Original Research

Title: Boolean Implication Analysis Improves Prediction Accuracy of In Silico Gene Reporting of Retinal Cell Types

Authors: Rohan Subramanian[1] and Debashis Sahoo[2,3]

Affiliations:

[1]*The International School Bangalore, Bengaluru, Karnataka, India.*

[2]*Department of Pediatrics, University of California San Diego, La Jolla, CA, USA.*

[3]*Department of Computer Science and Engineering, Jacobs School of Engineering, University of California San Diego, La Jolla, CA, USA.*

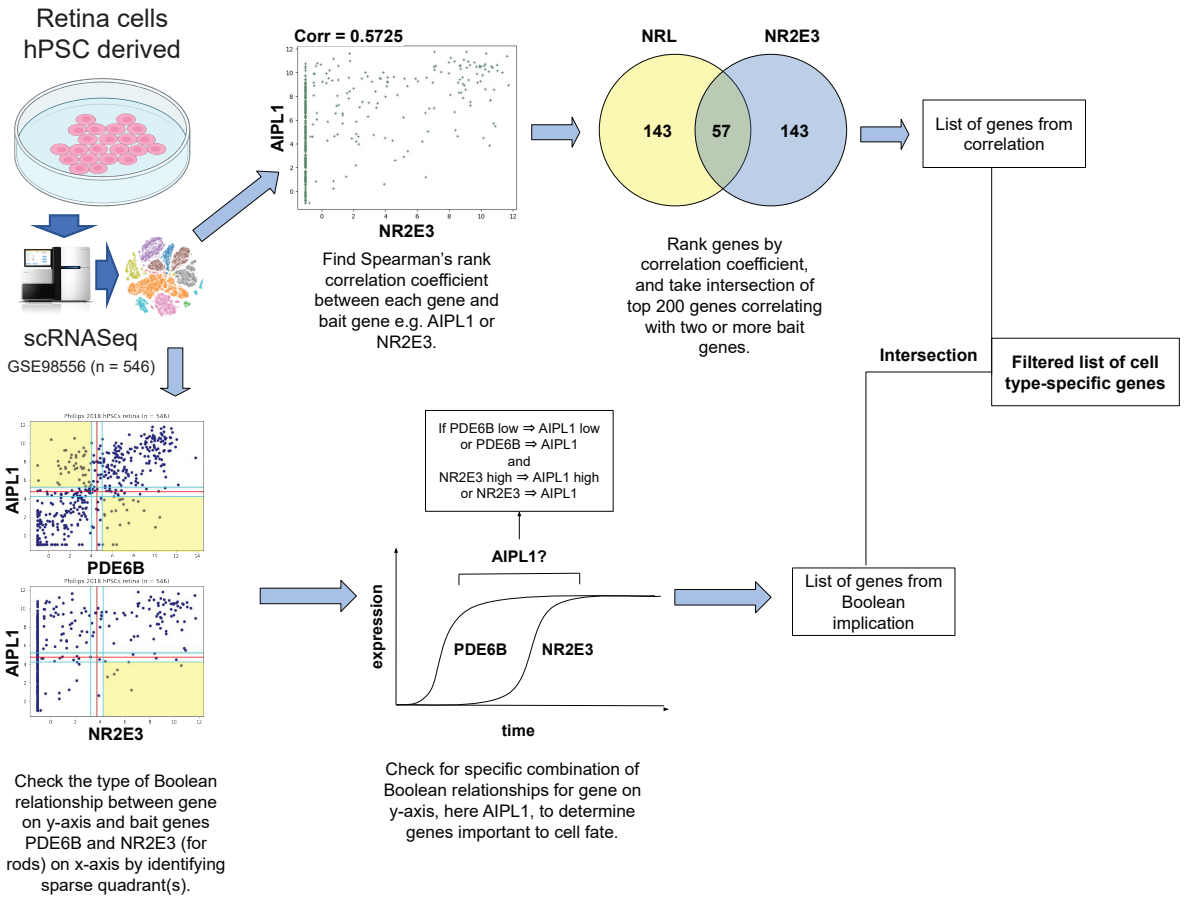
Corresponding Author:

Debashis Sahoo, Ph.D.; Assistant Professor, Department of Pediatrics, University of California San Diego; 9500 Gilman Drive, MC 0730, Leichtag Building 132; La Jolla, CA 92093-0831. **Phone:** 858-246-1803: **Fax:** 858-246-0019: **Email:** dsahoo@ucsd.edu

Author contributions:

Role	Authors
Conceptualization; Funding acquisition; Supervision; Project administration	D.S.
Investigation; Methodology; Data curation, Formal Analysis • Computational modeling • Data curation and analysis	R.S., D.S.
Software	R.S., D.S.
Visualization; Writing – original draft; Writing – review & editing	R.S., D.S.

Keywords: Retina, Single-cell RNA sequencing, Pluripotent stem cells, Boolean analysis, Bioinformatics


Retina cells
hPSC derived
scRNASeq
GSE98556 (n = 546)
Corr = 0.5725
AIPL1
NR2E3
Find Spearman's rank correlation coefficient between each gene and bait gene e.g. AIPL1 or NR2E3.
NRL
NR2E3
143
57
143
Rank genes by correlation coefficient, and take intersection of top 200 genes correlating with two or more bait genes.
List of genes from correlation
Intersection
Filtered list of cell type-specific genes
AIPL1
PDE6B
AIPL1
NR2E3
Check the type of Boolean relationship between gene on y-axis and bait genes PDE6B and NR2E3 (for rods) on x-axis by identifying sparse quadrant(s).
If PDE6B low ⇒ AIPL1 low
or PDE6B ⇒ AIPL1
and
NR2E3 high ⇒ AIPL1 high
or NR2E3 ⇒ AIPL1
AIPL1?
expression
PDE6B
NR2E3
time
Check for specific combination of Boolean relationships for gene on y-axis, here AIPL1, to determine genes important to cell fate.
List of genes from Boolean implication

Abstract

The retina is a complex tissue containing multiple cell types that is essential for vision. Understanding the gene expression patterns of various retinal cell types has potential applications in regenerative medicine. Retinal organoids (optic vesicles) derived from pluripotent stem cells have begun to yield insights into the transcriptomics of developing retinal cell types in humans through single cell RNA-sequencing studies. Previous methods of gene reporting have relied upon techniques in vivo using microarray data, or correlational and dimension reduction methods for analyzing single cell RNA-sequencing data in silico. Here, we present a bioinformatic approach using Boolean implication to discover retinal cell type-specific genes. We apply this approach to previously published retina and retinal organoid datasets and improve upon previously published correlational methods. Our method improves the prediction accuracy and reproducibility of marker genes of retinal cell types and discovers several new high confidence cone and rod-specific genes. Furthermore, our method is general and can impact all areas of gene expression analyses in cancer and other human diseases.

Significance Statement

Efforts to derive retinal cell types from pluripotent stem cells to the end of curing retinal disease require robust characterization of these cell types' gene expression patterns. The Boolean method described in this study improves prediction accuracy of earlier methods of gene reporting, and allows for the discovery and validation of retinal cell type-specific marker genes. The invariant nature of results from Boolean implication analysis can yield high-value molecular markers that can be used as biomarkers or drug targets.

Introduction

Characterization of retinal cell types is an important field of study with wide applications in ophthalmology and regenerative medicine. With the advent of single cell RNA-sequencing (scRNA-seq), methods for gene reporting in silico can yield valuable insights into genes that are important in determining cell fate.[1] Human pluripotent stem cells (hPSCs) can be used to generate retinal cell types in vitro with potential applications to cure age-related macular degeneration, retinitis pigmentosa and other retina-related causes of blindness. However, gene reporting and characterization of these cell types is difficult as they differentiate asynchronously in complex cultures.[2] Furthermore, there is a lack of human datasets. We propose using Boolean implication analysis to improve the prediction accuracy of existing correlational methods for in silico gene reporting.

Previous Methods In Vivo and In Vitro

One of the most common methods to study the effect of key genes on retinal development is the use of genetically modified "knockout" murine models, which are frequently used to validate differentially expressed genes from microarray data.[3-20] Fluorescent gene reporter lines are widely used to check for gene expression in single cells, or purified populations of a single cell type.[2, 21-25] Bulk RNA sequencing (RNA-seq) has helped define the transcriptomes of larger populations of retinal cell types.[3, 9, 14, 17, 21, 24, 26-35] To study the characteristics of isolated cells or droplets, flow cytometry was formerly a major method.[36, 37] Single-cell RNA sequencing (scRNA-seq) is increasingly common today and is one the most detailed methods to profile transcriptomes of retinal cell types and subtypes.[2, 8, 13, 22, 38-48]

Most studies on retinal cell types have relied upon murine models, but many increasingly study human donor retinas[6, 30, 31, 48-50], especially in order to profile retinal disease.[31, 43, 50-53] Glaucoma, age-related macular dystrophy and retinal light damage have also been studied in murine models.[7, 14, 29, 34, 35, 54, 55] Some studies have grown cell lines in vitro from fetal retina[49, 56], whereas other have used human pluripotent, induced pluripotent or embryonic stem cells to generate purified cell populations or retinal organoids.[2, 3, 8, 28, 38, 57-59] In order to study the development of retinal cell types over time, the lineage of stem cell progeny[58] and time course data from different time points (using PCR and RNA-seq) have been investigated.[39, 41, 54]

Previous Methods In Silico

Differential expression analysis is the most common method to identify retinal cell type-specific genes and biomarkers from microarray, RNA-seq and scRNA-seq data.[10, 13, 14, 17, 24, 29-31, 39, 41, 46, 47, 53, 56, 59] In single-cell analysis, dimension reduction through Principal Component Analysis to reduce the size of data and allow visualization is often performed before hierarchical clustering identify cell clusters.[2, 7, 30, 41, 42, 49, 56, 60] Cell clusters can be assigned to different cell types or subtypes based on the expression of key marker genes.[48] AI-guided identification of cell clusters has recently been investigated.[61]

scRNA-seq data provides opportunities for in depth analysis of the transcriptome of individual cells, and subsequent characterization of cell types, subtypes and regions of retina. However, scRNA-seq data is highly noisy, and contains large numbers of zeroes, among which true and false negatives are indistinguishable. Many of these zeroes are dropouts, caused by a failure to capture or amplify a transcript.[62]

Most studies up to date have been highly dependent on cell clustering, which is not always achievable, especially in datasets containing immature or developing cells.[1] Pseudo-time analysis, which aims to sort cells by their developmental stage, has been applied to retinal organoids, and takes into account transitory states rather than discrete clusters.[38] However, this approach is hindered by asynchronous differentiation of cell types in retina.[63] Correlational methods for ranking gene expression are also widely used, bypassing the need to discover cell clusters and identifying co-expressed genes in complex cultures, including developing retinal organoids.[2, 8, 23, 27, 49, 64]

Identifying relationships between genes has led towards broader goals of graph[47, 60] and network-based analysis.[9, 10, 17, 25, 27, 31, 60, 65] Gene expression networks can be used to identify transitions between phenotypes and disease states, paving the way for clinical target identification. Correlational analysis is traditionally used to derive co-expression networks, and knockout murine models are used to directly investigate the effect of one gene's absence of others. However, the symmetric nature of correlation can lead to loss of valuable information and does not provide insight into the expression of genes over time. Bayesian networks of gene regulation and expression in the retina mainly identify transcription factors and their targets.[60, 66] Hence, the motivation of our work was to develop a universally applicable state of the art method that filtered out noise, could be applied to a wide variety of datasets and lent insight into gene expression over differentiation.

A Boolean Approach

Boolean logic is a simple mathematical relationship between two values such as high/low or 1/0. We propose using Boolean implication ("if-then" relationships) to study the dependency between genes from scRNA-seq data. Research by Sahoo et al. has shown that analysis of Boolean implication relationships is better at filtering out noise than correlational approach.[67] Analysis of Boolean implication lends insight into asymmetric relationships disregarded by correlation.

While Boolean implication, like correlation, does not imply causation, asymmetric Boolean relationships can be thought of in terms of subsets. For example, the relationship Gene A high ⇒ Gene B high indicates that all cells with Gene B high are a subset of those with Gene A high. This allows for analysis of developmentally regulated genes using Boolean implication, first pioneered in the MiDReG tool published by Sahoo et al 2010.[68]

In previous research, Boolean methods have led to the discovery of prognostic biomarkers for bladder and colon cancer.[69-71] These methods have also led to characterization of hematopoietic stem cells and identification of B and T cell precursors.[72, 73] Our methods have not previously been applied to stem cell-derived retinal cell types, but have yielded insights into changes in transcriptional profiles of healthy retina and retinoblastoma.[74]

The StepMiner and BooleanNet algorithms were developed for microarray data by Sahoo et al. 2008 to identify Boolean implication relationships between genes, but have since been applied to a wide variety of high-throughput data, such as RNA-seq, and scRNA-seq.[68, 75, 76]

Methods

Data Normalization and Annotation

We applied $\log_2(v+1)$ transformation to TPM values from the 546 sequenced cells of the Phillips 2018 dataset (GSE98556, n = 546), as the log transformed RNA-seq data is closer to a normal distribution. Cells were annotated with clinical characteristics, and data were uploaded to Hegemon. In the Hegemon online tool, scatter plots between genes are generated, with each point representing the expression level of the genes in a single cell.[67-71]

Discovering Boolean Implications

StepMiner Algorithm

The StepMiner algorithm identifies thresholds to convert continuous expression values into discrete values by fitting a step function to sorted values. A step can be defined as the sharpest increase in sorted gene expression values over an interval. Having identified a threshold t, gene expression values greater than $t + 0.5$ are considered high, and those below $t - 0.5$ are considered low. Those between $t + 0.5$ and $t - 0.5$ are considered intermediate. These thresholds are used to divide the plot into four quadrants.[67, 77]

BooleanNet Algorithm

The BooleanNet algorithm identifies the type of Boolean implication relationship by identifying the sparse quadrant(s) using a statistic S and likelihood error rate p. There are six types of Boolean implication relationship: high ⇒ high, low ⇒ low, high ⇒ low, low ⇒ high, equivalent

and opposite. The first four are asymmetric and have only one sparse quadrant. The latter two are symmetric and have two sparse quadrants. Further information can be found in **Fig. S1**.[67, 77]

Thresholds for Analysis

Thresholds for S and p are applied to adjust the sensitivity of the analysis. While $S > 3$ and $p < 0.1$ are generally considered for microarray data, we decreased the threshold for S to 2.5 and increased the threshold for p to 0.25 for rods and 0.35 for all other analysis to account for the larger amount of noise in scRNA-seq data.

Boolean Approach to In Silico Gene Reporting of Retinal Cell Types

We propose using the method described in **Fig. 1** for in silico gene reporting of retinal cell types. We require two or more known genes for each cell type called "bait genes". We searched for genes which had a low ⇒ low or equivalent Boolean relationship with the first bait gene and high ⇒ high or equivalent Boolean relationship with the second bait gene.

This specific combination of Boolean relationships is akin to searching for genes which have an impact on cell fate. If a gene passes this analysis, the set of cells where Gene X is low is a subset of the cells where the first bait gene is low, and the set of cells where Gene X is high is a subset of the cells where the second bait gene is high. This method can allow us to infer genes which are expressed after the first bait gene, and before the second bait gene. Hence, the choice of bait genes plays an important role in determining the results. We chose bait genes which led to shorter gene lists compared to SRCCA, with a greater number of known markers of five retinal cell types. These were selected and verified from previous literature on rod and cone

photoreceptors[6, 78], retinal progenitor cells (RPCs)[79, 80], retinal ganglion cells (RGCs)[23, 24] and retinal pigment epithelium (RPE)[81-84].

More than two bait genes can be considered by searching for high ⇒ high, low ⇒ low or equivalent Boolean relationships in two out of three bait genes instead of one out of two. This allows for combination of multiple cell type-specific marker genes in the analysis.

Spearman's Rank Correlation Coefficient

Spearman's rank correlation coefficient (SRCC) is a nonparametric measure of the association between two ranked variables. We reviewed and reproduced the approach of Phillips et al. 2018, called Spearman's rank correlation coefficient analysis (SRCCA). The correlation coefficient between bait genes and all other genes are found and ranked. Then, the intersection between the top 200 correlating genes with each bait gene is taken.[2]

We combined both methods by taking the interaction of gene lists derived from both methods, hence filtering the list of correlating genes using Boolean implication as shown in **Fig. 1**. All analysis was performed using the Hegemon website, in Python 3 using the HegemonUtil tools and in R version 4.0.1.

Quantification of Results

Results were independently validated through differential expression. We evaluated whether genes were differentially expressed between rods and cones, and between photoreceptors and non-photoreceptor retinal cell types.

We selected and processed several validation datasets. Two were bulk RNA-seq datasets containing purified retinal cell types from Mus musculus: Hartl 2017 (GSE84589, n = 14) and Sarin 2018 (GSE98838, n = 22).[40, 46] The third was a similar human retina scRNA-seq dataset, Voigt 2020 (GSE130636 and GSE142449, n = 20,797).[48]

Using validation datasets with purified cell types, we checked for differential expression between retinal cell types by performing a one-tailed Welch's t-test between the groups of cells to determine whether there was a statistically significant difference between the means of the two groups. Using this method, we could evaluate the proportion of genes which were specific to the cell type in question, expressed equally throughout the retina, and expressed in a different, non-target cell type.

To evaluate the reproducibility of the genes, we directly repeated the analysis in GSE130636 and GSE142449 using common bait genes for SRCCA and Boolean implication. We compared the proportion of genes discovered by using a two proportion Z-test.

Results

Boolean Implication Enables Identification of Cell Type Specific Genes like SRCCA

Boolean Implication analysis explores both symmetric and asymmetric relationship between genes whereas SRCCA only focuses on symmetric relationships. We hypothesize that

application of asymmetric Boolean implication relationships may improve the accuracy of cell types specific genes identification (**Fig 1**).

Application of Boolean implication analysis led to shorter lists of genes compared to SRCCA **(Fig. 2)**. Selecting bait genes is crucial for both SRCCA and Boolean analysis. For Boolean analysis, a general marker and a more specific marker are ideal candidates. However, SRCCA relies only on specific bait genes. Because of these differences in specificity, we chose different set of bait genes for Boolean analysis from known marker genes for each retinal cell type.

Application of Boolean analysis for gene reporting of photoreceptors led to longer lists of genes than other cell types. The largest intersection between SRCCA and Boolean implication was observed in rod photoreceptors. The number of genes from Boolean implication in other retinal cell types such as RGCs, RPCs and RPE was far lower than photoreceptors.

For RPE, three bait genes were chosen due to the excessively small number of genes obtained from two bait genes. This is likely to be due to the smaller number of cells from these types present in the retina, compared to photoreceptors. The complete absence of intersection between genes from SRCCA and Boolean in RPE could also be explained by the very small number of RPE cells present in optic vesicle cultures produced by the method used by Phillips et al. 2018.

Filtering SRCCA using Boolean Implication Improves Prediction Accuracy

We independently validated the genes from SRCCA and Boolean implication using bulk RNA-seq datasets with purified retinal cell types. In **Fig. 3B**, there is a visible improvement in

proportion of rod-specific genes while taking the intersection of SRCCA and Boolean implication. Similarly, the majority of SRCCA genes not present in Boolean implication were not specific to rods, or specific to cones. We were able to show a statistically significant improvement in the proportion of rod PR-specific genes by filtering correlating genes using Boolean implication. The proportion of genes rod-specific genes from SRCCA, 29 out of 56 (0.517), was improved to 16 out of 19 (0.842) by filtering using Boolean implication. This proportion was shown to be statistically significant by performing a two-proportion Z-test, returning a p-value of 0.013.

Similarly, as shown in **Fig. 3C**, we were able to show a statistically significant improvement in photoreceptor-specificity of the rod genes using the combined correlational and Boolean approach. All 19 genes obtained by filtering SRCCA using Boolean implication were photoreceptor-specific, and the p-value from the two-proportion Z-test was 0.016.

As seen in **Fig. 3D**, prediction accuracy of both SRCCA and Boolean analysis was lower in cone photoreceptors. The proportion of cone-specific genes, 15 out of 30 (0.500), was still highest in Boolean implication. Here, the prediction accuracy of Boolean methods alone was not improved by taking the intersection with SRCCA. However, this result could not be shown to be statistically significant due to the larger number of total genes in SRCCA. Hence, we sought an additional method of evaluation for cone photoreceptors with better scope for comparison.

Filtering SRCCA using Boolean Implication Improves Reproducibility

We also performed SRCCA and Boolean implication analysis for cone photoreceptors using the same three bait genes: CRX, GNB3 and GNAT2. (**Fig. S2**) One major limitation of existing literature on characterization of retinal cell types is the lack of reproducibility of purported marker genes across datasets. Hence, we directly repeated the analysis in GSE130636 and GSE142449, which is also scRNA-seq of human retina. The main difference between this dataset and the Phillips dataset is the larger size (20,797 vs. 546 cells), and the larger proportion of adult, tissue-derived cells.

Fig. 3E displays the results from this method of quantification. Combining correlational and Boolean implication using common bait genes yielded highly reproducible results, as all 7 genes were reproduced. The two-proportion Z-test also returned a statistically significant p-value of 0.00082.

Boolean Implication Improves Prediction Accuracy of Novel High Confidence Genes

Considering the overall improvement in prediction accuracy through Boolean implication analysis, we also investigated several specific examples of new discoveries through this method.

Novel high confidence genes are an important contribution of gene reporting methods in silico. Identification of high confidence markers of retinal cell types using SRCCA alone may be arbitrary, but we show that Boolean implication can lend greater insight into the cell type-specific genes.

Boolean implication analysis identified WWC1 (WW domain containing protein-1) as a novel high confidence cone photoreceptor gene. This was validated independently in GSE84589 and GSE98838, with statistically significant overexpression in cone photoreceptors. (**Fig. 3A-B**) WWC1 has been described to have a broad function in the brain and memory by previous studies.[85, 86]

Boolean implication analysis of rods also identified two novel rod-specific genes: CASZ1 (Castor zinc finger 1) (**Fig. 3E-F**) and PPEF2 (Protein Phosphatase with EF-Hand Domain 2) (**Fig. 3G-H**). These showed rod specificity in both validation datasets. CASZ1 is known to play a role in cell differentiation, and may hence play a significant role in influencing rod cell fate. [87] PPEF2 has been documented in rods before, but has had several conflicting studies on its importance in rods.[88, 89] This is the first documentation of its rod-specific function in human or hPSC-derived retina. Boolean implication analysis has shed light on potential novel markers of cone and rod photoreceptors.

Boolean implication analysis refuted AKAP9 (A-kinase anchoring protein-9), identified to be a high confidence cone photoreceptor gene by Phillips et al. 2018 based on the results from SRCCA. **Fig. 3C-D** show that it is not differentially expressed in cones, and may be more rod-specific as per GSE84589.

Discussion

Boolean methods improved upon correlational methods by filtering out noise and identifying asymmetric relationships that lend insight into the specificity of genes. Filtering correlating

genes led to a statistically significant improvement in rod and photoreceptor-specificity for rod genes, and reproducibility for cone photoreceptor genes. Hence, we have shown that a combination of Boolean implication analysis and SRCCA improves the prediction accuracy of in silico gene reporting of retinal cell types.

Boolean implication analysis provided more accurate insight into high confidence genes, and led to the identification of WWC1 as a novel marker gene for cone photoreceptors. Previous attempts to identify high confidence genes from extensive gene lists obtained through SRCCA alone have no way to distinguish between noise and true cell type-specific genes. The asymmetric nature of Boolean relationships allows us to determine whether a gene is expressed more generally or specifically, which is not present in correlation.

Another advantage of Boolean implication is that the analysis can always be performed over the entire dataset. Boolean implication relationships between genes are best visible when there is a greater diversity of cell types, including those not expressing the gene. However, SRCCA generally requires the operator to choose a specific subset of the data (e.g. day 70) on which to perform the analysis, based on whether the cell type in question is present at that developmental stage. This choice has a significant effect on the result of SRCCA, and an inept choice of the subset may lead to false associations not generalizable over larger datasets. This issue can be solved using Boolean implication.

However, Boolean implication analysis was also not entirely free from error. The main source of error appears to be the dropouts, which lead to a greater density of points in quadrants a_{10}, a_{00}

and a_{01} in many cases. This, along with the slightly relaxed thresholds adapted for scRNA-seq, led to false discoveries of Boolean relationships. This issue likely reduced the improvement in quality of analysis in cone photoreceptors. Even so, a combination of correlational and Boolean implication analysis could lead to completely error-free results in some cases. (**Fig. 3C**)

The method of independent validation considered several datasets to evaluate specificity and reproducibility. These high-quality datasets provided reliable results for most genes, as human and mouse retina are very similar. However, it was not infallible due to small variations between the species. We compared the results in mouse datasets with the Kim 2019 dataset (GSE119343, n = 1346) containing cone-enriched optic vesicles. There, we found small differences in expression patterns in the mouse vs. human retina, such as CERKL, a gene specific to human cone photoreceptors, but expressed in both cone and rod photoreceptors in mice.

There were differences in the performance of our methods between different cell types. In cell types present in smaller numbers in the retina, we can observe that the number of genes from Boolean analysis alone and combined with SRCCA is also smaller. The analysis performed best in rods, the most numerous neural retina cell type.[90] In RPE, which is rarely present in the optic vesicle culture protocol employed by Phillips et al. 2018, there was no intersection between Boolean analysis and SRCCA, indicating that the results in that case are likely to be mainly noise. However, there is no link between the number of genes obtained from SRCCA and the population of cell type, as a result of always considering a fixed number of top correlating genes. Hence, Boolean analysis can lend insight into the cell types for which the data is comprehensive enough to provide accurate resolution.

Boolean implication analysis provides all the advantages offered previously by SRCCA including efficiency, ability to combine multiple bait genes, and improved prediction accuracy compared to earlier methods. Our method can allow researchers to analyze single-cell data even when cell clusters cannot be identified, a common issue in datasets containing developing cells. Combining both methods provides statistically significant improvements in specificity and reproducibility of genes. Boolean implication can be easily inferred from scatter plots on the Hegemon online tool, making it an intuitive option for biologists and computer scientists alike.[76]

Conclusion

In this work, we have developed a novel approach for analysis of scRNA-seq data based on Boolean implication. We have shown a statistically significant improvement in the prediction accuracy and reproducibility of retinal cell-type specific genes, as compared to earlier approaches based solely on correlation. Application of our method to retinal organoid datasets identified novel high confidence cell type-specific genes such as WWC1 for cones and CASZ1 and PPEF2 for rods. This Boolean approach allows for analysis and characterization of cell types in complex cultures, even when cell clustering cannot be achieved. Considering asymmetric relationships has allowed us to effectively filter out noise, lending insight into genes with potential importance in regenerative medicine.

Acknowledgements

This work was supported by the National Institutes for Health (NIH) grants R00-CA151673, R01-GM138385, UG3 TR003355, R01-AI155696 (to DS), UCOP-RGPO (R00RG2628 &

R00RG2642 to DS), The Sanford Stem Cell Clinical Center at UCSD (to DS), Padres Pedal the Cause / Rady Children's Hospital Translational PEDIATRIC Cancer Research Award (Padres Pedal the Cause/RADY #PTC2017) to DS, 2017, Padres Pedal the Cause /C3 Collaborative Translational Cancer Research Award (San Diego NCI Cancer Centers Council (C3) #PTC2017) to DS.

Competing interests: The authors declare no competing interests.

Data and materials availability

All data is available in public repository and the relevant accession number are provided in the text and the supplementary materials.

References

1. Zerti, D., et al., *Understanding the complexity of retina and pluripotent stem cell derived retinal organoids with single cell RNA sequencing: current progress, remaining challenges and future prospective.* Curr Eye Res, 2020. **45**(3): p. 385-396.
2. Phillips, M.J., et al., *A Novel Approach to Single Cell RNA-Sequence Analysis Facilitates In Silico Gene Reporting of Human Pluripotent Stem Cell-Derived Retinal Cell Types.* Stem Cells, 2018. **36**(3): p. 313-324.
3. Brooks, M.J., et al., *Improved Retinal Organoid Differentiation by Modulating Signaling Pathways Revealed by Comparative Transcriptome Analyses with Development In Vivo.* Stem Cell Reports, 2019. **13**(5): p. 891-905.
4. Brooks, M.J., et al., *Next-generation sequencing facilitates quantitative analysis of wild-type and Nrl(-/-) retinal transcriptomes.* Mol Vis, 2011. **17**: p. 3034-54.
5. Cheng, H., et al., *In vivo function of the orphan nuclear receptor NR2E3 in establishing photoreceptor identity during mammalian retinal development.* Hum Mol Genet, 2006. **15**(17): p. 2588-602.
6. Corbo, J.C., et al., *A typology of photoreceptor gene expression patterns in the mouse.* Proc Natl Acad Sci U S A, 2007. **104**(29): p. 12069-74.
7. Howell, G.R., et al., *Molecular clustering identifies complement and endothelin induction as early events in a mouse model of glaucoma.* J Clin Invest, 2011. **121**(4): p. 1429-44.
8. Kallman, A., et al., *Investigating cone photoreceptor development using patient-derived NRL null retinal organoids.* Commun Biol, 2020. **3**(1): p. 82.

9. Kim, J.W., et al., *NRL-Regulated Transcriptome Dynamics of Developing Rod Photoreceptors.* Cell Rep, 2016. **17**(9): p. 2460-2473.
10. Ma, H., et al., *Loss of cone cyclic nucleotide-gated channel leads to alterations in light response modulating system and cellular stress response pathways: a gene expression profiling study.* Hum Mol Genet, 2013. **22**(19): p. 3906-19.
11. Mizeracka, K., C.R. DeMaso, and C.L. Cepko, *Notch1 is required in newly postmitotic cells to inhibit the rod photoreceptor fate.* Development, 2013. **140**(15): p. 3188-97.
12. Montana, C.L., et al., *Reprogramming of adult rod photoreceptors prevents retinal degeneration.* Proc Natl Acad Sci U S A, 2013. **110**(5): p. 1732-7.
13. Mustafi, D., et al., *Transcriptome analysis reveals rod/cone photoreceptor specific signatures across mammalian retinas.* Hum Mol Genet, 2016. **25**(20): p. 4376-4388.
14. Mustafi, D., et al., *Defective photoreceptor phagocytosis in a mouse model of enhanced S-cone syndrome causes progressive retinal degeneration.* FASEB J, 2011. **25**(9): p. 3157-76.
15. Onishi, A., et al., *The orphan nuclear hormone receptor ERRbeta controls rod photoreceptor survival.* Proc Natl Acad Sci U S A, 2010. **107**(25): p. 11579-84.
16. Palczewska, G., et al., *Receptor MER Tyrosine Kinase Proto-oncogene (MERTK) Is Not Required for Transfer of Bis-retinoids to the Retinal Pigmented Epithelium.* J Biol Chem, 2016. **291**(52): p. 26937-26949.
17. Perez-Cervantes, C., et al., *Enhancer transcription identifies cis-regulatory elements for photoreceptor cell types.* Development, 2020. **147**(3).
18. Roger, J.E., et al., *Preservation of cone photoreceptors after a rapid yet transient degeneration and remodeling in cone-only Nrl-/- mouse retina.* J Neurosci, 2012. **32**(2): p. 528-41.
19. Sundermeier, T.R., et al., *DICER1 is essential for survival of postmitotic rod photoreceptor cells in mice.* FASEB J, 2014. **28**(8): p. 3780-91.
20. Yoshida, S., et al., *Expression profiling of the developing and mature Nrl-/- mouse retina: identification of retinal disease candidates and transcriptional regulatory targets of Nrl.* Hum Mol Genet, 2004. **13**(14): p. 1487-503.
21. Buenaventura, D.F., A. Corseri, and M.M. Emerson, *Identification of Genes With Enriched Expression in Early Developing Mouse Cone Photoreceptors.* Invest Ophthalmol Vis Sci, 2019. **60**(8): p. 2787-2799.
22. Cherry, T.J., et al., *Development and diversification of retinal amacrine interneurons at single cell resolution.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9495-500.
23. Langer, K.B., et al., *Retinal Ganglion Cell Diversity and Subtype Specification from Human Pluripotent Stem Cells.* Stem Cell Reports, 2018. **10**(4): p. 1282-1293.
24. Sajgo, S., et al., *Molecular codes for cell type specification in Brn3 retinal ganglion cells.* Proc Natl Acad Sci U S A, 2017. **114**(20): p. E3974-E3983.
25. Siegert, S., et al., *Transcriptional code and disease map for adult retinal cell types.* Nat Neurosci, 2012. **15**(3): p. 487-95, S1-2.
26. Cherry, T.J., et al., *Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease.* Proc Natl Acad Sci U S A, 2020. **117**(16): p. 9001-9012.
27. Dorrell, M.I., et al., *Global gene expression analysis of the developing postnatal mouse retina.* Invest Ophthalmol Vis Sci, 2004. **45**(3): p. 1009-19.

28. Gill, K.P., et al., *Enriched retinal ganglion cells derived from human embryonic stem cells.* Sci Rep, 2016. **6**: p. 30552.
29. Harder, J.M., et al., *Jnk2 deficiency increases the rate of glaucomatous neurodegeneration in ocular hypertensive DBA/2J mice.* Cell Death Dis, 2018. **9**(6): p. 705.
30. Li, M., et al., *Comprehensive analysis of gene expression in human retina and supporting tissues.* Hum Mol Genet, 2014. **23**(15): p. 4001-14.
31. Newman, A.M., et al., *Systems-level analysis of age-related macular degeneration reveals global biomarkers and phenotype-specific functional networks.* Genome Med, 2012. **4**(2): p. 16.
32. Ratnapriya, R., et al., *Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration.* Nat Genet, 2019. **51**(4): p. 606-610.
33. Sugino, K., et al., *Mapping the transcriptional diversity of genetically and anatomically defined cell populations in the mouse brain.* Elife, 2019. **8**.
34. Williams, P.A., et al., *Nicotinamide and WLD(S) Act Together to Prevent Neurodegeneration in Glaucoma.* Front Neurosci, 2017. **11**: p. 232.
35. Williams, P.A., et al., *Vitamin B3 modulates mitochondrial vulnerability and prevents glaucoma in aged mice.* Science, 2017. **355**(6326): p. 756-760.
36. Carter, D.A., A.D. Dick, and E.J. Mayer, *CD133+ adult human retinal cells remain undifferentiated in Leukaemia Inhibitory Factor (LIF).* BMC Ophthalmol, 2009. **9**: p. 1.
37. Portillo, J.A., et al., *Identification of primary retinal cells and ex vivo detection of proinflammatory molecules using flow cytometry.* Mol Vis, 2009. **15**: p. 1383-9.
38. Collin, J., et al., *Deconstructing Retinal Organoids: Single Cell RNA-Seq Reveals the Cellular Components of Human Pluripotent Stem Cell-Derived Retina.* Stem Cells, 2019. **37**(5): p. 593-598.
39. Daum, J.M., et al., *The formation of the light-sensing compartment of cone photoreceptors coincides with a transcriptional switch.* Elife, 2017. **6**.
40. Hartl, D., et al., *Cis-regulatory landscapes of four cell types of the retina.* Nucleic Acids Res, 2017. **45**(20): p. 11607-11621.
41. Lu, Y., et al., *Single-Cell Analysis of Human Retina Identifies Evolutionarily Conserved and Species-Specific Mechanisms Controlling Development.* Dev Cell, 2020. **53**(4): p. 473-491 e9.
42. Macosko, E.Z., et al., *Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.* Cell, 2015. **161**(5): p. 1202-1214.
43. Orozco, L.D., et al., *Integration of eQTL and a Single-Cell Atlas in the Human Eye Identifies Causal Genes for Age-Related Macular Degeneration.* Cell Rep, 2020. **30**(4): p. 1246-1259 e6.
44. Rheaume, B.A., et al., *Single cell transcriptome profiling of retinal ganglion cells identifies cellular subtypes.* Nat Commun, 2018. **9**(1): p. 2759.
45. Roesch, K., M.B. Stadler, and C.L. Cepko, *Gene expression changes within Müller glial cells in retinitis pigmentosa.* Molecular vision, 2012. **18**: p. 1197-1214.
46. Sarin, S., et al., *Role for Wnt Signaling in Retinal Neuropil Development: Analysis via RNA-Seq and In Vivo Somatic CRISPR Mutagenesis.* Neuron, 2018. **98**(1): p. 109-126 e8.
47. Shekhar, K., et al., *Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics.* Cell, 2016. **166**(5): p. 1308-1323 e30.

48. Voigt, A.P., et al., *Molecular characterization of foveal versus peripheral human retina by single-cell RNA sequencing.* Exp Eye Res, 2019. **184**: p. 234-242.
49. Cui, Z., et al., *Transcriptomic Analysis of the Developmental Similarities and Differences Between the Native Retina and Retinal Organoids.* Invest Ophthalmol Vis Sci, 2020. **61**(3): p. 6.
50. Kirwan, R.P., et al., *Differential global and extra-cellular matrix focused gene expression patterns between normal and glaucomatous human lamina cribrosa cells.* Mol Vis, 2009. **15**: p. 76-88.
51. Bennis, A., et al., *Comparison of Mouse and Human Retinal Pigment Epithelium Gene Expression Profiles: Potential Implications for Age-Related Macular Degeneration.* PLOS ONE, 2015. **10**(10): p. e0141597.
52. Charish, J., et al., *Neogenin neutralization prevents photoreceptor loss in inherited retinal degeneration.* The Journal of Clinical Investigation, 2020. **130**(4): p. 2054-2068.
53. Galvao, J., et al., *The Kruppel-Like Factor Gene Target Dusp14 Regulates Axon Growth and Regeneration.* Invest Ophthalmol Vis Sci, 2018. **59**(7): p. 2736-2747.
54. Agudo, M., et al., *Time course profiling of the retinal transcriptome after optic nerve transection and optic nerve crush.* Mol Vis, 2008. **14**: p. 1050-63.
55. Hadziahmetovic, M., et al., *Microarray analysis of murine retinal light damage reveals changes in iron regulatory, complement, and antioxidant genes in the neurosensory retina and isolated RPE.* Invest Ophthalmol Vis Sci, 2012. **53**(9): p. 5231-41.
56. Strunnikova, N.V., et al., *Transcriptome analysis and molecular signature of human retinal pigment epithelium.* Hum Mol Genet, 2010. **19**(12): p. 2468-86.
57. Kuroda, T., et al., *Identification of a Gene Encoding Slow Skeletal Muscle Troponin T as a Novel Marker for Immortalization of Retinal Pigment Epithelial Cells.* Sci Rep, 2017. **7**(1): p. 8163.
58. Hafler, B.P., et al., *Transcription factor Olig2 defines subpopulations of retinal progenitor cells biased toward specific cell fates.* Proc Natl Acad Sci U S A, 2012. **109**(20): p. 7882-7.
59. Chuang, J.H., et al., *Expression profiling of cell-intrinsic regulators in the process of differentiation of human iPSCs into retinal lineages.* Stem Cell Res Ther, 2018. **9**(1): p. 140.
60. Hu, J., et al., *Computational analysis of tissue-specific gene networks: application to murine retinal functional studies.* Bioinformatics, 2010. **26**(18): p. 2289-97.
61. Chen, L., et al., *Integrating Deep Supervised, Self-Supervised and Unsupervised Learning for Single-Cell RNA-seq Clustering and Annotation.* Genes (Basel), 2020. **11**(7).
62. Haque, A., et al., *A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications.* Genome Medicine, 2017. **9**(1): p. 75.
63. Saelens, W., et al., *A comparison of single-cell trajectory inference methods.* Nat Biotechnol, 2019. **37**(5): p. 547-554.
64. Zhang, S.S., et al., *A biphasic pattern of gene expression during mouse retina development.* BMC Dev Biol, 2006. **6**: p. 48.
65. Howell, G.R., et al., *Datgan, a reusable software system for facile interrogation and visualization of complex transcription profiling data.* BMC Genomics, 2011. **12**: p. 429.

66. Qian, J., et al., *Identification of regulatory targets of tissue-specific transcription factors: application to retina-specific gene regulation.* Nucleic Acids Res, 2005. **33**(11): p. 3479-91.
67. Sahoo, D., et al., *Boolean implication networks derived from large scale, whole genome microarray datasets.* Genome Biol, 2008. **9**(10): p. R157.
68. Sahoo, D., et al., *MiDReG: a method of mining developmentally regulated genes using Boolean implications.* Proc Natl Acad Sci U S A, 2010. **107**(13): p. 5732-7.
69. Dalerba, P., et al., *Single-cell dissection of transcriptional heterogeneity in human colon tumors.* Nat Biotechnol, 2011. **29**(12): p. 1120-7.
70. Dalerba, P., et al., *CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer.* N Engl J Med, 2016. **374**(3): p. 211-22.
71. Volkmer, J.P., et al., *Three differentiation states risk-stratify bladder cancer into distinct subtypes.* Proc Natl Acad Sci U S A, 2012. **109**(6): p. 2078-83.
72. Inlay, M.A., et al., *Ly6d marks the earliest stage of B-cell specification and identifies the branchpoint between B-cell and T-cell development.* Genes Dev, 2009. **23**(20): p. 2376-81.
73. Pang, W.W., et al., *Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age.* Proc Natl Acad Sci U S A, 2011. **108**(50): p. 20012-7.
74. Rajasekaran, S., et al., *Non-coding and Coding Transcriptional Profiles Are Significantly Altered in Pediatric Retinoblastoma Tumors.* Front Oncol, 2019. **9**: p. 221.
75. Dabydeen, S.A., A. Desai, and D. Sahoo, *Unbiased Boolean analysis of public gene expression data for cell cycle gene identification.* Mol Biol Cell, 2019. **30**(14): p. 1770-1779.
76. Pandey, S. and D. Sahoo, *Identification of gene expression logical invariants in Arabidopsis.* Plant Direct, 2019. **3**(3): p. e00123.
77. Dang, D., et al., *Computational Approach to Identifying Universal Macrophage Biomarkers.* Frontiers in Physiology, 2020. **11**(275).
78. de Melo, J., et al., *The Spalt family transcription factor Sall3 regulates the development of cone photoreceptors and retinal horizontal interneurons.* Development, 2011. **138**(11): p. 2325-36.
79. Agathocleous, M. and W.A. Harris, *From Progenitors to Differentiated Cells in the Vertebrate Retina.* Annual Review of Cell and Developmental Biology, 2009. **25**(1): p. 45-69.
80. Trimarchi, J.M., M.B. Stadler, and C.L. Cepko, *Individual retinal progenitor cells display extensive heterogeneity of gene expression.* PLoS One, 2008. **3**(2): p. e1588.
81. Bennis, A., et al., *Stem Cell Derived Retinal Pigment Epithelium: The Role of Pigmentation as Maturation Marker and Gene Expression Profile Comparison with Human Endogenous Retinal Pigment Epithelium.* Stem Cell Rev Rep, 2017. **13**(5): p. 659-669.
82. Brandl, C., et al., *In-depth characterisation of Retinal Pigment Epithelium (RPE) cells derived from human induced pluripotent stem cells (hiPSC).* Neuromolecular Med, 2014. **16**(3): p. 551-64.
83. Liao, J.L., et al., *Molecular signature of primary retinal pigment epithelium and stem-cell-derived RPE cells.* Hum Mol Genet, 2010. **19**(21): p. 4229-38.

84. Plaza Reyes, A., et al., *Identification of cell surface markers and establishment of monolayer differentiation to retinal pigment epithelial cells.* Nat Commun, 2020. **11**(1): p. 1609.
85. Kremerskothen, J., et al., *Characterization of KIBRA, a novel WW domain-containing protein.* Biochem Biophys Res Commun, 2003. **300**(4): p. 862-7.
86. Papassotiropoulos, A., et al., *Common Kibra alleles are associated with human memory performance.* Science, 2006. **314**(5798): p. 475-8.
87. Liu, Z., et al., *Molecular cloning and characterization of human Castor, a novel human gene upregulated during cell differentiation.* Biochem Biophys Res Commun, 2006. **344**(3): p. 834-44.
88. Ramulu, P., et al., *Normal light response, photoreceptor integrity, and rhodopsin dephosphorylation in mice lacking both protein phosphatases with EF hands (PPEF-1 and PPEF-2).* Mol Cell Biol, 2001. **21**(24): p. 8605-14.
89. Sherman, P.M., et al., *Identification and characterization of a conserved family of protein serine/threonine phosphatases homologous to Drosophila retinal degeneration C.* Proc Natl Acad Sci U S A, 1997. **94**(21): p. 11639-44.
90. Reese, B.E. and P.W. Keeley, *Genomic control of neuronal demographics in the retina.* Prog Retin Eye Res, 2016. **55**: p. 246-259.

Figure Legends

Figure 1. Schematic Algorithm

Schematic algorithm to discover cell type-specific genes from scRNA-seq data by combining correlational and Boolean implication analysis. Boolean implication analysis uses one general and one specific bait gene to identify cell type-specific biomarkers. Spearman's rank correlation coefficient analysis (SRCCA) uses one or more genes specific to a cell type as bait genes to identify other genes expressed in the same cell type. Boolean analysis is directly compared to SRCCA and improvement is tested using two proportion Z-test.

Figure 2. Results

Results of Boolean implication analysis of Phillips 2018 scRNA-seq dataset using two or more bait genes, for 5 retinal cell types. **Abbreviations**: SRCCA - Spearman's Rank Correlation Coefficient Analysis: PR - photoreceptors; RPC - retinal progenitor cell; RPE - retinal pigment epithelium; RGC - retinal ganglion cell.

Figure 3. Independent Validation of Results

(A): Validation bulk RNA-seq datasets such as GSE84589 containing purified rods and cones from Mus musculus were used to validate rod and cone gene lists through differential expression. **(B):** Rod cell type-specificity of rod gene lists from 4 methods: Boolean implication, SRCCA, SRCCA filtered using Boolean implication and SRCCA without Boolean implication. **(C):** Photoreceptor-specificity of rod gene lists from 4 methods. **(D):** Cone cell type-specificity of rod gene lists from 4 methods. **(E):** Proportion of genes from SRCCA and SRCCA filtered using

Boolean implication using a common set of bait genes CRX, GNB3 and GNAT2 directly reproducible in GSE130636 and GSE142449. **Abbreviations:** Corr. - Correlation; Bool. - Boolean; SRCCA - Spearman's rank correlation coefficient analysis. **Note:** P-values are from two-proportion Z-test between proportion of cell type-specific genes in lists from SRCCA and SRCCA filtered using Boolean implication.

Figure 4. Specific Examples

Analysis of high confidence candidate PR genes from Boolean implication analysis and SRCCA. **(A-B):** High confidence cone PR gene WWC1 from Boolean implication analysis shows statistically significant overexpression in cones compared to rods in both datasets. **(C-D):** High confidence cone PR gene AKAP9 from SRCCA does not show cone specificity in either dataset. Cone PR group labelled in blue and rod PR group labelled in red on boxplots. **(E-H):** High confidence rod PR genes CASZ1 and PPEF2 from Boolean implication analysis show statistically significant overexpression in rods compared to cones in both datasets. **Note**: p-values reported from Welch's t-test (unequal variances). **Abbreviations:** PR - photoreceptor; SRCCA - Spearman's rank correlation coefficient analysis.

Figure S1. Discovery of Boolean Implication Relationships

Method for discovering and applying Boolean implication relationships in single cell RNA sequencing data. **(A-F):** Six types of Boolean implication relationships are visible on scatter plots. Two are symmetric with two sparse quadrants (A-B) and four are asymmetric with one sparse quadrant (C-F). **(G):** This plot is divided into four quadrants based on thresholds identified by the StepMiner algorithm. **(H-I):** The BooleanNet algorithm identifies the sparse quadrants

using a statistic *S* and a likelihood error rate *p* and applying thresholds of 2.5 and 0.35, respectively. **(J):** Analysis of Boolean implication relationships was used to find genes involved in cell fate determination using bait genes (A and B)**.**

Figure S2. Additional Gene Lists

(A): Additional gene lists from SRCCA and Boolean analysis using common bait genes for cones CRX, GNAT2 and GNB3.

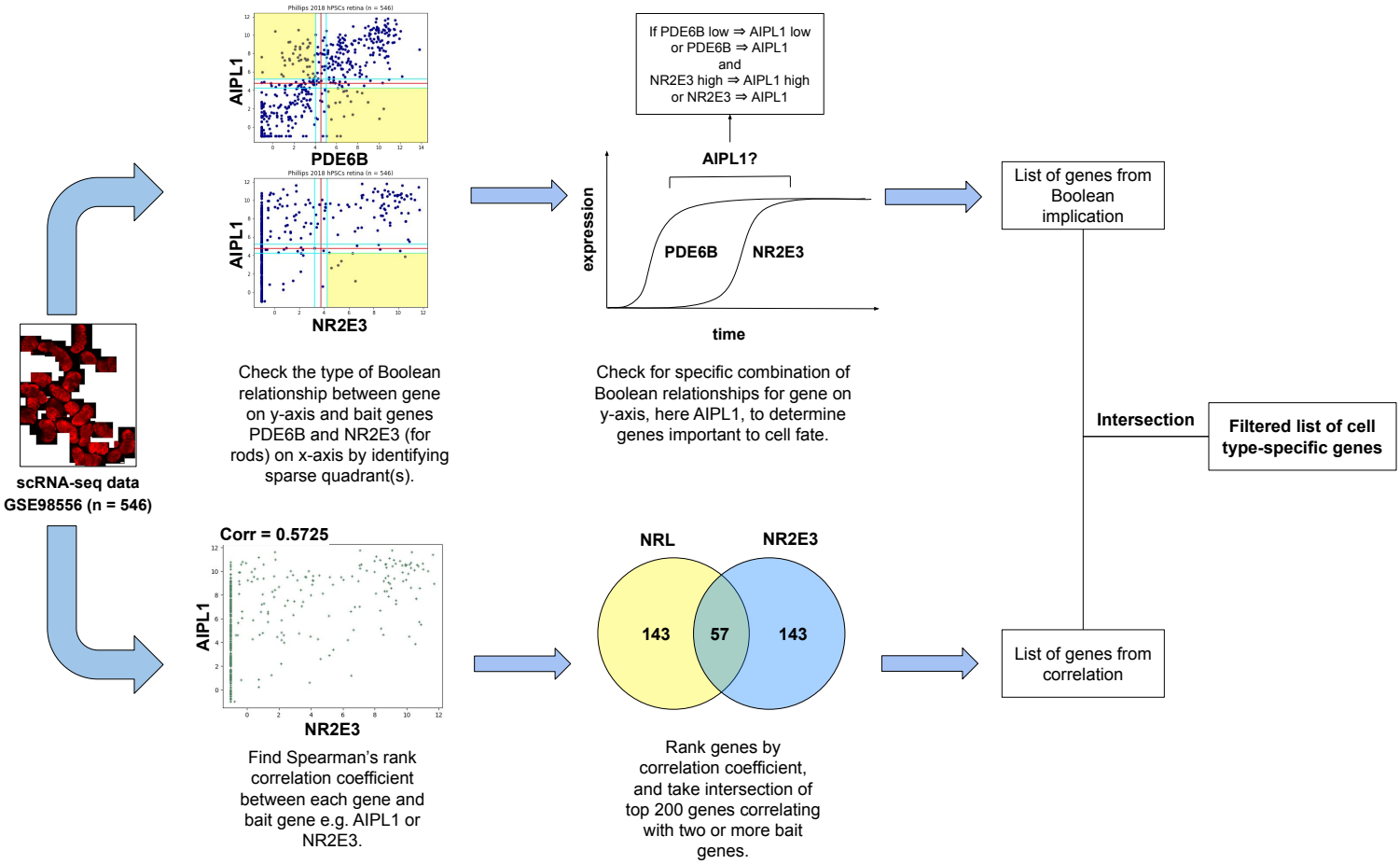


Figure 1. Schematic algorithm to discover cell type-specific genes from scRNA-seq data by combining correlational and Boolean implication analysis. Boolean implication analysis uses one general and one specific bait gene to identify cell type-specific biomarkers. Spearman's rank correlation coefficient analysis (SRCCA) uses one or more genes specific to a cell type as bait genes to identify other genes expressed in the same cell type. Boolean analysis is directly compared to SRCCA and improvement is tested using two proportion Z-test.

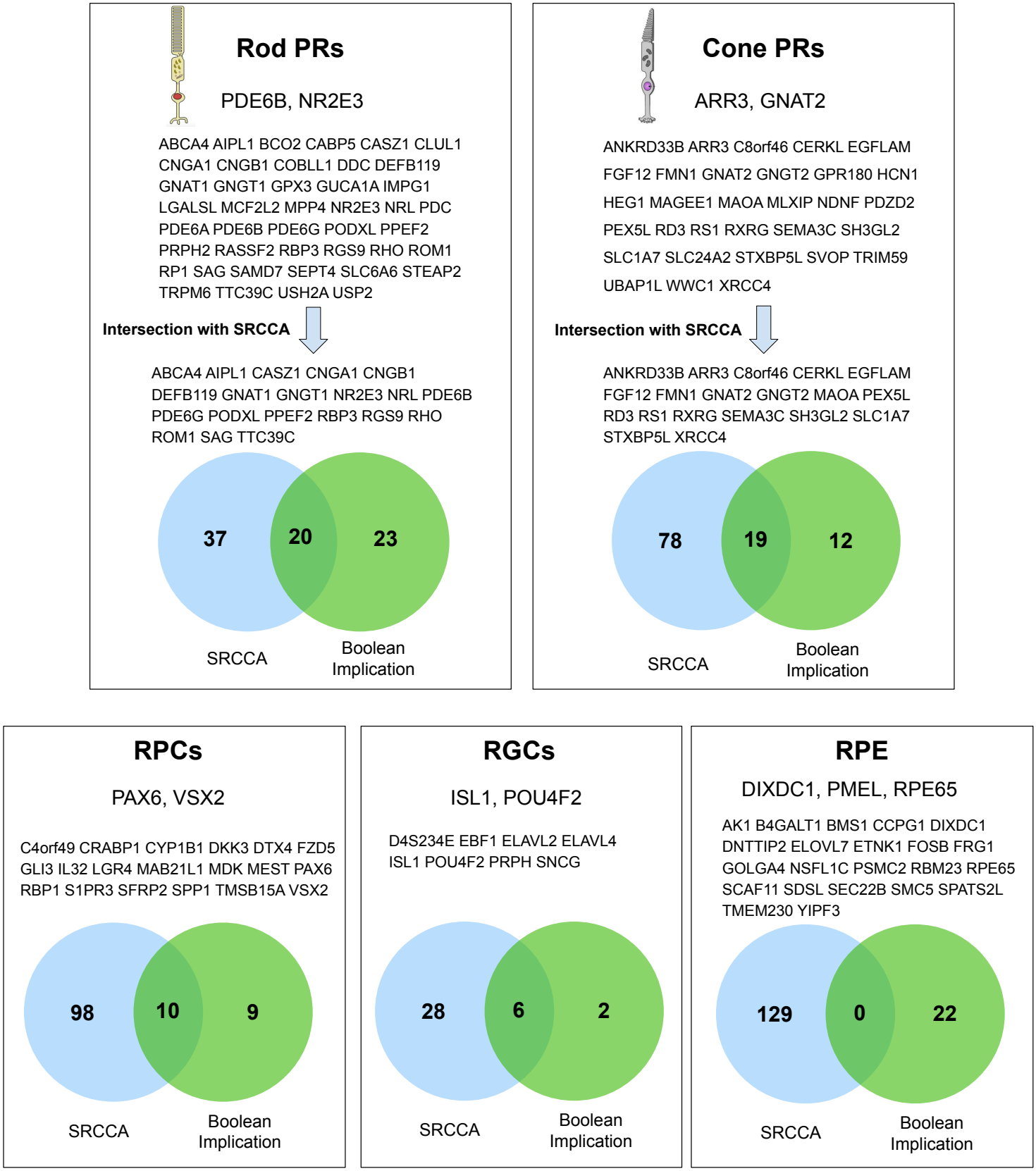


Figure 2. Results of Boolean implication analysis of Phillips 2018 scRNA-seq dataset using two or more bait genes, for 5 retinal cell types. **Abbreviations**: SRCCA - Spearman's Rank Correlation Coefficient Analysis: PR - photoreceptors; RPC - retinal progenitor cell; RPE - retinal pigment epithelium; RGC - retinal ganglion cell.

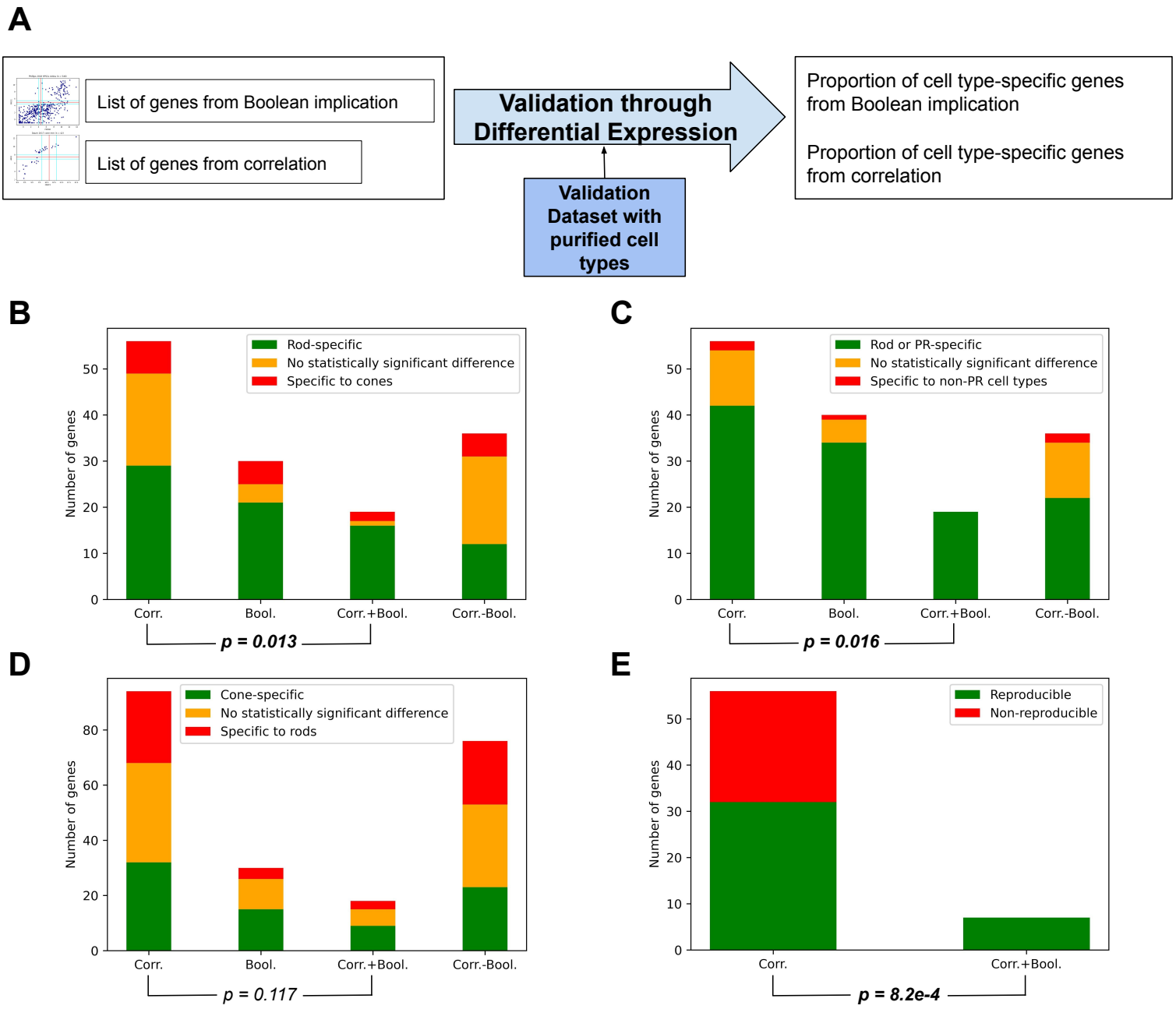


Figure 3. (A): Validation bulk RNA-seq datasets such as GSE84589 containing purified rods and cones from Mus musculus were used to validate rod and cone gene lists through differential expression. **(B):** Rod cell type-specificity of rod gene lists from 4 methods: Boolean implication, SRCCA, SRCCA filtered using Boolean implication and SRCCA without Boolean implication. **(C):** Photoreceptor-specificity of rod gene lists from 4 methods. **(D):** Cone cell type-specificity of rod gene lists from 4 methods. **(E):** Proportion of genes from SRCCA and SRCCA filtered using Boolean implication using a common set of bait genes CRX, GNB3 and GNAT2 directly reproducible in GSE130636 and GSE142449. **Abbreviations:** Corr. - Correlation; Bool. - Boolean; SRCCA - Spearman's rank correlation coefficient analysis. **Note:** P-values are from two-proportion Z-test between proportion of cell type-specific genes in lists from SRCCA and SRCCA filtered using Boolean implication.

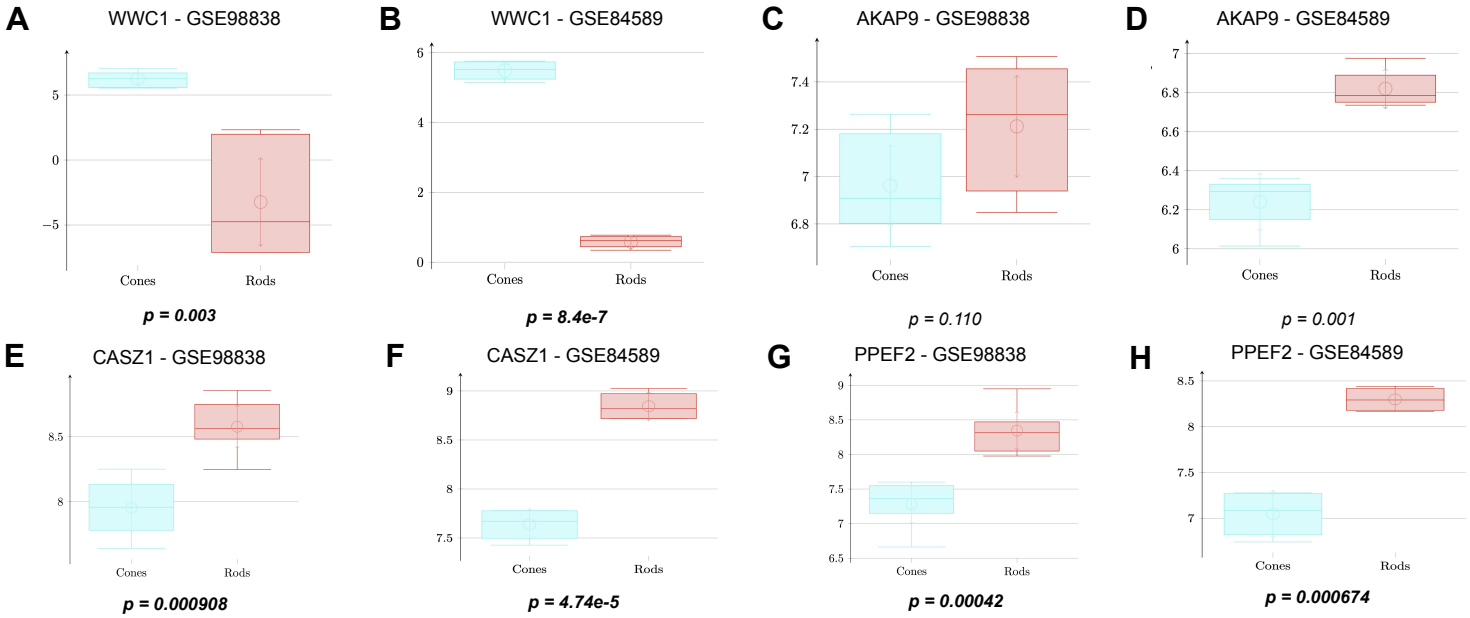
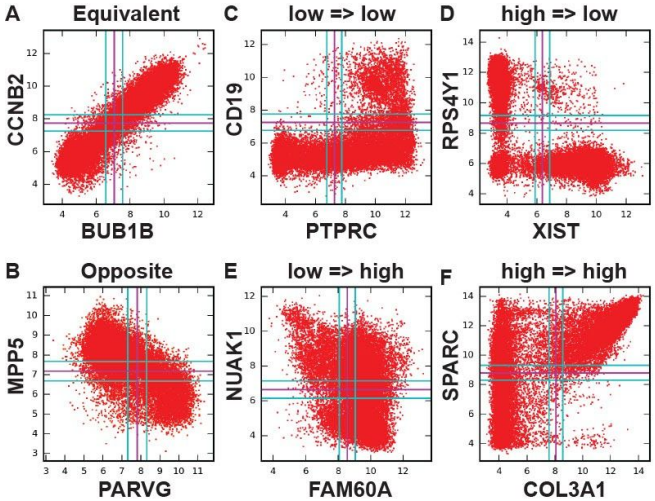
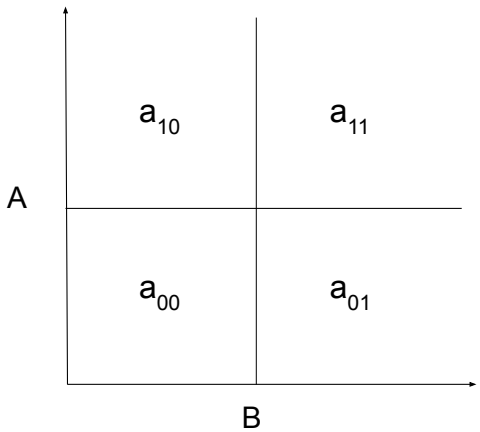


Figure 4. Analysis of high confidence candidate PR genes from Boolean implication analysis and SRCCA. **(A-B):** High confidence cone PR gene WWC1 from Boolean implication analysis shows statistically significant overexpression in cones compared to rods in both datasets. **(C-D):** High confidence cone PR gene AKAP9 from SRCCA does not show cone specificity in either dataset. Cone PR group labelled in blue and rod PR group labelled in red on boxplots. **(E-H):** High confidence rod PR genes CASZ1 and PPEF2 from Boolean implication analysis show statistically significant overexpression in rods compared to cones in both datasets. **Note**: p-values reported from Welch's t-test (unequal variances). **Abbreviations:** PR - photoreceptor; SRCCA - Spearman's rank correlation coefficient analysis.



G



H

$$nA_{low} = (a_{00} + a_{01})$$
$$nB_{low} = (a_{00} + a_{10})$$
$$\text{total} = a_{00} + a_{01} + a_{10} + a_{11}$$

$$\hat{e} = \frac{\frac{nA_{low}}{\text{total}} \cdot \frac{nB_{low}}{\text{total}}}{\text{total}}$$

$$S_{00} = \frac{\hat{e} - n}{\sqrt{\hat{e}}}$$

$$p_{00} = \frac{1}{2}\left(\frac{a_{00}}{(a_{00} + a_{01})} + \frac{a_{00}}{(a_{00} + a_{10})}\right)$$

$$\text{Boolean Implication} = S > 2.5, p < 0.35$$

I

$$\text{Equivalent} = S_{01} > 2.5, p_{01} < 0.35, S_{10} > 2.5, p_{10} < 0.35$$

$$\text{Opposite} = S_{00} > 2.5, p_{00} < 0.35, S_{11} > 2.5, p_{11} < 0.35$$

$$A \text{ low} \implies B \text{ high} = (S_{00} > 2.5, p_{00} < 0.35)$$

$$A \text{ low} \implies B \text{ low} = (S_{01} > 2.5, p_{01} < 0.35)$$

$$A \text{ high} \implies B \text{ high} = (S_{10} > 2.5, p_{10} < 0.35)$$

$$A \text{ high} \implies B \text{ low} = (S_{11} > 2.5, p_{11} < 0.35)$$

J

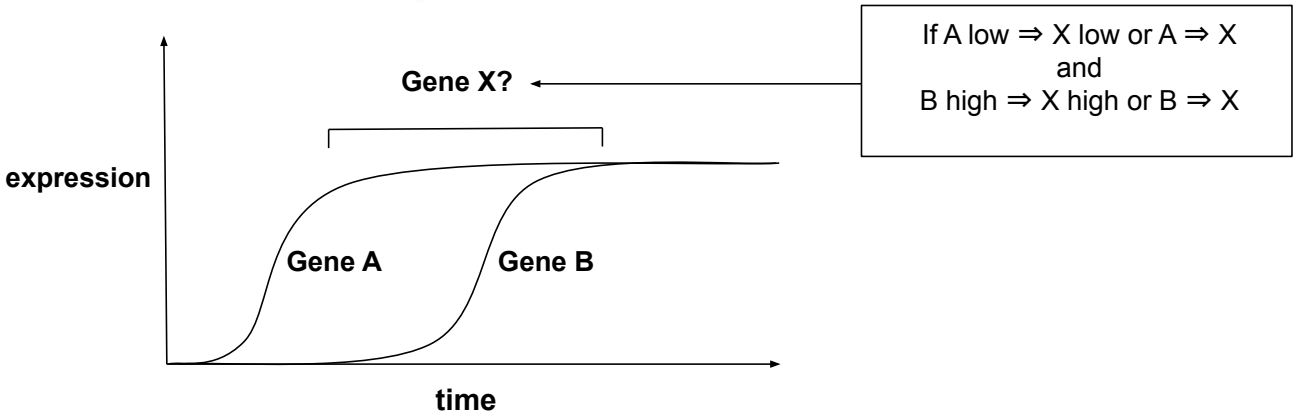


Figure S1. Method for discovering and applying Boolean implication relationships in single cell RNA sequencing data. **(A-F):** Six types of Boolean implication relationships are visible on scatter plots. Two are symmetric with two sparse quadrants (A,B) and four are asymmetric with one sparse quadrant (C-F). **(G):** This plot is divided into four quadrants based on thresholds identified by the StepMiner algorithm. **(H-I):** The BooleanNet algorithm identifies the sparse quadrants using a statistic *S* and a likelihood error rate *p* and applying thresholds of 2.5 and 0.35, respectively. **(J):** Analysis of Boolean implication relationships was used to find genes involved in cell fate determination using bait genes (A and B).

A

CRX, GNB3, GNAT2

SRCCA	Boolean Implication
ANKRD33B AP1S3 ATP1B2 CACNB2 CDHR1 CPLX4 CYP1A2 DUSP19 EPHA10 EYS FSTL5 GPR155 GRM6 HCN1 IBA57 IMPG2 KCNV2 MAK MPL MPP4 MREG ORAI2 PARD6B PNPO RAX2 RCVRN RD3 RP1 RPGRIP1 SLC14A2 SLC24A2 SLC24A4 TMEM120B USH2A ZNF716	C8orf84 CDHR1 CRX DCDC2 FAM161A FAM161B FSTL5 GALNTL2 GNAT2 GNB3 GPKOW IMPG2 MAK MPP4 MYOZ3 NFAM1 PAFAH2 PDC PLXDC1 RCVRN RIMS2 RP1 TMEM220 TPD52 TRIM72 VGLL3 ZNF577

Figure S2. (A): Additional gene lists from SRCCA and Boolean analysis using common bait genes for cones CRX, GNAT2 and GNB3.