose the following model of CELF1-mediated regulation on this transcript: in the presence of CELF1, exon 22 is included in about half of the synthesized transcripts. Hence, both *Ablim1* isoforms with and without exon 22 are present (Figure 16B). In the absence of CELF1, exon 22 is excluded and only the mRNA isoform of *Ablim1* without exon 22 is present (Figure B’). Hence, the binding of CELF1 to intron 22 would stimulate the inclusion of exon 22. Because it meets all the above criteria, *Ablim1* is considered as a high-confidence candidate in the rest of the studies.

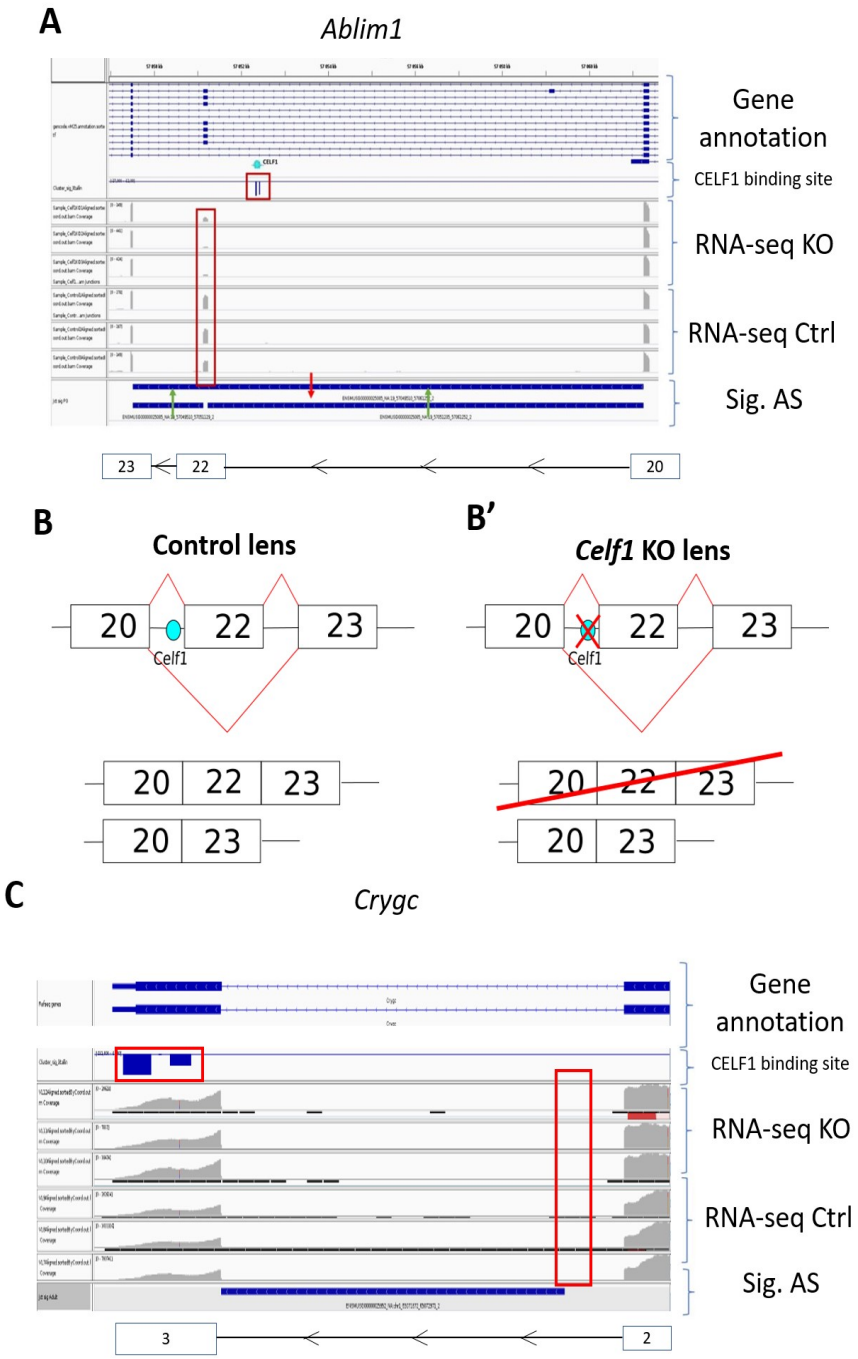


Figure 16: Example of selection of confident candidate gene whose alternative splicing is regulated by CELF1.*(A) Visualization of the iCLIP-seq and RNA-seq results on the Ablim1 gene. The CELF1 binding site and the alternative exon are highlight by the red boxes.(B) A schematic of the predicted regulation of alternative splicing of Ablim1 by CELF1. In the control lenses, exon 22 is included in half of the transcripts leading to the production of the two mRNA isoforms with or without the exon 22 ; (B')in the absence of CELF1 the exon 22 is excluded leading to the production of only the mRNA isoform without exon 22. (C) Visualization of the iCLIP-seq and RNA-seq results on the gene Crygc. The CELF1 binding site and the absence of the supposed alternative exon are highlighted by the red boxes.*

The situation of some other candidate genes is not so clear. For example, in the gene *Crygc*, the binding site of CELF1 is in the last exon, but the alternative junction that we identified during our study was not previously annotated. In addition, the predicted cryptic exon was not found in the RNA-seq data (Figure 16C). The regulation of *Crygc* AS by CELF1 is hard to identify, thus *Crygc* was not selected for the rest of the study.

In total, I prioritized 22 high-confidence candidate genes (Table 9). Six of them are identified as mis-spliced in both newborn and adult datasets. *Klc1, Bfsp1, Pxdn, Ywhae, Ank2* and *Maf* were already associated with cataract pathology. It is worth noting that *Celf1* transcript itself is also present in both newborn and adult datasets, with significant alternative junctions. This suggests that CELF1 could regulate the alternative splicing of its own transcript, as already described[382]. Since the aim of this study was to identify CELF1 targets that could lead to cataract when mis-spliced, *Celf1* was not kept in the following analysis.

Among the 22 high-confidence genes, 10 are related to the cytoskeleton. These genes are: *Ablim1, Klc1, Sptbn1, Clta, Ctnna2, Bfsp1, Nupr1, Septin8, Ywhae, Ank2* and *Ivns1abp*. This cytoskeleton related genes enrichment is particularly interesting regarding the *Celf1* KO lens defects. Indeed, in the absence of CELF1, the lens presents a major structural defect associated with an abnormal F-actin pattern[140]. Thus, we focused on the study of the mis-regulation of these 10 genes.

Table 9: 22 high-confidence candidate genes. *SigAS:Significant Alternative Splice junctions*

Gene	SigAS identified in newborn RNA-seq	SigAS identified in adult RNA-seq	Associated with cataract (CatMap)
Ablim1	**Yes**	**Yes**	No
P4ha1	**Yes**	**Yes**	No
Klc1	**Yes**	**Yes**	**Yes**
Sptbn1	**Yes**	No	No
Clta	**Yes**	**Yes**	No
Eif5b	**Yes**	**Yes**	No
Gls	**Yes**	**Yes**	No
Ctnna2	No	**Yes**	No
Bfsp1	**Yes**	No	**Yes**
Cdc37l1	**Yes**	No	No
Hnrnpdl	**Yes**	No	No
Nupr1	**Yes**	No	No
Psme4	**Yes**	No	No
Pxdn	**Yes**	No	**Yes**
Rnf180	**Yes**	No	No
Septin8	**Yes**	No	No
Smco3	**Yes**	No	No
Txn2	**Yes**	No	No
Ywhae (14-3-3 ε)	**Yes**	No	**Yes**
Ank2	No	**Yes**	**Yes**
Ivns1abp	No	**Yes**	No
Maf	No	**Yes**	**Yes**

C.4. Validation of candidate genes by RT-PCR

I next attempted to validate by RT-PCR the alternative splice events initially observed by RNA-seq. As matrices I used RNA extracted from control or KO adult mouse lenses. Primers were designed to amplify the different isoforms observed in the RNA-seq. For example, as exon 22 of Ablim1 is differentially included in the presence and the absence of CELF1 (see above), the primers for RT-PCR were in exons 21 and 23 (Figure 17A). Here, I will show the results for the 7 mRNAs out of 10 for which the RT-PCR results are in accordance with the RNA-seq results.

C.4.a Ablim1

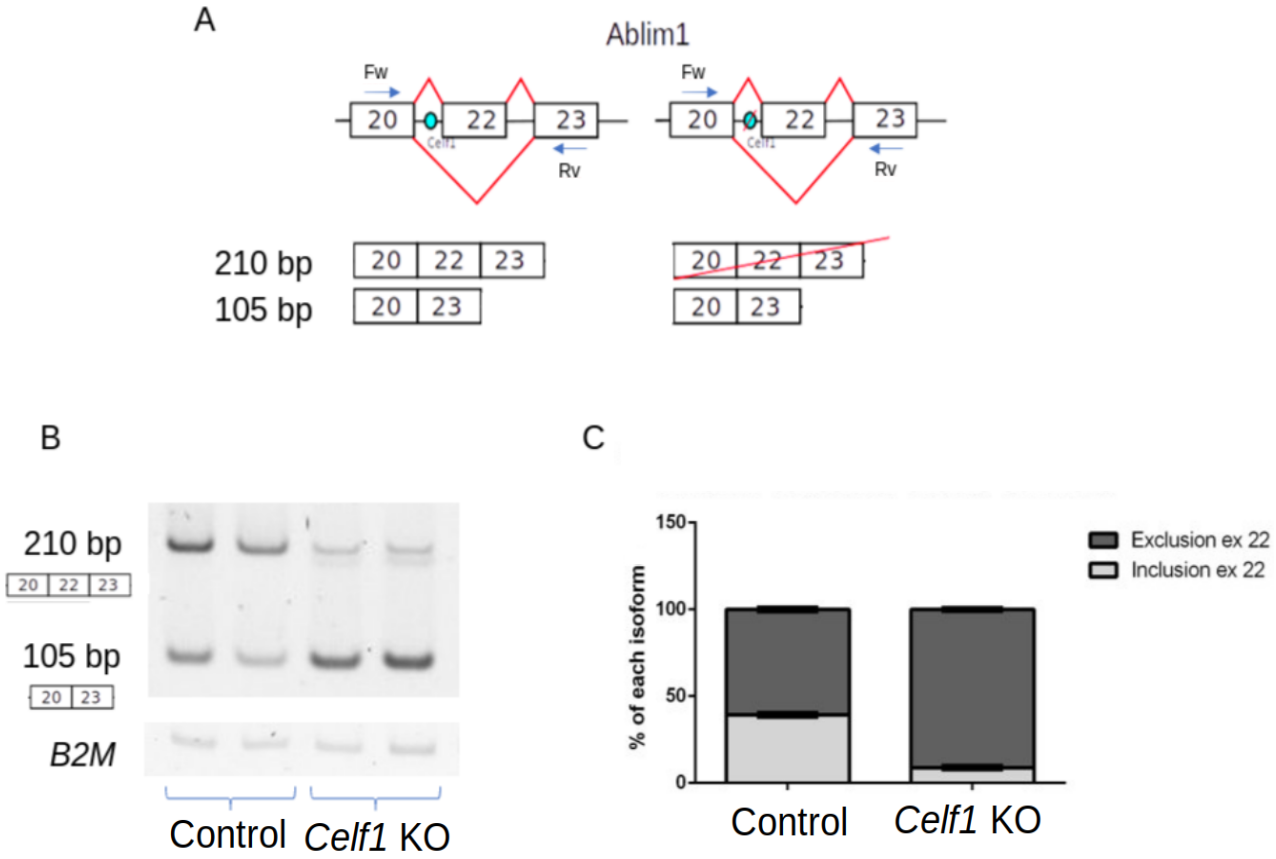


Figure 17: Splicing patterns of Ablim1 in control and Celf1 KO lenses. *(A) Alternative splicing event of Ablim1 predicted by RNA-seq. Exon 22 can be included or skipped, and RNA-seq data indicate that CELF1 stimulates its inclusion. RT-PCR primers position, in exons 20 and 23, are indicated by arrows The blue dot indicates the position of the CELF1 binding site, as revealed by iCLIP-seq. (B) RT-PCR on lens RNA from WT or KO Celf1mice. B2M, beta-2-microglobuline, PCR control. (C) Proportions of Ablim1 mRNA isoforms, means and s.d. of two different mice for each genotype.*

In control lenses, both isoforms with and without exon 22 of *Ablim1* were present. About 40% of the *Ablim1* transcript contained exon 22 (Figure 17B. Please note that while the signal for the isoform with exon 22 is stronger than the signal for the isoform without exon 22, it corresponds to only 40% of the total signal following normalization by transcript length). In *Celf1* KO lenses, this isoform is only found at less than 10% (Figure 17B). For *Ablim1*, the RT-PCR results are fully in accordance with the RNAseq results (see above, Figure 16).

The inclusion of exon 22 codes results in adding 35 amino acids to the protein without shifting the open read frame.

C.4.b Clta

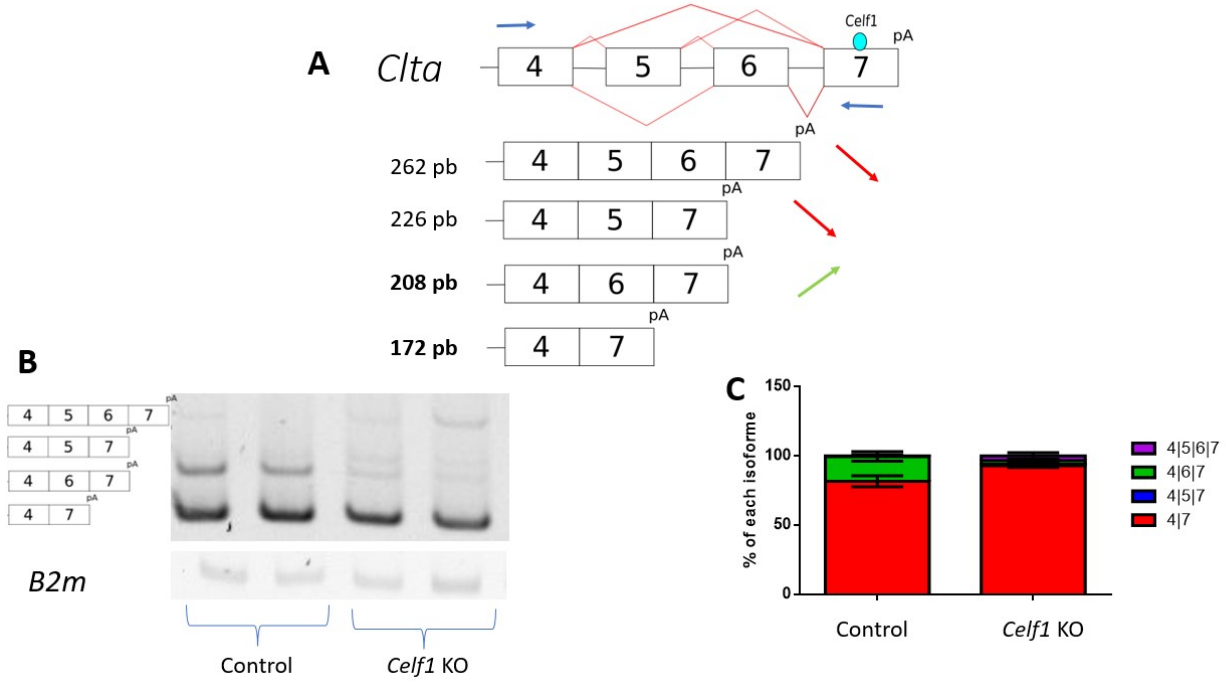


Figure 18: Splicing patterns of Clta in control and Celf1 KO lenses.*(A) Alternative splicing event of Clta predicted by RNA-seq. Exon 5 and exon 6 can be included or skipped, and RNA-seq data indicate that CELF1 stimulates exon 6 inclusion. RT-PCR primers position, in exons 4 and 7, are indicated by arrows. (B) RT-PCR on lens RNA from WT or KO Celf1 mice. B2M, PCR control.(C) Proportions of Clta mRNA isoforms, means and s.d. of two different mice for each genotype.*

Similarly, I validated by RT-PCR the mis-splicing events of the pre-mRNA of *Clta*. The iCLIP-seq identifies a CELF1 cluster on the last exon of the genes (Exon 7), and sigAS are observed between exons 5 and 6. The mouse genome annotation already identify alternative isoforms related to these exons. The RNA-seq data from adult and newborn mice predict that the mRNA isoforms including the exon 6, but not the exon 5 are promoted by the presence of CELF1 (Figure 18A). The RT-PCR confirmed that the isoform without the exon 5 and 6 (noted 4|7) is the main *Clta* isoform in the lens (Figure 18B). However in control lenses where CELF1 is present, an isoform containing exon 6 (noted 4|6|7) is also present. Inclusion of exon 6 results in an mRNA isoform longer by 36 nucleotides coding for 12 additional amino acids. In the presence of CELF1 this isoforms correspond to 18% of *Clta* mRNA, while in *Celf1* KO mice, only 1% of the *Clta* mRNA includes exon 6 (Figure 18C).

It is worth noting that mRNA isoforms including exon 5 (noted 4|5|7 and 4|5|6|7) are only present in small traces in the lens from the control and KO mice, with no indication that exon 5 inclusion is controlled by CELF1 (Figure 18B,C). Together, these data indicate that CELF1 bound to exon 7 stimulates the inclusion of exon 6.

C.4.c Ctnna2

I also validated the mis-splicing event occurring in the absence of CELF1 on the pre-mRNA of *Ctnna2*. The identified CELF1 cluster is in intron 1. The RNA-seq data predict that CELF1 promotes the exclusion of exon2 (Figure 19A). Accordingly, the RT-PCR validates the lower proportion of mRNA isoform without exon 2 (noted 1|3) in control mouse lenses compared to the KO mouse lenses (Figure 19B,C). Two alternative translation start sites are present in *Ctnna2* mRNA, in exon 2 when it is present and in exon 3 otherwise. Both translation start sites have the same reading frame. Hence, in the absence of exon 2, the mRNA 1|3 is translated into a protein with a N-terminal region that is shortened by 13 amino acids.

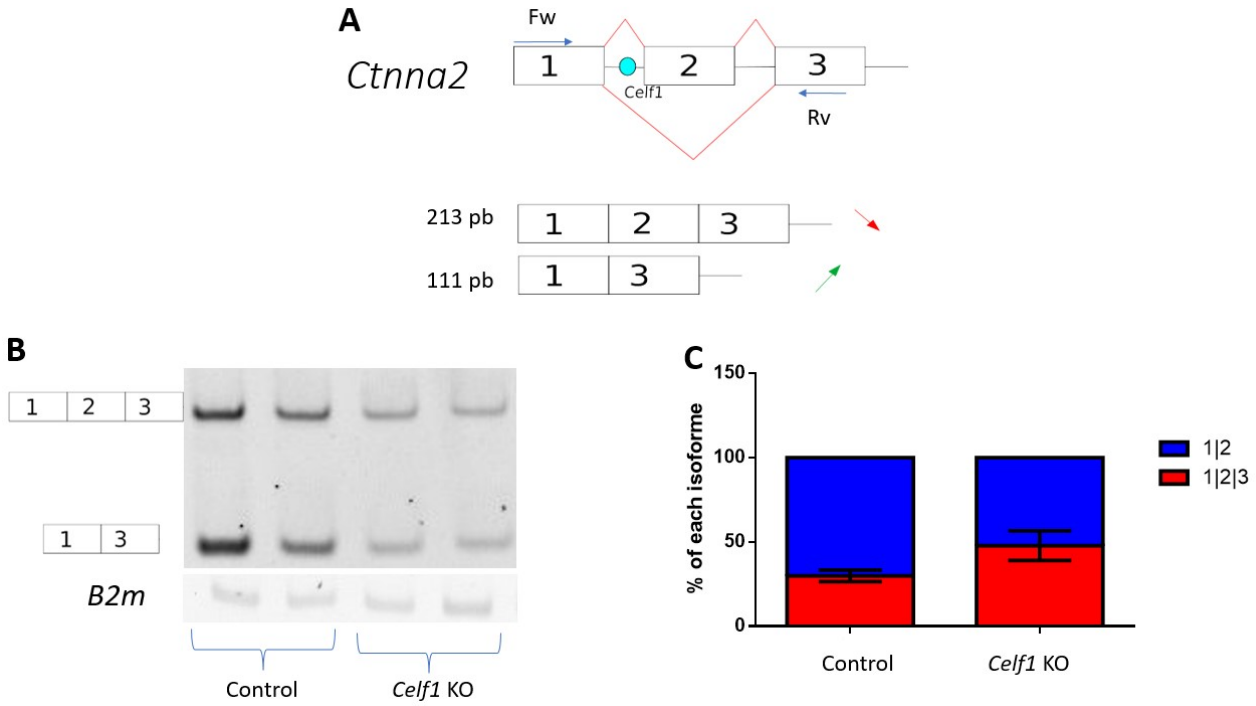


Figure 19: Splicing patterns of Ctnna2 in control and Celf1 KO lenses.*(A) Alternative splicing events of Ctnna2 predicted by RNA-seq. Exon 2 can be included or skipped, and RNA-seq data indicate that CELF1 stimulates exon 2 exclusion. RT-PCR primers position, in exons 1 and 3, are indicated by arrows. (B) RT-PCR on lens RNA from WT or KO Celf1 mice. B2M, PCR control.(C) Proportions of Ctnna2 mRNA isoforms, means and s.d. of two different mice for each genotype.*

C.4.d Ywhae

Regarding *Ywhae* (also named 14-3-3 epsilon), the CELF1 cluster is present in the last exon of the RNA and is predicted from RNA-seq data to inhibit the inclusion of a new cryptic exon (never described in the mouse genome annotation), between exons 5 and 6 (Figure 20A).

This cryptic exon will be referred to as exon 5'. The RT-PCR confirm that CELF1 represses the inclusion of exon 5'. The proportion of the mRNA

containing exon 5’ (named 5|5’|6) in the presence of CELF1 is 4.65%, whereas in the absence of CELF1 it increases to 11,13% (Figure 20B,C). Hence, the RT-PCR data are in accordance with the RNA-seq data. Exon 5’ contains an in-frame stop codon, and the C-terminal domain of the protein encoded by the mRNA isoform including exon 5' is shortened by 16 amino acids replace by 2 new amino acids.

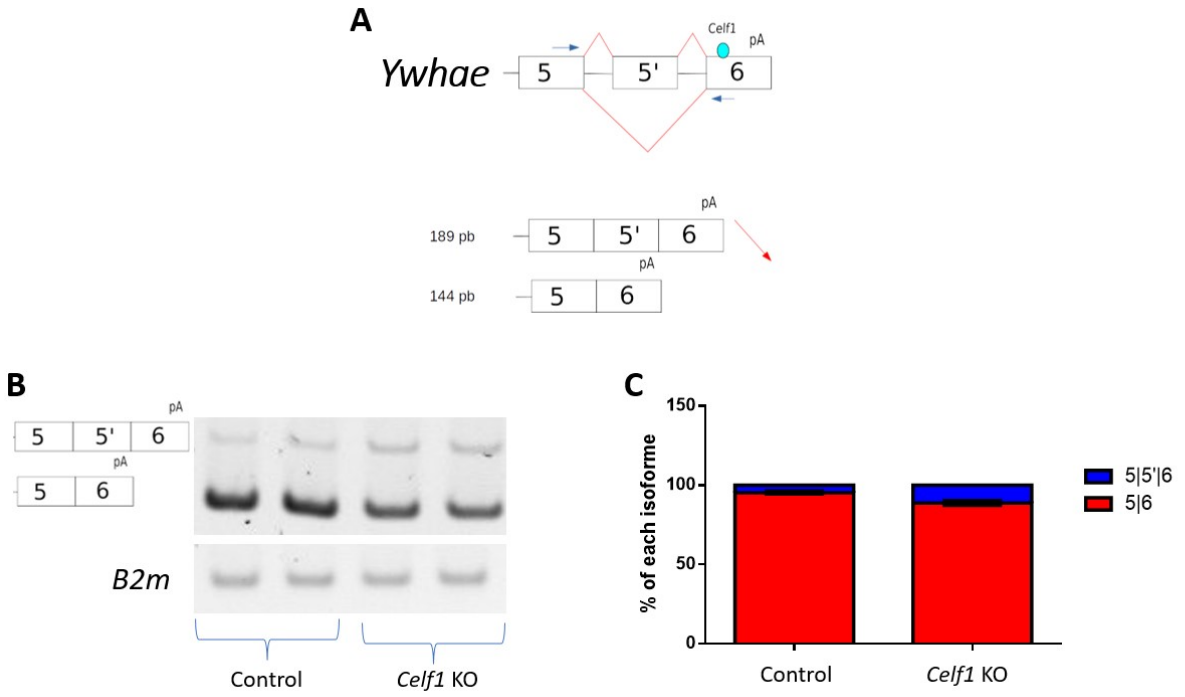


Figure 20: Splicing patterns of Ywhae in control and Celf1 KO lenses.*(A) Alternative splicing event of Ywhae predicted by RNA-seq. The cryptic exon 5’ can be included or skipped, and RNA-seq data indicate that CELF1 stimulates exon 5’ exclusion. RT-PCR primers position, in exons 5 and 6, are indicated by arrows. (B) RT-PCR on lens RNA from WT or KO Celf1 mice. B2M, PCR control.(C) Proportions of Ywhae mRNA isoforms, means and s.d. of two different mice for each genotype.*

C.4.e Ank2

Ank2 RNA contains multiple CELF1 clusters in exons 45 and 47. The RNA-seq predicts that CELF1 repress the inclusion of exon 46 (Fig21A). The RT-PCR validates the nearly complete exclusion of exon 46 in control lenses (1.5%) whereas the KO mouse lenses have a higher proportion of mRNA including exon 46 (14.0%) (Figure 21B,C).

The inclusion of the exon 46 introduces 92 nucleotides which shift the ORF leading to a different STOP codon in the last exon (exon 47). This will result in the addition of 93 amino acids at the C-terminal region of ANK2 that will replace 4 amino acids usually coded by the last exon in the absence of the exon 46.

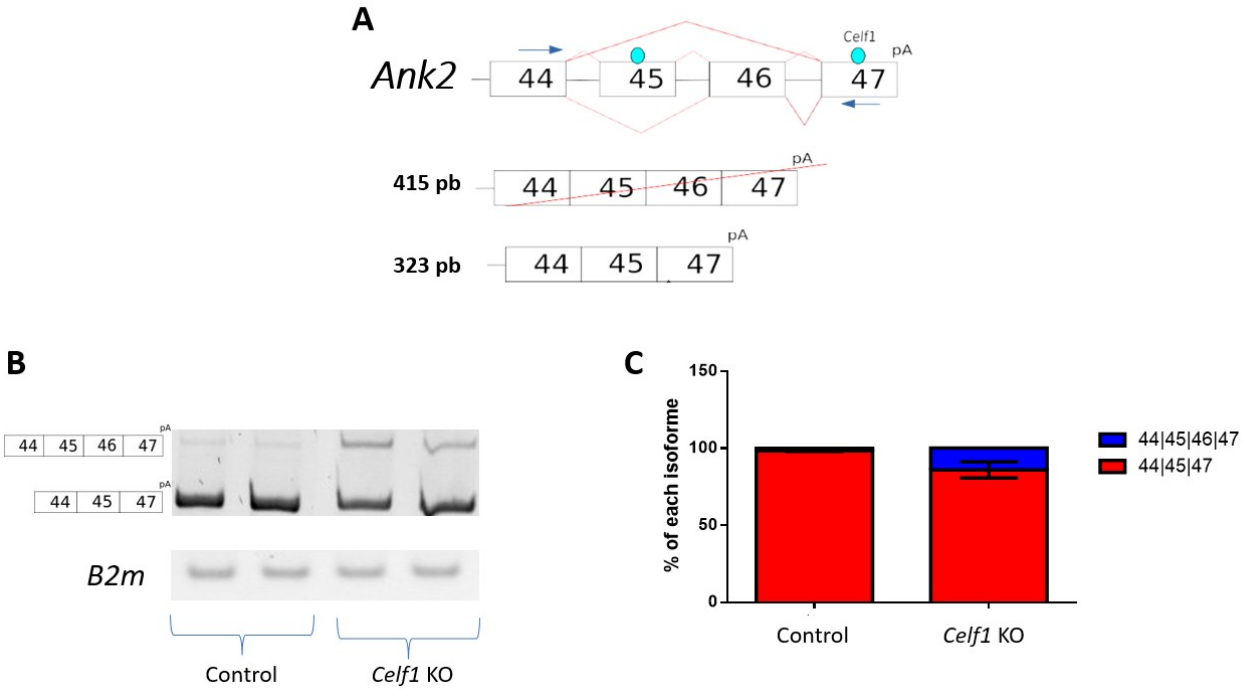


Figure 21: Splicing patterns of Ank2 in control and Celf1 KO lenses.(*A) Alternative splicing event of Ank2 predicted by RNA-seq. Exon 46 can be included or skipped, and RNA-seq data indicate that CELF1 stimulates exon 46 exclusion. RT-PCR primers position, in exons 44 and 47, are indicated by arrows. (B) RT-PCR on lens RNA from WT or KO Celf1 mice. B2M, PCR control.(C) Proportions of Ank2 mRNA isoforms, means and s.d. of two different mice for each genotype.*

C.4.f Septin 8

Septin8 mRNA presents numerous isoforms. Exon10 has 3 different acceptor splice sites. The longest possible exon 10 (which results from using the 5' most acceptor site) is indicated as 10-10'-10" in Figure 22A. The second longest one is indicated as 10'-10" and the shortest one as 10". As the regions between the alternative splice sites contain stop codons, using either of these sites results in mRNA isoforms with distinct 3' untranslated regions and protein isoforms with different C-terminal regions. In addition, exon 10 can also be skipped and exon 9 directly joined to exon 11 (Figure 22A). The RNA-seq predicts that CELF1 promotes the inclusion of the exon 11 instead of exon 10. The CELF1 clusters are present in exon 10 (Figure 22A).

I validated by RT-PCR the higher inclusion of the exon 11 in the control lens compared to the KO lenses (Figure 22B,C). Apparently, essentially isoform 9-10" is increased in the absence of CELF1, and the levels of the two other isoforms containing exon 10 (9-10-10'-10" and 9-10'-10") are not modified. This suggests that CELF1 could inhibit the usage of the 3' most acceptor site, between exonic regions 10' and 10".

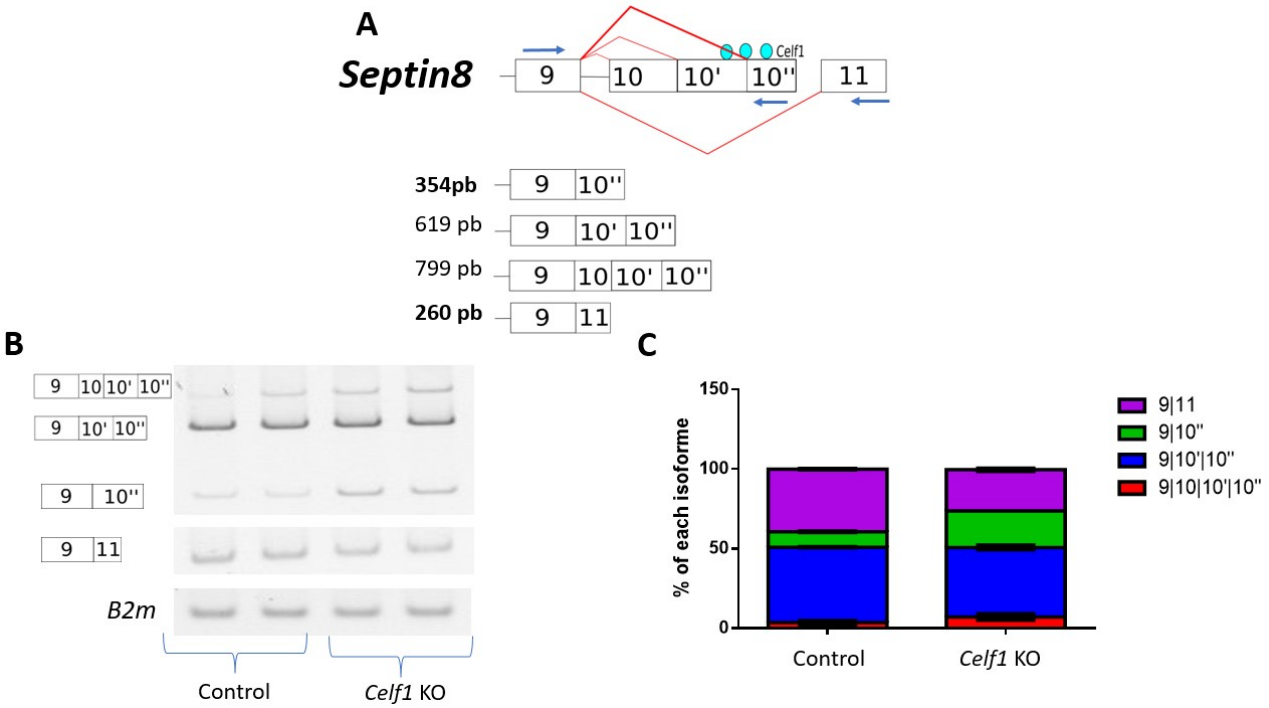


Figure 22: Splicing patterns of Septin8 in control and Celf1 KO lenses.(*A) Alternative splicing event of Septin8 predicted by RNA-seq. Exon 10 or exon 11 can be included, and RNA-seq data indicate that CELF1 stimulates exon 11 inclusion. RT-PCR primers position, in exons 9, 10 and 11, are indicated by arrows. (B) RT-PCR on lens RNA from WT or KO Celf1 mice. B2M, PCR control. (C) Proportions of Septin8 mRNA isoforms, means and s.d. of two different mice for each genotype.*

C.4.g Sptbn1

The last validate gene is *Sptbn1*. Exon 32 contains both a splice donor site, making an internal exon joined to exon 33, and a more distal cleavage and polyadenylation site, making it a terminal exon when the internal splice donor site is not used. In Figure 23A, the region that is 5' to the splice donor site is indicated as 32 and the region that is between the splice donor site and the cleavage and polyadenylation site is indicated as 32'.. The CELF1 cluster is present in region 32'. The RNA-seq predicted that the isoform resulting from using the cleavage and polyadenylation site (with exonic region 32') is repressed by CELF1(Figure 23A). Accordingly, in RT-PCR, isoform (31|32|32') is barely detectable in control lenses and a bit more abundant (albeit still a minority) in KO lenses (Figure 23B,C). The mRNA isoform with exon 33 and following exons (up to exon 35) codes for a protein with a longer C-terminal region with 224 additional amino acids. The additional 224 amino acids contain a predicted pleckstrin homology domain. We will discus later the implication of this domain for SPTBN1 function (see Discussion).

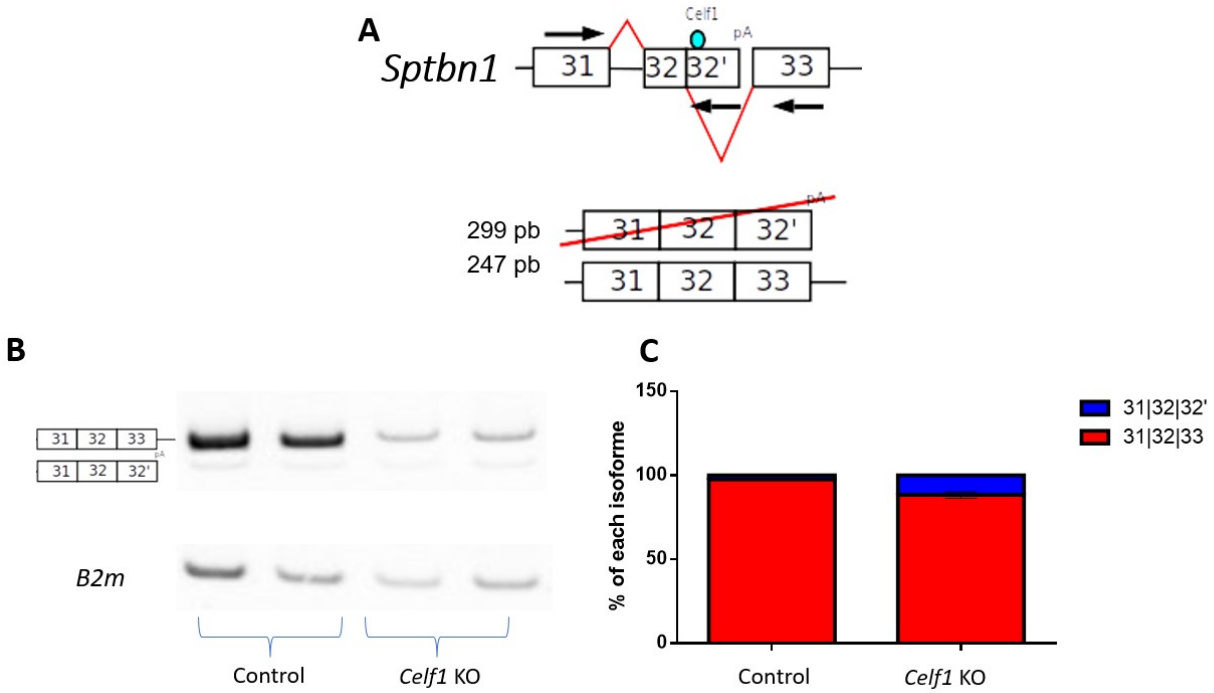


Figure 23: Splicing patterns of Sptbn1 in control and Celf1 KO lenses.*(A) Alternative splicing event of Sptbn1 predicted by RNA-seq. Exons 33 to35 can be included, and RNA-seq data indicate that CELF1 stimulates their exclusion. RT-PCR primers position, in exons 31, 32 and 33, are indicated by arrows. (B) RT-PCR on lens RNA from WT or KO Celf1 mice. B2M, PCR control.(C) Proportions of Sptbn1 mRNA isoforms.*

To summarize, we confirmed by RT-PCR that 7 candidate genes have different splicing patterns in control and *Celf1* KO lenses: *Ablim1*, *Clta*, *Ctnna2*, *Sptbn1*, *Septin8*, *Ywhae* and *Ank2*. For the three remaining candidate genes (*Bfsp1*, *Klc1* and *Ivns1abp*), the RT-PCR results were not in accordance with the RNA-seq.

C.5 Identifying the structural differences between the protein isoforms

Now the main question is if and how the mis-splicing events observed in the absence of CELF1 could participate in the formation of the congenital cataract observed in mice deficient *for Celf1.* As a first step in providing an answer, I used the AlphaFold2 (AF2) software. AF2 is a highly accurate protein structure prediction tool[380]. I used it to predict the protein conformation of each isoforms of the 7 cytoskeleton-related validated genes.

For example, for the candidate gene *Clta*, the RNA-seq data and the RT-PCR validation show that the presence of CELF1 in the lens leads to the production of an isoform including exon 6 (Figure 24A). This isoform is less produced in the absence of CELF1. AF2 predicts that the inclusion of exon 6 changes the structure of the protein CLTA, with the main alpha helice

presenting a blending not present in the protein isoform that is produced without CELF1 (Figure 24B).

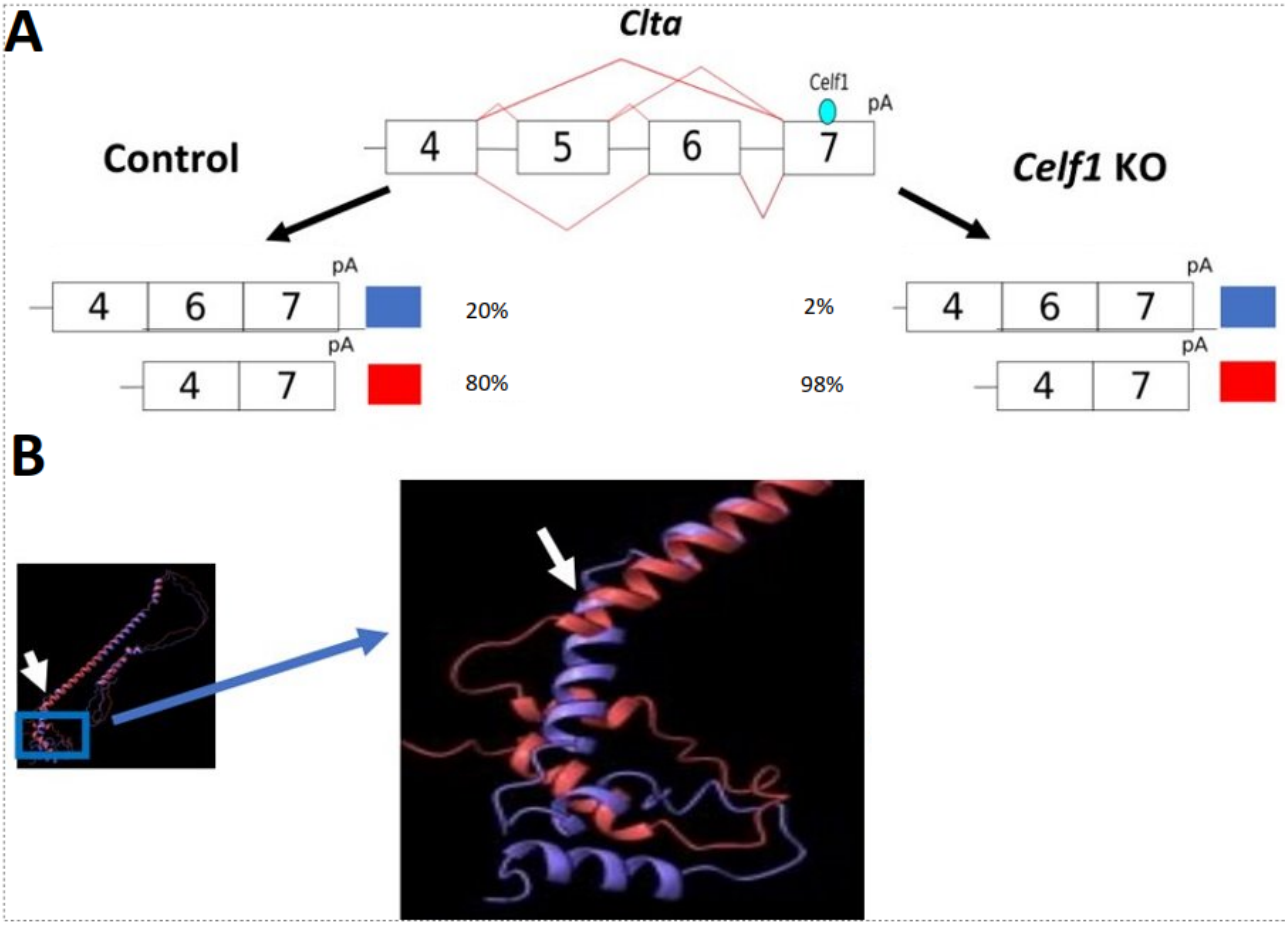


Figure 24: CLTA protein isoforms conformation in mouse lenses. *(A)The validated regulation of the mRNA isoforms in the mouse lens in control or Celf1 KO lens. The presence of CELF1 in the lens increases the proportion of exon 6 inclusion. (B) CLTA proteins conformation predicted by AF2 highlights a structural difference between the isoforms. In red the protein isoform coded by the mRNA without exon 6. In blue the protein isoform coded by the mRNA with exon 6. The protein isoform coded by the mRNA with exon 6 presents a blending on the main alpha helice of the protein (highlighted by the white arrow).*

A summary of the predicted impact of alternative splicing in the 7 cytoskeleton-related, high-confidence genes, is given in Table 10. As shown above, for the gene *Sptbn1*, the absence of CELF1 leads to the production of a short mRNA with a cleavage and polyadenylation site within exon 32 instead of exon 36. This short mRNA codes for a SPTBN1 protein isoform without its C-terminal containing Pleckstrin homology domain. For the other candidate genes the absence of CELF1 increases or reduces the production of protein isoforms with an Intrinsically Disordered Region (IDR). For CTNNA2 and ANK2, the absence of CELF1 increases the production of the isoform with the IDR. Conversely, for ABLIM1, YWHAE and SEPTIN8, the isoform with the IDR is reduced in the absence of CELF1. The structure of these region is given with a low confidence score by AF2. Generally, those IDR do not acquire their tertiary

structures by their own, and may necessitate the interaction with another protein to correctly fold. Thus these IDR may modulate protein-protein interactions[383]. It may be therefore that the major consequence of the splicing changes observed in the absence of CELF1 for these proteins is a modification of their interactome. Of course, the modifications of the protein-protein interactions of cytoskeleton-related proteins is expected to have a great impact on cytoskeleton and can contribute to cataract. However, a considerable amount of work with biochemical approaches is required to explore this possibility.

Table 10: ***Splicing and structural changes of the validated candidate genes.*** *IDR: Intrinsically disorder region*

Validated cytoskeleton gene	mRNA isoform changes in the absence of CELF1	Protein structural changes in the absence of CELF1
Ablim1	Exclusion of exon 22	Loss of an IDR
Sptbn1	Premature polyadenylation site in exon 32	Loss of the pleckstrin homology domain
Clta	Exclusion of exon 6	Blending of the main alpha helice
Ctnna2	Exclusion of exon 2	Addition of an IDR
Ywhae	Inclusion of a cryptic exon (5')	Loss of an IDR
Ank2	Inclusion of exon 46	Addition of an IDR
Septin8	Inclusion of exon 11 instead of exon 10	Loss of an IDR

D. Discussion

CELF1 is an RBP that post-transcriptionally regulates other genes in the lens. *Cefl1* deficiency in the mouse lens causes the formation of a congenital cataract with a deficient nuclear degradation and global cytoskeleton defects. This suggests that CELF1 deficiency disrupts the fiber cell (FC) differentiation process[59,140]. In a previous study, we identified by RNA-seq a large set of mis-regulated genes in conditionally KO new born mice. It is unlikely that all those genes are directly regulated by CELF1 itself. CELF1 could directly modulate the expression of transcription factors, RBP or other genetic regulators whose mis-regulation could lead to the large genetic mis-regulations that are observed[371]. Few direct targets have been identified such as *Dnase2b*, *Cdkn1b, Pax6* and *Prox1*[59,140]. Those previous studies only focused on the cytoplasmic role of CELF1, the impact of CELF1 on the stability and/or translation of the mRNA. However, during the development of the lens the localization of CELF1 differs, depending on the lens cell-type. CELF1 is highly cytoplasmic in the lens epithelial cells (LEC). It is also found in the cytoplasm of the FC, but it is highly enriched in the nucleus of the FC during their differentiation[140]. Thus CELF1 could also have an important nuclear role for proper FC differentiation and thus lens development.

In the nucleus, CELF1 modulates the alternative splicing of pre-mRNA into mRNA and regulates the inclusion or exclusion of exons. Up to now, the splicing of *Sptb* was the only splicing event reported to be potentially regulated by CELF1 in mouse lens, and it was not known if this regulation was directly regulated by CELF1[140]. Another study tried to identify nuclear targets of CELF1 in the lens[308]. However the authors used a LEC cell line to identify mis-spliced genes in the context of *CELF1* over-expression. As CELF1 is only cytoplasmic in the LEC, these cells may not present the specific splicing regulations that take place in the differentiating FC. Furthermore, they used a CELF1 RIP-seq dataset from HeLa cells to identify the direct targets of CELF1. The HeLa cells are cervical cancerous cells and thus present a context different from the differentiating FC: different levels of gene expression and presence of other RBP that may compete or interact with CELF1[308].

In our study, a CELF1 iCLIP-seq assay made on adult mice lens allowed the identification of 1,728 genes whose RNAs directly interact with CELF1. Since this iCLIP-seq was made on whole lenses, the RNA targeted by CELF1 correspond to both lens cell-type LEC and FC. Thus, these data can be highly valuable for the studies on the both cytoplasmic and nuclear role of CELF1 in the lens. Notably, a minoritary but non negligible proportion of the binding sites

are present in the intronic regions of the targeted genes (5%), confirming the nuclear activity of CELF1 in the lens since CELF1 can interact with those pre-mRNA region only in the nucleus.

We chose to focus our work on the nuclear functions of CELF1, mainly during FC differentiation. In this purpose two RNA-seq datasets were used to identify the mis-spliced gene in lenses deficient for CELF1 in adult and newborn mice. The RNA-seq data on newborn lenses used cKO mice and was already studied but only in the context of mis-expressed genes[371]. The RNA-seq data on adult lenses correspond to constitutionally KO mice. Those data have been generated for this study to have transcriptomic data from lenses the same age (adult mice) as the iCLIP-seq data. Mis-spliced junctions have been identified in 438 and 1,061 genes in respectively newborn and adult mice. Fewer mis-spliced genes are observed in newborn than in adult mice. This could be due to differences in development stages, as the mis-spliced mRNA could accumulate in the adult lens compacted to the newborn lens. Thus more mis-spliced genes would reach the significancy threshold. However, this also could be due to the KO strategy that was used. As it has been show in the newborn RNA-seq study, the cKO is not totally effective. A small amount of CELF1 protein is still present probably in a subset of cells, even if it is not enough to prevent the formation of the congenital cataract in mice[371]. This small amount of CELF1 could reduce the defect in mis-splicing and thus reduce the number of detected AS junctions.

By integrating the iCLIP-seq data with the two RNA-seq datasets and performing a manual selection process, we have identified 22 high-confidence genes (refer to Table 9). Considering that the *Celf1* deficient lens have cytoskeleton defects, we prioritized 10 genes related to the cytoskeleton among these 22 genes. These genes are likely to play crucial roles in mediating the cytoskeletal abnormalities observed in the absence of *Celf1*. We validated by RT-PCR the differential splicing of *Ablim1, Ctnna2, Clta, Septin8, Sptbn1, Ywhae* and *Ank2.* While prioritizing cytoskeleton-related genes is important given the major cytoskeleton defects observed in the *Celf1*-deficient lens phenotype, it is also worth noting that the mis-splicing events occurring in the 12 non-cytoskeleton candidate genes could serve as valuable subjects for future studies.

With the new protein structure prediction tool AF2[380], I predicted the structural differences of the proteins isoforms of ABLIM1, CTNNA2, CLTA, SEPTIN8, SPTBN1, YWHAE and ANK2. The summary of the structural changes of those proteins in the absence of CELF1 is given in Table 10

Sptbn1 codes for a spectrin highly enriched in the lens[31,33]. The spectrins associate in multimeres to form long filaments on the cytoplasmic side of the cell membrane. They interact with the ankyrin and the F-actin and thus participate actively to the regulation of the cell shape[384]. In the absence of CELF1 in the lens, a shortened mRNA isoform is expressed. It codes for a SPTBN1 isoform lacking the Pleckstrin Homology (PH) domain in the C-terminal region of the protein. The function of the PH domain in SPTBN1 specifically in the lens is unknown. However, regarding the function of the PH domain in other proteins, we can speculate that the PH domain can help the localization of SPTBN1 at the cell membrane[385,386]. In the absence of CELF1, the mis-regulation of this gene could lead to the mis-localization of SPTBN1 and its cytoskeleton partners. Careful immunofluorescence studies are required to compare the location of SPTBN1 in control and *Celf1* cKO lenses. However, the isoform lacking the PH domain is by far the most abundant one in both control and cKO lenses (see Figure 23). Hence, in the absence of isoform-specific antibody, it might be very tricky to detect differences in SPTBN1 location between control and cKO lenses. Nevertheless, the defective control of *Sptbn1* splicing in the absence of CELF1 may participate to the global cytoskeleton defect observed in the *Celf1* KO mice lens[140].

Clta codes for the Clathrin light chain a. This protein interacts with the clathrin heavy chain to mediate the formation of clathrin vesicles. In the neurons, the inclusion or exclusion of exons 5 and 6 have been shown to modulate the number and the size of formed clathrin vesicles[387]. The regulation of *Clta* splicing by CELF1 was already reported in the heart where the over-expression of *CELF1* represses the inclusion of exon 6. However, the opposite regulation is observed in the lens where it is the absence of CELF1 that prevents the inclusion of exon 6[388]. The reasons for the opposite directions of regulation of the same mRNA by the same RBP remain to be investigated. During FC development, a cohesive network of cells is formed to prevent any intercellular space that would result in light scattering. Toward this goal, the FC membranes form interdigitation domains presenting a ball-and-socket form. The formation of those protrusions is associated with the presence of Clathrin protein at the surface of the protrusion. Thus the formation of the interdigitations has some similarity with the formation of clathrin vesicles[39]. Hence, the mis-splicing of *Clta* exon 6 could impact the formation of these FC protrusions. Indeed, in *Celf1* cKO lenses, the interdigitations are not properly formed[140].

For *Ablim1, Ctnna2, Septin8, Ywhae* and *Ank2*, the observed mis-splicing changes in the absence of CELF1 in the lens result in the translation of protein isoforms that either lose or gain an IDR (intrinsically disordered region). The IDR are regions that necessitate to interact with other proteins to obtain their

correct conformation. Thus, these IDR are involved in protein-protein interactions. It is important to note that several studies have demonstrated that a significant number of alternative splicing events, such as tissue-specific alternative splicing, do not necessarily alter protein domains but instead introduce or remove an IDR. By adding or removing IDRs through alternative splicing, these events can modulate the protein's interactome, leading to distinct protein-protein interaction networks[383,389]. Thus we hypothesize that the IDRs of ABLIM1, CTNNA2, SEPTIN8, YWHAE and ANK2 regulated by CELF1 within the lens may result in different interactomes for these proteins. In future studies, the identification of the differences between the isoform-specific interactomes for each of these genes will be crucial to understand the molecular mechanisms underlying the formation of cataract in *Celf1* deficient lenses. This will allow the identification of new cataract associated genes.

Chapter III: Characterization of a new lens organoid model

Abstract

The ocular lens is a complex organ. In addition to lens epithelial cells (LEC), the lens essentially consists of specialized fiber cells (FC). The FC undergo elongation, and they lose their nuclei and organelles during their differentiation. Consequently, the FC cannot be cultured in vitro following traditional cell culture methods. Only the LEC, which are able to divide, can be grown using standard two-dimensional (2D) methods. This makes FC hard to study in vitro.

As a result, various research teams have created three-dimensional (3D) lens organoid models to overcome these limitations and to enable more accurate *in vitro* lens studies. Nevertheless, these models often require the utilization of embryonic stem cells or induced pluripotent stem cells, which are cultivated with specific growth factors during long culture times. The complexity and the cost of these cultures limits their usage.

Here, we present a novel lens organoid model derived from lens epithelial cells. They do not require specific growth factors. This makes it user-friendly, cost-effective, and suitable for mass production. These organoids exhibit lens-like optical properties, being transparent and capable of focusing light.

To characterize this model, I made a transcriptomic analysis to identify the gene expression pattern during organoid formation. Similar to lens development, the expression of some lens-specific genes increased during organoid formation. This includes transcription factors, RNA-binding proteins, and key signaling pathway members.

Histological and regional transcriptomic analyses revealed the presence of three distinct regions in the organoid. Laminin, a protein associated with lens capsule, is present in the external region, capable of proliferation. This region also expressed lens epithelial cell-specific genes. The intermediate region is characterized by elongated nuclei. Finally, the internal region exhibited compact nuclei and contained fiber cell-specific proteins like CRYAB, nuclear PROX1, and HSF4. Mitochondria start being degraded. Moreover, this region showed enrichment in fiber cell-specific genes, suggesting that organoid differentiation partly recapitulates fiber cell differentiation.

Finally, due to its optical characteristics, this model can be used to study the formation of cataracts caused by environmental or genetic factors. This model loses its transparency and light-focusing capacity due to mis-expression

of genes known to cause cataracts (e.g., *Celf1*) or exposure to compounds that can induce cataracts.

In conclusion, this model successfully mimics certain aspects of lens formation and of fiber cell differentiation. This model can support the investigation of novel candidate genes implicated in cataract formation. By providing this novel organoid model, we propose the lens research community an accessible tool for studying lens biology, at a cost-effective price, and capable of being used to screen potential anti-cataract compounds.

Duot et al, submitted

Preprint available on bioRxiv

Eye lens organoids going simple: characterization of a new 3-dimensional organoid model for lens development and pathology
Matthieu Duot, Roselyne Viel, Justine Viet, Catherine Le Goff-Gaillard, Luc Paillard, Salil A. Lachke, Carole Gautier-Courteille, David Reboutier
bioRxiv 2023.07.12.548679; doi: https://doi.org/10.1101/2023.07.12.548679

Eye lens organoids going simple: characterization of a new 3-dimensional organoid model for lens development and pathology

Matthieu Duot[‡1,2], Roselyne Viel[‡3], Justine Viet[1], Catherine Le Goff-Gaillard[1], Luc Paillard[1], Salil A. Lachke[#2,4], Carole Gautier-Courteille[#*1], David Reboutier[#*1]

[1] CNRS, UMR 6290, IGDR (Institut de Génétique et Développement de Rennes), Université de Rennes, Rennes, France
[2] Department of Biological Sciences, University of Delaware, Newark, DE 19716, USA
[3] CNRS, Inserm UMS Biosit, H2P2 Core Facility, Université de Rennes, Rennes, France
[4] Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE 19716, USA
* Correspondence: carole.gautier-courteille@univ-rennes1.fr; 33 2 23 23 47 81 (CGC); david.reboutier@univ-rennes1.fr; 33 2 23 23 46 10 (DR).
‡ These authors contributed equally to this work.
# SAL, CGC and DR are joint senior authors on this work.

Abstract: Cataract, the opacification of the lens, is the leading cause of blindness worldwide. Although effective, cataract surgery is costly and can lead to complications. Toward identifying alternate treatments, it is imperative to develop organoid models relevant for lens studies and drug screening. Here, we demonstrate that by culturing mouse lens epithelial cells under defined 3-dimensional (3D) culture conditions, it is possible to generate organoids that display optical properties and recapitulate many aspects of lens organization and biology. These organoids can be rapidly produced in large amounts. High-throughput RNA-sequencing (RNA-seq) on specific organoid regions isolated by laser capture microdissection (LCM) and immunofluorescence assays demonstrate that these lens organoids display spatiotemporal expression of key lens genes, e.g., *Jag1*, *Pax6*, *Prox1*, *Hsf4* and *Cryab*. Further, these lens organoids are amenable to induction of opacities. Finally, knockdown of a cataract-linked RNA-binding protein encoding gene, *Celf1*, induces opacities in these organoids, indicating their use in rapidly screening for genes functionally relevant to lens biology and cataract. In sum, this lens organoid model represents a compelling new tool to advance the understanding of lens biology and pathology, and can find future use in the rapid screening of compounds aimed at preventing and/or treating cataract.

Keywords: cataract; eye lens; organoid; pathophysiology; *Celf1*.

1. Introduction

The lens, in conjunction with the cornea, is responsible for the focusing of light onto the retina, thus creating a clear image. It is a fully transparent biological tissue that involves extreme cell differentiation processes. At the histological level, the lens is composed of an anterior monolayered epithelium containing proliferating cells in the equatorial region that later exit the cell cycle and progressively differentiate into fiber cells. These latter cells form the lens cortex, once the differentiation process is complete. To achieve lens transparency, fiber cells lengthen extensively (~1000X), produce large amounts of refractive proteins called crystallins, and eliminate their organelles, including their nuclei [1,2]. Lens clouding or cataract is the leading cause of blindness worldwide.

While the primary reason for the development of cataracts is aging, they can also be induced by environmental factors or have a congenital origin, often triggered by genetic predispositions or abnormalities [3,4]. To date, the only treatment for cataracts is surgery, which consists of replacing the clouded lens with an artificial implant. Although it is effective, surgery is costly and can have side effects that are far from harmless [5]. Therefore, efforts to develop drugs to treat cataracts have been initiated [6–8]. Animal models such as zebrafish [9], Xenopus [10], chicken or mammals, namely, rodents, dogs or macaques [11] are used for the study of lens pathophysiology. However, a major bottleneck toward developing anti-cataract drugs remains the lack of an adequate biological model for intensive drug screening.

In recent years, biology and medicine have undergone a revolution with the advent of particular 3-dimensional (3D) cultures called organoids [12]. These are in vitro cellular models that mimic several aspects of the structure and function of the corresponding organ. Lens epithelium explants were a first generation of 3D lens cultures [13]. Later on, lentoid bodies, which are 3D cellular structures emerging from various types of 2D stem cells cultures [14–16], or individual micro-lenses grown from lens epithelial cells or pluripotent stem cells [17,18] were described. Although these models have very interesting molecular and/or optical characteristics, they do not exhibit any particular organization reminiscent of the histology of the lens [19]. Moreover, they often require sequential treatments by individual or combined growth factors and remain tricky and time-consuming to implement. Consequently, they generally (except for the paper from Murphy and colleagues [17]) do not allow for high-throughput studies.

The goal of the present study was to develop a mammalian organoid lens model that could be generated rapidly and is more convenient to use. As a starting point, we considered a previous paper which shows that lens epithelium can regenerate a functional lens after its ablation in several vertebrate models [20]. This capacity relies on the presence of lens stem or progenitor cells that sustain self-renewal. Characterization of these cells demonstrated that they express *Pax6* and *Bmi1* and that they are able to spontaneously generate lentoid bodies. The 21EM15 mouse lens epithelial cell (LEC) line expresses *Pax6* and *Bmi1* and can spontaneously forms lentoid bodies in vitro [21,22]. However, these lentoid bodies have not been characterized and the culture conditions to controllably induce 21EM15 cells to become such 3D structures have not been defined.

Therefore, in the present study, we sought to derive the culture conditions that could generate lens organoids from 21EM15 cells *en masse*. Further, we sought to undertake their detailed characterization to evaluate their utility in studying genes and pathways relevant to lens biology and pathology. Our work indicates that using simple 3D culture conditions, we can generate numerous lens organoids in a short period of time. These organoids show very interesting optical properties and recapitulate lens physiology at the molecular, histological and cellular levels. In addition, our results demonstrate the possibility to induce various types of opacities, thus mimicking cataract, in these organoids. As a whole, the 21EM15 organoids should therefore provide the lens community with a compelling new model to advance the understanding of lens biology

and pathology. From a clinical point of view, although derived from mice, these organoids can potentially be used to screen compounds that could have an effect on the prevention and/or treatment of cataracts.

2. Materials and Methods

Cell Culture

21EM15 (obtained from Dr. John Reddan, Oakland University, Michigan) and HaCat cells (ATCC) were cultured in DMEM with 4.5 g/L glucose, L-glutamine, and sodium pyruvate included (Life Technologies), 10% Fetal Bovine Serum (Eurobio), and 1% penicillin-streptomycin (Life Technologies). A253 cells (ATCC) were cultured in McCoy 5A with 4.5 g/L glucose, L-glutamine, and sodium pyruvate included (Life Technologies), 10% Fetal Bovine Serum (Eurobio), and 1% penicillin-streptomycin (Life Technologies). All three cell lines were cultured in 100 mm cell culture treated Petri dishes (Corning) with 10 mL of culture medium. The cells were grown at 37 °C in a water saturated atmosphere with 5% CO2. These cells grow well in these conditions and usually are 80% confluent after three days in culture (after 10% seeding). Cells were passaged three times per week.

Organoid culture

Round-bottom 96-well plates were coated with poly(2-hydroxyethylmethacrylate) (Polyhema) (Sigma) 1 week before cell culture. Polyhema was first dissolved at 50 mg/mL in 95% EtOH. This stock solution was then diluted to 30 mg/mL with absolute ethanol. For coating, each well, except for the outermost ones, was filled with 50 µL of a 30 mg/mL solution of Polyhema before the plate was allowed to dry overnight. For culture, the outermost wells were filled with 200 µL PBS to avoid evaporation. The remaining wells were seeded with 10,000 cells and filled with 200 µL culture medium for 10 days prior to experiments.

Histology and image acquisition

For histological analyses, organoids were washed with 1X PBS, then fixed 24 hr in 4% pH 7 buffered formalin and processed for paraffin wax embedding in an Excelsior ES automaton (Thermo Scientific). Paraffin-embedded tissue was sectioned at 4 µm, mounted on positively charged slides and dried at 58°C for 60 minutes. Immunohistochemical staining was performed on the Discovery ULTRA Automated IHC stainer (ROCHE) using the Ventana detection kit (Ventana Medical Systems, Tucson, Ariz). For fluorescent labeling, following deparaffination with Discovery wash solution (Ventana) at 75°C for 8 minutes, antigen retrieval was performed using Ventana Tris-based buffer solution pH8 at 95°C to 100°C for 40 minutes. Endogen peroxidase was blocked with 3% H2O2 for 12 minutes. After rinsing, slides were incubated at 37°C for 60 minutes with primary antibody. Signal enhancement was performed using a secondary HRP-conjugated antibody at 37°C for 16 minutes and DISCOVERY Rhodamine Kit (Roche) for 8 minutes. For fluorescence multiplex labelling, slides were prepared as follow. First Sequence: following deparaffination with Discovery wash solution (Ventana) at 75°C for 8 minutes,

antigen retrieval was performed using Ventana proprietary, Tris-based buffer solution pH8, at 95°C to 100°C for 40 minutes. Endogen peroxidase was blocked with 3% H2O2 for 12 minutes. After rinsing, slides were incubated at 37°C for 60 minutes with primary antibody : rabbit anti-PROX1. Signal enhancement was performed using a Goat anti-Rabbit HRP at 37°C for 16 minutes and DISCOVERY Rhodamine Kit (542-568nm) for 8 minutes. Second Sequence: slides were neutralized with Discovery inhibitor (Ventana) for 8 minutes. After rinsing, slides were incubated at 37°C for 60 minutes with primary antibody mouse anti-JAG1. Signal enhancement was performed using a Goat anti-mouse HRP at 37°C for 16 minutes and DISCOVERY cy5 Kit for 8 minutes. Third Sequence: slides were neutralized with Discovery inhibitor (Ventana) for 8 minutes. After rinsing, slides were incubated at 37°C for 60 minutes with primary antibody Rabbit anti PAX6 Signal enhancement was performed using a Goat anti- rabbit HRP at 37°C for 16 minutes and DISCOVERY Fam Kit for 8 minutes. DAPI staining was used to visualize DNA/nucleus. For chromogenic labeling, following deparaffination with EZ Prep (Roche) at 75 °C for 8 minutes, antigen retrieval was performed using CC1 buffer (Roche) pH 8.0 at 95°C to 100°C for 40 minutes. Endogen peroxidase was blocked with 3% H2O2 for 12 minutes. After rinsing, slides were incubated at 37°C for 60 minutes with primary Antibody. Signal enhancement was performed using a secondary HRP-conjugated antibody at 37°C for 16 minutes and revealed using the OmniMap DAB kit (Roche). The slides were counterstained with the Mayer's hematoxylin. Antibodies were as follows: Ki67, ab16667 dilution 1/200 (Abcam); Cleaved Caspase 3, #9661 dilution 1/250 (Cell Signaling Technology); Lamin B1, A16909 dilution 1/200 (Abclonal); Laminin Z0097 dilution 1/200 (Dako); PAX6 AB2237 dilution 1/200 (Sigma-Aldrich); JAG1 sc390177 dilution 1/200 (Santa Cruz); PROX1 925202 dilution 1/200 (BioLegend); CRYAB ADI-SPA-223 dilution 1/200 (Enzo Life Science); HSF4 HPA048584, dilution 1/200 (Atlas Antibodies). TOMM20 ab186735 dilution 1/5000 (Abcam); Fibrillarin 32639 dilution 1/100 H2P2 (Cell Signaling Technology). HES staining was realized on a ST 5020 automaton (Leica). Bright field images were acquired using a digital slide scanner Nanozoomer (Hamamatsu), while fluorescence microscopy images were acquired using a DeltaVision Elite setup equipped with a Nikon IX71 microscope and a CoolSnap HQ camera (AppliedPrecision).

Transparency, light focalization

Transparency was assessed by placing the organoids for 30 seconds on an electron microscopy copper grid (mesh 300) and imaging them with an AZ100 macroscope (Nikon). Light focalization was quantified using an Axio Observer inverted microscope (Zeiss). Briefly, a stack of images starting from the focus and progressively lowering the objective under the sample was acquired. For each image of the z-stack, the maximum light intensity at the center of the spheroid and the mean intensity in the field around the spheroid were quantified. The ratio of the maximum light intensity to the mean intensity was then calculated and plotted to give a graph.

21EM15 2D and 3D RNA isolation

RNAs were isolated using the Nucleospin kit (Macherey-Nagel) from a 100 mm Petri dish of 21EM15 cell culture at 80% confluence (2D) or 60 organoids (3D), for each replicate.

Laser capture microdissection

For each replicate, 120 organoids were washed with 1X PBS and pelleted before being included in OCT and then snap frozen using a SnapFrost 80 deep freezer (Excilone). The frozen OCT block was then mounted onto a cryostat (Leica) and cut to obtain 10 µm sections. The sections were then deposited on polyethylene naphthalate (PEN) membrane frame slides. OCT was removed by multiple washes: 2 washes in 70° Ethanol (-20°C; 5 min), 1 wash in 90% Ethanol (RT; 20 min), 1 wash in 100% Ethanol (RT; 20 min) and 3 washes in 100 % Xylene (RT; 1 min). The internal or external regions of the spheroids were microdissected using a XT laser capture microdissection setup (Arcturus). The RNA was isolated from these samples using the Arcturus PicoPure kit (ThermoFisher).

3'-end RNA-sequencing (RNA-seq) and analysis

Libraries were prepared from the extracted RNAs using the QuantSeq 3' mRNA-Seq library Kit (Lexogen). The 3'-end seq library were sequence (strand-specific, 150 bp) by the Illumina NovaSeq 6000. Quality of the sequence were validated by FastaQC, and only sense reads were used for the analysis. The RNA-seq data are available on the NCBI Gene Expression Omnibus (GEO) database under series GSE228547 (reviewer token to access the data gdejkmiwznsxxgl). The adaptor sequence and the poly(A) tail were trimmed from the raw sequences, and only read length superior to 20 nucleotides were retained. Trimmed sequences were aligned by the STAR software (STAR(v2.7.8a)) [71] onto the mouse genome (GRCm38.p6). Only uniquely mapped reads were conserved for the analysis. Reads were associated to genes by FeatureCounts (v1.6.0) [72]. For differential gene expression analysis, only genes with an expression >0.2 cpm (counts per million) were considered. The R package edgeR (v3.32.1) [73] was used to identify significantly differentially expressed genes (DEGs), with as cut-offs: |Fold Change| (FC) >2 (|logFC| > 1) and a False Discovery Rate (FDR) < 0.05.

Gene expression data analysis

To determine the pattern of expression of the 2D or 3D DEGs in the mice embryonic lens, we used microarray data from the iSyTE 2.0 database [23] to identify genes that exhibit lens-enriched expression in normal lens development across stages E14.5 to P0. As described previously, comparison of global gene expression data between lens and whole embryonic body tissue (WB) allows estimation of lens-enriched expression. To compare the regional transcriptomic profile between organoid and lens we used previously generated RNA-seq data to identify DEGs with a expression profile specific to isolated FC or LEC [74]. These data correspond to WT mice at stage E14.5, E16.5, E18,5 and P0.5. The

identification of genes with an expression profile specific to FC or LEC was based on cut-offs of P-value adjusted < 0.05 and |FC| > 2.

3. Results

3.1. 21EM15 spheroids are transparent and have the ability to focus light

To test their ability to grow under 3D culture condition, we seeded 21EM15 cells in 96-well culture plate wells coated with polyhema. Twenty-four hours after seeding, the vast majority of the cells were assembled into round spheroids with no isolated cell being observed. Upon subsequent culture, these spheroids grew in size (Supplemental Figure 1A), acquiring an ovoid asymmetric shape between days 3 and 7. Thereafter, culturing the spheroids beyond day 10 only resulted in limited changes in their overall appearance (Supplemental Figure 1A). At day 10, 21EM15 spheroids are transparent contrary to the spheroids generated with two other epithelial cell lines grown in the same conditions, namely, human epithelial keratinocytes (HaCaT cells) or head and neck squamous cell carcinoma A253 cells (Figure 1A). We then tested the capacity of the 21EM15 spheroids to focus light following a previously described approach [17]. Briefly, we imaged the spheroids at different z-positions starting from the focus (Figure 1B). A very bright light spot is observed at the center of the 21EM15 spheroids at a specific z-position (Figure 1C), suggesting that they had acquired properties to focus light. We quantified the light focusing ability as the ratio of the maximum light intensity at the center of the spheroid to the mean intensity around the spheroid (Figures 1C and D). This ratio reaches values well above 1 at z-positions below the focus in 21EM15 spheroids, confirming their capacity to focus light (Figure 1D), not observed in HaCaT spheroids (Supplemental Figure 1B).

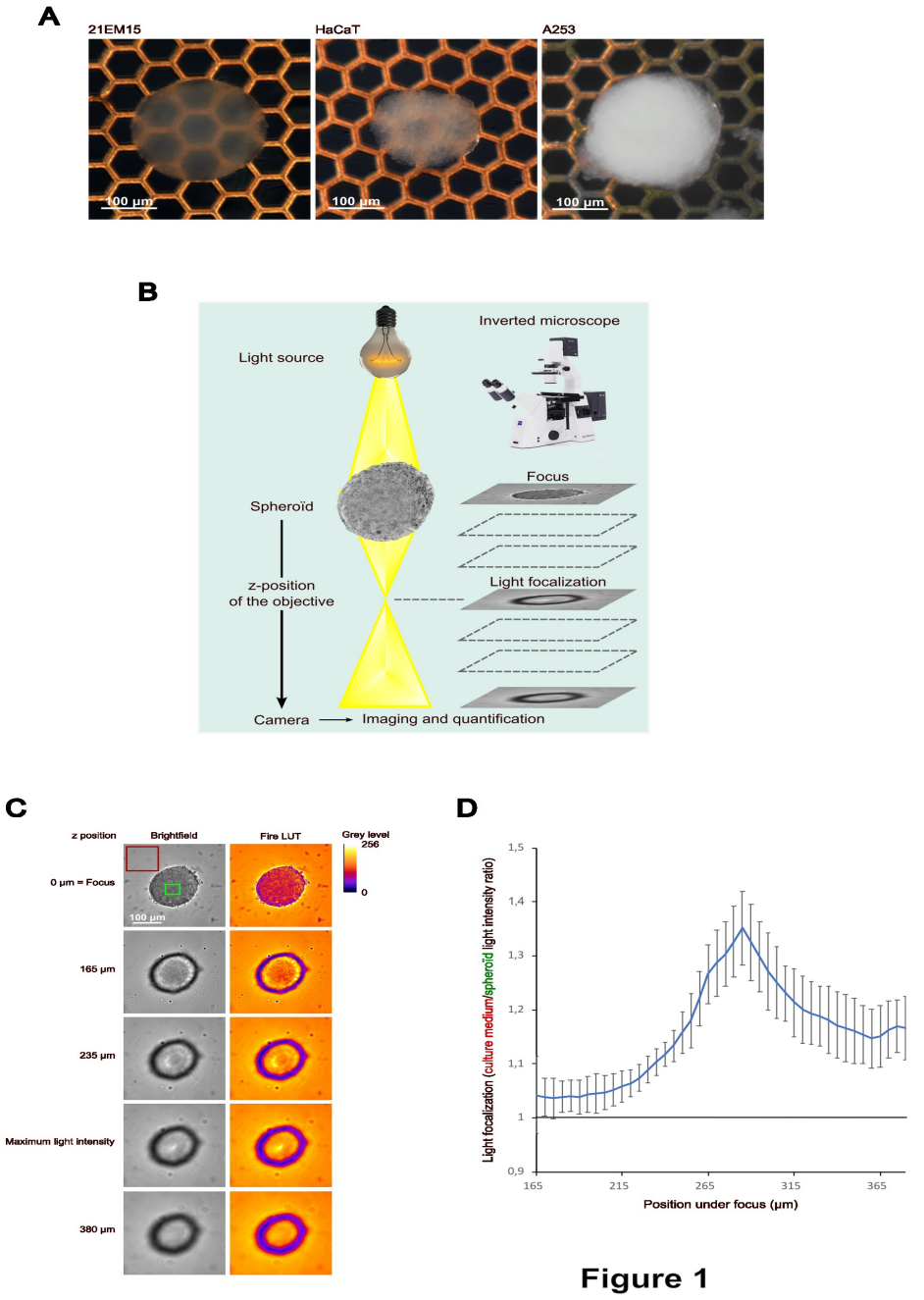


Figure 1

Figure 1. 21EM15 spheroids are transparent and can focus light. A, Macroscopic views of 21EM15, HaCaT or A253 10-day old spheroids formed with 10,000 cells. The electron microscopy grid allows evaluation of transparency. B, Schematic of the imaging setup for quantifying the light focusing ability of spheroids. C, Microscopic images of a 21EM15 spheroid showing its ability to transmit and focus light. The mean intensity of light transmitted by the medium and the maximum intensity of light transmitted by the spheroid used for quantification are respectively indicated by the red and the green squares. D, Graph showing the light focusing ability of the 21EM15 spheroids calculated as the ratio between the mean intensity measured in the red square and the maximum intensity measured in the green square. This graph is representative of 5 independent experiments with n=12 spheroids for each experiment. Error bars represent standard deviations. All spheroids were generated from 10,000 cells.

3.2. Transcriptome analysis of 21EM15 spheroids reveals strong similarities with lens development

During their growth, the 21EM15 spheroids acquire asymmetric shape and optical properties to focus light. To gain insights into the molecular modifications associated with these morphological changes and assess whether they may be relevant to lens development, we compared the transcriptome landscape of 21EM15 spheroids with those of mouse lenses at several stages of development. We profiled gene expression in 21EM15 spheroids by 3' end RNA-sequencing (Supplemental table 1) and we retrieved gene expression levels of mouse lenses from the iSyTE 2.0 database [23]. In iSyTE 2.0, "lens-enrichment" is estimated as the log-ratio of gene expression in the lens to gene expression in the whole embryonic body (WB). Using WB comparative analysis, we similarly estimated gene enrichment in 3D 21EM15 cultures. We compared the 10% most enriched genes in 3D cultures (N = 1032, 10,320 genes in total) with the 10% most enriched genes in E14.5 lenses (N = 1032). We found that 198 genes are present at the overlap of the two datasets, which is far above what was expected by chance ($p = 8.7x10^{-22}$, hypergeometric test). Irrespective of the threshold set to classify the genes as top-enriched (between 0 and 10%), the number of genes observed in the overlap of top-enriched genes in 3D cultures and top-enriched genes in E14.5 lenses largely exceeds that expected (Figure 2A). Nine genes are present in the overlap of top-1% most enriched genes in E14.5 lenses and the top-1% most enriched genes in 3D cultures, whereas only one was expected by chance. Among these genes are the *Cryab*, *Six3*, *Adamtsl4*, *Cp* (encoding ceruloplasmin), *Crim1*, *Dkk3* and *Nupr1*, all of which are known to be directly linked to lens pathophysiology (Figure 2A). We obtained similar results for all lens development stages present in iSyTE 2.0 (E10.5, E12.5, E14.5, E16.5, E17.5, E19.5, data not shown). Together, these results reveal an overlap in expression of key genes between 21EM15 spheroids and normal lenses.

We next wanted to assess the contributions of the culture conditions (3D vs. 2D) on gene expression, especially as it relates to the lens. To do so, we profiled gene expression in 21EM15 2D cultures by 3' end RNA-sequencing in the same conditions as 3D spheroids (Supplemental Table 1). Principal component analysis and hierarchical clustering analysis showed that the 2D and 3D samples cluster separately from each other (Supplemental Figures 2A and B). We therefore used these datasets to identify differentially expressed genes (DEGs). At FDR = 0.01 and log2(Fold Change) > 1 in absolute value, this analysis uncovered 291 genes that are elevated and 191 genes that are reduced in 3D spheroids (Figure 2B). The elevated 291 genes are referred to as "3D genes" and the reduced 191 genes as "2D genes". As expected, the heat map of these 482 DEGs shows a clear separation between these two sets of genes (Figure 2C). Several genes elevated under 2D conditions of growth relate to the cell cycle, such as *Ccnb1*, *Ccnb2*, *Ccnb1*, *Ccne2*, *Cdc20* (Figure 2B). This indicates that culturing 21EM15 in 3D conditions reduces the expression of cell cycle genes comparted to growth under 2D conditions. Conversely, several genes elevated under 3D growth conditions are relevant to lens development (Figure 2B). These include *Aqp1, Wls, Cdkn1, Cxcr4, Apoe, Lama4, Lamp2, Ctsl, Notch3, Wnt6, Cp, Psen2* and *Maf*.

Finally, we used iSyTE 2.0 to examine the expression of these DEGs in normal lens development [23]. On average, 2D genes are more expressed than 3D genes in early lens developmental stages (*e.g.* E10.5, when the lens placode has invaginated into a "lens pit"), and their expression decreases as the lens progresses in development (Figure 2D). Conversely, the mean expression of 3D genes increases progressively in normal lens development (Figure 2D). Hence, switching the culture conditions of 21EM15 cells from 2D to 3D spheroids partly recapitulates gene expression changes in normal lens development. Together, these data show that culturing 21EM15 in 3D conditions reinforces their similarity with normal mouse lenses, and in particular at later stages of development.

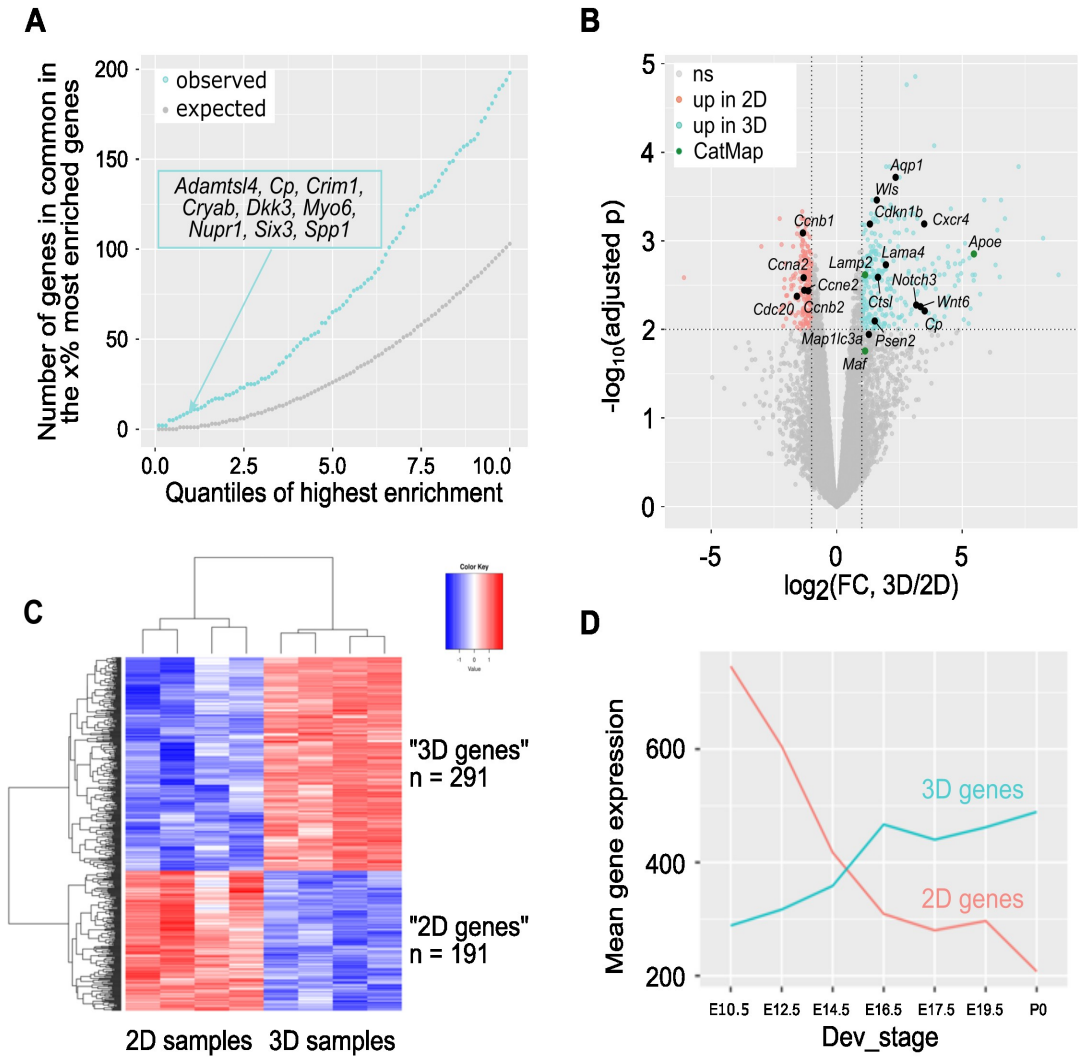


Figure 2. Growing 21EM15 cells in 3D culture conditions captures aspects of gene expression in lens development. A, We ranked all the genes expressed in mouse E14.5 lens and in 21EM15 3D cultures (N = 10,320) based on lens-enrichment (expression in the lens compared to whole body) as described in [23]. We separately listed the percent (x%) of the most enriched genes in the lens and in 3D cultures for several x values ranging from 0 to 10 (X-axis). For each x value, we retrieved the genes in common between the x% most enriched genes in lens and the x% most enriched genes in

3D cultures. The Y-axis shows the number of shared genes for each x value. Blue, the observed number. Grey, the number expected if enrichment in 3D 21EM15 cultures is independent from enrichment in E14.5 lenses. The inset is the list of genes that are in both the 1% most enriched genes in 3D 21EM15 cultures and in the 1% most enriched genes in E14.5 lenses. B, Volcano plot showing the statistical significance (Benjamini-Hochberg adjusted p-value, -log10 scale) against the fold change (log2 scale) of the expression in 2D and 3D samples. This analysis identifies 482 differentially expressed genes (FDR < 0.01 and log2(FC) > 1 in absolute value), among which 291 are upregulated in 3D samples and 191 are upregulated in 2D samples. Genes that are found to be linked to cataract in the database Cat-Map are indicated. C, Heat map of differentially expressed genes in 21EM15 spheroids (3D samples) and 21EM15 2D cultures. D, Mean expression of the 191 "2D genes" and the 291 "3D genes" throughout lens development. In A and D, data for gene expression in lens are from iSyTE 2.0 [23].

3.3. 21EM15 3D cultured cells form multilayered lens organoids

The lens is surrounded by a basal lamina called the capsule. Under the capsule and starting from the anterior pole of the lens, there is a quiescent epithelium, containing cells that proliferate in the “germinal zone” and exit the cell cycle at the “transition zone” to initiate differentiation into fiber cells that contribute to the bulk of the lens. Differentiation into fibers cells is characterized by the remodeling of the cytoskeleton leading to elongation of the cells, accompanied by high levels of expression of key lens proteins (e.g., crystallins, membrane proteins, etc.) and the progressive loss of cellular organelles [1,2]. As 21EM15 cells cultured as 3D spheroids express genes associated with lens differentiation, we wanted to characterize their structure to identify whether different cell types emerge through a process of differentiation. Histological analysis shows that twelve hours after seeding under 3D conditions, 21EM15 cells cluster to form a non-organized flattened structure (Figure 3A). Twenty-four hours after seeding, the structure is spherical with a rather homogenous cellular content. Ten days after seeding, the spheroids show a different appearance, with distinct zones: the external zone with round nuclei, the intermediate zone with elongated nuclei, and the internal zone with cells characterized by an intense pink cytoplasm and small compacted nuclei (Figure 3A). Importantly, the internal zone is off-centered, revealing that the initial central symmetry of the spheroid was broken during the 10-day culture (Figure 3A). This is in line with previous observations that the spheroid acquires an ovoid shape after a few days of culture (Supplemental Figure 1A).

To evaluate cell compartmentalization in the core of the spheroid we visualized cells boundaries, using WGA (wheat germ agglutinin), a membrane marker. The core of the spheroid is composed of highly compacted cells when compared to the cortex (Figure 3B). This indicates that the cells located at the central region of the spheroid undergo a phenomenon of packing. As controls, histological sections of HaCaT and A253 spheroids were generated (Figure 3C). After 10 days of culture, specific organization of cells was not observed in these controls. HaCaT spheroids are made of cells roughly aggregated and cavities likely filled with extracellular matrix, while A253 spheroids are made of homogenously distributed cells with a clear pink cytoplasm and some spots of necrotic cells (Figure 3C). As 21EM15 spheroids constantly grow over a period of more than 10 days (see Supplemental Figure 1B), we wanted to determine which cells were responsible

for the growth. Staining with the KI67 antigen, an established marker of proliferation, shows that proliferating cells are essentially localized in the external region of the spheroid and that cells stop proliferating once their nuclei are elongated (Figure 3D). The observation that only a few cells are able to proliferate is consistent with the finding that many cell cycle-related genes are down-regulated in 3D cultures (Figure 2B). No cell undergoing apoptosis was observed after 10 days of culture (Figure 3E).

The above data indicate that 21EM15 spheroids are made up of at least 3 distinct regions. We next sought to determine the gene expression landscape within the different layers of the spheroids. Toward this goal, the most internal and external regions were isolated by laser capture microdissection (LCM) and subjected to 3' end RNA-sequencing (Supplemental table 2). Due to technical limitations, we were unable to microdissect the intermediate region. We retained n=3 external region samples and n=4 internal region samples based on PCA (Supplemental Figure 3A). We identified 793 DEG (FDR = 0.01, and log2(Fold Change) > 1 in absolute value). Of these, 465 exhibit enriched expression in the external region and 328 exhibit enriched expression in the internal region (Figure 3F). The heat map of these 793 DEGs separates "external genes" from "internal genes" (Supplemental Figure 3B). Among the genes that are overexpressed in the internal region, we found genes known to be expressed in fiber cells like *Ank2*, *Atp1b1*, *Cap2*, *Eya1*, *Fundc1*, *Fzd6*, *Hsf4*, *Jag1*, *Maf*, *Meis1*, *Prox1*, *Tdrd7*, *Wls*. Among the genes that are overexpressed in the external region, we found genes known to be expressed in lens epithelial cells like *Ccna2*, *Ccnb1*, *Ccnb2*, *Ccnd1*, *Ccne2*, *Cdc20*, *Cdk1*.

To globally assess the resemblance of the internal and external regions of 21EM15 spheroids with lens fiber and epithelial cells, respectively, we retrieved the transcriptomic data from microdissected E14.5 epithelial cells and lens fiber cells [24]. Of the 793 DEG, 378 are also enriched either in E14.5 lens epithelium or fiber cells. The "external genes" (negative $\log_2$(FC) in Figure 3G), are enriched in lens epithelial genes (green spots), whereas "internal genes" (positive $\log_2$(FC) in Figure 3G) are enriched in lens fiber cells (orange spots). The contingency table shown in Figure 3H confirm this bias ($p = 2.1*10^{-6}$, chi-square test). These data confirm that the transcriptome of the internal region resembles that of lens fiber cells and the transcriptome of the external region that of lens epithelial cells. Further, while 2D cultures of 21EM15 cells have overlapping expression with lens epithelial cells, growing the cells in 3D commits the internal cells toward a differentiation program overlapping with lens fiber cells. Taken together, our results show that the 3D 21EM15 cultures self-organize and establish different cell types expressing specific gene sets, thus mimicking certain aspects of lens development. Moreover, these structures are able to break their original central symmetry to establish axial symmetry, which is characteristic of organoid development [25–27]. Thus, henceforth, the 3D 21EM15 cultures will be referred to as 21EM15 “organoids”.

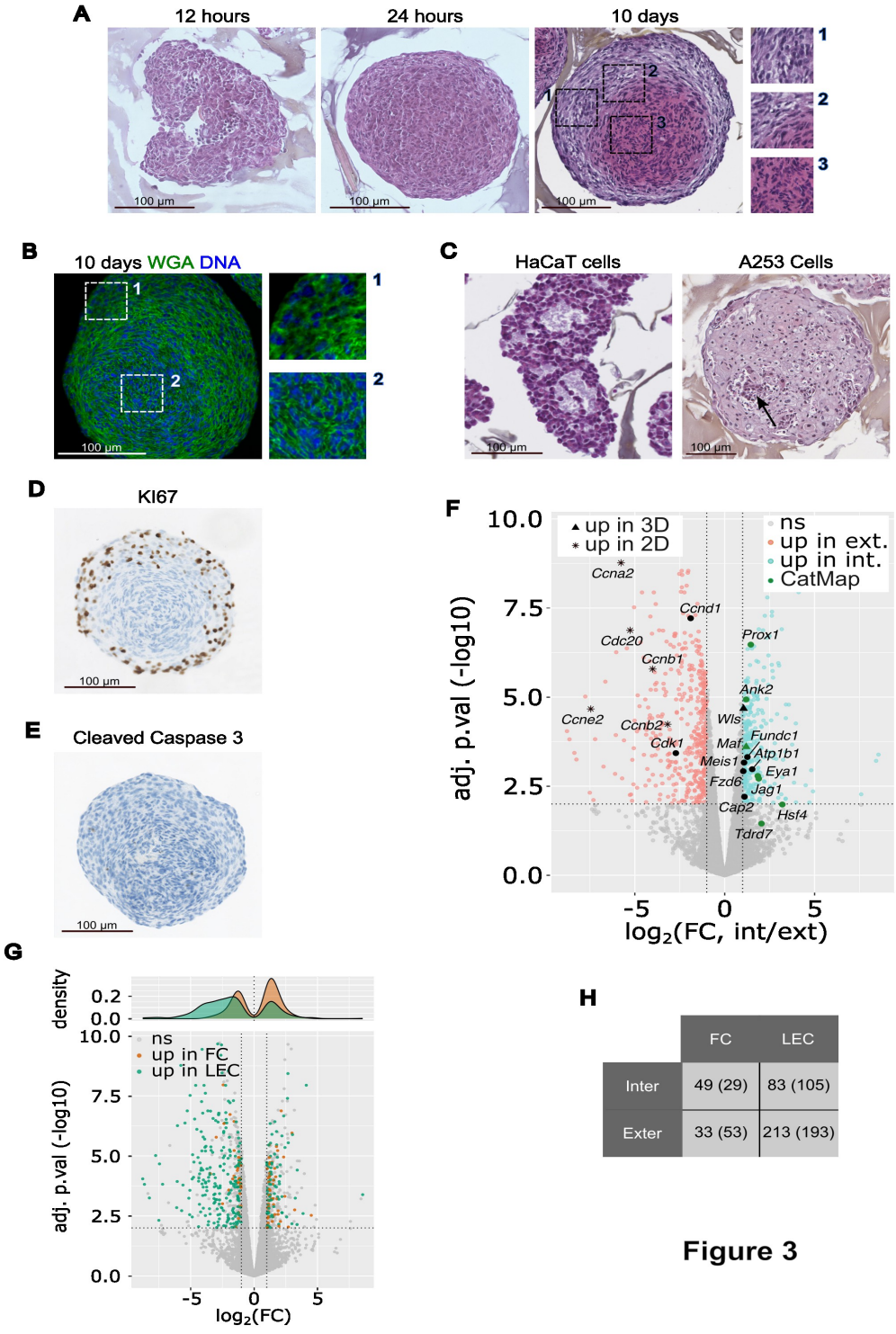


Figure 3

Figure 3. 21EM15 LECs 3D cultures differentiate to form multilayered lens organoids. A, Histological analysis of 21EM15 spheroids grown for 12 hours, 24 hours or 10 days, stained with Hematoxylin, Eosin and Safran (HES). The insets show the three different histological regions. B, Microscopic image of a 21EM15 spheroid stained with Wheat Germ Agglutinin (WGA) to show distinct cellular boundaries. Insets show details of the outer (1) or inner regions (2). C, Histological analysis of HaCaT or A253 spheroids grown for 10 days (HES staining). The arrow shows a necrotic region. D, Histological analysis showing the localization of KI67 in a 10-day old 21EM15 spheroid. E, Cleaved Caspase 3 staining of 10-day old 21EM15 spheroid, revealing an absence of cells undergoing apoptosis. For figures A-E, data are

representative of at least 3 independent experiments, n=30 organoids. F, Volcano plot showing the statistical significance (Benjamini-Hochberg adjusted p-value, -log10 scale) against the fold change (FC, log2 scale) of the expression in internal and external regions microdissected from 21EM15 spheroids. This analysis identifies 793 differentially expressed genes (FDR < 0.01 and log2(FC) > 1 in absolute values), among which 328 are upregulated in internal regions and 465 are upregulated in external regions. "Up in 2D" and "Up in 3D" correspond to genes respectively up in 2D or up in 3D in figure 2B. Black solid symbols correspond to genes that are of particular significance regarding the lens. G, Lower panel, same volcano plot as in F, with genes that are overexpressed in microdissected lens epithelial cells (LEC) and fiber cells (FC) [24] colored in green and orange, respectively. Higher panel, density plot showing the distribution of fold changes (log2 scale) for FC and LEC genes. H, Contingency table showing the number of genes that are enriched in FC or LEC compared to internal and external regions. The numbers in the brackets represent the value that is expected in the event that the enrichment in FC or LEC was unrelated to the enrichment in internal or external regions, respectively. The difference between the expected and the observed value indicates that a higher number of genes than expected are found to be enriched both in FC and internal regions, as well as in LEC and external regions.

3.4. Morphological organization of 21EM15 organoids partially recapitulates lens patterning

Different regions or cell types within the lens can be characterized by expression of specific markers. From the transcriptomic studies described above, we identified a subset of key genes involved in lens development to be differentially expressed between the inner and outer regions (see Supplemental Table 2). Our next objective was to confirm that key lens genes have specific expression patterns relevant to the organization of a whole lens. For this purpose, we examined the expression and the localization of structural components such as Laminin, a major component of basal lamina including the lens capsule, and αB-Crystallin (CRYAB), a major component of lens fiber cells in later developmental stages [28,29]. We also examined transcription factors such as PAX6, PROX1 (elevated in fiber cells) and signaling molecules as JAG1 [30–32]. Laminin was found to be present in 2 or 3 layers of the outermost cells (Figure 4A). It is also present in the most peripheral part of the lens, but only in the form of a single outer basal lamina [29]. Immunostaining showed that nuclear PAX6 is present in an internal region surrounding the central core of the organoid composed of highly compacted cells (Figure 4B). Consistent with symmetry breaking, it then extends toward the central axis. JAG1 is also found in an asymmetric distribution, as it is enriched in the membranes of cells localized in two lateral areas surrounding the central axis of the organoid (Figure 4B). PROX1 is present in two regions: in the cytoplasm of cells that most strongly express *Jag1*, and in the nuclei of cells that more weakly express *Jag1* and are localized along the central axis (Figure 4B and C). Finally, αB-Crystallin is low/absent in the external cell layers but high in the cortex and the core of the organoid with both cytoplasmic and nuclear localization (Figure 4D).

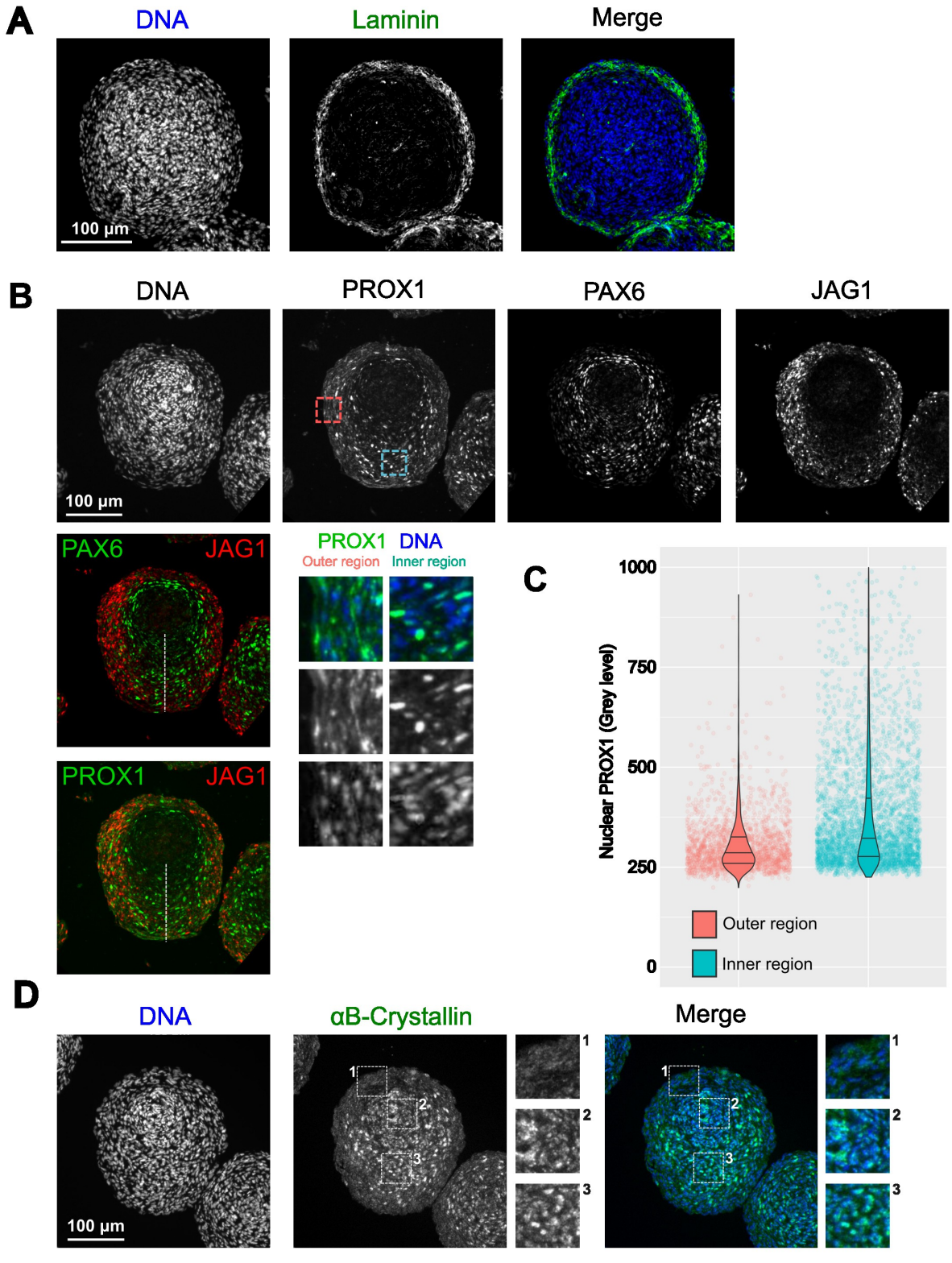


Figure 4. Morphological organization of 21EM15 organoids partially recapitulates eye lens patterning. A, Immunofluorescence (IF) microscopy reveals the localization of Laminin on the outer region of the organoid. B, Multiplex microscopic images indicates the localization of the lens expressed proteins PAX6, PROX1 and JAG1. The red and blue squares correspond to the typical areas (respectively outer and inner regions) used to display the magnification

insets presented below the grey level image of PROX1 and to quantify the PROX1 nuclear labeling shown in C. The white dashed line symbolizes the central axis of the organoid. Violin plots combined with jittered scattered plots showing that PROX1 is enriched in nuclei of cells located in the inner region. This graph is representative of three independent experiments, n=30 organoids. $p < 2.2 \times 10^{-16}$, Wilcoxon rank sum test with continuity correction. D, IF images showing the localization of αB-Crystallin. Insets show the enlargement of the outer region, the central axis and the core of the organoid. Data are representative of at least three independent experiments, n=30 organoids.

One of the most striking characteristics of fiber cell differentiation is the progressive degradation of cellular organelles such as nuclei and mitochondria [33]. In the lens, these various events can be highlighted by the observation of components of the nuclear envelope, of the mitochondria or by the expression of specific transcription factors [34–37]. We sought to examine whether similar cellular changes occurred in 21EM15 organoids. We find that Lamin-B1 (LMNB1), a component of the nuclear envelope, progressively disappears from the exterior to the interior of the organoid (Figure 5A). While Lamin-B1 labeling surrounds the nucleus in a continuous manner in the outermost cells, this labeling becomes more and more discontinuous toward the inner region of the organoid, until it completely disappears. Concomitantly, we observed a gradual change of nuclei shape accompanied by chromatin compaction evoking pyknosis (Figure 5B). These nuclei are not transcriptionally active in cells located in the center of the organoids (Figure 5C) as indicated by the progressive loss of nuclear Fibrillarin (FBL), which is considered as a marker of transcriptional status of fiber cell nuclei [38]. Mitochondria are also in a process of degradation as indicated by the decrease in TOMM20 labelling, a constituent of the mitochondria external membrane, toward the center of the organoid (Figure 5D). Conversely, the expression of *Hsf4,* a gene encoding a transcription factor whose downstream targets are considered to be involved in organelle degradation [35,39], increases in the core region relative to the outer region (Figures 5E and F).

Taken together, these results suggest that similar to the cellular and morphological changes accompanying lens development, the 21EM15 organoids are organized into specific expression domains for key lens proteins like αB-Crystallin, PAX6, PROX1 and JAG1. Moreover, the cells lying in the internal-most region commit to a process of organelle degradation reminiscent of what is observed in the whole lens. Our results thus show that 21EM15 organoids recapitulates specific molecular aspects and morphological organization of the lens.

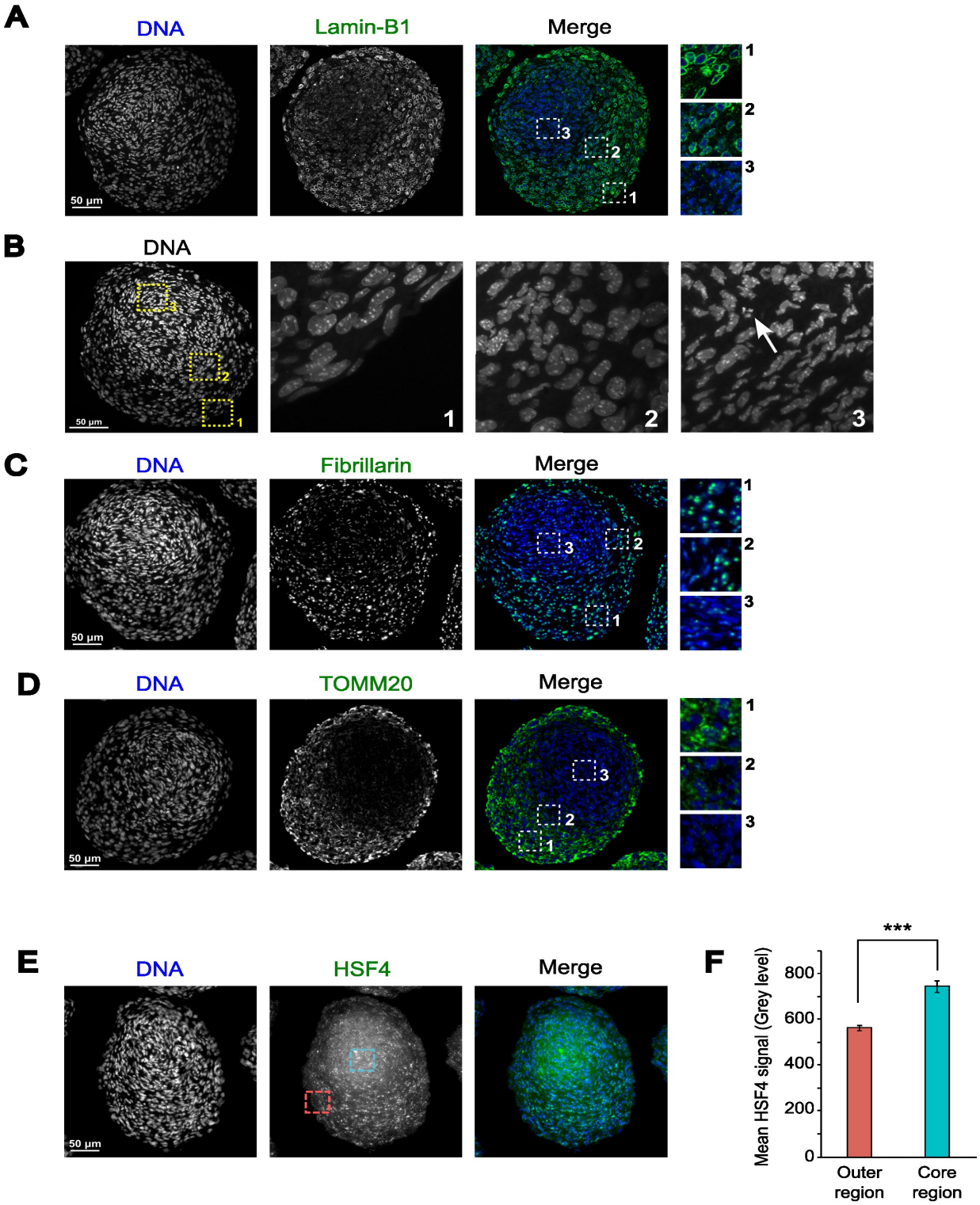


Figure 5. Cells located in the core of lens organoid are engaged in a process of organelle degradation. A, IF microscopy images of 10-day old lens organoid showing the localization of Lamin-B1. B, Microscopy images showing the appearance of nuclei in three different areas of an organoid. The arrow points out a nucleus with pyknotic features. C, Nuclei located in the core of the organoid show reduced transcriptional activity, as evidenced by decreased levels of Fibrillarin, which has previously been used as a marker of the transcriptional state of lens fiber cells. D, The decrease in TOMM20 labeling indicates that the organoid core cells are engaged in mitochondrial degradation. E, IF microscopy images showing the localization of HSF4. The red and blue squares correspond to the typical areas (respectively outer and inner regions) used to quantify the HSF4 mean signal quantified in F. F, Histogram presenting the quantification of the mean HSF4 signal in the outer and the core regions of the organoids. This graph is representative of three independent experiments, n=30 organoids. Asterisks indicate a *p*-value < 0.001. Error bars represent standard deviations.

3.5. 21EM15 organoids model lens cataract

As the organoids described above present interesting optical, morphological, histological, molecular and functional characteristics, we explored if they could be utilized as a model to uncover the pathophysiology of the lens. Cataract can be induced by H_2O_2 or hypertonic NaCl treatments in dissected lens [40]. Therefore, we incubated 8-day old 21EM15 organoids for 48 hours with these compounds (at previously used concentrations) and evaluated their transparency and light focusing ability. We found that H_2O_2 does not trigger changes in transparency or light focusing ability for concentrations ranging from 0 to 350 µM, but organoids become opaque and cease to transmit light at concentrations above 500 µM (Figures 6A and B). Increasing concentrations of NaCl (from 1.25% to 1.7%) gradually reduce organoid transparency (Figure 6A), but only the highest concentration has a significant impact on light focusing ability (Figure 6C). We next sought to examine whether the 21EM15 organoid model could be applied for testing function of genes associated with cataract. We previously showed that *Celf1* deletion in a germline or lens conditional manner causes early-onset cataract in mice [41]. We therefore tested the transparency and light focusing ability of organoids made from 21EM15 cells stably expressing a shRNA targeting the *Celf1* gene [41,42]. Interestingly, *Celf1* knockdown reduces organoids transparency, as observed in mice deficient for *Celf1* (Figure 6D). It also reduces the light focusing property of 21EM15 organoids (Figure 6E). All together, these results demonstrate that 21EM15 organoids are a new model that can be used to study cataract.

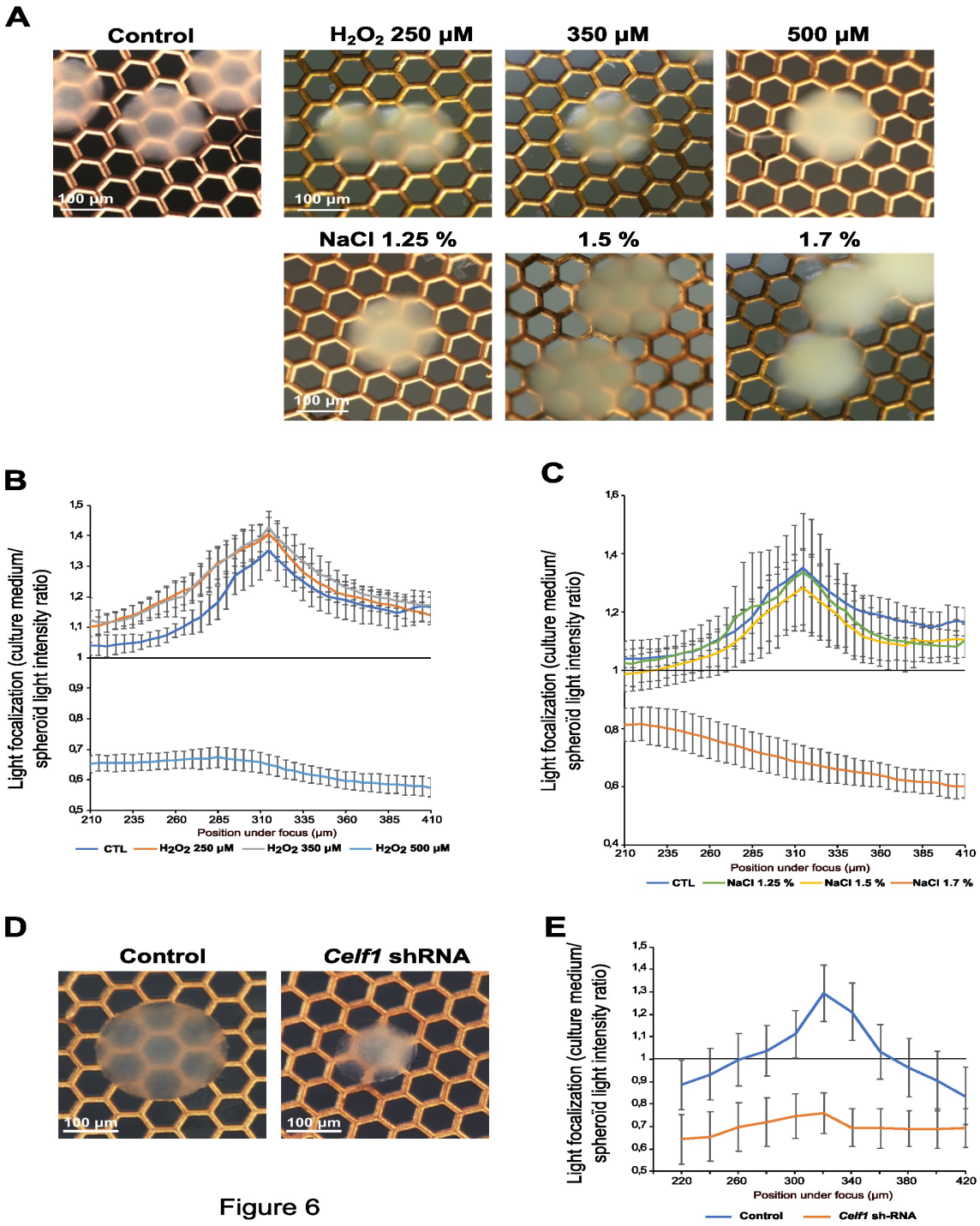


Figure 6

Figure 6. 21EM15 lens organoids can be induced to develop opacity. A, Macroscopic views of 10-day old lens organoids treated with the indicated concentrations of H_2O_2 or NaCl. The electron microscopy grid allows evaluation of transparency. B, Graph showing the light-focusing ability of the 21EM15 lens organoids treated with increasing concentrations of H_2O_2. C, Graph showing the light-focusing ability of the 21EM15 lens organoids treated with increasing concentrations of NaCl. D, Macroscopic views of lens organoids expressing control or *Celf1*-targeting shRNA. The electron microscopy grid allows evaluation of transparency. E, Graph showing the light-focusing ability of the 21EM15

lens organoids expressing control or *Celf1*-targetting shRNA. Graphs B, C and E are representative of three independent experiments with n=12 organoids 10-day old for each experiment. Error bars represent standard deviations.

4. Discussion

In the present study, we have developed a mouse 3D lens model that can be rapidly generated and can be applied for studying processes relevant to lens biology. It has specific optical properties, including transparency and light-focusing ability (Figure 1). We characterized this model by a combination of histological, transcriptomic and immunohistochemical/immunofluorescence approaches. This analysis revealed that, similarly to the lens, 21EM15 lens organoids comprises three main regions, a peripheral layer, an intermediate part and a core region. Transparency and light-focusing ability indicate that even though the histological organization of the 21EM15 organoids is simplified compared to a true lens, it is sufficient to bring relevant optical properties. In addition to transparency and light-focusing ability, refractive index measurement, as described by Young and colleagues [43], could also be very useful in the future to assess the influence of cellular organization on organoid optics.

When grown in 2D culture conditions, 21EM15 cells express typical LEC genes [21]. In the present study, we used 3'end RNA-seq to profile the transcriptomes of 21EM15 2D and 3D cultures, and of laser-micro-dissected internal and external regions of the organoids. These analyses show that, regardless of how they are grown (2D or 3D), 21EM15 cells express large amounts of various crystallins, as *Cryab*, *Cryba4*, *Crybg1*, *Crybg3*, *Cryz*, *Cryzl2*, *Cryl1* (See supplemental table 1 and 2). Surprisingly, the organoids did not express the *Crygs* gene, although it had previously been shown to be expressed in 21EM15 cells cultured in 2D. This may be related to the fact that 21EM15 only express weak amount of *Crygs* and to the different methods used (Microarray *vs* 3'end RNA-seq) [21]. More importantly, these results also indicate that a subset of genes associated with lens development and fiber cell differentiation are induced under 3D culture conditions (Figure 2D). Among this set of genes we found *Six3*, a well characterized lens development gene [44], *Cp* that is typically expressed in lens epithelium [45], *Cryab*, *Crim1* and *Nupr1* that are expressed in lens fiber cells [23,28,46,47] and *Dkk3* that is a component of the Wnt signaling pathway involved in lens development [48] (Figure 2B). RNA-seq analysis also shows that the genes elevated in the external region of the lens organoids are enriched for candidates that are related to the cell cycle function (*Ccna2*, *Ccnb1*, *Ccnb2*, *Ccnd1*, *Ccne2*, *Cdc20*, *Cdk1*), whereas those elevated in the internal region are enriched for candidates involved in the Pax6 regulatory pathway (*Cap2*, *Meis1*), Notch and Wnt signaling (*Jag1*, *Wls*, *Fzd6*), Maf pathway (*Maf*, *Eya1*), lens fiber cell morphology and physiology (*Ank2*, *Atp1b1*, *Prox1*, *Tdrd7*) and organelle degradation (*Hsf4*, *Fundc1*) (Figure 3F and Supplemental table 2). These results are relevant to lens organization, as the anterior epithelium is involved in lens growth and the cortex is the place where fiber cells progressively differentiate [1,2].

Accordingly, KI67 staining shows that only the external cells are proliferative (Figure 3D). We also found the outermost region to be positive for Laminin staining (Figure 4A).

Laminin is a glycoprotein specific of basal lamina. It is present in the capsule, a cell-free structure around the lens [29]. In 21EM15 organoids, laminin appears to be secreted by cells in the first two or three layers that are encompassed into this extracellular matrix. We were unable to detect E-cadherin or FOXE3 labeling in these external-most cell layers, suggesting that organoid epithelial cells do not fully recapitulate properties of lens anterior epithelium, which express E-cadherin or FOXE3 [49,50] (data not shown). This suggests that while 21EM15 cells are able to proliferate or enter the early stages of fiber cell differentiation, they cannot become true epithelial cells in the culture conditions that we used. One possible explanation relates to the fact that 21EM15 cells were likely selected based on their ability to proliferate rapidly [21]. Although it is now established that the anterior epithelium contains stem cells [20,51], the 21EM15 cells probably originate from the germinative zone and are therefore engaged in the early stages of fiber cell differentiation process, preventing them from committing to the epithelial fate. Interestingly, they express *Pax6* and *Bmi1*, which are required for lens epithelial cells to regenerate a functional lens [20], but also typical stem cell marker genes like *Nes*, *Chrd*, *Sox4, Sox9, Sox12* and *Klf4* [21]. They are also able to spontaneously form lentoid bodies [21,22]. Finally, we show here, based on histological analysis, that symmetry is broken in 3D 21EM15 cell cultures (Figures 3A, 4, 5). Symmetry breaking is a general hallmark of organoids development [25–27]. Taken together, these observations suggest that 21EM15 cells possess stem cell-like properties accounting for their ability to form lens organoids.

A key property of lens fiber cells is the elimination of organelles, as they are potential sources of light scattering. We tested if this also applies to the core region of 21EM15 organoids. Several gene regulatory networks involved in autophagy or nuclear degradation in the lens have been identified [33,52]. One of them involves the transcription factor HSF4, a major regulator of membrane organelle degradation in the lens [35,39,53,54]. Accordingly, we found in immunostaining experiments that HSF4 is more abundant in the core than in the peripheral region of the organoid (Figure 5E). At the RNA level, *Bnip3l*, *Lamp1*, *Fundc1*, and *Smurf1*, which are also involved in organelle degradation [33,52], also exhibit elevated expression in the core region than in the outer region of organoids (Supplemental Table 2). The high expression of these genes is accompanied by an apparently complete degradation of the mitochondria, as inferred from the loss of TOMM20 staining, a mitochondrial marker commonly used to assess mitophagy in various cell types including lens fiber cells [34,37,55] (Figure 5D). The shape of the nuclei is also strongly affected in the organoid core, with some nuclei clearly showing a pyknosis-like appearance (Figure 5B). Pyknotic nuclei are characteristic of various types of terminal cell differentiation processes requiring nucleus degradation, as in red cells or lens fiber cells [56]. In addition to organelle degradation, fiber cell differentiation is also featured by a strong expression of crystallins, which involves PAX6 and to some extent, HSF4 [28]. αb-Crystallin (*Cryab*) is poorly expressed in the outer region whereas it is enriched in the central region where *Pax6* is expressed, as revealed by both 3'end RNAseq on laser-microdissected regions (Supplemental Table 2, FDR = 0.03) and IF (Figure 4D). This is similar to the expression pattern of αb-Crystallin in wild-type

lenses, wherein it is high in fiber cells compared to epithelial cells in later stages of development [24]. Together, these data show that, while the outermost layer of 21EM15 organoids resembles lens epithelial cells, their core region resembles fiber cells.

In the lens, the transition zone is the area where cells of the anterior epithelium exit the cell cycle and begin differentiation into fiber cells that contribute to the bulk of the lens tissue. Beyond the transition zone, cells sequentially elongate and degrade their organelles while expressing key lens development master genes. We were unable to profile the transcriptome of the intermediate region of 21EM15 organoids due to technical limitations in laser microdissection. However, we gained significant insights into this zone by immunostaining of PAX6, JAG1 and PROX1 (Figure 4). These key genes function to orchestrate the development of these different regions of the lens [30,57]. The function of JAG1 is to keep epithelial cells from the germinative zone undifferentiated and proliferating, by activating Notch signaling [57–59]. PAX6 and PROX1 respectively trigger cell cycle exit and are associated with expression of crystallins and cell elongation during secondary fiber cells differentiation [60–64]. PROX1 exhibits dynamic expression and localization in lens cells. PROX1 is first located in the cytoplasm in the anterior epithelium, and then becomes nuclear in the transition zone where it is involved in orchestrating fiber cell elongation and differentiation [31,63]. 21EM15 organoids show an interesting distribution of these 3 key genes. JAG1 is consistently localized in the lateral regions and severely reduced near the central axis, in contrast to PAX6, which is highly expressed in a ring-like manner surrounding the organoid core, and also in the central axis. The changes in PROX1 localization seem to recapitulate its endogenous lens expression. While there is diffused cytoplasmic labeling of PROX1 in the *Jag1*-expressing region, PROX1 becomes nuclear along the central axis, a region where *Pax6* is quite highly expressed. It is interesting to note that the cells that exhibit nuclear PROX1 are located in the area where the cells are most elongated, indicating that the organoids recapitulate this functional aspect of PROX1 similar to endogenous lens development (Figure 3A right and Figure 4B).

From a developmental point of view, the 21EM15 organoids do not form the lens vesicle typical of mammals and their development is more reminiscent of what happens in the fish eye. This is probably due to the fact that the environment of the organoid is very simplified compared with what happens in the whole eye. Nevertheless, the fact that differentiation events occur and that the expression of key lens genes is spatially regulated, accompanied by significant cellular changes (nuclei, mitochondria), makes this system a valuable tool for more in-depth studies. In particular, it could be used to understand the signaling pathways responsible for the cell polarity of the lens, its morphological asymmetry and the acquisition of its optical properties.

Overall, these results show that 21EM15 organoids recapitulate several aspects of lens organization, gene expression profiles, biological processes or optical properties. We therefore summarized these data in a model (Figure 7). In this model, we were unable to formally identify any canonical epithelium. We termed the outermost zone comprising the first layers of cells that are proliferative and are embedded in Laminin as the “capsuloid”. The capsuloid likely corresponds to the fusion of the capsule and the germinative zone of

the lens epithelium. The lateral region expressing Jag1 and Ki67 is probably a mix of more or less proliferating cells (as indicated by their expression of Ki67) and early differentiating cells (expressing *Jag1*). Although cells from the germinative zone do not express *Jag1*, it is tempting to compare this region of the organoid to the germinative zone in the lens, where cells that express *Jag1* would prevent cell cycle exit and fiber cell differentiation of proliferating cells. In a more internal region, between the germinative zone and the central axis of the organoid, cells begin to express *Pax6*, *Cryab*, and PROX1 becomes nuclear. This region is characteristic of the transition zone in endogenous lens development, where the cells progressively engage in the process of fiber cell differentiation. Interestingly, along the central axis and around the organoid nucleus, the cells no longer express *Jag1*. However, these cells still express *Pax6*, PROX1 is nuclear, and most importantly, exhibit elongation (Figure 4B). Cells in the central-most region exhibit a very intensely stained "pink" cytoplasm in histological analysis, reminiscent of the lens, with pyknotic nuclei and strong expression of *Hsf4* and *Cryab*. They also show features consistent with the degradation of their organelles (loss of Lamin B, TOMM20 and fibrillarin markers). All of these data suggest ongoing cellular and molecular processes in the lens organoids that contribute to transparency.

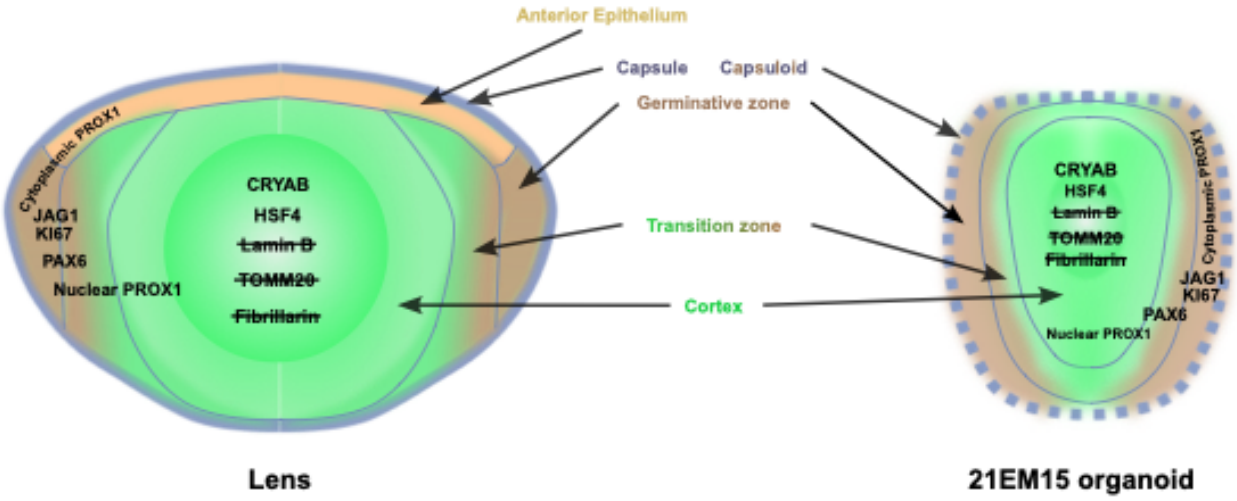


Figure 7. Model for 21EM15 lens organoid organization. This model shows that, unlike the lens, 21EM15 organoids lack the typical lens capsule and anterior epithelium. Instead, there is a zone that we term "capsuloid", comprising the outer layers of proliferative cells embedded in a laminin layer. Further inwards and from the outside in, 21EM15 organoids recapitulate several aspects of the organization of the lens, with a region corresponding to the germinal zone, where *Ki67* and *Jag1* are expressed, followed by a transition zone that expresses *Pax6* and where PROX1 progressively becomes nuclear. Along the central axis of the organoid, PROX1 is predominantly nuclear and cells progressively degrade their organelles, such as nuclei (loss of Lamin B and Fibrillarin) and mitochondria (loss of TOMM20 signal), and begin to express *Hsf4* and *Cryab*. These events correspond closely to those described in the normal lens cortex.

Finally, we addressed the suitability of 21EM15 organoids as a tool for studying lens opacity or cataracts (Figure 6). We observed that two previously established treatments to

induce cataract, namely exposure to H_2O_2 and NaCl, also are able to induce opacity in 21EM15 organoids. This is significant as previously, H_2O_2 was shown to model the events leading to age-related cataract [65–67]. Further, treatment with hyperosmotic NaCl was shown to trigger osmotic stress and disrupts fluid balance of lens, frequently associated with dry eye disease and diabetic cataracts [68–70]. Thus, 21EM15 organoids could find application in further understanding how these processes induce cataract. Finally, we also observed that knockdown of the cataract-linked RNA-binding protein encoding gene, *Celf1*, reduces 21EM15 organoids transparency and light focusing properties. CELF1 is involved in regulation of mRNA splicing, stability or translation [71], and its inactivation in mice leads to misregulation of post-transcriptional gene expression control in the lens and cataract [41,42,72]. These results indicate that 21EM15 organoids respond to cataractogenic conditions representative of a wide range of lens-related etiologies (*e.g.*, age-related cataract, diabetic cataract, genetic cataract) and thus can be used in further advancing knowledge on lens pathology.

5. Conclusions

In conclusion, our work presents a new mouse organoid model, that is effective and easy to set up, and does not require development of technical skills in stem cell culture. For a limited investment, both in terms of technique and time, this allows to relatively rapidly obtain lens organoids that recapitulate specific aspects of lens biology, with the added possibility of performing functional genetic analysis in a cost-effective manner. The 21EM15 organoids provide therefore the lens community with a compelling new model to improve the understanding of lens biology. Cataract remains a major public health problem that is currently only treated by surgery. It would therefore be interesting to develop drug treatments, particularly in order to offer alternatives to populations in countries that do not have ready access to surgery, or to prevent cataract formation in populations exposed to cataractogenic conditions. From a clinical point of view, these lens organoids should make it possible to develop screens for identifying compounds that impact the prevention and/or treatment of cataract.

Author Contributions: Conceptualization, Carole Gautier-Courteille, David Reboutier; data curation, Matthieu Duot, Luc Paillard; formal analysis, Matthieu Duot, Luc Paillard; methodology, Matthieu Duot, Carole Gautier-Courteille, David Reboutier; resources, Salil A. Lachke; investigation, Matthieu Duot, Carole Gautier-Courteille, David Reboutier, Roselyne Viel, Justine Viet, Catherine Le Goff-Gaillard; writing—original draft preparation, David Reboutier; writing—review and editing, Carole Gautier-Courteille, Catherine Le Goff-Gaillard, Salil A. Lachke, Luc Paillard, David Reboutier; visualization, Carole Gautier-Courteille, David Reboutier; supervision, Carole Gautier-Courteille, Salil A. Lachke, Luc Paillard, David Reboutier; project administration, Carole Gautier-Courteille, Luc Paillard; funding acquisition, Luc Paillard, Salil A. Lachke. All authors have read and agreed to the published version of the manuscript.

Funding: L.P. was supported by a grant from Retina France. S.A.L. was supported by National Institutes of Health / National Eye Institute [R01 EY021505 and R01 EY029770].

Acknowledgments: We thank the Biosit platforms (Rennes, France) for access to their facilities, and notably MRic (microscopy) and H2P2 (histology). We particularly thank Alain Fautrel, Nicolas Mouchet (H2P2 facility) for fruitful discussions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bassnett, S.; Shi, Y.; Vrensen, G.F.J.M. Biological Glass: Structural Determinants of Eye Lens Transparency. *Philos Trans R Soc Lond B Biol Sci* **2011**, *366*, 1250–1264, doi:10.1098/rstb.2010.0302.
2. Liu, Z.; Huang, S.; Zheng, Y.; Zhou, T.; Hu, L.; Xiong, L.; Li, D.W.-C.; Liu, Y. The Lens Epithelium as a Major Determinant in the Development, Maintenance, and Regeneration of the Crystalline Lens. *Prog Retin Eye Res* **2022**, 101112, doi:10.1016/j.preteyeres.2022.101112.
3. Shiels, A.; Hejtmancik, J.F. Inherited Cataracts: Genetic Mechanisms and Pathways New and Old. *Exp Eye Res* **2021**, *209*, 108662, doi:10.1016/j.exer.2021.108662.
4. Shiels, A.; Hejtmancik, J.F. Biology of Inherited Cataracts and Opportunities for Treatment. *Annu Rev Vis Sci* **2019**, *5*, 123–149, doi:10.1146/annurev-vision-091517-034346.
5. Liu, Y.-C.; Wilkins, M.; Kim, T.; Malyugin, B.; Mehta, J.S. Cataracts. *Lancet* **2017**, *390*, 600–612, doi:10.1016/S0140-6736(17)30544-5.
6. Daszynski, D.M.; Santhoshkumar, P.; Phadte, A.S.; Sharma, K.K.; Zhong, H.A.; Lou, M.F.; Kador, P.F. Failure of Oxysterols Such as Lanosterol to Restore Lens Clarity from Cataracts. *Sci Rep* **2019**, *9*, 8459, doi:10.1038/s41598-019-44676-4.
7. Makley, L.N.; McMenimen, K.A.; DeVree, B.T.; Goldman, J.W.; McGlasson, B.N.; Rajagopal, P.; Dunyak, B.M.; McQuade, T.J.; Thompson, A.D.; Sunahara, R.; et al. Pharmacological Chaperone for Alpha-Crystallin Partially Restores Transparency in Cataract Models. *Science* **2015**, *350*, 674–677, doi:10.1126/science.aac9145.
8. Zhao, L.; Chen, X.J.; Zhu, J.; Xi, Y.B.; Yang, X.; Hu, L.D.; Ouyang, H.; Patel, S.H.; Jin, X.; Lin, D.; et al. Lanosterol Reverses Protein Aggregation in Cataracts. *Nature* **2015**, *523*, 607–611, doi:10.1038/nature14650.
9. Goishi, K.; Shimizu, A.; Najarro, G.; Watanabe, S.; Rogers, R.; Zon, L.I.; Klagsbrun, M. AlphaA-Crystallin Expression Prevents Gamma-Crystallin Insolubility and Cataract Formation in the Zebrafish Cloche Mutant Lens. *Development* **2006**, *133*, 2585–2593, doi:10.1242/dev.02424.
10. Viet, J.; Reboutier, D.; Hardy, S.; Lachke, S.A.; Paillard, L.; Gautier-Courteille, C. Modeling Ocular Lens Disease in Xenopus. *Dev. Dyn.* **2019**, doi:10.1002/dvdy.147.
11. Chang, C.-C. *Animal Models of Ophthalmic Diseases*; Springer International Publishing: Switzerland, 2016;
12. Zhao, Z.; Chen, X.; Dowbaj, A.M.; Sljukic, A.; Bratlie, K.; Lin, L.; Fong, E.L.S.; Balachander, G.M.; Chen, Z.; Soragni, A.; et al. Organoids. *Nat Rev Methods Primers* **2022**, *2*, 1–21, doi:10.1038/s43586-022-00174-y.
13. O'Connor, M.D.; McAvoy, J.W. In Vitro Generation of Functional Lens-like Structures with Relevance to Age-Related Nuclear Cataract. *Invest Ophthalmol Vis Sci* **2007**, *48*, 1245–1252, doi:10.1167/iovs.06-0949.
14. Anchan, R.M.; Lachke, S.A.; Gerami-Naini, B.; Lindsey, J.; Ng, N.; Naber, C.; Nickerson, M.; Cavallesco, R.; Rowan, S.; Eaton, J.L.; et al. Pax6- and Six3-Mediated Induction of Lens Cell Fate in Mouse and Human ES Cells. *PLoS One* **2014**, *9*, e115106, doi:10.1371/journal.pone.0115106.
15. Fu, Q.; Qin, Z.; Jin, X.; Zhang, L.; Chen, Z.; He, J.; Ji, J.; Yao, K. Generation of Functional Lentoid Bodies From Human Induced Pluripotent Stem Cells Derived From Urinary Cells. *Invest. Ophthalmol. Vis. Sci.* **2017**, *58*, 517–527, doi:10.1167/iovs.16-20504.
16. Yang, C.; Yang, Y.; Brennan, L.; Bouhassira, E.E.; Kantorow, M.; Cvekl, A. Efficient Generation of Lens Progenitor Cells and Lentoid Bodies from Human Embryonic Stem Cells in Chemically Defined Conditions. *FASEB J.* **2010**, *24*, 3274–3283, doi:10.1096/fj.10-157255.
17. Murphy, P.; Kabir, M.H.; Srivastava, T.; Mason, M.E.; Dewi, C.U.; Lim, S.; Yang, A.; Djordjevic, D.; Killingsworth, M.C.; Ho, J.W.K.; et al. Light-Focusing Human Micro-Lenses Generated from Pluripotent Stem Cells Model Lens Development and Drug-Induced Cataract in Vitro. *Development* **2018**, *145*, dev155838, doi:10.1242/dev.155838.
18. Plüss, C.J.; Kustermann, S. A Human Three-Dimensional In Vitro Model of Lens Epithelial Cells as a Model to Study Mechanisms of Drug-Induced Posterior Subcapsular Cataracts. *J Ocul Pharmacol Ther* **2019**, doi:10.1089/jop.2019.0010.
19. Cvekl, A.; Camerino, M.J. Generation of Lens Progenitor Cells and Lentoid Bodies from Pluripotent Stem Cells: Novel Tools for Human Lens Development and Ocular Disease Etiology. *Cells* **2022**, *11*, 3516, doi:10.3390/cells11213516.
20. Lin, H.; Ouyang, H.; Zhu, J.; Huang, S.; Liu, Z.; Chen, S.; Cao, G.; Li, G.; Signer, R.A.J.; Xu, Y.; et al. Lens Regeneration Using Endogenous Stem Cells with Gain of Visual Function. *Nature* **2016**, *531*, 323–328, doi:10.1038/nature17181.
21. Terrell, A.M.; Anand, D.; Smith, S.F.; Dang, C.A.; Waters, S.M.; Pathania, M.; Beebe, D.C.; Lachke, S.A. Molecular Characterization of Mouse Lens Epithelial Cell Lines and Their Suitability to Study RNA Granules and Cataract Associated Genes. *Exp. Eye Res.* **2015**, *131*, 42–55, doi:10.1016/j.exer.2014.12.011.
22. Weatherbee, B.A.T.; Barton, J.R.; Siddam, A.D.; Anand, D.; Lachke, S.A. Molecular Characterization of the Human Lens Epithelium-Derived Cell Line SRA01/04. *Exp. Eye Res.* **2019**, 107787, doi:10.1016/j.exer.2019.107787.
23. Kakrana, A.; Yang, A.; Anand, D.; Djordjevic, D.; Ramachandruni, D.; Singh, A.; Huang, H.; Ho, J.W.K.; Lachke, S.A. ISyTE 2.0: A Database for Expression-Based Gene Discovery in the Eye. *Nucleic Acids Res* **2018**, *46*, D875–D885, doi:10.1093/nar/gkx837.
24. Zhao, Y.; Zheng, D.; Cvekl, A. A Comprehensive Spatial-Temporal Transcriptomic Analysis of Differentiating Nascent Mouse Lens Epithelial and Fiber Cells. *Exp Eye Res* **2018**, *175*, 56–72, doi:10.1016/j.exer.2018.06.004.
25. Anand, G.M.; Megale, H.C.; Murphy, S.H.; Weis, T.; Lin, Z.; He, Y.; Wang, X.; Liu, J.; Ramanathan, S. Controlling Organoid Symmetry Breaking Uncovers an Excitable System Underlying Human Axial Elongation. *Cell* **2023**, S0092-8674(22)01586-0, doi:10.1016/j.cell.2022.12.043.
26. Brassard, J.A.; Lutolf, M.P. Engineering Stem Cell Self-Organization to Build Better Organoids. *Cell Stem Cell* **2019**, *24*, 860–876, doi:10.1016/j.stem.2019.05.005.

27. Serra, D.; Mayr, U.; Boni, A.; Lukonin, I.; Rempfler, M.; Challet Meylan, L.; Stadler, M.B.; Strnad, P.; Papasaikas, P.; Vischi, D.; et al. Self-Organization and Symmetry Breaking in Intestinal Organoid Development. *Nature* **2019**, *569*, 66–72, doi:10.1038/s41586-019-1146-y.
28. Cvekl, A.; McGreal, R.; Liu, W. Lens Development and Crystallin Gene Expression. *Prog Mol Biol Transl Sci* **2015**, *134*, 129–167, doi:10.1016/bs.pmbts.2015.05.001.
29. DeDreu, J.; Walker, J.L.; Menko, A.S. Dynamics of the Lens Basement Membrane Capsule and Its Interaction with Connective Tissue-like Extracapsular Matrix Proteins. *Matrix Biol* **2021**, *96*, 18–46, doi:10.1016/j.matbio.2020.12.005.
30. Cvekl, A.; Zhang, X. Signaling and Gene Regulatory Networks in Mammalian Lens Development. *Trends Genet* **2017**, *33*, 677–702, doi:10.1016/j.tig.2017.08.001.
31. Duncan, M.K.; Cui, W.; Oh, D.-J.; Tomarev, S.I. Prox1 Is Differentially Localized during Lens Development. *Mech Dev* **2002**, *112*, 195–198, doi:10.1016/s0925-4773(01)00645-1.
32. Garg, A.; Hannan, A.; Wang, Q.; Makrides, N.; Zhong, J.; Li, H.; Yoon, S.; Mao, Y.; Zhang, X. Etv Transcription Factors Functionally Diverge from Their Upstream FGF Signaling in Lens Development. *Elife* **2020**, *9*, e51915, doi:10.7554/eLife.51915.
33. Brennan, L.; Disatham, J.; Kantorow, M. Mechanisms of Organelle Elimination for Lens Development and Differentiation. *Exp Eye Res* **2021**, *209*, 108682, doi:10.1016/j.exer.2021.108682.
34. Costello, M.J.; Brennan, L.A.; Basu, S.; Chauss, D.; Mohamed, A.; Gilliland, K.O.; Johnsen, S.; Menko, S.; Kantorow, M. Autophagy and Mitophagy Participate in Ocular Lens Organelle Degradation. *Exp Eye Res* **2013**, *116*, 141–150, doi:10.1016/j.exer.2013.08.017.
35. Cui, X.; Liu, H.; Li, J.; Guo, K.; Han, W.; Dong, Y.; Wan, S.; Wang, X.; Jia, P.; Li, S.; et al. Heat Shock Factor 4 Regulates Lens Epithelial Cell Homeostasis by Working with Lysosome and Anti-Apoptosis Pathways. *Int J Biochem Cell Biol* **2016**, *79*, 118–127, doi:10.1016/j.biocel.2016.08.022.
36. Imai, F.; Yoshizawa, A.; Fujimori-Tonou, N.; Kawakami, K.; Masai, I. The Ubiquitin Proteasome System Is Required for Cell Proliferation of the Lens Epithelium and for Differentiation of Lens Fiber Cells in Zebrafish. *Development* **2010**, *137*, 3257–3268, doi:10.1242/dev.053124.
37. Morishita, H.; Eguchi, S.; Kimura, H.; Sasaki, J.; Sakamaki, Y.; Robinson, M.L.; Sasaki, T.; Mizushima, N. Deletion of Autophagy-Related 5 (Atg5) and Pik3c3 Genes in the Lens Causes Cataract Independent of Programmed Organelle Degradation. *J Biol Chem* **2013**, *288*, 11436–11447, doi:10.1074/jbc.M112.437103.
38. Sandilands, A.; Hutcheson, A.M.; Long, H.A.; Prescott, A.R.; Vrensen, G.; Löster, J.; Klopp, N.; Lutz, R.B.; Graw, J.; Masaki, S.; et al. Altered Aggregation Properties of Mutant Gamma-Crystallins Cause Inherited Cataract. *EMBO J* **2002**, *21*, 6005–6014, doi:10.1093/emboj/cdf609.
39. Cui, X.; Wang, L.; Zhang, J.; Du, R.; Liao, S.; Li, D.; Li, C.; Ke, T.; Li, D.W.-C.; Huang, H.; et al. HSF4 Regulates DLAD Expression and Promotes Lens De-Nucleation. *Biochim Biophys Acta* **2013**, *1832*, 1167–1172, doi:10.1016/j.bbadis.2013.03.007.
40. Ruiss, M.; Kronschläger, M.; Schlatter, A.; Dechat, T.; Findl, O. Comparison of Methods to Experimentally Induce Opacification and Elasticity Change in Ex Vivo Porcine Lenses. *Sci Rep* **2021**, *11*, 23406, doi:10.1038/s41598-021-02851-6.
41. Siddam, A.D.; Gautier-Courteille, C.; Perez-Campos, L.; Anand, D.; Kakrana, A.; Dang, C.A.; Legagneux, V.; Méreau, A.; Viet, J.; Gross, J.M.; et al. The RNA-Binding Protein Celf1 Post-Transcriptionally Regulates P27Kip1 and Dnase2b to Control Fiber Cell Nuclear Degradation in Lens Development. *PLoS Genet.* **2018**, *14*, e1007278, doi:10.1371/journal.pgen.1007278.
42. Aryal, S.; Viet, J.; Weatherbee, B.A.T.; Siddam, A.D.; Hernandez, F.G.; Gautier-Courteille, C.; Paillard, L.; Lachke, S.A. The Cataract-Linked RNA-Binding Protein Celf1 Post-Transcriptionally Controls the Spatiotemporal Expression of the Key Homeodomain Transcription Factors Pax6 and Prox1 in Lens Development. *Hum Genet* **2020**, *139*, 1541–1554, doi:10.1007/s00439-020-02195-7.
43. Young, L.K.; Jarrin, M.; Saunter, C.D.; Quinlan, R.A.; Girkin, J.M. Non-Invasive in Vivo Quantification of the Developing Optical Properties and Graded Index of the Embryonic Eye Lens Using SPIM. *Biomed Opt Express* **2018**, *9*, 2176–2188, doi:10.1364/BOE.9.002176.
44. Oliver, G.; Loosli, F.; Köster, R.; Wittbrodt, J.; Gruss, P. Ectopic Lens Induction in Fish in Response to the Murine Homeobox Gene Six3. *Mech Dev* **1996**, *60*, 233–239, doi:10.1016/s0925-4773(96)00632-6.
45. Wang, Y.; Mahesh, P.; Wang, Y.; Novo, S.G.; Shihan, M.H.; Hayward-Piatkovskyi, B.; Duncan, M.K. Spatiotemporal Dynamics of Canonical Wnt Signaling during Embryonic Eye Development and Posterior Capsular Opacification (PCO). *Exp Eye Res* **2018**, *175*, 148–158, doi:10.1016/j.exer.2018.06.020.
46. Anand, D.; Kakrana, A.; Siddam, A.D.; Huang, H.; Saadi, I.; Lachke, S.A. RNA Sequencing-Based Transcriptomic Profiles of Embryonic Lens Development for Cataract Gene Discovery. *Hum Genet* **2018**, *137*, 941–954, doi:10.1007/s00439-018-1958-0.
47. Tam, O.H.; Pennisi, D.; Wilkinson, L.; Little, M.H.; Wazin, F.; Wan, V.L.; Lovicu, F.J. Crim1 Is Required for Maintenance of the Ocular Lens Epithelium. *Exp Eye Res* **2018**, *170*, 58–66, doi:10.1016/j.exer.2018.02.012.
48. Harned, J.; Fleisher, L.N.; McGahan, M.C. Lens Epithelial Cells Synthesize and Secrete Ceruloplasmin: Effects of Ceruloplasmin and Transferrin on Iron Efflux and Intracellular Iron Dynamics. *Exp Eye Res* **2006**, *83*, 721–727, doi:10.1016/j.exer.2006.01.018.
49. Pontoriero, G.F.; Smith, A.N.; Miller, L.-A.D.; Radice, G.L.; West-Mays, J.A.; Lang, R.A. Co-Operative Roles for E-Cadherin and N-Cadherin during Lens Vesicle Separation and Lens Epithelial Cell Survival. *Dev Biol* **2009**, *326*, 403–417, doi:10.1016/j.ydbio.2008.10.011.
50. Landgren, H.; Blixt, A.; Carlsson, P. Persistent FoxE3 Expression Blocks Cytoskeletal Remodeling and Organelle Degradation during Lens Fiber Differentiation. *Invest Ophthalmol Vis Sci* **2008**, *49*, 4269–4277, doi:10.1167/iovs.08-2243.

51. Bassnett, S.; Šikić, H. The Lens Growth Process. *Prog Retin Eye Res* **2017**, *60*, 181–200, doi:10.1016/j.preteyeres.2017.04.001.
52. Bassnett, S. On the Mechanism of Organelle Degradation in the Vertebrate Lens. *Exp Eye Res* **2009**, *88*, 133–139, doi:10.1016/j.exer.2008.08.017.
53. Fujimoto, M.; Izu, H.; Seki, K.; Fukuda, K.; Nishida, T.; Yamada, S.-I.; Kato, K.; Yonemura, S.; Inouye, S.; Nakai, A. HSF4 Is Required for Normal Cell Growth and Differentiation during Mouse Lens Development. *EMBO J* **2004**, *23*, 4297–4306, doi:10.1038/sj.emboj.7600435.
54. Huang, M.; Li, D.; Huang, Y.; Cui, X.; Liao, S.; Wang, J.; Liu, F.; Li, C.; Gao, M.; Chen, J.; et al. HSF4 Promotes G1/S Arrest in Human Lens Epithelial Cells by Stabilizing P53. *Biochim Biophys Acta* **2015**, *1853*, 1808–1817, doi:10.1016/j.bbamcr.2015.04.018.
55. Chan, N.C.; Salazar, A.M.; Pham, A.H.; Sweredoski, M.J.; Kolawa, N.J.; Graham, R.L.J.; Hess, S.; Chan, D.C. Broad Activation of the Ubiquitin-Proteasome System by Parkin Is Critical for Mitophagy. *Hum Mol Genet* **2011**, *20*, 1726–1737, doi:10.1093/hmg/ddr048.
56. McAvoy, J.W.; Richardson, N.A. Nuclear Pyknosis during Lens Fibre Differentiation in Epithelial Explants. *Curr Eye Res* **1986**, *5*, 711–715, doi:10.3109/02713688609015139.
57. Mochizuki, T.; Masai, I. The Lens Equator: A Platform for Molecular Machinery That Regulates the Switch from Cell Proliferation to Differentiation in the Vertebrate Lens. *Dev Growth Differ* **2014**, *56*, 387–401, doi:10.1111/dgd.12128.
58. Rowan, S.; Conley, K.W.; Le, T.T.; Donner, A.L.; Maas, R.L.; Brown, N.L. Notch Signaling Regulates Growth and Differentiation in the Mammalian Lens. *Dev Biol* **2008**, *321*, 111–122, doi:10.1016/j.ydbio.2008.06.002.
59. Saravanamuthu, S.S.; Gao, C.Y.; Zelenka, P.S. Notch Signaling Is Required for Lateral Induction of Jagged1 during FGF-Induced Lens Fiber Differentiation. *Dev Biol* **2009**, *332*, 166–176, doi:10.1016/j.ydbio.2009.05.566.
60. Shaham, O.; Smith, A.N.; Robinson, M.L.; Taketo, M.M.; Lang, R.A.; Ashery-Padan, R. Pax6 Is Essential for Lens Fiber Cell Differentiation. *Development* **2009**, *136*, 2567–2578, doi:10.1242/dev.032888.
61. Shaham, O.; Menuchin, Y.; Farhy, C.; Ashery-Padan, R. Pax6: A Multi-Level Regulator of Ocular Development. *Prog Retin Eye Res* **2012**, *31*, 351–376, doi:10.1016/j.preteyeres.2012.04.002.
62. Cvekl, A.; Ashery-Padan, R. The Cellular and Molecular Mechanisms of Vertebrate Lens Development. *Development* **2014**, *141*, 4432–4447, doi:10.1242/dev.107953.
63. Wigle, J.T.; Chowdhury, K.; Gruss, P.; Oliver, G. Prox1 Function Is Crucial for Mouse Lens-Fibre Elongation. *Nat Genet* **1999**, *21*, 318–322, doi:10.1038/6844.
64. Audette, D.S.; Anand, D.; So, T.; Rubenstein, T.B.; Lachke, S.A.; Lovicu, F.J.; Duncan, M.K. Prox1 and Fibroblast Growth Factor Receptors Form a Novel Regulatory Loop Controlling Lens Fiber Differentiation and Gene Expression. *Development* **2016**, *143*, 318–328, doi:10.1242/dev.127860.
65. Spector, A.; Wang, G.M.; Wang, R.R.; Garner, W.H.; Moll, H. The Prevention of Cataract Caused by Oxidative Stress in Cultured Rat Lenses. I. H2O2 and Photochemically Induced Cataract. *Curr Eye Res* **1993**, *12*, 163–179, doi:10.3109/02713689308999484.
66. Spector, A. Oxidative Stress-Induced Cataract: Mechanism of Action. *FASEB J* **1995**, *9*, 1173–1182.
67. Hernebring, M.; Adelöf, J.; Wiseman, J.; Petersen, A.; Zetterberg, M. H2O2-Induced Cataract as a Model of Age-Related Cataract: Lessons Learned from Overexpressing the Proteasome Activator PA28αβ in Mouse Eye Lens. *Exp Eye Res* **2021**, *203*, 108395, doi:10.1016/j.exer.2020.108395.
68. Gu, Y.; Xu, B.; Feng, C.; Ni, Y.; Hong, N.; Wang, J.; Jiang, B. Topical Use of NaCl Solution with Different Concentration Affects Lens Transparency in Anesthetized Mice. *Curr Eye Res* **2016**, *41*, 943–950, doi:10.3109/02713683.2015.1080280.
69. Butler, P.A. Reversible Cataracts in Diabetes Mellitus. *J Am Optom Assoc* **1994**, *65*, 559–563.
70. Obrosova, I.G.; Chung, S.S.M.; Kador, P.F. Diabetic Cataracts: Mechanisms and Management. *Diabetes Metab Res Rev* **2010**, *26*, 172–180, doi:10.1002/dmrr.1075.
71. Dasgupta, T.; Ladd, A.N. The Importance of CELF Control: Molecular and Biological Roles of the CUG-BP, Elav-like Family of RNA-Binding Proteins. *Wiley Interdiscip Rev RNA* **2012**, *3*, 104–121, doi:10.1002/wrna.107.
72. Siddam, A.D.; Duot, M.; Coomson, S.Y.; Anand, D.; Aryal, S.; Weatherbee, B.A.T.; Audic, Y.; Paillard, L.; Lachke, S.A. High-Throughput Transcriptomics of Celf1 Conditional Knockout Lens Identifies Downstream Networks Linked to Cataract Pathology. *Cells* **2023**, *12*, 1070, doi:10.3390/cells12071070.

General discussion

Throughout my PhD research, I aimed at uncovering the complex post-transcriptional regulations that underly the congenital cataract observed in mice lacking CELF1. When I joined my host laboratory, they had already established two *Celf1*-deficient mice models: one was constitutionally knocked out (KO) for *Celf1*, and the other was conditionally inactivated: the *Celf1* gene is floxed, and *Celf1* is disrupted in the eyes by the Cre recombinase controlled by the Pax6 promoter. Both models display congenital cataracts, characterized by an absence of fiber cell (FC) karyolysis, defects in the cytoskeleton, and defects in FC packing[59,140]. While several genes regulated by CELF1 had been identified in the lens, a comprehensive understanding of the complete gene set directly or indirectly controlled by CELF1 for proper lens development remained to be fully elucidated.

In my work, I analysed transcriptomic data of cKO newborn lenses. Subsequently, I integrated this transcriptomic data with new RNA-seq datasets from adult lenses, and with an iCLIP-seq dataset. This approach allowed me to identify a group of differentially spliced genes. The splicing of these genes is presumably directly controlled by CELF1, and the defective splicing of these genes in the absence of CELF1 probably contributes to the lens defects observed in *Celf1*-deficient mice. Lastly, I characterized a novel lens organoid model derived from lens epithelial cells (LEC) that mimics many aspects of lens formation. This model is a new tool for further investigations about the functions of CELF1 in the lens, but it also has a significant potentialfor other conducted by the lens research community.

Global transcriptomic landscape in *Celf1* deficient lenses

The RNA-seq analysis conducted on the cKO *Celf1* lens has identified a substantial set of genes with altered expression levels in the absence of CELF1. Among them, the expression of 327 genes is reduced, while the expression of 660 genes is increased, as compared to controls.

Considering the observed changes in RNA expression levels of these genes, it is plausible that CELF1 directly regulates their expression. This regulatory process might involve CELF1's cytoplasmic activity, influencing mRNA stability. Moreover, CELF1's nuclear activity could also contribute to the modulation of RNA levels for these genes. By regulating their alternative splicing events, CELF1 could modify the coding sequences (CDS) and/or untranslated regions (UTRs) of these mRNA molecules. These altered

sequences could potentially become regulated by other RNA-binding proteins (RBPs) or microRNAs that lead to mRNAs levels changes.

However, among the genes that are misregulated, a considerable subset consists of transcription factors and RBPs, notably the key lens regulators PAX6[60,61,64,65] and PROX1[100–102]. Hence, it is plausible that a significant portion of the misexpressed genes results from the aberrant expression of *Pax6* and *Prox1* and are not directly under the control of CELF1.

How is it possible to sort direct from indirect targets of CELF1? It is interesting that roughly two thirds of the misexpressed genes are up-regulated in the absence of CELF1. Given that the predominant role of CELF1 is to promote the decay of its targets[279,291–298], we can suppose that a subset of these genes are directly controlled by CELF1. Conversely, the misexpression of the 327 genes that are under-expressed is more likely to be an indirect consequence of the absence of CELF1. This assumption is further supported by the integration of CLIP-seq data obtained from HeLa cells[257]. Although HeLa cells exhibit a distinct transcriptome from lenses, the majority of the RNA molecules bound by CELF1 in HeLa cells are also probably bound by CELF1 in lenses. Among the 987 genes that are differentially expressed in *Celf1* cKO lenses, 318 are CELF1 ligands of which 83% (264/318) are up-regulated in the absence of CELF1. As this proportion is higher than the proportion of up-regulated genes, we conclude that direct CELF1 ligands are enriched in up-regulated genes, consistent with a direct regulation by CELF1.

To gain a better insights into the role of these misexpressed genes, I incorporated the lens enrichment score from the iSyTE database[30]. This score reflects the embryonic expression of these genes in the lens compared to their expression in the entire embryonic mouse body. This metric aids in prioritizing genes crucial for lens development. Consequently, an intriguing pattern emerged: a majority of the genes that are downregulated in the absence of CELF1 exhibited high enrichment scores, implying that they are important in lens development. Their reduced expression potentially contributes to the onset of congenital cataracts. Among these genes, numerous encode crystallins, DNASE2B (critical for fiber cell denucleation[18]), and lens-specific cytoskeletal proteins like BFSP1[390]. GO term enrichment analysis of these genes underscores their involvement in eye and lens development, notably marked by the significant enrichment of the GO term for lens fiber cell differentiation.

Conversely, genes showing elevated expression in the *Celf1* cKO lens are not enriched in the lens. This suggests that in control lenses, CELF1 reduces their expression as they may not be necessary for lens development or could even be detrimental to it. These genes are enriched in GO terms associated with vesicle membranes and ATPase complexes. The association of these genes with the vesicle membranes is interesting, since during the FC differentiation

their membrane present important transformations. Notably they develop interdigitation structures. These structures present a resemblance to aspects of vesicle formation, particularly the clathrin-coated vesicle[39]. Consequently, the elevated expression of genes linked to vesicle could potentially disrupt the formation of interdigitation structures in FC. This phenomenon has indeed been observed in the context of *Celf1* cKO lenses, where FC lack the expected interdigitation structure[140]. Regarding the enrichment of genes related to the ATPase complex, these genes encode proteins found in the mitochondrial membrane. Mitochondria are expected to be degraded during FC differentiation[24,25]. However, the increased expression of these genes in the *Celf1* cKO lens suggests that the degradation of mitochondria in FC might not have occurred completely.

Additionally, I compared the mis-expressed genes identified in this RNA-seq analysis and those identified through two different microarray data on newborn and P6 (6 days post-natal) lens mice *Celf1* cKO. This comparative analysis, considering the distinct strengths and limitations of each transcriptomic approach, allowed the identification of highly reliable mis-regulated genes in *Celf1* deficient lenses. These genes emerge as strong candidates for investigating the impact of their mis-regulation on lens development, not only in the context of *Celf1* cKO but also in other lens-related defects. Notably, among these candidate genes, *Ell2* and *Prdm16* could be interesting candidates for further studies. ELL2, as an elongation factor highly enriched in the lens, might play a crucial role in enhancing the transcription of highly expressed lens-specific genes, including crystallins. On the other hand, the transcription factor PRDM16 has not been extensively studied in the context of the lens. As it is down-regulated in the absence of CELF1, it could have a significant impact on the global transcriptomic mis-regulation observed in *Celf1*-deficient lenses.

Alternatively spliced RNA candidate to be directly controlled by CELF1

CELF1 controls gene expression through different post-transcriptional mechanisms. When localized in the nucleus, it controls the alternative splicing events of its target pre-mRNAs[247,259,265,275,306,308,314]. This regulation changes the mRNA sequence, affecting either their untranslated regions or their coding sequences, subsequently influencing the protein isoforms produced. Notably, in the lens, CELF1 is found in the nucleus of the differentiating FC prior tontheir complete denucleation[140]. This nuclear localization strongly implies a critical role for CELF1 in ensuring proper FC differentiation, through the regulation of the alternative splicing events. Based on these insights, we decided to look further into the mis-spliced genes in lenses devoid of CELF1.

For this study, I reanalysed the previously described RNA-seq data obtained from lenses of newborn *Celf1* cKO mice[371]. Additionally, I also analysed an RNA-seq dataset from adult mouse lensesconstitutionally disrupted for *Celf1*, along with iCLIP-seq data.

iCLIP-seq, a technique derived from ChIP-seq, enables the identification and precise localization of RNA targets (mRNA and pre-mRNA) bound by a specific RBP[372]. I analyzed the result of an iCLIP-seq experiment targeting CELF1 in the lens of adult mice. I identified and localized 4,653 binding site between CELF1 and RNAs. Only 5% of these binding sites are present within intronic regions. It is less than in previous iCLIPseq of CELF1 made in other tissues or cells, e.g. in HeLa cells where 65 % of the binding sites are intronic[257]. A likely reason for this is that CELF1 interacts with intronic regions only in the nuclei, where pre-mRNAs are present, while the bulk of lens cells consist of fiber cells that contain no nuclei. Nevertheless, finding 243 intronic binding sites provides further evidence of CELF1's nuclear activity in the lens, and notably in epithelial cells and cells in the transition zone.

The two RNA-seq datasets I used to identify alternatively used junctions in CELF1-deficient lenses are from adult and newborn mice, respectively. The RNA-seq data for newborn lenses was previously examined in the context of mis-expressed genes, specifically using cKO mice. In contrast, the RNA-seq data for adult lenses was obtained from constitutively inactivated mice, ensuring age consistency with iCLIP-seq data. We identified mis-spliced junctions in 438 and 1,061 genes in newborn and adult mice, respectively.

The integration of these multi-omic datasets has enabled the identification of genes whose alternative splicing (AS) is likely to be directly regulated by CELF1. After meticulous manual curation of the identified genes, a subset of 22 high-confidence candidates emerged. Notably, 10 of these genes are associated with the cytoskeleton. Given the observed disruption of the actin network organization and the morphological and packing defects of fiber cells in *Celf1*-deficient mouse lenses, these genes have been prioritized for further investigation. Out of these 10 prioritized genes, differential splicing has been substantiated in 7 of them through RT-PCR analysis (*Ablim1, Ctnna2, Clta, Septin8, Sptbn1, Ywhae*, and *Ank2*).

All these AS events modify the coding sequence of the genes, resulting in the production of protein isoforms with distinct domains. To understand the implications of these changes, we employed AlphaFold2 (AF2), a state-of-the-art protein structure prediction tool[380].

AF2 predicted intriguing result for SPTBN1. It appears that CELF1 represses the formation of isoforms that lack the Pleckstrin Homology (PH) domain. Although the exact function of the PH domain remains unknown in the context of the lens, its role in other organs is related to protein localization to

the membrane. Consequently, the mis-localization of the isoform lacking PH could potentially contribute to the broader cytoskeletal abnormalities observed in the lens of *Celf1* deficient mice.

AF2's predictions suggest a conformational change in the primary alpha helix of CLTA. This alteration is involved in the modulation of clathrin-coated vesicle size, in the neurons[387]. In the lens context, this conformational change could potentially disrupt the formation of the FC specific interdigitation structure. This structural disturbance might have implications for FC packing, which can participate to the formation of a cataract[39].

For the other validated candidates genes *(Ablim1, Ctnna2, Septin8, Ywhae*, and *Ank2*) AF2 has not been able to predict the conformation of the lost or gained domain. This suggests that these domains are intrinsically disordered regions (IDR). IDRs are segments of proteins that require interactions with other proteins to adopt their proper conformation. These IDRs play a crucial role in mediating protein-protein interactions[383,389]. Based on this, we propose that the IDRs in ABLIM1, CTNNA2, SEPTIN8, YWHAE, and ANK2 whose presence within the mature proteins is controlled by CELF1 results in different protein-protein interaction networks.

In the future, the identification of isoform-specific interactomes for each of these genes will help understand the molecular mechanisms behind cataract formation in *Celf1*-deficient lenses. This could also potentially unveil novel genes critical for lens development.

Characterization of a new lens organoid model

The study of lens pathologies often requires the use of animal models. Since one of the two lens cell types, the FC, lacks the ability to proliferate, most of the *in cellulo* investigations are centered around LEC. While two-dimensional (2D) cultured LEC can partially mimic certain aspects of FC differentiation, the *in cellulo* study of FC differentiation remains limited[355]. This limitation restricts researchers' understanding of lens development and cataract formation, as the differentiation of FCs is critical in maintaining lens transparency due to their predominance within the lens structure. Several three-dimensional (3D) models have emerged in recent years, exhibiting distinct characteristics reminiscent of those found in the lens (as discussed in the introduction chapter D.2). However, these models require the culture of embryonic stem cells or induced pluripotent stem cells and specific growth factors for a lengthy duration, resulting in higher costs and challenges in scalability[361–365]. In contrast, we have successfully established a novel lens organoid model using the murine LEC cell line 21EM15. This innovative approach offers comparable lens-like properties while being more time-efficient and straightforward to generate.

Our proposed protocol can generate organoids with interesting optical properties within just 10 days, without any specific growth factor. These organoids are transparent and, similarly to the lens, are capable of light focusing. We conducted RNA-seq to discern the transcriptomic modifications during the transition from 2D to 3D. Interestingly, genes overexpressed in 3D organoids also displayed elevated expression during mice lens development. Among these genes were those encoding lens-specific transcription factors like *Eya1, Meis1, Prox1, Hsf4* and *Maf*. Additionally, known lens RBPs such as *Tdrd7* and *Caprin2*, as well as members of signaling pathways crucial in lens development (*Jag1, Notch3, Tgfb2, Tgfb3, Fgfrl1, Ephb6, Epha7, Efna1*) are also up-regulated in 3D cultures, compard to 2D cultures. These changes support the notion that our model undergoes, at least partially, the same developmental alterations as those observed in the lens.

Similar to the lens's structural organization, our organoid model exhibits three distinct histological regions: (i) an external region with round nuclei, (ii) an intermediate region with elongated nuclei, and (iii) an internal region with compact nuclei. Likewise to the lens, only cells within the outer region can proliferate, while those in the other regions cease proliferation to initiate differentiation, leading to altered morphology and transcriptomes. It is worth noting that, in contrast to the lens, the cells in the internal region of these organoids did not undergo complete nuclear degradation. However, despite this difference, the model still maintains its transparency.

To characterize the transcriptomic differences between the different regions, we employed Laser Capture Micro-dissection to separate the internal and external regions before sequencing their respective transcriptomes. As expected, the external region exhibited an enrichment of cell cycle genes, as these cells are the only ones capable of proliferation. However, some genes associated with the capsule were also identified (e.g., *Lmnb1*). On the other hand, the internal region showed enrichment in fiber cell-specific genes, including transcription factors like *Prox1* and *Maf*, cytoskeletal proteins like *Ank2*, and crystallin proteins like *Cryab*. Immunofluorescence staining validated this regionalization: the capsule protein laminin was predominantly located in the external region, whereas the internal region exhibited high levels of HSF4 and CRYAB. An interesting observation is that while PROX1 was found in both the intermediate and internal regions, its subcellular localization differed. PROX1 was predominantly nuclear in the internal region, implying that its activity as a transcrition factor might be limited to this area. Furthermore, a staining of TOMM20, a mitochondrial outer membrane protein, revealed a reduced presence of mitochondria in the cells of the internal region. This suggests that these cells initiate mitochondrial degradation as part of their differentiation process.

Importantly, this novel organoid model holds promise for studying cataract formation. When exposed to drugs known to induce cataracts, they lose their characteristic transparency and light-focusing ability, similar to the lens. Moreover, by inactivating *Celf1* by shRNA expression, the organoids become opaque and lose their ability to focus light, like we observed in *Celf1* deficient mouse lenses. This makes the model a valuable tool for investigating the molecular changes that occur in cataract development due to environmental stresses or mutations in cataract-associated genes. Notably, the cost-effectiveness of this model will allow for the development of compound screening assays. This approach could be employed to test the impact of various compounds for the prevention of cataract.

Conclusion

In summary, my PhD research highlights the transcriptomic alterations taking place in the lens due to the absence of CELF1. I successfully identified genes that are likely to be under direct regulation by CELF1, through modulation of either mRNA stability or pre-mRNA alternative splicing. To validate the significance of these regulatory changes on the formation of the cataract, I contributed to develop a novel organoid model that proves useful for investigating various types of cataracts, including those occurring from *Celf1* deficiency. This novel model not only offers insights into mechanisms underlying *Celf1*-deficient-cataract, but will also serve as a versatile resource for the lens research community to study other types of cataract.

References

1. Cvekl, A. & Ashery-Padan, R. The cellular and molecular mechanisms of vertebrate lens development. *Development* **141**, 4432–4447 (2014).
2. Zhang, G., Wei, Q., Lu, L., Lin, A. L. & Qu, C. The evolution of mechanism of accommodation and a novel hypothesis. *Graefes Arch Clin Exp Ophthalmol* (2023) doi:10.1007/s00417-023-06045-w.
3. Ott, M. Visual accommodation in vertebrates: mechanisms, physiological response and stimuli. *J Comp Physiol A* **192**, 97–111 (2006).
4. Artal, P., Tejada, P. H. de, Tedó, C. M. & Green, D. G. Retinal image quality in the rodent eye. *Visual Neuroscience* **15**, 597–605 (1998).
5. Woolf, D. A comparative cytological study of the ciliary muscle. *Anat Rec* **124**, 145–163 (1956).
6. Shi, Y., Maria, A. D., Lubura, S., Šikić, H. & Bassnett, S. The Penny Pusher: A Cellular Model of Lens Growth. *Investigative Ophthalmology & Visual Science* **56**, 799 (2015).
7. Papaconstantinou, J. Molecular Aspects of Lens Cell Differentiation. *Science* **156**, 338–346 (1967).
8. COULOMBRE, J. L. & COULOMBRE, A. J. Lens Development. IV. Size, Shape, and Orientation. *Investigative Ophthalmology & Visual Science* **8**, 251–257 (1969).
9. Coulombre, J. L. & Coulombre, A. J. LENS DEVELOPMENT: FIBER ELONGATION AND LENS ORIENTATION. *Science (New York, N.Y.)* **142**, 1489–1490 (1963).
10. Chamberlain, C. G. & McAvoy, J. W. Fibre differentiation and polarity in the mammalian lens: a key role for FGF. *Progress in Retinal and Eye Research* **16**, 443–478 (1997).

11. Zhao, Y., Zheng, D. & Cvekl, A. A comprehensive spatial-temporal transcriptomic analysis of differentiating nascent mouse lens epithelial and fiber cells. *Exp Eye Res* **175**, 56–72 (2018).
12. Hoang, T. V. *et al.* Comparative transcriptome analysis of epithelial and fiber cells in newborn mouse lenses with RNA sequencing. *Mol Vis* **20**, 1491–1517 (2014).
13. Danysh, B. P. & Duncan, M. K. The Lens Capsule. *Experimental eye research* **88**, 151 (2009).
14. Lovicu, F. J. & McAvoy, J. W. Localization of acidic fibroblast growth factor, basic fibroblast growth factor, and heparan sulphate proteoglycan in rat lens: implications for lens polarity and growth patterns. *Invest Ophthalmol Vis Sci* **34**, 3355–3365 (1993).
15. Mascarelli, F., Fuhrmann, G. & Courtois, Y. aFGF binding to low and high affinity receptors induces both aFGF and aFGF receptors dimerization. *Growth Factors* **8**, 211–233 (1993).
16. Pendergrass, W., Penn, P., Possin, D. & Wolf, N. Accumulation of DNA, Nuclear and Mitochondrial Debris, and ROS at Sites of Age-Related Cortical Cataract in Mice. *Investigative Ophthalmology & Visual Science* **46**, 4661–4670 (2005).
17. Nishimoto, S. *et al.* Nuclear cataract caused by a lack of DNA degradation in the mouse eye lens. *Nature* **424**, 1071–1074 (2003).
18. Nakahara, M. *et al.* Degradation of nuclear DNA by DNase II-like acid DNase in cortical fiber cells of mouse eye lens. *The FEBS Journal* **274**, 3055–3064 (2007).
19. Youle, R. J. & Narendra, D. P. Mechanisms of mitophagy. *Nat Rev Mol Cell Biol* **12**, 9–14 (2011).
20. Brennan, L. A. *et al.* Spatial expression patterns of autophagy genes in the eye lens and induction of autophagy in lens cells. *Mol Vis* **18**, 1773–1786 (2012).

21. Chen, J. *et al.* Mutations in FYCO1 cause autosomal-recessive congenital cataracts. *Am J Hum Genet* **88**, 827–838 (2011).
22. Khan, S. Y. *et al.* The role of FYCO1-dependent autophagy in lens fiber cell differentiation. *Autophagy* **18**, 2198–2215.
23. Schweers, R. L. *et al.* NIX is required for programmed mitochondrial clearance during reticulocyte maturation. *Proc Natl Acad Sci U S A* **104**, 19500–19505 (2007).
24. Brennan, L. A. *et al.* BNIP3L/NIX is required for elimination of mitochondria, endoplasmic reticulum and Golgi apparatus during eye lens organelle-free zone formation. *Exp Eye Res* **174**, 173–184 (2018).
25. Costello, M. J. *et al.* Autophagy and mitophagy participate in ocular lens organelle degradation. *Experimental eye research* **116**, (2013).
26. Morishita, H. *et al.* Organelle degradation in the lens by PLAAT phospholipases. *Nature* **592**, 634–638 (2021).
27. Audette, D. S., Scheiblin, D. A. & Duncan, M. K. The molecular mechanisms underlying lens fiber elongation. *Exp Eye Res* **156**, 41–49 (2017).
28. Lo, W.-K., Wen, X.-J. & Zhou, C.-J. Microtubule configuration and membranous vesicle transport in elongating fiber cells of the rat lens. *Exp Eye Res* **77**, 615–626 (2003).
29. Beebe, D. C., Feagans, D. E., Blanchette-Mackie, E. J. & Nau, M. E. Lens epithelial cell elongation in the absence of microtubules: evidence for a new effect of colchicine. *Science* **206**, 836–838 (1979).
30. Kakrana, A. *et al.* iSyTE 2.0: a database for expression-based gene discovery in the eye. *Nucleic Acids Res* **46**, D875–D885 (2018).
31. Bassnett, S., Wilmarth, P. A. & David, L. L. The membrane proteome of the mouse lens fiber cell. *Mol Vis* **15**, 2448–2463 (2009).
32. Mousa, G. Y. & Trevithick, J. R. Differentiation of rat lens epithelial cells in tissue culture: II. Effects of cytochalasins B and D on actin organization and differentiation. *Developmental Biology* **60**, 14–25 (1977).

33. Zhao, Y. *et al.* Proteome-transcriptome analysis and proteome remodeling in mouse lens epithelium and fibers. *Exp Eye Res* **179**, 32–46 (2019).
34. Blankenship, T. N., Hess, J. F. & FitzGerald, P. G. Development- and Differentiation-Dependent Reorganization of Intermediate Filaments in Fiber Cells. *Investigative Ophthalmology & Visual Science* **42**, 735–742 (2001).
35. Logan, C. M. *et al.* N-cadherin regulates signaling mechanisms required for lens fiber cell elongation and lens morphogenesis. *Dev Biol* **428**, 118–134 (2017).
36. Bassnett, S. & Šikić, H. The Lens Growth Process. *Progress in retinal and eye research* **60**, 181 (2017).
37. Kumari, S. S., Gandhi, J., Mustehsan, M. H., Eren, S. & Varadaraj, K. Functional characterization of an AQP0 missense mutation, R33C, that causes dominant congenital lens cataract, reveals impaired cell-to-cell adhesion. *Exp Eye Res* **116**, 371–385 (2013).
38. Wang, E., Geng, A., Maniar, A. M., Mui, B. W. H. & Gong, X. Connexin 50 Regulates Surface Ball-and-Socket Structures and Fiber Cell Organization. *Invest Ophthalmol Vis Sci* **57**, 3039–3046 (2016).
39. Zhou, C.-J. & Lo, W.-K. Association of clathrin, AP-2 adaptor and actin cytoskeleton with developing interlocking membrane domains of lens fibre cells. *Exp Eye Res* **77**, 423–432 (2003).
40. Biswas, S. K., Lee, J. E., Brako, L., Jiang, J. X. & Lo, W.-K. Gap junctions are selectively associated with interlocking ball-and-sockets but not protrusions in the lens. *Mol Vis* **16**, 2328–2341 (2010).
41. Shiels, A., Bennett, T. M. & Hejtmancik, J. F. Cat-Map: putting cataract on the map. *Molecular Vision* **16**, 2007 (2010).
42. Jones, J. L. *et al.* Pathogenic genetic variants identified in Australian families with paediatric cataract. *BMJ Open Ophthalmol* **7**, e001064 (2022).

43. Xia, C. hong *et al.* Absence of alpha3 (Cx46) and alpha8 (Cx50) connexins leads to cataracts by affecting lens inner fiber cells. *Experimental eye research* **83**, 688–696 (2006).
44. Shi, Y., Li, X. & Yang, J. Mutations of CX46/CX50 and Cataract Development. *Front Mol Biosci* **9**, 842399 (2022).
45. Cheng, Z. *et al.* Anterior Umbilication of Lens in a Family with Congenital Cataracts Associated with a Missense Mutation of MIP Gene. *Genes (Basel)* **13**, 1987 (2022).
46. Bloemendal, H. The vertebrate eye lens. *Science (New York, N.Y.)* **197**, 127–138 (1977).
47. Haslbeck, M., Peschek, J., Buchner, J. & Weinkauf, S. Structure and function of α-crystallins: Traversing from in vitro to in vivo. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1860**, 149–166 (2016).
48. Shiels, A. & Hejtmancik, J. F. Mutations and mechanisms in congenital and age-related cataracts. *Exp Eye Res* **156**, 95–102 (2017).
49. Rajini, B. *et al.* Calcium Binding Properties of γ-Crystallin: CALCIUM ION BINDS AT THE GREEK KEY βγ-CRYSTALLIN FOLD *. *Journal of Biological Chemistry* **276**, 38464–38471 (2001).
50. Duncan, G. & Wormstone, I. M. Calcium cell signalling and cataract: Role of the endoplasmic reticulum. *Eye* **13**, 480–483 (1999).
51. Gao, J. *et al.* Connections between connexins, calcium, and cataracts in the lens. *J Gen Physiol* **124**, 289–300 (2004).
52. Litt, M. *et al.* Autosomal dominant cerulean cataract is associated with a chain termination mutation in the human beta-crystallin gene CRYBB2. *Hum Mol Genet* **6**, 665–668 (1997).
53. Pei, Y. F. & Rhodin, J. A. G. The prenatal development of the mouse eye. *The Anatomical record* **168**, 105–125 (1970).
54. Augusteyn, R. C. Growth of the eye lens: I. Weight accumulation in multiple species. *Molecular Vision* **20**, 410 (2014).

55. Augusteyn, R. C. Growth of the human eye lens. *Molecular Vision* **13**, 252 (2007).
56. Halder, G., Callaerts, P. & Gehring, W. J. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science (New York, N.Y.)* **267**, 1788–1792 (1995).
57. Lang, R. A. Pathways regulating lens induction in the mouse. *The International Journal of Developmental Biology* **48**, 783–791 (2004).
58. Verma, A. S. & FitzPatrick, D. R. Anophthalmia and microphthalmia. *Orphanet J Rare Dis* **2**, 47 (2007).
59. Aryal, S. *et al.* The cataract-linked RNA-binding protein Celf1 post-transcriptionally controls the spatiotemporal expression of the key homeodomain transcription factors Pax6 and Prox1 in lens development. *Human Genetics* **139**, 1541–1554 (2020).
60. Duncan, M. K., Haynes, J. I., Cvekl, A. & Piatigorsky, J. Dual roles for Pax-6: a transcriptional repressor of lens fiber cell-specific beta-crystallin genes. *Mol Cell Biol* **18**, 5579–5586 (1998).
61. Shaham, O. *et al.* Pax6 is essential for lens fiber cell differentiation. *Development* **136**, 2567–2578 (2009).
62. Goudreau, G. *et al.* Mutually regulated expression of Pax6 and Six3 and its implications for the Pax6 haploinsufficient lens phenotype. *Proc Natl Acad Sci U S A* **99**, 8719–8724 (2002).
63. CVEKL, A., YANG, Y., CHAUHAN, B. K. & CVEKLOVA, K. Regulation of gene expression by Pax6 in ocular cells. *Int J Dev Biol* **48**, 829–844 (2004).
64. Cvekl, A. & Callaerts, P. PAX6: 25th anniversary and more to learn. *Experimental Eye Research* **156**, 10–21 (2017).
65. Cvekl, A. & Piatigorsky, J. Lens development and crystallin gene expression: many roles for Pax-6. *Bioessays* **18**, 621–630 (1996).
66. Anchan, R. M. *et al.* Pax6- and Six3-mediated induction of lens cell fate in mouse and human ES cells. *PLoS One* **9**, e115106 (2014).

67. Lagutin, O. *et al.* Six3 promotes the formation of ectopic optic vesicle-like structures in mouse embryos. *Dev Dyn* **221**, 342–349 (2001).
68. Lagutin, O. V. *et al.* Six3 repression of Wnt signaling in the anterior neuroectoderm is essential for vertebrate forebrain development. *Genes Dev* **17**, 368–379 (2003).
69. Liu, W., Lagutin, O. V., Mende, M., Streit, A. & Oliver, G. Six3 activation of Pax6 expression is essential for mammalian lens induction and specification. *The EMBO Journal* **25**, 5383 (2006).
70. Oliver, G., Loosli, F., Köster, R., Wittbrodt, J. & Gruss, P. Ectopic lens induction in fish in response to the murine homeobox gene Six3. *Mech Dev* **60**, 233–239 (1996).
71. Dvorakova, M. *et al.* Early ear neuronal development, but not olfactory or lens development, can proceed without SOX2. *Dev Biol* **457**, 43–56 (2020).
72. Fantes, J. *et al.* Mutations in SOX2 cause anophthalmia. *Nat Genet* **33**, 461–463 (2003).
73. Wang, P., Liang, X., Yi, J. & Zhang, Q. Novel SOX2 Mutation Associated With Ocular Coloboma in a Chinese Family. *Archives of Ophthalmology* **126**, 709–713 (2008).
74. Yonova-Doing, E. *et al.* Common variants in SOX-2 and congenital cataract genes contribute to age-related nuclear cataract. *Commun Biol* **3**, 755 (2020).
75. Shimada, N., Aya-Murata, T., Reza, H. M. & Yasuda, K. Cooperative action between L-Maf and Sox2 on δ-crystallin gene expression during chick lens development. *Mechanisms of Development* **120**, 455–465 (2003).
76. Kondoh, H., Uchikawa, M. & Kamachi, Y. Interplay of Pax6 and SOX2 in lens development as a paradigm of genetic switch mechanisms for cell differentiation. *The International Journal of Developmental Biology* **48**, 819–827 (2004).

77. Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R. & Kondoh, H. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes Dev* **15**, 1272–1286 (2001).
78. Holmes, Z. E. *et al.* The Sox2 transcription factor binds RNA. *Nat Commun* **11**, 1805 (2020).
79. Fujino, M. *et al.* c-MAF deletion in adult C57BL/6J mice induces cataract formation and abnormal differentiation of lens fiber cells. *Exp Anim* **69**, 242–249 (2020).
80. Reza, H. M., Ogino, H. & Yasuda, K. L-Maf, a downstream target of Pax6, is essential for chick lens development. *Mech Dev* **116**, 61–73 (2002).
81. Reza, H. M., Urano, A., Shimada, N. & Yasuda, K. Sequential and combinatorial roles of maf family genes define proper lens development. *Mol Vis* **13**, 18–30 (2007).
82. Kawauchi, S. *et al.* Regulation of Lens Fiber Cell Differentiation by Transcription Factor c-Maf*. *Journal of Biological Chemistry* **274**, 19254–19260 (1999).
83. Yoshida, T. & Yasuda, K. Characterization of the chicken L-Maf, MafB and c-Maf in crystallin gene regulation and lens differentiation. *Genes Cells* **7**, 693–706 (2002).
84. Reza, H. M. & Yasuda, K. Roles of maf family proteins in lens development. *Developmental Dynamics* **229**, 440–448 (2004).
85. Valleix, S. *et al.* Homozygous nonsense mutation in the FOXE3 gene as a cause of congenital primary aphakia in humans. *Am J Hum Genet* **79**, 358–364 (2006).
86. Ahmad, N. *et al.* Pitx3 directly regulates <I>Foxe3</I> during early lens development. *The International Journal of Developmental Biology* **57**, 741–751 (2013).

87. Wang, Y., Li, W., Wang, Y. & Huang, Y. Growth inhibition of human lens epithelial cells by short hairpin RNA in transcription factor forkhead box E3 (FOXE3). *Graefes Arch Clin Exp Ophthalmol* **250**, 999–1007 (2012).
88. Blixt, Å. *et al.* A forkhead gene, FoxE3, is essential for lens epithelial proliferation and closure of the lens vesicle. *Genes Dev* **14**, 245–254 (2000).
89. Blixt, Å., Landgren, H., Johansson, B. R. & Carlsson, P. Foxe3 is required for morphogenesis and differentiation of the anterior segment of the eye and is sensitive to Pax6 gene dosage. *Developmental Biology* **302**, 218–229 (2007).
90. Medina-Martinez, O. *et al.* Severe defects in proliferation and differentiation of lens cells in Foxe3 null mice. *Mol Cell Biol* **25**, 8854–8863 (2005).
91. Landgren, H., Blixt, Å. & Carlsson, P. Persistent FoxE3 Expression Blocks Cytoskeletal Remodeling and Organelle Degradation during Lens Fiber Differentiation. *Investigative Ophthalmology & Visual Science* **49**, 4269–4277 (2008).
92. Krall, M. *et al.* A zebrafish model of foxe3 deficiency demonstrates lens and eye defects with dysregulation of key genes involved in cataract formation in humans. *Hum Genet* **137**, 315–328 (2018).
93. Wada, K. *et al.* Expression of Truncated PITX3 in the Developing Lens Leads to Microphthalmia and Aphakia in Mice. *PLoS One* **9**, e111432 (2014).
94. Sorokina, E. A., Muheisen, S., Mlodik, N. & Semina, E. V. MIP/Aquaporin 0 Represents a Direct Transcriptional Target of PITX3 in the Developing Lens. *PLoS One* **6**, e21122 (2011).
95. Verdin, H. *et al.* Novel and recurrent PITX3 mutations in Belgian families with autosomal dominant congenital cataract and anterior segment dysgenesis have similar phenotypic and functional characteristics. *Orphanet J Rare Dis* **9**, 26 (2014).

96. Wu, Z. *et al.* PITX3 mutations associated with autosomal dominant congenital cataract in the Chinese population. *Mol Med Rep* **19**, 3123–3131 (2019).
97. Medina-Martinez, O., Shah, R. & Jamrich, M. Pitx3 controls multiple aspects of lens development. *Dev Dyn* **238**, 2193–2201 (2009).
98. Ho, H.-Y., Chang, K.-H., Nichols, J. & Li, M. Homeodomain protein Pitx3 maintains the mitotic activity of lens epithelial cells. *Mechanisms of Development* **126**, 18–29 (2009).
99. Duncan, M. K., Cui, W., Oh, D.-J. & Tomarev, S. I. Prox1 is differentially localized during lens development. *Mechanisms of Development* **112**, 195–198 (2002).
100. Wigle, J. T., Chowdhury, K., Gruss, P. & Oliver, G. Prox1 function is crucial for mouse lens-fibre elongation. *Nat Genet* **21**, 318–322 (1999).
101. Audette, D. S. *et al.* Prox1 and fibroblast growth factor receptors form a novel regulatory loop controlling lens fiber differentiation and gene expression. *Development* **143**, 318–328 (2016).
102. Yu, Z.-Y. *et al.* RNA-seq reveals transcriptome changes of the embryonic lens cells in Prox1 tissue specific knockout mice. *Eur Rev Med Pharmacol Sci* **23**, 7740–7748 (2019).
103. Dawes, L. J., Shelley, E. J., McAvoy, J. W. & Lovicu, F. J. A role for Hippo/YAP-signaling in FGF-induced lens epithelial cell proliferation and fibre differentiation. *Experimental Eye Research* **169**, 122–133 (2018).
104. Lu, Q. *et al.* Heterozygous Loss of Yap1 in Mice Causes Progressive Cataracts. *Investigative Ophthalmology & Visual Science* **61**, (2020).
105. He, Q. *et al.* Deficiency of Yes-Associated Protein Induces Cataract in Mice. *Aging Dis* **10**, 293–306 (2019).
106. Song, J. Y. *et al.* Dual function of Yap in the regulation of lens progenitor cells and cellular polarity. *Dev Biol* **386**, 281–290 (2014).

107. Zhao, H. *et al.* Fibroblast growth factor receptor signaling is essential for lens fiber cell differentiation. *Developmental biology* **318**, 276–288 (2008).

108. Roskoski, R. ERK1/2 MAP kinases: Structure, function, and regulation. *Pharmacological Research* **66**, 105–143 (2012).

109. Lovicu, F. J. & McAvoy, J. W. FGF-induced lens cell proliferation and differentiation is dependent on MAPK (ERK1/2) signalling. *Development* **128**, 5075–5084 (2001).

110. Garcia, C. M. *et al.* The Function of FGF Signaling in the Lens Placode. *Dev Biol* **351**, 176–185 (2011).

111. Wang, D. *et al.* Roles of TGFβ and FGF signals during growth and differentiation of mouse lens epithelial cell in vitro. *Sci Rep* **7**, 7274 (2017).

112. Dawes, L. *et al.* Interactions between lens epithelial and fiber cells reveal an intrinsic self-assembly mechanism. *Dev Biol* **385**, 291–303 (2014).

113. Saravanamuthu, S. S., Gao, C. Y. & Zelenka, P. S. Notch signaling is required for lateral induction of Jagged1 during FGF-induced lens fiber differentiation. *Dev Biol* **332**, 166–176 (2009).

114. Saravanamuthu, S. S. *et al.* Conditional ablation of the Notch2 receptor in the ocular lens. *Developmental Biology* **362**, 219 (2012).

115. Jia, J., Lin, M., Zhang, L., York, J. P. & Zhang, P. The Notch Signaling Pathway Controls the Size of the Ocular Lens by Directly Suppressing p57Kip2 Expression. *Mol Cell Biol* **27**, 7236–7247 (2007).

116. Le, T. T., Conley, K. W. & Brown, N. L. Jagged 1 is necessary for normal mouse lens formation. *Developmental biology* **328**, 118–126 (2009).

117. Rowan, S. *et al.* Notch signaling regulates growth and differentiation in the mammalian lens. *Dev Biol* **321**, 111–122 (2008).

118. Wang, D., Yao, Y., Zhang, M. & Chen, J. Genetic and phenotypic investigation of a Chinese pedigree with lattice corneal dystrophy IIIB subtype. *Eye Sci* **28**, 144–147 (2013).

119. Frances, R., Rodriguez Benitez, A. M. & Cohen, D. R. Arrhythmogenic right ventricular dysplasia and anterior polar cataract. *Am J Med Genet* **73**, 125–126 (1997).

120. Sanford, L. P. *et al.* TGFβ2 knockout mice have multiple developmental defects that are non-overlapping with other TGFβ knockout phenotypes. *Development* **124**, 2659–2670 (1997).

121. Huai, B., Huang, C. & Hu, L. Curcumin Suppresses TGF-β2-Induced Proliferation, Migration, and Invasion in Lens Epithelial Cells by Targeting KCNQ1OT1/miR-377-3p/COL1A2 Axis in Posterior Capsule Opacification. *Curr Eye Res* **47**, 715–726 (2022).

122. Huang, P., Hu, Y. & Duan, Y. TGF-β2-induced circ-PRDM5 regulates migration, invasion, and EMT through the miR-92b-3p/COL1A2 pathway in human lens epithelial cells. *J Mol Histol* **53**, 309–320 (2022).

123. Wang, J. *et al.* TGF-β regulation of microRNA miR-497-5p and ocular lens epithelial cell mesenchymal transition. *Sci China Life Sci* **63**, 1928–1937 (2020).

124. Furuta, Y. & Hogan, B. L. M. BMP4 is essential for lens induction in the mouse embryo. *Genes Dev* **12**, 3764–3775 (1998).

125. Wawersik, S. *et al.* BMP7 Acts in Murine Lens Placode Development. *Developmental Biology* **207**, 176–188 (1999).

126. Rajagopal, R. *et al.* The type I BMP receptors, Bmpr1a and Acvr1, activate multiple signaling pathways to regulate lens formation. *Dev Biol* **335**, 305–316 (2009).

127. Faber, S. C., Robinson, M. L., Makarenkova, H. P. & Lang, R. A. Bmp signaling is required for development of primary lens fiber cells. *Development* **129**, 3727–3737 (2002).

128. Vidya, N. G., Vasavada, A. R. & Rajkumar, S. Evaluating the association of bone morphogenetic protein 4-V152A and SIX homeobox 6-H141N

polymorphisms with congenital cataract and microphthalmia in Western Indian population. *J Postgrad Med* **64**, 86–91 (2018).

129. Arvanitis, D. & Davy, A. Eph/ephrin signaling: networks. *Genes Dev* **22**, 416–429 (2008).

130. Park, J. E. *et al.* Human cataract mutations in EPHA2 SAM domain alter receptor stability and function. *PLoS One* **7**, e36564 (2012).

131. Zhang, T. *et al.* Mutations of the EPHA2 receptor tyrosine kinase gene cause autosomal dominant congenital cataract. *Hum Mutat* **30**, E603-611 (2009).

132. Sundaresan, P. *et al.* EPHA2 polymorphisms and age-related cataract in India. *PLoS One* **7**, e33001 (2012).

133. Cooper, M. A. *et al.* Loss of ephrin-A5 function disrupts lens fiber cell packing and leads to cataract. *Proc Natl Acad Sci U S A* **105**, 16620–16625 (2008).

134. Cheng, C. & Gong, X. Diverse Roles of Eph/ephrin Signaling in the Mouse Lens. *PLoS One* **6**, e28147 (2011).

135. Cheng, C., Ansari, M. M., Cooper, J. A. & Gong, X. EphA2 and Src regulate equatorial cell morphogenesis during lens development. *Development* **140**, 4237–4245 (2013).

136. Pan, D. Hippo signaling in organ size control. *Genes Dev* **21**, 886–897 (2007).

137. Lachke, S. A. *et al.* iSyTE: integrated Systems Tool for Eye gene discovery. *Invest Ophthalmol Vis Sci* **53**, 1617–1627 (2012).

138. Lachke, S. A. *et al.* Mutations in the RNA Granule Component TDRD7 Cause Cataract and Glaucoma. *Science (New York, N.y.)* **331**, 1571 (2011).

139. Dash, S., Dang, C. A., Beebe, D. C. & Lachke, S. A. Deficiency of the RNA binding protein Caprin2 causes lens defects and features of Peters anomaly. *Developmental dynamics : an official publication of the American Association of Anatomists* **244**, 1313 (2015).

140. Siddam, A. D. *et al.* The RNA-binding protein Celf1 post-transcriptionally regulates p27Kip1 and Dnase2b to control fiber cell nuclear degradation in lens development. *PLoS Genet* **14**, e1007278 (2018).

141. Wu, X., Long, E., Lin, H. & Liu, Y. Prevalence and epidemiological characteristics of congenital cataract: a systematic review and meta-analysis. *Sci Rep* **6**, 28564 (2016).

142. Zhou, Z. *et al.* Genetic variations in GJA3, GJA8, LIM2, and age-related cataract in the Chinese population: a mutation screening study. *Mol Vis* **17**, 621–626 (2011).

143. Ceroni, F. *et al.* New GJA8 variants and phenotypes highlight its critical role in a broad spectrum of eye anomalies. *Hum Genet* **138**, 1027–1042 (2019).

144. Lambert, S. R., Taylor, D., Kriss, A., Holzel, H. & Heard, S. Ocular Manifestations of the Congenital Varicella Syndrome. *Archives of Ophthalmology* **107**, 52–56 (1989).

145. Bouthry, E. *et al.* Rubella and pregnancy: diagnosis, management and outcomes. *Prenat Diagn* **34**, 1246–1253 (2014).

146. Matalia, J. & Shirke, S. Congenital Rubella. *N Engl J Med* **375**, 1468 (2016).

147. de Oliveira Dias, J. R. *et al.* Zika and the Eye: Pieces of a Puzzle. *Progress in Retinal and Eye Research* **66**, 85–106 (2018).

148. de Paula Freitas, B. *et al.* Anterior-Segment Ocular Findings and Microphthalmia in Congenital Zika Syndrome. *Ophthalmology* **124**, 1876–1878 (2017).

149. Kiziltoprak, H., Tekin, K., Inanc, M. & Goker, Y. S. Cataract in diabetes mellitus. *World J Diabetes* **10**, 140–153 (2019).

150. Ang, M. J. & Afshari, N. A. Cataract and systemic disease: A review. *Clinical & Experimental Ophthalmology* **49**, 118–127 (2021).

151. Li, Y., Jiang, S.-H., Liu, S. & Wang, Q. Role of lncRNA NEAT1 mediated by YY1 in the development of diabetic cataract via targeting the microRNA-205-3p/MMP16 axis. *Eur Rev Med Pharmacol Sci* **24**, 5863–5870 (2020).

152. Zhang, L., Cheng, R. & Huang, Y. MiR-30a inhibits BECN1-mediated autophagy in diabetic cataract. *Oncotarget* **8**, 77360–77368 (2017).

153. Delcourt, C. *et al.* Lifetime Exposure to Ambient Ultraviolet Radiation and the Risk for Cataract Extraction and Age-Related Macular Degeneration: The Alienor Study. *Investigative Ophthalmology & Visual Science* **55**, 7619–7627 (2014).

154. Löfgren, S. Solar ultraviolet radiation cataract. *Experimental Eye Research* **156**, 112–116 (2017).

155. Davidson, R. S. *et al.* Surgical correction of presbyopia. *Journal of Cataract & Refractive Surgery* **42**, 920 (2016).

156. Wolffsohn, J. S. & Davies, L. N. Presbyopia: Effectiveness of correction strategies. *Progress in Retinal and Eye Research* **68**, 124–143 (2019).

157. Cooper, J. & Tkatchenko, A. V. A Review of Current Concepts of the Etiology and Treatment of Myopia. *Eye Contact Lens* **44**, 231–247 (2018).

158. Delbarre, M., Le, H. M., Boucenna, W. & Froussart-Maille, F. Hypermétropie et chirurgie réfractive. *Journal Français d'Ophtalmologie* **44**, 723–729 (2021).

159. Li, Y. & Piatigorsky, J. Targeted deletion of Dicer disrupts lens morphogenesis, corneal epithelium stratification, and whole eye development. *Dev Dyn* **238**, 2388–2400 (2009).

160. Wu, C.-R., Ye, M., Qin, L., Yin, Y. & Pei, C. Expression of lens-related microRNAs in transparent infant lenses and congenital cataract. *Int J Ophthalmol* **10**, 361–365 (2017).

161. Yao, L. & Yan, H. MiR-182 inhibits oxidative stress and epithelial cell apoptosis in lens of cataract rats through PI3K/Akt signaling pathway. *Eur Rev Med Pharmacol Sci* **24**, 12001–12008 (2020).

162. Li, Z.-N., Ge, M.-X. & Yuan, Z.-F. MicroRNA-182-5p protects human lens epithelial cells against oxidative stress-induced apoptosis by inhibiting NOX4 and p38 MAPK signalling. *BMC Ophthalmol* **20**, 233 (2020).

163. Spector, A. Oxidative stress-induced cataract: mechanism of action. *FASEB J* **9**, 1173–1182 (1995).

164. Conte, I. *et al.* miR-204 is required for lens and retinal development via Meis2 targeting. *Proc Natl Acad Sci U S A* **107**, 15491–15496 (2010).

165. Avellino, R. *et al.* miR-204 Targeting of Ankrd13A Controls Both Mesenchymal Neural Crest and Lens Cell Migration. *PLoS One* **8**, e61099 (2013).

166. Wu, C. *et al.* MiRNAs regulate oxidative stress related genes via binding to the 3’ UTR and TATA-box regions: a new hypothesis for cataract pathogenesis. *BMC Ophthalmol* **17**, 142 (2017).

167. Li, X., Sun, M., Cheng, A. & Zheng, G. LncRNA GAS5 regulates migration and epithelial-to-mesenchymal transition in lens epithelial cells via the miR-204-3p/TGFBR1 axis. *Laboratory Investigation* **102**, 452–460 (2022).

168. Fan, C. *et al.* Circular RNA circ KMT2E is up-regulated in diabetic cataract lenses and is associated with miR-204-5p sponge function. *Gene* **710**, 170–177 (2019).

169. Wang, Y. *et al.* MicroRNA-204-5p regulates epithelial-to-mesenchymal transition during human posterior capsule opacification by targeting SMAD4. *Invest Ophthalmol Vis Sci* **54**, 323–332 (2013).

170. Hoffmann, A. *et al.* Implication of the miR-184 and miR-204 competitive RNA network in control of mouse secondary cataract. *Mol Med* **18**, 528–538 (2012).

171. Dong, N. Long Noncoding RNA NEAT1 Regulates TGF-β2-Induced Epithelial-Mesenchymal Transition of Lens Epithelial Cells through the miR-34a/Snail1 and miR-204/Zeb1 Pathways. *Biomed Res Int* **2020**, 8352579 (2020).

172. Tang, S. *et al.* AQP5 regulates vimentin expression via miR-124–3p.1 to protect lens transparency. *Experimental Eye Research* **205**, 108485 (2021).

173. Guo, X., Li, C., Wang, Y., Jiang, C. & Yang, L. Long non-coding RNA nuclear paraspeckle assembly transcript 1 downregulation protects lens epithelial cells from oxidative stress-induced apoptosis by regulating the microRNA-124-3p/death-associated protein kinase 1 axis in age-related cataract. *Int Ophthalmol* (2023) doi:10.1007/s10792-023-02749-4.

174. Xu, Y., Zheng, Y., Shen, P. & Zhou, L. Role of long noncoding RNA KCNQ1 overlapping transcript 1/microRNA-124-3p/BCL-2-like 11 axis in hydrogen peroxide (H2O2)-stimulated human lens epithelial cells. *Bioengineered* **13**, 5035–5045 (2022).

175. Liu, Y., Li, S., Liu, Y., Lv, X. & Zhou, Q. MicroRNA-124 facilitates lens epithelial cell apoptosis by inhibiting SPRY2 and MMP-2. *Mol Med Rep* **23**, 381 (2021).

176. Anand, D. *et al.* Genome-Wide Analysis of Differentially Expressed miRNAs and Their Associated Regulatory Networks in Lenses Deficient for the Congenital Cataract-Linked Tudor Domain Containing Protein TDRD7. *Frontiers in Cell and Developmental Biology* **9**, (2021).

177. Ryan, D. G., Oliveira-Fernandes, M. & Lavker, R. M. MicroRNAs of the mammalian eye display distinct and overlapping tissue specificity. *Mol Vis* **12**, 1175–1184 (2006).

178. Dismuke, W. M., Challa, P., Navarro, I., Stamer, W. D. & Liu, Y. Human Aqueous Humor Exosomes. *Exp Eye Res* **132**, 73–77 (2015).

179. Hughes, A. E. *et al.* Mutation altering the miR-184 seed region causes familial keratoconus with cataract. *Am J Hum Genet* **89**, 628–633 (2011).

180. Bykhovskaya, Y., Caiado Canedo, A. L., Wright, K. W. & Rabinowitz, Y. S. C.57 C > T Mutation in MIR 184 is Responsible for Congenital Cataracts and Corneal Abnormalities in a Five-generation Family from Galicia, Spain. *Ophthalmic Genet* **36**, 244–247 (2015).

181. Lechner, J. *et al.* Mutational analysis of MIR184 in sporadic keratoconus and myopia. *Invest Ophthalmol Vis Sci* **54**, 5266–5272 (2013).

182. Luo, Y., Liu, S. & Yao, K. Transcriptome-wide Investigation of mRNA/circRNA in miR-184 and Its r.57c > u Mutant Type Treatment of Human Lens Epithelial Cells. *Mol Ther Nucleic Acids* **7**, 71–80 (2017).

183. Iliff, B. W., Riazuddin, S. A. & Gottsch, J. D. A single-base substitution in the seed region of miR-184 causes EDICT syndrome. *Invest Ophthalmol Vis Sci* **53**, 348–353 (2012).

184. Zhang, J. *et al.* Knockout of miR-184 in zebrafish leads to ocular abnormalities by elevating p21 levels. *FASEB J* **37**, e22927 (2023).

185. Cheng, T. *et al.* lncRNA H19 contributes to oxidative damage repair in the early age-related cataract by regulating miR-29a/TDG axis. *J Cell Mol Med* **23**, 6131–6139 (2019).

186. Sun, M. *et al.* lncRNA TUG1 regulates Smac/DIABLO expression by competitively inhibiting miR-29b and modulates the apoptosis of lens epithelial cells in age-related cataracts. *Chin Med J (Engl)* (2023) doi:10.1097/CM9.0000000000002530.

187. Liu, S.-J. *et al.* miR-15a-3p affects the proliferation, migration and apoptosis of lens epithelial cells. *Mol Med Rep* **19**, 1110–1116 (2019).

188. Fan, F., Zhuang, J., Zhou, P., Liu, X. & Luo, Y. MicroRNA-34a promotes mitochondrial dysfunction-induced apoptosis in human lens epithelial cells by targeting Notch2. *Oncotarget* **8**, 110209–110220 (2017).

189. Lu, B. *et al.* miR-211 regulates the antioxidant function of lens epithelial cells affected by age-related cataracts. *Int J Ophthalmol* **11**, 349–353 (2018).

190. Shen, Q. & Zhou, T. Knockdown of lncRNA TUG1 protects lens epithelial cells from oxidative stress-induced injury by regulating miR-196a-5p expression in age-related cataracts. *Exp Ther Med* **22**, 1286 (2021).

191. Gao, W., Zhou, X. & Lin, R. miR-378a-5p and miR-630 induce lens epithelial cell apoptosis in cataract via suppression of E2F3. *Braz J Med Biol Res* **53**, e9608 (2020).

192. Yao, P. *et al.* miR-23a-3p regulates the proliferation and apoptosis of human lens epithelial cells by targeting Bcl-2 in an in vitro model of cataracts. *Exp Ther Med* **21**, 436 (2021).

193. Kim, Y. J. *et al.* Investigation of MicroRNA Expression in Anterior Lens Capsules of Senile Cataract Patients and MicroRNA Differences According to the Cataract Type. *Transl Vis Sci Technol* **10**, 14 (2021).

194. Ye, W. *et al.* LncRNA MALAT1 Regulates miR-144-3p to Facilitate Epithelial-Mesenchymal Transition of Lens Epithelial Cells via the ROS/NRF2/Notch1/Snail Pathway. *Oxid Med Cell Longev* **2020**, 8184314 (2020).

195. Zhang, L. *et al.* MicroRNA-30a Regulation of Epithelial-Mesenchymal Transition in Diabetic Cataracts Through Targeting SNAI1. *Sci Rep* **7**, 1117 (2017).

196. Han, R. *et al.* MicroRNA-34a inhibits epithelial-mesenchymal transition of lens epithelial cells by targeting Notch1. *Exp Eye Res* **185**, 107684 (2019).

197. Feng, D., Zhu, N., Yu, C. & Lou, D. MicroRNA-34a suppresses human lens epithelial cell proliferation and migration via downregulation of c-Met. *Clin Chim Acta* **495**, 326–330 (2019).

198. Shaham, O. *et al.* Pax6 regulates gene expression in the vertebrate lens through miR-204. *PLoS Genet* **9**, e1003357 (2013).

199. Bitel, C. L., Perrone-Bizzozero, N. I. & Frederikse, P. H. HuB/C/D, nPTB, REST4, and miR-124 regulators of neuronal cell identity are also utilized in the lens. *Mol Vis* **16**, 2301–2316 (2010).

200. Wang, C., Zhao, R. & Zhang, S. lncRNA XIST knockdown suppresses cell proliferation and promotes apoptosis in diabetic cataracts through the miR-34a/SMAD2 axis. *Mol Med Rep* **25**, 7 (2022).

201. Gao, C. *et al.* Exosomal miR-29b found in aqueous humour mediates calcium signaling in diabetic patients with cataract. *Int J Ophthalmol* **14**, 1484–1491 (2021).

202. Tao, D. *et al.* CircPAG1 interacts with miR-211-5p to promote the E2F3 expression and inhibit the high glucose-induced cell apoptosis and oxidative stress in diabetic cataract. *Cell Cycle* **21**, 708–719 (2022).

203. Zeng, K., Feng, Q.-G., Lin, B.-T., Ma, D.-H. & Liu, C.-M. Effects of microRNA-211 on proliferation and apoptosis of lens epithelial cells by targeting SIRT1 gene in diabetic cataract mice. *Biosci Rep* **37**, BSR20170695 (2017).

204. Yang, J., Zhao, S. & Tian, F. SP1-mediated lncRNA PVT1 modulates the proliferation and apoptosis of lens epithelial cells in diabetic cataract via miR-214-3p/MMP2 axis. *J Cell Mol Med* **24**, 554–561 (2020).

205. Liu, X. *et al.* microRNA-199a-5p regulates epithelial-to-mesenchymal transition in diabetic cataract by targeting SP1 gene. *Mol Med* **26**, 122 (2020).

206. Liu, Y., Chen, T. & Zheng, G. Exosome-transmitted circ-CARD6 facilitates posterior capsule opacification development by miR-31/FGF7 axis. *Exp Eye Res* **207**, 108572 (2021).

207. Wang, X. *et al.* MiR-22-3p inhibits fibrotic cataract through inactivation of HDAC6 and increase of α-tubulin acetylation. *Cell Prolif* **53**, e12911 (2020).

208. Liu, J. *et al.* Circ-MKLN1/miR-377-3p/CTGF Axis Regulates the TGF-β2-induced Posterior Capsular Opacification in SRA01/04 Cells. *Curr Eye Res* **47**, 372–381 (2022).

209. Dong, N., Tang, X. & Xu, B. miRNA-181a inhibits the proliferation, migration, and epithelial-mesenchymal transition of lens epithelial cells. *Invest Ophthalmol Vis Sci* **56**, 993–1001 (2015).

210. Wang, C., Zhao, B., Fang, J. & Shi, Z. IGF-1 Promotes Epithelial-Mesenchymal Transition of Lens Epithelial Cells That Is Conferred by miR-3666 Loss. *Contrast Media Mol Imaging* **2022**, 5383146 (2022).

211. Liu, J. *et al.* LncRNA KCNQ1OT1 knockdown inhibits viability, migration and epithelial-mesenchymal transition in human lens epithelial cells via miR-26a-5p/ITGAV/TGF-beta/Smad3 axis. *Exp Eye Res* **200**, 108251 (2020).

212. Yu, X.-H., Liu, S.-Y. & Li, C.-F. TGF-β2-induced NEAT1 regulates lens epithelial cell proliferation, migration and EMT by the miR-26a-5p/FANCE axis. *Int J Ophthalmol* **14**, 1674–1682 (2021).

213. Eiberg, H. *et al.* A splice-site variant in the lncRNA gene RP1-140A9.1 cosegregates in the large Volkmann cataract family. *Mol Vis* **25**, 1–11 (2019).

214. Dong, N. Long Noncoding RNA MALAT1 Acts as a Competing Endogenous RNA to Regulate TGF-β2 Induced Epithelial-Mesenchymal Transition of Lens Epithelial Cells by a MicroRNA-26a-Dependent Mechanism. *Biomed Res Int* **2019**, 1569638 (2019).

215. Peng, C. *et al.* LncRNA-MALAT1/miRNA-204-5p/Smad4 Axis Regulates Epithelial-Mesenchymal Transition, Proliferation and Migration of Lens Epithelial Cells. *Curr Eye Res* **46**, 1137–1147 (2021).

216. Hu, Y. *et al.* Regulation of the Inflammatory Response, Proliferation, Migration, and Epithelial-Mesenchymal Transition of Human Lens Epithelial Cells by the lncRNA-MALAT1/miR-26a-5p/TET1 Signaling Axis. *J Ophthalmol* **2023**, 9942880 (2023).

217. Xia, F. *et al.* LncRNA KCNQ1OT1: Molecular mechanisms and pathogenic roles in human diseases. *Genes Dis* **9**, 1556–1565 (2022).

218. Yao, L., Yang, L., Song, H., Liu, T. & Yan, H. MicroRNA miR-29c-3p modulates FOS expression to repress EMT and cell proliferation while induces apoptosis in TGF-β2-treated lens epithelial cells regulated by lncRNA KCNQ1OT1. *Biomed Pharmacother* **129**, 110290 (2020).

219. Xiang, J. *et al.* LncRNA PLCD3-OT1 Functions as a CeRNA to Prevent Age-Related Cataract by Sponging miR-224-5p and Regulating PLCD3 Expression. *Invest Ophthalmol Vis Sci* **60**, 4670–4680 (2019).

220. Chen, J. *et al.* Molecular Genetic Analysis of Pakistani Families With Autosomal Recessive Congenital Cataracts by Homozygosity Screening. *Investigative Ophthalmology & Visual Science* **58**, 2207 (2017).

221. Tan, Y. Q. *et al.* Loss-of-function mutations in TDRD7 lead to a rare novel syndrome combining congenital cataract and nonobstructive azoospermia in humans. *Genetics in Medicine* **21**, 1209–1217 (2019).

222. Zheng, C. *et al.* RNA granule component TDRD7 gene polymorphisms in a Han Chinese population with age-related cataract. *Journal of International Medical Research* **42**, 153–163 (2014).

223. Tu, C. *et al.* TDRD7 participates in lens development and spermiogenesis by mediating autophagosome maturation. *Autophagy* **17**, 3848 (2021).

224. Raju, I. & Abraham, E. C. Mutants of Human αB-Crystallin Cause Enhanced Protein Aggregation and Apoptosis in Mammalian Cells: Influence of co-expression of HspB1. *Biochem Biophys Res Commun* **430**, 107–112 (2013).

225. Barnum, C. E. *et al.* The Tudor-domain protein TDRD7, mutated in congenital cataract, controls the heat shock protein HSPB1 (HSP27) and lens fiber cell morphology. *Human Molecular Genetics* **29**, 2076 (2020).

226. Nakazawa, K., Shichino, Y., Iwasaki, S. & Shiina, N. Implications of RNG140 (caprin2)-mediated translational regulation in eye lens differentiation. *The Journal of Biological Chemistry* **295**, 15029 (2020).

227. Choquet, H. *et al.* A large multiethnic GWAS meta-analysis of cataract identifies new risk loci and sex-specific effects. *Nature Communications* **12**, (2021).

228. Ma, B. *et al.* Polymorphisms in TRIB2 and CAPRIN2 Genes Contribute to the Susceptibility to High Myopia-Induced Cataract in Han Chinese

Population. *Medical science monitor : international medical journal of experimental and clinical research* **29**, (2023).

229. Al-Ghoul, K. J. *et al.* Structural Evidence of Human Nuclear Fiber Compaction as a Function of Ageing and Cataractogenesis. *Experimental Eye Research* **72**, 199–214 (2001).
230. Grifone, R., Shao, M., Saquet, A. & Shi, D. L. RNA-Binding Protein Rbm24 as a Multifaceted Post-Transcriptional Regulator of Embryonic Lineage Differentiation and Cellular Homeostasis. *Cells* **9**, (2020).
231. Wang, Y., Li, W., Zhang, C., Peng, W. & Xu, Z. RBM24 is localized to stress granules in cells under various stress conditions. *Biochemical and Biophysical Research Communications* **608**, 96–101 (2022).
232. Dash, S. *et al.* The master transcription factor SOX2, mutated in anophthalmia/microphthalmia, is post-transcriptionally regulated by the conserved RNA-binding protein RBM24 in vertebrate eye development. *Human Molecular Genetics* **29**, 591 (2020).
233. Shao, M. *et al.* Rbm24 controls poly(A) tail length and translation efficiency of crystallin mRNAs in the lens via cytoplasmic polyadenylation. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 7245–7254 (2020).
234. Dash, S. *et al.* The master transcription factor SOX2, mutated in anophthalmia/microphthalmia, is post-transcriptionally regulated by the conserved RNA-binding protein RBM24 in vertebrate eye development. *Human Molecular Genetics* **29**, 591 (2020).
235. Frederikse, P., Nandanoor, A. & Kasinathan, C. PTBP-dependent PSD-95 and CamKIIα alternative splicing in the lens. *Molecular Vision* **20**, 1660 (2014).
236. Cockburn, D. M. *et al.* The double-stranded RNA-binding protein Staufen 2 regulates eye size. *Molecular and Cellular Neuroscience* **51**, 101–111 (2012).

237. Cao, W., McMahon, M., Wang, B., O'Connor, R. & Clarkson, M. A case report of spontaneous mutation (C33>U) in the iron-responsive element of l-ferritin causing hyperferritinemia-cataract syndrome. *Blood Cells, Molecules, and Diseases* **44**, 22–27 (2010).

238. Turunen, J. J., Niemelä, E. H., Verma, B. & Frilander, M. J. The significant other: splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA* **4**, 61–76 (2013).

239. Bezen, D. *et al.* A homozygous Y443C variant in the RNPC3 is associated with severe syndromic congenital hypopituitarism and diffuse brain atrophy. *Am J Med Genet A* **188**, 2701–2706 (2022).

240. Verberne, E. A., Faries, S., Mannens, M. M. A. M., Postma, A. V. & van Haelst, M. M. Expanding the phenotype of biallelic RNPC3 variants associated with growth hormone deficiency. *Am J Med Genet A* **182**, 1952–1956 (2020).

241. Neumann, D. P., Goodall, G. J. & Gregory, P. A. The Quaking RNA-binding proteins as regulators of cell differentiation. *Wiley Interdiscip Rev RNA* **13**, e1724 (2022).

242. Shin, S. *et al.* Qki activates Srebp2-mediated cholesterol biosynthesis for maintenance of eye lens transparency. *Nat Commun* **12**, 3005 (2021).

243. Zhao, L. *et al.* Lanosterol reverses protein aggregation in cataracts. *Nature* **523**, 607–611 (2015).

244. Dasgupta, T. & Ladd, A. N. The importance of CELF control: molecular and biological roles of the CUG-BP, Elav-like family of RNA binding proteins. *Wiley interdisciplinary reviews. RNA* **3**, 104–104 (2012).

245. The Human Protein Atlas. https://www.proteinatlas.org/.

246. Ladd, A. N., Stenberg, M. G., Swanson, M. S. & Cooper, T. A. Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. *Developmental Dynamics* **233**, 783–793 (2005).

247. Kalsotra, A. *et al.* A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20333–20333 (2008).

248. Blech-Hermoni, Y., Stillwagon, S. J. & Ladd, A. N. Diversity and conservation of CELF1 and CELF2 RNA and protein expression patterns during embryonic development. *Dev Dyn* **242**, 767–777 (2013).

249. Suzuki, H. *et al.* Vegetal localization of the maternal mRNA encoding an EDEN-BP/Bruno-like protein in zebrafish. *Mech Dev* **93**, 205–209 (2000).

250. Schoser, B. & Timchenko, L. Myotonic Dystrophies 1 and 2: Complex Diseases with Complex Mechanisms. *Current Genomics* **11**, 77–77 (2010).

251. Davis, B. M., McCurrach, M. E., Taneja, K. L., Singer, R. H. & Housman, D. E. Expansion of a CUG trinucleotide repeat in the 3’ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc Natl Acad Sci U S A* **94**, 7388–7393 (1997).

252. Taneja, K. L., McCurrach, M., Schalling, M., Housman, D. & Singer, R. H. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol* **128**, 995–1002 (1995).

253. Carrell, S. T. *et al.* Dmpk gene deletion or antisense knockdown does not compromise cardiac or skeletal muscle function in mice. *Hum Mol Genet* **25**, 4328–4338 (2016).

254. Mahadevan, M. S. *et al.* Reversible model of RNA toxicity and cardiac conduction defects in myotonic dystrophy. *Nat Genet* **38**, 1066–1070 (2006).

255. Rao, A. N. *et al.* Reversible cardiac disease features in an inducible CUG repeat RNA-expressing mouse model of myotonic dystrophy. *JCI Insight* **6**, e143465, 143465 (2021).

256. Marquis, J. *et al.* CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding. *Biochemical Journal* **400**, 291–291 (2006).

257. Le Tonquèze, O., Gschloessl, B., Legagneux, V., Paillard, L. & Audic, Y. Identification of CELF1 RNA targets by CLIP-seq in human HeLa cells. *Genomics Data* **8**, 97–97 (2016).

258. Lambert, N. *et al.* RNA Bind-n-Seq: quantitative assessment of the sequence and structural binding specificity of RNA binding proteins. *Mol Cell* **54**, 887–900 (2014).

259. Wang, E. T. *et al.* Antagonistic regulation of mRNA expression and splicing by CELF and MBNL proteins. *Genome Res* **25**, 858–871 (2015).

260. Miller, J. W. *et al.* Recruitment of human muscleblind proteins to (CUG)n expansions associated with myotonic dystrophy. *The EMBO Journal* **19**, 4439–4439 (2000).

261. Kanadia, R. N. *et al.* A muscleblind knockout model for myotonic dystrophy. *Science* **302**, 1978–1980 (2003).

262. López-Martínez, A., Soblechero-Martín, P., de-la-Puente-Ovejero, L., Nogales-Gadea, G. & Arechavala-Gomeza, V. An Overview of Alternative Splicing Defects Implicated in Myotonic Dystrophy Type I. *Genes (Basel)* **11**, 1109 (2020).

263. Timchenko, L. T. *et al.* Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic acids research* **24**, 4407–4414 (1996).

264. Timchenko, N. A. *et al.* Overexpression of CUG Triplet Repeat-binding Protein, CUGBP1, in Mice Inhibits Myogenesis. *Journal of Biological Chemistry* **279**, 13129–13139 (2004).

265. Ho, T. H., Bundman, D., Armstrong, D. L. & Cooper, T. A. Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Human Molecular Genetics* **14**, 1539–1547 (2005).

266. Ohsawa, N., Koebis, M., Mitsuhashi, H., Nishino, I. & Ishiura, S. ABLIM1 splicing is abnormal in skeletal muscle of patients with DM1 and regulated

by MBNL, CELF and PTBP1. *Genes to cells : devoted to molecular & cellular mechanisms* **20**, 121–134 (2015).

267. Masuda, A. *et al.* CUGBP1 and MBNL1 preferentially bind to 3′ UTRs and facilitate mRNA decay. *Sci Rep* **2**, 209 (2012).

268. Verma, S. K. *et al.* Reactivation of Fetal Splicing Programs in Diabetic Hearts Is Mediated by Protein Kinase C Signaling*. *Journal of Biological Chemistry* **288**, 35372–35386 (2013).

269. Belanger, K. A. *et al.* CELF1 contributes to aberrant alternative splicing patterns in the Type 1 diabetic heart. *Biochemical and biophysical research communications* **503**, 3205–3205 (2018).

270. Hu, X., Wu, P., Liu, B., Lang, Y. & Li, T. RNA-binding protein CELF1 promotes cardiac hypertrophy via interaction with PEBP1 in cardiomyocytes. *Cell Tissue Res* **387**, 111–121 (2022).

271. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* **45**, 1452–1458 (2013).

272. Wang, X., Wang, H., Ji, F., Zhao, S. & Fang, X. Lentivirus-Mediated Knockdown of CUGBP1 Suppresses Gastric Cancer Cell Proliferation In Vitro. *Appl Biochem Biotechnol* **173**, 1529–1536 (2014).

273. Xia, L. *et al.* CELF1 is Up-Regulated in Glioma and Promotes Glioma Cell Proliferation by Suppression of CDKN1B. *Int J Biol Sci* **11**, 1314–1324 (2015).

274. Wang, H. *et al.* MiR-330-3p functions as a tumor suppressor that regulates glioma cell proliferation and migration by targeting CELF1. *Arch Med Sci* **16**, 1166–1175 (2020).

275. David, G. *et al.* The RNA-binding proteins CELF1 and ELAVL1 cooperatively control the alternative splicing of CD44. *Biochemical and Biophysical Research Communications* **626**, 79–84 (2022).

276. Cifdaloz, M. *et al.* Systems analysis identifies melanoma-enriched pro-oncogenic networks controlled by the RNA binding protein CELF1. *Nat Commun* **8**, 2249 (2017).

277. Tan, Y. *et al.* Small molecule targeting CELF1 RNA-binding activity to control HSC activation and liver fibrosis. *Nucleic Acids Res* **50**, 2440–2451 (2022).

278. Wu, L.-N., Xue, Y.-J., Zhang, L.-J., Ma, X.-M. & Chen, J.-F. Si-RNA mediated knockdown of CELF1 gene suppressed the proliferation of human lung cancer cells. *Cancer Cell Int* **13**, 115 (2013).

279. Talwar, S. *et al.* Overexpression of RNA-binding protein CELF1 prevents apoptosis and destabilizes pro-apoptotic mRNAs in oral cancer cells. *RNA Biol* **10**, 277–286 (2013).

280. Wang, H. *et al.* RNA-binding protein CELF1 enhances cell migration, invasion, and chemoresistance by targeting ETS2 in colorectal cancer. *Clin Sci (Lond)* **134**, 1973–1990 (2020).

281. Milne, C. A. & Hodgkin, J. ETR-1, a homologue of a protein linked to myotonic dystrophy, is essential for muscle development in Caenorhabditis elegans. *Current Biology* **9**, 1243–1246 (1999).

282. Chang, K.-T., Cheng, C.-F., King, P.-C., Liu, S.-Y. & Wang, G.-S. CELF1 Mediates Connexin 43 mRNA Degradation in Dilated Cardiomyopathy. *Circ Res* **121**, 1140–1152 (2017).

283. Gautier-Courteille, C. *et al.* EDEN-BP-dependent post-transcriptional regulation of gene expression in Xenopus somitic segmentation. *Development* **131**, 6107–6117 (2004).

284. Kress, C., Gautier-Courteille, C., Osborne, H. B., Babinet, C. & Paillard, L. Inactivation of CUG-BP1/CELF1 causes growth, viability, and spermatogenesis defects in mice. *Mol Cell Biol* **27**, 1146–1157 (2007).

285. Matsui, T., Sasaki, A., Akazawa, N., Otani, H. & Bessho, Y. Celf1 regulation of dmrt2a is required for somite symmetry and left-right patterning during zebrafish development. *Development* **139**, 3553–3560 (2012).

286. Tahara, N., Bessho, Y. & Matsui, T. Celf1 Is Required for Formation of Endoderm-Derived Organs in Zebrafish. *Int J Mol Sci* **14**, 18009–18023 (2013).

287. Kim, Y. K., Mandal, M., Yadava, R. S., Paillard, L. & Mahadevan, M. S. Evaluating the effects of CELF1 deficiency in a mouse model of RNA toxicity. *Hum Mol Genet* **23**, 293–302 (2014).

288. Charlet-B, N., Singh, G., Cooper, T. A. & Logan, P. Dynamic Antagonism between ETR-3 and PTB Regulates Cell Type-Specific Alternative Splicing. *Molecular Cell* **9**, 649–658 (2002).

289. Liu, Y. *et al.* CUGBP1, a crucial factor for heart regeneration in mice. *Cell death & disease* **13**, (2022).

290. CELF1 | gnomAD v2.1.1 | gnomAD. https://gnomad.broadinstitute.org/gene/ENSG00000149187.

291. Paillard, L. *et al.* EDEN and EDEN-BP, a cis element and an associated factor that mediate sequence-specific mRNA deadenylation in Xenopus embryos. *The EMBO Journal* **17**, 278–278 (1998).

292. Paillard, L., Legagneux, V., Maniey, D. & Osborne, H. B. c-Jun ARE targets mRNA deadenylation by an EDEN-BP (embryo deadenylation element-binding protein)-dependent pathway. *Journal of Biological Chemistry* **277**, 3232–3235 (2002).

293. Paillard, L., Legagneux, V. & Osborne, H. B. A functional deadenylation assay identifies human CUG-BP as a deadenylation factor. *Biology of the Cell* **95**, 107–113 (2003).

294. Zhang, L., Lee, J. E., Wilusz, J. & Wilusz, C. J. The RNA-binding Protein CUGBP1 Regulates Stability of Tumor Necrosis Factor mRNA in Muscle Cells:

IMPLICATIONS FOR MYOTONIC DYSTROPHY*S. *The Journal of Biological Chemistry* **283**, 22457–22457 (2008).

295. Lee, J. E., Lee, J. Y., Wilusz, J., Tian, B. & Wilusz, C. J. Systematic Analysis of Cis-Elements in Unstable mRNAs Demonstrates that CUGBP1 Is a Key Regulator of mRNA Decay in Muscle Cells. *PLoS ONE* **5**, (2010).

296. Moraes, K. C. M., Wilusz, C. J. & Wilusz, J. CUG-BP binds to RNA substrates and recruits PARN deadenylase. *RNA (New York, N.Y.)* **12**, 1084–1091 (2006).

297. Vlasova, I. A. *et al.* Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1. *Molecular cell* **29**, 263–270 (2008).

298. Rattenbacher, B. *et al.* Analysis of CUGBP1 Targets Identifies GU-Repeat Sequences That Mediate Rapid mRNA Decay. *Molecular and Cellular Biology* **30**, 3970–3970 (2010).

299. Fox, J. T. & Stover, P. J. Mechanism of the Internal Ribosome Entry Site-mediated Translation of Serine Hydroxymethyltransferase 1. *The Journal of Biological Chemistry* **284**, 31085–31085 (2009).

300. Tirta, Y. K. *et al.* CELF1 represses Doublesex1 expression via its 5’ UTR in the crustacean Daphnia magna. *PloS one* **17**, (2022).

301. Liu, L. *et al.* Competition between RNA-binding proteins CELF1 and HuR modulates MYC translation and intestinal epithelium renewal. *Mol Biol Cell* **26**, 1797–1810 (2015).

302. Chaudhury, A. *et al.* CELF1 is a central node in post-transcriptional regulatory programmes underlying EMT. *Nature Communications 2016 7:1* **7**, 1–15 (2016).

303. Philips, A. V., Timchenko, L. T. & Cooper, T. A. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**, 737–741 (1998).

304. Wang, G. S., Kearney, D. L., De Biasi, M., Taffet, G. & Cooper, T. A. Elevation of RNA-binding protein CUGBP1 is an early event in an inducible

heart-specific mouse model of myotonic dystrophy. *The Journal of Clinical Investigation* **117**, 2802–2802 (2007).

305. Giudice, J., Xia, Z., Li, W. & Cooper, T. A. Neonatal cardiac dysfunction and transcriptome changes caused by the absence of Celf1. *Sci Rep* **6**, 35550 (2016).

306. Sen, S., Talukdar, I. & Webster, N. J. G. SRp20 and CUG-BP1 Modulate Insulin Receptor Exon 11 Alternative Splicing. *Mol Cell Biol* **29**, 871–880 (2009).

307. Cheng, C., Nowak, R. B. & Fowler, V. M. The lens actin filament cytoskeleton: Diverse structures for complex functions. *Experimental Eye Research* **156**, 58–71 (2017).

308. Xiao, J., Jin, S., Wang, X., Huang, J. & Zou, H. CELF1 Selectively Regulates Alternative Splicing of DNA Repair Genes Associated With Cataract in Human Lens Cell Line. *Biochemical genetics* (2022) doi:10.1007/S10528-022-10324-2.

309. Daughters, R. S. *et al.* RNA Gain-of-Function in Spinocerebellar Ataxia Type 8. *PLoS Genet* **5**, e1000600 (2009).

310. Ule, J. *et al.* CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302**, 1212–1215 (2003).

311. Hafner, M. *et al.* CLIP and complementary methods. *Nat Rev Methods Primers* **1**, 1–23 (2021).

312. Masuda, A. *et al.* CUGBP1 and MBNL1 preferentially bind to 3′ UTRs and facilitate mRNA decay. *Sci Rep* **2**, 209 (2012).

313. Graindorge, A. *et al.* Identification of CUG-BP1/EDEN-BP target mRNAs in Xenopus tropicalis. *Nucleic Acids Research* **36**, 1861–1861 (2008).

314. Xia, H. *et al.* CELF1 preferentially binds to exon-intron boundary and regulates alternative splicing in HeLa cells. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1860**, 911–921 (2017).

315. Coulombre, A. J. & Coulombre, J. L. LENS DEVELOPMENT. I. ROLE OF THE LENS IN EYE GROWTH. *The Journal of experimental zoology* **156**, 39–47 (1964).

316. Brastrom, L. K., Scott, C. A., Dawson, D. V. & Slusarski, D. C. A High-Throughput Assay for Congenital and Age-Related Eye Diseases in Zebrafish. *Biomedicines* **7**, (2019).

317. Liu, C.-F. *et al.* Zebrafish (Danio rerio) Is an Economical and Efficient Animal Model for Screening Potential Anti-cataract Compounds. *Transl Vis Sci Technol* **11**, 21 (2022).

318. Glass, A. S. & Dahm, R. The Zebrafish as a Model Organism for Eye Development. *Ophthalmic Research* **36**, 4–24 (2004).

319. Maher, G. J. *et al.* The cataract-associated protein TMEM114, and TMEM235, are glycosylated transmembrane proteins that are distinct from claudin family members. *FEBS Lett* **585**, 2187–2192 (2011).

320. Pfirrmann, T. *et al.* Molecular mechanism of CHRDL1-mediated X-linked megalocornea in humans and in Xenopus model. *Hum Mol Genet* **24**, 3119–3132 (2015).

321. Rothe, M. *et al.* An Epha4/Sipa1l3/Wnt pathway regulates eye development and lens maturation. *Development* **144**, 321–333 (2017).

322. Viet, J. *et al.* MODELING OCULAR LENS DISEASE IN XENOPUS. *Dev Dyn* **249**, 610–621 (2020).

323. De Jong, W. W. Eye Lens Proteins and Vertebrate Phylogeny. in *Macromolecular Sequences in Systematic and Evolutionary Biology* (ed. Goodman, M.) 75–114 (Springer US, 1982). doi:10.1007/978-1-4684-4283-0_3.

324. Litt, M. *et al.* Autosomal Dominant Congenital Cataract Associated with a Missense Mutation in the Human Alpha Crystallin Gene CRYAA. *Human Molecular Genetics* **7**, 471–474 (1998).

325. Brady, J. P. *et al.* Targeted disruption of the mouse αA-crystallin gene induces cataract and cytoplasmic inclusion bodies containing the small heat shock protein αB-crystallin. *Proc Natl Acad Sci U S A* **94**, 884–889 (1997).

326. Xia, C. *et al.* Arginine 54 and Tyrosine 118 residues of {alpha}A-crystallin are crucial for lens formation and transparency. *Invest Ophthalmol Vis Sci* **47**, 3004–3010 (2006).

327. Cartier, M., Breitman, M. L. & Tsui, L.-C. A frameshift mutation in the γE–crystallin gene of the Elo mouse. *Nat Genet* **2**, 42–45 (1992).

328. Boyle, D. L., Takemoto, L., Brady, J. P. & Wawrousek, E. F. Morphological characterization of the Alpha A- and Alpha B-crystallin double knockout mouse lens. *BMC Ophthalmol* **3**, 3 (2003).

329. Graw, J. Cataract mutations and lens development. *Progress in Retinal and Eye Research* **18**, 235–267 (1999).

330. Puk, O., Ahmad, N., Wagner, S., Hrabé de Angelis, M. & Graw, J. First mutation in the βA2-crystallin encoding gene is associated with small lenses and age-related cataracts. *Invest Ophthalmol Vis Sci* **52**, 2571–2576 (2011).

331. Chambers, C. & Russell, P. Deletion mutation in an eye lens beta-crystallin. An animal model for inherited cataracts. *J Biol Chem* **266**, 6742–6746 (1991).

332. Klopp, N. *et al.* Three murine cataract mutants (Cat2) are defective in different gamma-crystallin genes. *Genomics* **52**, 152–158 (1998).

333. Graw, J., Neuhäuser-Klaus, A., Löster, J. & Favor, J. A 6-bp deletion in the Crygc gene leading to a nuclear and radial cataract in the mouse. *Invest Ophthalmol Vis Sci* **43**, 236–240 (2002).

334. Gao, Y. & Spray, D. C. Structural changes in lenses of mice lacking the gap junction protein connexin43. *Invest Ophthalmol Vis Sci* **39**, 1198–1209 (1998).

335. Shiels, A. & Bassnett, S. Mutations in the founder of the MIP gene family underlie cataract development in the mouse. *Nat Genet* **12**, 212–215 (1996).

336. Jun, G. *et al.* EPHA2 is associated with age-related cortical cataract in mice and humans. *PLoS Genet* **5**, e1000584 (2009).

337. Li, Y. *et al.* CP49 and filensin intermediate filaments are essential for formation of cold cataract. *Mol Vis* **26**, 603–612 (2020).

338. Capetanaki, Y., Smith, S. & Heath, J. P. Overexpression of the vimentin gene in transgenic mice inhibits normal lens cell differentiation. *J Cell Biol* **109**, 1653–1664 (1989).

339. Moré, M. I., Kirsch, F. P. & Rathjen, F. G. Targeted ablation of NrCAM or ankyrin-B results in disorganized lens fibers leading to cataract formation. *J Cell Biol* **154**, 187–196 (2001).

340. Hill, R. E. *et al.* Mouse small eye results from mutations in a paired-like homeobox-containing gene. *Nature* **354**, 522–525 (1991).

341. Satoh, K., Takemura, Y., Satoh, M., Ozaki, K. & Kubota, S. Loss of FYCO1 leads to cataract formation. *Sci Rep* **11**, 13771 (2021).

342. Van Agtmael, T. *et al.* Dominant mutations of Col4a1 result in basement membrane defects which lead to anterior segment dysgenesis and glomerulopathy. *Hum Mol Genet* **14**, 3161–3168 (2005).

343. Fujimoto, M. *et al.* HSF4 is required for normal cell growth and differentiation during mouse lens development. *EMBO J* **23**, 4297–4306 (2004).

344. Lassen, N. *et al.* Multiple and additive functions of ALDH3A1 and ALDH1A1: cataract phenotype and ocular oxidative damage in Aldh3a1(-/-)/Aldh1a1(-/-) knock-out mice. *J Biol Chem* **282**, 25668–25676 (2007).

345. Kervestin, S. & Jacobson, A. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol* **13**, 700–712 (2012).

346. Lam, P. T. *et al.* Considerations for the use of Cre recombinase for conditional gene deletion in the mouse lens. *Hum Genomics* **13**, 10 (2019).

347. Sinha, D. *et al.* A spontaneous mutation affects programmed cell death during development of the rat eye. *Experimental Eye Research* **80**, 323–335 (2005).

348. Yoshida, M. *et al.* New genetic model rat for congenital cataracts due to a connexin 46 (Gja3) mutation. *Pathol Int* **55**, 732–737 (2005).

349. West-Mays, J. & Bowman, S. Animal Models of Cataracts. in *Animal Models of Ophthalmic Diseases* (ed. Chan, C.-C.) 11–29 (Springer International Publishing, 2016). doi:10.1007/978-3-319-19434-9_2.

350. Östenson, C.-G. & Efendic, S. Islet gene expression and function in type 2 diabetes; studies in the Goto-Kakizaki rat and humans. *Diabetes, Obesity and Metabolism* **9**, 180–186 (2007).

351. Pfahler, S. *et al.* Prevalence and formation of primary cataracts and persistent hyperplastic tunica vasculosa lentis in the German Pinscher population in Germany. *Vet Ophthalmol* **18**, 135–140 (2015).

352. Oberbauer, A. M., Hollingsworth, S. R., Belanger, J. M., Regan, K. R. & Famula, T. R. Inheritance of cataracts and primary lens luxation in Jack Russell Terriers. *Am J Vet Res* **69**, 222–227 (2008).

353. Mellersh, C. S., Pettitt, L., Forman, O. P., Vaudin, M. & Barnett, K. C. Identification of mutations in HSF4 in dogs of three different breeds with hereditary cataracts. *Veterinary Ophthalmology* **9**, 369–378 (2006).

354. Rudd Garces, G., Christen, M., Loechel, R., Jagannathan, V. & Leeb, T. FYCO1 Frameshift Deletion in Wirehaired Pointing Griffon Dogs with Juvenile Cataract. *Genes (Basel)* **13**, 334 (2022).

355. Liu, J., Hales, A. M., Chamberlain, C. G. & McAvoy, J. W. Induction of cataract-like changes in rat lens epithelial explants by transforming growth factor beta. *Invest Ophthalmol Vis Sci* **35**, 388–401 (1994).

356. Hales, A. M., Schulz, M. W., Chamberlain, C. G. & McAvoy, J. W. TGF-beta 1 induces lens cells to accumulate alpha-smooth muscle actin, a marker for subcapsular cataracts. *Curr Eye Res* **13**, 885–890 (1994).

357. Taliana, L., Evans, M. D. M., Ang, S. & McAvoy, J. W. Vitronectin is present in epithelial cells of the intact lens and promotes epithelial mesenchymal transition in lens epithelial explants. *Mol Vis* **12**, 1233–1242 (2006).

358. O'Connor, M. D. & McAvoy, J. W. In Vitro Generation of Functional Lens-Like Structures with Relevance to Age-Related Nuclear Cataract. *Investigative Ophthalmology & Visual Science* **48**, 1245–1252 (2007).

359. Arita, T., Murata, Y., Lin, L. R., Tsuji, T. & Reddy, V. N. Synthesis of lens capsule in long-term culture of human lens epithelial cells. *Investigative Ophthalmology & Visual Science* **34**, 355–362 (1993).

360. Hirano, M. *et al.* Generation of structures formed by lens and retinal cells differentiating from embryonic stem cells. *Dev Dyn* **228**, 664–671 (2003).

361. Yang, C. *et al.* Efficient generation of lens progenitor cells and lentoid bodies from human embryonic stem cells in chemically defined conditions. *The FASEB Journal* **24**, 3274–3283 (2010).

362. Murphy, P. *et al.* Light-focusing human micro-lenses generated from pluripotent stem cells model lens development and drug-induced cataract in vitro. *Development* **145**, dev155838 (2018).

363. Dewi, C. U. *et al.* A simplified method for producing human lens epithelial cells and light-focusing micro-lenses from pluripotent stem cells. *Exp Eye Res* **202**, 108317 (2021).

364. Hiramatsu, N. *et al.* Formation of three-dimensional cell aggregates expressing lens-specific proteins in various cultures of human iris-derived tissue cells and iPS cells. *Exp Ther Med* **24**, 539 (2022).

365. Fu, Q. *et al.* Generation of Functional Lentoid Bodies From Human Induced Pluripotent Stem Cells Derived From Urinary Cells. *Investigative Ophthalmology & Visual Science* **58**, 517–527 (2017).

366. Ali, M. *et al.* Generation and proteome profiling of PBMC-originated, iPSC-derived lentoid bodies. *Stem Cell Research* **46**, 101813 (2020).

367. Qin, Z. *et al.* Opacification of lentoid bodies derived from human induced pluripotent stem cells is accelerated by hydrogen peroxide and involves protein aggregation. *Journal of Cellular Physiology* **234**, 23750–23762 (2019).

368. Zhang, L. *et al.* Postponement of the opacification of lentoid bodies derived from human induced pluripotent stem cells after lanosterol treatment—the first use of the lens aging model in vitro in cataract drug screening. *Front Pharmacol* **13**, 959978 (2022).

369. Daszynski, D. M. *et al.* Failure of Oxysterols Such as Lanosterol to Restore Lens Clarity from Cataracts. *Sci Rep* **9**, 8459 (2019).

370. Lyu, D. *et al.* Modeling congenital cataract in vitro using patient-specific induced pluripotent stem cells. *NPJ Regen Med* **6**, 60 (2021).

371. Siddam, A. D. *et al.* High-Throughput Transcriptomics of Celf1 Conditional Knockout Lens Identifies Downstream Networks Linked to Cataract Pathology. *Cells* **12**, 1070 (2023).

372. Huppertz, I. *et al.* iCLIP: protein-RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).

373. Curk, T. tomazc/iCount. (2020).

374. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

375. Yeo, G. W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16**, 130–137 (2009).

376. Kwon, T. Benchmarking Transcriptome Quantification Methods for Duplicated Genes in Xenopus laevis. *Cytogenetic and Genome Research* **145**, 253–264 (2015).

377. Noiret, M. *et al.* Robust identification of Ptbp1-dependent splicing events by a junction-centric approach in Xenopus laevis. *Developmental Biology* **426**, 449 (2017).

378. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res* **22**, 2008–2017 (2012).

379. Quinlan, A. bedtools - the swiss army knife for genome arithmetic. (2023).

380. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

381. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70–82 (2021).

382. Srivastava, R., Budak, G., Dash, S., Lachke, S. A. & Janga, S. C. Transcriptome analysis of developing lens reveals abundance of novel transcripts and extensive splicing alterations. *Sci Rep* **7**, 11572 (2017).

383. Buljan, M. *et al.* Alternative splicing of intrinsically disordered regions and rewiring of protein interactions. *Current Opinion in Structural Biology* **23**, 443–450 (2013).

384. Morrow, J. S. *et al.* Of Membrane Stability and Mosaics: The Spectrin Cytoskeleton. in *Comprehensive Physiology* 485–540 (John Wiley & Sons, Ltd, 2011). doi:10.1002/cphy.cp140111.

385. Kaminska, J. The PH-like Domain of VPS13 Proteins - a Determinant of Localization to the Golgi Apparatus or to the Plasma Membrane. *Contact (Thousand Oaks)* **5**, 25152564221106024 (2022).

386. Powis, G., Meuillet, E. J., Indarte, M., Booher, G. & Kirkpatrick, L. Pleckstrin Homology [PH] domain, structure, mechanism, and contribution to human disease. *Biomedicine & Pharmacotherapy* **165**, 115024 (2023).

387. Redlingshöfer, L. *et al.* Clathrin light chain diversity regulates membrane deformation in vitro and synaptic vesicle formation in vivo. *Proc Natl Acad Sci U S A* **117**, 23527–23538 (2020).

388. Giudice, J. *et al.* Alternative splicing regulates vesicular trafficking genes in cardiomyocytes during postnatal heart development. *Nat Commun* **5**, 3603 (2014).

389. Buljan, M. *et al.* Tissue-Specific Splicing of Disordered Segments that Embed Binding Motifs Rewires Protein Interaction Networks. *Molecular Cell* **46**, 871–883 (2012).

390. Wang, Z., Han, J., David, L. L. & Schey, K. L. Proteomics and phosphoproteomics analysis of human lens fiber cell membranes. *Investigative Ophthalmology and Visual Science* **54**, 1135–1143 (2013).

COLLEGE SCIENCES
DOCTORAL DE LA VIE
BRETAGNE ET DE LA SANTE

Université de Rennes

Contrôle post-transcriptionnel de l'expression génétique par CELF1 dans le développement et les pathologies du cristallin.

Mots clés : Cristallin, Cataracte, Celf1, Épissage alternatif, iCLIP-seq, Organoïde

Résumé :

Le cristallin permet à la lumière de se focaliser sur la rétine. Son opacification due à l'âge, à des facteurs environnementaux ou à certaines variations génétiques provoque une cataracte pouvant conduire à une cécité. Le contrôle post-transcriptionnel exercé par la protéine de liaison à l'ARN CELF1 est critique pendant le développement du cristallin. Au cours de ma thèse, j'ai identifié les cibles ARN de CELF1 dans le cristallin afin de comprendre les régulations post-transcriptionnelles qu'elle exerce et les causes de la cataracte observée en absence de cette protéine.

J'ai d'abord mené une analyse transcriptomique par RNA-seq sur des cristallins de souris nouveau-nées pour décrire à l'échelle globale les perturbations transcriptomiques qui se produisent dans le cristallin déficient en CELF1.

J'ai ensuite recherché les ARN dont l'épissage alternatif est régulé par CELF1 dans le cristallin. Dans ce but, j'ai intégré différentes approches omiques: profilage d'expression par RNA-seq, et identification des sites de liaison de CELF1 sur ses ARN ligands par iCLIP-seq. J'ai ainsi notamment démontré que CELF1 contrôle l'épissage alternatif de sept ARNm codant des protéines associées au cytosquelette: *Ablim1*, *Ctnna2*, *Clta*, *Ywhae*, *Septin8*, *Sptbn1* et *Ank2*.

Enfin, j'ai caractérisé un nouveau modèle d'organoïde de cristallin qui peut reproduire certains processus du développement du cristallin. Ce modèle pourrait être un outil précieux pour la communauté de recherche sur le cristallin.

Post-transcriptional control of gene expression by CELF1 in lens development and pathology

Keywords : Lens, Cataract, Celf1, Alternative splicing, iCLIP-seq, Organoid

Abstract:

The ocular lens allows light to focus on the retina. Its oppacification due to age, environmental factors or genetic risks causes cataract that can lead to blindness. The post-transcriptional controle exerted by the RNA-binding protein CELF1 is critical during lens development. During my thesis, I identified the RNA targets of CELF1 in the lens to uncover the post-transcriptional regulations it exerts and the causes of the cataract observed in the absence of this protein.

I first carried out an RNA-seq transcriptomic analysis of newborn mouse lenses to describe on a global scale the transcriptomic perturbations that occur in *Celf1*-deficient lenses.

I then looked for RNAs whose alternative splicing is regulated by CELF1. To this end, I integrated different omics approaches: expression profiling by RNA-seq and identification of CELF1 binding sites on its RNA ligands by iCLIP-seq. Notably, I demonstrated that CELF1 controls the alternative splicing of seven mRNAs encoding cytoskeleton-associated proteins: *Ablim1, Ctnna2, Clta, Ywhae, Septin8, Spbn1* and *Ank2*.

Finally, I characterized a new lens organoid model that can mimick some process of lens development. This model could be a valuable tool for the lens research community