

Modeling human monogenic diseases using the choanoflagellate *Salpingoeca rosetta*

We applied a decision-making framework for identifying tractable genes from our organism selection dataset for pilot experiments in the choanoflagellate *Salpingoeca rosetta*.

Contributors (A-Z)

Prachee Avasthi, Audrey Bell, David S. Booth, Keith Cheveralls, Seemay Chou, Megan L. Hochstrasser, Dennis A. Sun, Ryan York

Version 2 · Aug 12, 2025

Purpose

We developed a framework for identifying actionable disease–gene pairs from organism selection data [1]. Here, we apply that framework to genes in the choanoflagellate *Salpingoeca rosetta* and identify seven genes to pursue for experimental testing. We share information about *S. rosetta*, including its life history and currently available genetic tools. We also provide examples of genes we rejected or were particularly excited about in our decision-making process.

Scientists interested in using choanoflagellates or other microbial eukaryotes to model human disease may find our analyses and experimental design process useful in their work. We include scripts to help visualize gene expression in *S. rosetta* and a summary of the state of technology, which can aid in scientific decision-making.

- All **code** related to this pub is available in [this GitHub repository](#). Code specific to *S. rosetta* is in [this analysis notebook](#) and [these visualization scripts](#).
- **Data**, including Foldseek and ProteinCartography results for certain *S. rosetta* proteins, is available on [Zenodo](#).

About the organism

Salpingoeca rosetta is a choanoflagellate and an emerging genetic model organism [2]. Choanoflagellates are a group of microbial eukaryotes that share a common ancestor with the last common ancestor of animals [3]. Research groups often study this diverse group of organisms to understand the origins of multicellularity.

Choanoflagellates often have transient multicellular life stages generated through cell division [4]. Cooperativity in these colonies provides insight into the transition from unicellular to multicellular life.

Free-swimming single *S. rosetta* cells have a stereotypical cell shape characterized by a flagellum circumscribed by an array of microvilli referred to as a “collar” [5] (Figure 1, A). *S. rosetta* cells can divide to form fragile “chains” of related cells attached by cell-cell adhesion [5]. *S. rosetta* also forms spherical colonies referred to as “rosettes” in response to a sulfonolipid produced by the bacteria *Algorphiagus machipongonensis* [6]. In addition to its multicellular life stages, *S. rosetta* also has diverse unicellular life stages, including slow swimmers, fast swimmers, and thecate cells (Figure 1, A) [5]. *S. rosetta* can even form amoeboid cells in response to physical compression [7].

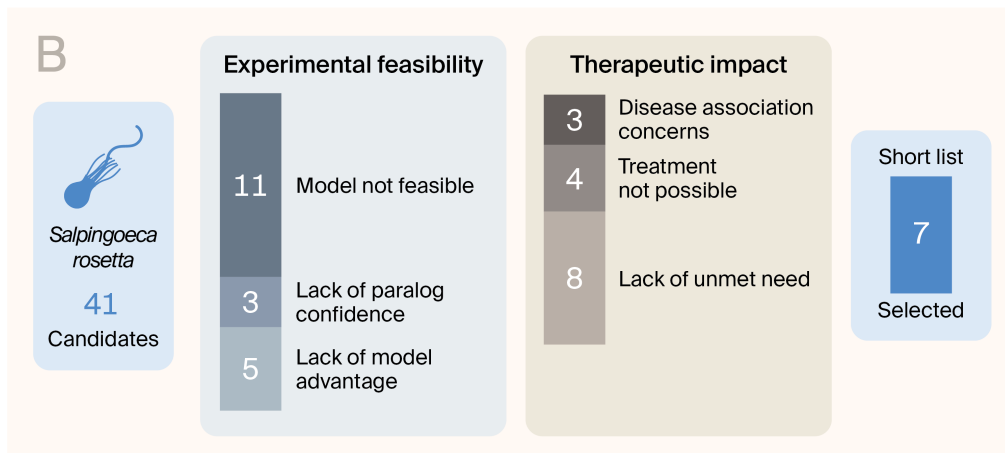
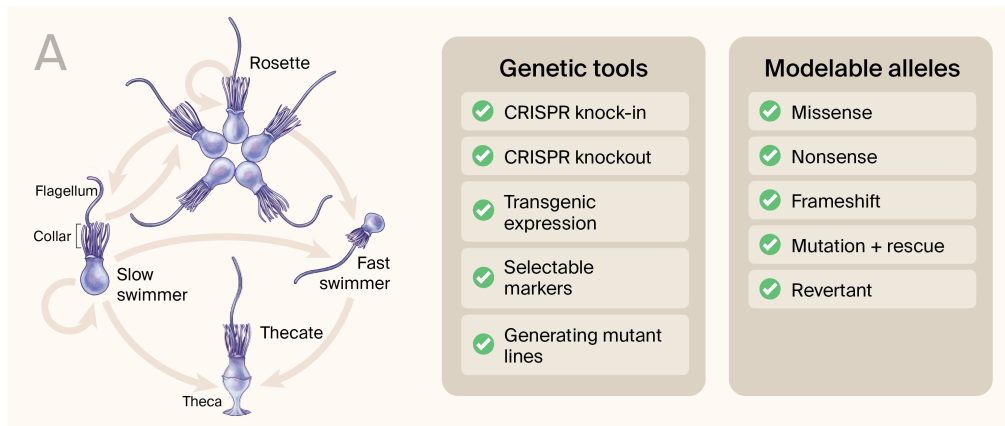


Figure 1

Overview of diligence on *Salpingoeca rosetta*.

(A) Illustration of different cell types of *S. rosetta* and schematic of available genetic tools and types of human alleles that we expect to be able to model, given available tools. Cell types include slow swimmers, rosettes, fast swimmers, and thecate cells.

(B) Illustration of manual diligencing outcomes listing the number of genes eliminated by each failure mode. The initial short list was seven genes, all of which we selected for experimental testing.

State of technology

S. rosetta has robust genetic tools and genomic resources.

- **Genetic tools.** *S. rosetta* has robust protocols for generating targeted genetic variation. CRISPR-Cas9 mutagenesis is feasible in these organisms through a marker-based approach, and it's possible to generate and maintain stable mutant lines [8]. Transgenic expression of genes through episomal markers is also possible, with some size limitations [9].
- **Phenotypic assays.** *S. rosetta* cells are around 2.5–4 µm in diameter [5], and can be grown in diverse formats, from flasks to 96-well plates, allowing for robust phenotypic assays across scales. Recent work has shown that it's even possible to screen drug molecules for activity in *S. rosetta* [10]. *S. rosetta* cultures are amenable to bulk phenotyping through growth curves and time course imaging. Established protocols for immunofluorescence [11], scanning and transmission electron microscopy [5], and other approaches make it possible to comprehensively identify diverse phenotypes in *S. rosetta* cells.
- **Gene expression.** Bulk RNA-seq data for *S. rosetta* are available for various cell types and stages. We visualized gene expression results from Leon et al. 2025 [12], who profiled bulk gene expression across slow swimmers, fast swimmers, rosettes, and thecate cells.

You can **generate gene expression visualizations** from *S. rosetta* data by following the examples in [this Jupyter notebook](#).

- **Feasible genetic experiments.** Given the wealth of gene-editing tools available in *S. rosetta*, it should be possible to model a variety of different kinds of human disease alleles, including missense, nonsense, and frameshift mutations. The ability to transgenically express genes also makes it possible to test dominant-negative mutations, or to complement mutations with wild-type protein. Conventionally, researchers perform rescue experiments to validate the genetic basis of observed phenotypes by generating CRISPR revertants – mutations that change a previously mutated allele back to the wild-type allele – due to a lack of episomal replication.

Diligence results

We manually diligenced 41 different genes in *Salpingoeca rosetta* to arrive at seven potentially actionable genes for pilot experiments ([Figure 1](#), B).

You can see a copy of the **working document** we used to catalog our thoughts for each gene we diligenced in *S. rosetta* [here](#).

Download **Foldseek results** for the *S. rosetta* plastin 1 ([F2TWP3](#)) and coronin 1a ([F2TZV2](#)) homologs, plus **ProteinCartography data** for the *S. rosetta* homolog of the protein encoded by *DDC* ([F2UCR6](#)) from [Zenodo](#) (DOI: [10.5281/zenodo.15724261](#)).

Example genes we rejected

Here, we describe our specific reasoning for rejecting a handful of select genes from our short list for *S. rosetta* to give you a sense of the variety of reasons we might have passed on a given gene.

***DDC* (dopamine decarboxylase)**

Lack of homolog confidence

The *DDC* gene is implicated in deficiency of aromatic-L-amino-acid decarboxylase, which leads to severe postembryonic developmental phenotypes in patients **[13]**. *S. rosetta* is a top hit for this gene in our organism selection analysis. However, the *S. rosetta* protein (UniProt ID [F2UCR6](#)) is putatively annotated as “glutamate decarboxylase,” and FoldSeek searches of this protein don’t return the protein encoded by *DDC* as a top hit.

To understand the broader scope of related gene families, we ran sequence- and structure-based searching and clustering using ProteinCartography **[14]** and observed that the *S. rosetta* protein is found in a different Leiden cluster (LC00) than dopamine decarboxylases from vertebrates (mostly in LC12) and invertebrates (LC05) ([Figure 2](#); download inputs and results from [Zenodo](#)). This suggests that the *S. rosetta* protein may not be a dopamine decarboxylase. Due to this uncertainty, we rejected this candidate gene.

Figure 2

Interactive scatter plot that ProteinCartography generated for the choanoflagellate homolog of the protein encoded by *DDC* identified in the organism selection dataset.

***TREH* (α,α -trehalase)**

Lack of unmet need

Trehalose is a disaccharide commonly found in mushrooms and yeast. Patients with a deficiency in α,α -trehalase are unable to process this sugar, resulting in gastrointestinal symptoms such as vomiting and diarrhea [15]. While unpleasant for patients, this disease is treatable by omitting food products containing mushrooms and yeast in patient diets. Due to the low severity of the disease's symptoms and ease of treatment, we didn't consider this disease to have a high unmet need, and therefore, we rejected it.

***HYLS1* (centriolar and ciliogenesis-associated protein HYLS1)**

Treatment not possible

Homozygous loss of function of *HYLS1* causes a lethal malformation syndrome with defects across organ systems, resulting in stillbirth or death shortly after birth [16]. Due to the lethal and developmental nature of this disease, we didn't believe it would be possible to develop a therapeutic, and therefore rejected this gene.

Our initial short list

Here are some of the best-supported case studies from our short list in *S. rosetta*.

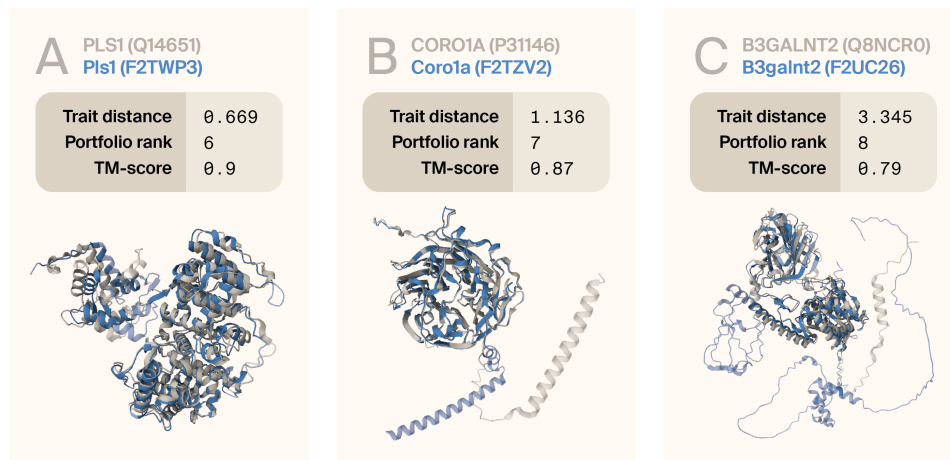


Figure 3

Structural comparisons of top-selected short-list proteins to human proteins.

(A) Plastin 1

(B) Coronin 1a

(C) Beta-1,3-N-acetylgalactosaminyltransferase 2

Tan proteins are human structures; blue proteins are predicted *S. rosetta* homolog structures. “Trait distance” is a multivariate Mahalanobis distance between pairs of proteins calculated based on 10 physicochemical properties (see the “Methods” section of our previous pub for more details [17]). “Portfolio rank” is the relative rank of this protein compared to all other proteins found in the organism selection/the Zoogle portal. TM-score is a standard measure of solid-body protein structural similarity [18]. In general, TM-scores of 0.5 or above indicate the same fold.

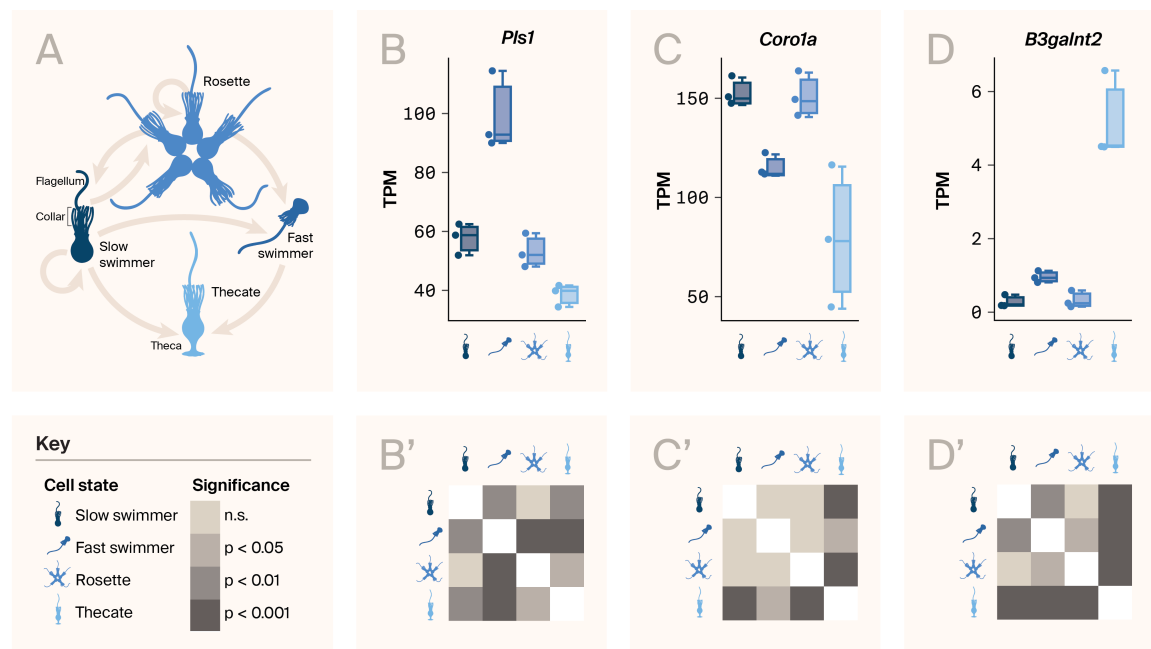
PLS1 (plastin 1)

Plastin 1, also known as fimbrin, is an actin-bundling protein implicated in autosomal-dominant hearing loss [19]. Missense mutations in this protein lead to destabilization of plastin binding to actin and progressive or non-progressive hearing loss with variable age of onset [19]. While hearing loss in humans is reported as a missense

allele, mouse knockouts have shown comparable modeling of human disease. *Pls1* knockout mice show defects in the morphology of stereocilia on ear hair cells [20], which are the mechanosensitive cells in the cochlea responsible for transducing sound into an electrical signal. Knockout mice also show an increasing loss of hearing as they age [20].

Choanoflagellate cells normally exhibit a stereociliar structure, the collar, which resembles ear hair cell stereocilia [21]. This structure is involved in feeding and possibly in mechanosensation [22]. The choanoflagellate plastin homolog shows high structural similarity (TM-score = 0.9) to plastin 1 and slightly lower structural similarity to two other major human homologs of plastin, LCP1 (also known as plastin 2, TM-score = 0.88) and plastin 3 (TM-score = 0.85) (Figure 3) (download Foldseek results from [Zenodo](#)). The choanoflagellate gene also appears to be expressed in all four cell states in RNA-seq data, with particularly high expression in fast swimmers (Figure 4).

Given the high degree of structural similarity between the human and choanoflagellate plastins, we were interested in generating knockouts of choanoflagellate plastin and assaying for possible phenotypes. Modeling plastin function in choanoflagellates also poses certain advantages over modeling the same function in mice or in differentiated human stem cells. Protocols for differentiating iPSC-derived hair-like cells take 50–60 days to complete [23], while assaying for strong phenotypes in mice can take months, due to an increase in hearing loss signal as the animals age [20]. In contrast, choanoflagellate cells exhibit collar morphologies at all life stages. Generating knockouts, identifying possible phenotypes, and screening for molecules that rescue plastin function can all be performed with greater speed in choanoflagellates than in existing models. For these reasons, we were particularly excited to model *PLS1* function using *S. rosetta*.



CORO1A (coronin 1a)

Coronin 1a is associated with severe combined immunodeficiency, also known as immunodeficiency 8 (IMD8). Patients with loss of coronin have early-childhood onset of recurrent infections, often associated with Epstein-Barr virus [24]. Mice with missense mutations in *Coro1a* show defects in actin localization in the leading edge of T cells [25].

Choanoflagellate cells produce a variety of actin-mediated cell structures, such as ruffles, filopodia, exocytic cups, and others [22][5][26][7]. The choanoflagellate

homolog of coronin 1a also has a high structural similarity to multiple coronins (the top seven hits in our Foldseek search included CORO6, CORO1C, CORO2B, CORO1A, CORO1B, CORO2A, and CORO7) ([Figure 3](#)) (download Foldseek results from [Zenodo](#)). This gene is also expressed across cell states with three different expression classes: slow swimmers and rosettes, fast swimmers, and thecate cells ([Figure 4](#)). Due to the high number of different phenotypes that could be captured in *S. rosetta*, we felt confident that we could identify some defect in cells with a mutant version of this gene.

***B3GALNT2* (beta-1,3-N-acetylgalactosaminyltransferase 2)**

Missense, nonsense, and frameshift mutations in this glycosylation enzyme are associated with a form of muscular dystrophy caused by congenital brain and eye abnormalities [27][28]. Morpholino knockdown of the zebrafish homolog of beta-1,3-N-acetylgalactosaminyltransferase 2 results in severe morphological defects, including those of the eye, brain, and spinal cord [28]. Mutations in *B3GALNT2* also appear to cause hydrocephalus in horses [29].

The *S. rosetta* homolog of beta-1,3-N-acetylgalactosaminyltransferase 2 is a strong structural match (TM-score = 0.79) to the human protein ([Figure 3](#)) and is expressed at low levels in all cells and slightly higher levels in thecate cells. ([Figure 4](#)).

Glycosylation plays an important role in the proper development of choanoflagellate rosettes. Recent work has shown that loss of function of glycosylation enzymes can lead to “clumpy” choanoflagellate cells that stick together [30]. This can be assayed by mixing non-fluorescent and fluorescent choanoflagellate cells and looking for aggregates across strains. We expect that loss of function in this particular glycosylation enzyme may produce similar phenotypes, which can be easily measured using this assay. For these reasons, we decided to pursue this gene.

Other hits

Several other genes – *CLTC*, *UNC13D*, *GALC*, and *B3GALNT2* – also made our short list ([Table 1](#)).

HGNC gene symbol (UniProt ID)	Human protein name	Associated human disease (OMIM)	S. rosetta protein UniProt ID
<i>PLS1</i> (Q14651)	Plastin 1	<u>Deafness, autosomal dominant 76</u>	<u>F2TWP3</u>
<i>CORO1A</i> (P31146)	Coronin 1a	<u>Immunodeficiency 8 with lymphoproliferation</u>	<u>F2TZV2</u>
<i>B4GALT7</i> (Q9UBV7)	Beta-1,4-galactosyltransferase 7	<u>Ehlers-Danlos syndrome, spondylodysplastic type, 1</u>	<u>F2UGX2</u>
<i>CLTC</i> (Q00610)	Clathrin heavy chain	<u>Intellectual developmental disorder, autosomal dominant 56</u>	<u>F2U3P4</u>
<i>UNC13D</i> (Q70J99)	Unc-13 homolog D	<u>Hemophagocytic lymphohistiocytosis, familial, 3</u>	<u>F2U085</u>
<i>GALC</i> (P54803)	Galactosylceramidase	<u>Krabbe disease</u>	<u>F2TVZ9</u>
<i>B3GALNT2</i> (Q8NCRO)	Beta-1,3-N-acetylgalactosaminyltransferase 2	<u>Muscular dystrophy-dystroglycanopathy (congenital with brain and eye anomalies), type A, 11</u>	<u>F2UC26</u>

Table 1

Summary of the relevant diseases and our initial guess for phenotypes for each short-list gene.

HGNC: HUGO gene nomenclature committee.

Expert feedback and our final short list

We sought expert feedback on our short list from David Booth, a choanoflagellate researcher at UCSF, one of the external scientists we consulted as part of this work. While we didn't have strong phenotypic hypotheses for all seven genes on our short list, given the ease of performing simple phenotypic assays such as growth, morphology, cell state transitions, and cell clumping, we decided to pursue all of our hits. Below is a rough overview of our projected experimental plan in *S. rosetta*.

Experimental plan

1. **Generate mutants and characterize growth phenotypes.**
2. **Phenotype cell state transitions.** For example, the ability to transition from slow swimmers to rosettes.
3. **Investigate other cellular phenotypes.** For example, cell motility or morphology. Such phenotypes are interesting because they are high-dimensional, quantitative, and easy to assay in high throughput. We can then follow up to perform downstream, finer-grained analyses using lower-throughput phenotyping methods.
4. **Perform genetic rescue experiments using complementation with wild-type protein or synonymous revertants.**

Methods

For details about the technical analyses described in this pub, see the [companion pub describing our evaluation framework](#).

All **code** related to this pub is available in [this GitHub repo](#) (DOI: [10.5281/zenodo.15707938](https://doi.org/10.5281/zenodo.15707938)). Code specific to *S. rosetta* is in [this analysis notebook](#) and [these visualization scripts](#).

Salpingoeca bulk cell type gene expression

We downloaded the *S. rosetta* cell type-specific RNA-seq data from [this FigShare deposition](#) and used a [custom script](#) found in our GitHub repository to visualize gene expression counts from bulk data for four cell types: slow swimmers, fast swimmers, thecate cells, and rosettes. We also visualized the computed pairwise differential expression significance p -values, such as in [Figure 4](#).

AI tool usage

We used Claude to help write code, clean up code, comment our code, draft text that we edited, suggest wording ideas for small phrases and sentence structures, and clarify and streamline text that we wrote. We also used Cursor to access Claude.

Visualization

We used plotly (v5.17.0) [31] arcadia-pycolor (v0.6.3) [32] to generate figures before manual adjustment.

Key takeaways

Advantages of *S. rosetta* as a disease model

During our diligence process, we came to realize a variety of advantages that make *S. rosetta* a valuable potential model for human diseases, including:

- **Multiple cell types.** In contrast to many microbial eukaryotes, which largely have a single cell type, *S. rosetta* has the ability to “differentiate” into different states, each of which may have different phenotypes. This allows for a broader variety of phenotypes that may be identified in mutants.
- **Persistent and dynamic cytoskeletal structures.** *S. rosetta* displays stereocilia and flagella in a variety of cell states. These structures aren’t readily found in all

human cell types, requiring potentially expensive differentiation protocols to obtain cells with relevant structures. *S. rosetta* also displays some transient structures, such as filopodia and exocytic cups, which could expand the search space of possible phenotypes.

- **Ease of genetic experimentation.** The state of genetic manipulations in *S. rosetta* is more robust and precise than in many other microbial eukaryotes. The lack of complex multinuclear structures, such as those found in organisms like *Stentor coeruleus* or *Tetrahymena thermophila*, means that genetic manipulation is more likely to result in an observable phenotype.

Next steps

Our diligence process uncovered just seven of many possible disease-causing genes that researchers can feasibly model using *S. rosetta*. Moreover, we sought to identify “low-hanging fruit” among our candidates. Scientists with greater experience with particular types of assays or a willingness to generate targeted mutations might be able to pursue disease–gene pairs that we rejected in our analysis. We encourage others to review our [working diligence document](#) and see whether their evaluation of feasibility differs from ours.

We’re funding downstream work by researchers in the [David Booth lab](#) at UCSF. They’ll share their results openly during their experiments. Stay tuned for more!

References

- 1 10.57844/arcadia-ugg5-emyd
- 2 Booth DS, King N. (2022). The history of *Salpingoeca rosetta* as a model for reconstructing animal origins. <https://doi.org/10.1016/bs.ctdb.2022.01.001>
- 3 King N. (2005). Choanoflagellates. <https://doi.org/10.1016/j.cub.2005.02.004>

- 4 Fairclough SR, Dayel MJ, King N. (2010). Multicellular development in a choanoflagellate. <https://doi.org/10.1016/j.cub.2010.09.014>
- 5 Dayel MJ, Alegado RA, Fairclough SR, Levin TC, Nichols SA, McDonald K, King N. (2011). Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. <https://doi.org/10.1016/j.ydbio.2011.06.003>
- 6 Woznica A, Cantley AM, Beemelmans C, Freinkman E, Clardy J, King N. (2016). Bacterial lipids activate, synergize, and inhibit a developmental switch in choanoflagellates. <https://doi.org/10.1073/pnas.1605015113>
- 7 Brunet T, Albert M, Roman W, Coyle MC, Spitzer DC, King N. (2021). A flagellate-to-amoeboid switch in the closest living relatives of animals. <https://doi.org/10.7554/elife.61037>
- 8 Booth DS, King N. (2020). Genome editing enables reverse genetics of multicellular development in the choanoflagellate *Salpingoeca rosetta*. <https://doi.org/10.7554/elife.56193>
- 9 Booth DS, Szmidt-Middleton H, King N. (2018). Transfection of choanoflagellates illuminates their cell biology and the ancestry of animal septins. <https://doi.org/10.1091/mbc.e18-08-0514>
- 10 Rutaganira FU, Coyle MC, Nguyen MHT, Hernandez I, Scompton AP, Dar AC, King N. (2022). A stress-responsive p38 signaling axis in choanoflagellates. <https://doi.org/10.1101/2022.08.26.505350>
- 11 Levin TC, Greaney AJ, Wetzel L, King N. (2014). The rosetteless gene controls development in the choanoflagellate *S. rosetta*. <https://doi.org/10.7554/elife.04070>
- 12 Leon F, Espinoza-Esparza JM, Deng V, Coyle MC, Espinoza S, Booth DS. (2025). Cell differentiation controls iron assimilation in the choanoflagellate *Salpingoeca rosetta*. <https://doi.org/10.1128/msphere.00917-24>
- 13 Abeling NGGM, Bräutigam C, Hoffmann GF, Barth PG, Wevers RA, Jaeken J, Fiumara A, Knust A, van Gennip AH. (2000). Pathobiochemical implications of hyperdopaminuria in patients with aromatic L-amino acid decarboxylase deficiency. <https://doi.org/10.1023/a:1005650325003>
- 14 Avasthi P, Bigge BM, Celebi FM, Cheveralls K, Gehring J, McGeever E, Mishne G, Radkov A, Sun DA. (2024). ProteinCartography: Comparing proteins with structure-based maps for interactive exploration. <https://doi.org/10.57844/ARCADIA-A5A6-1068>

- 15 Murray IA, Coupland K, Smith JA, Ansell ID, Long RG. (2000). Intestinal trehalase activity in a UK population: establishing a normal range and the effect of disease. <https://doi.org/10.1017/s0007114500000313>
- 16 Mee L, Honkala H, Kopra O, Vesa J, Finnilä S, Visapää I, Sang T-K, Jackson GR, Salonen R, Kestilä M, Peltonen L. (2005). Hydroletharus syndrome is caused by a missense mutation in a novel gene HYL1. <https://doi.org/10.1093/hmg/ddi157>
- 17 Avasthi P, McGeever E, Patton AH, York R. (2024). Leveraging evolution to identify novel organismal models of human biology. <https://doi.org/10.57844/ARCADIA-33B4-4DC5>
- 18 Zhang Y. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. <https://doi.org/10.1093/nar/gki524>
- 19 Morgan A, Koboldt DC, Barrie ES, Crist ER, García García G, Mezzavilla M, Faletra F, Mihalic Mosher T, Wilson RK, Blanchet C, Manickam K, Roux A, Gasparini P, Dell'Orco D, Girotto G. (2019). Mutations in *PLS1*, encoding fimbrin, cause autosomal dominant nonsyndromic hearing loss. <https://doi.org/10.1002/humu.23891>
- 20 Taylor R, Bullen A, Johnson SL, Grimm-Günter E-M, Rivero F, Marcotti W, Forge A, Daudet N. (2014). Absence of plastin 1 causes abnormal maintenance of hair cell stereocilia and a moderate form of hearing loss in mice. <https://doi.org/10.1093/hmg/ddu417>
- 21 Ros-Rocher N, Brunet T. (2023). What is it like to be a choanoflagellate? Sensation, processing and behavior in the closest unicellular relatives of animals. <https://doi.org/10.1007/s10071-023-01776-z>
- 22 Laundon D, Larson BT, McDonald K, King N, Burkhardt P. (2019). The architecture of cell differentiation in choanoflagellates and sponge choanocytes. <https://doi.org/10.1371/journal.pbio.3000226>
- 23 Li H, Roblin G, Liu H, Heller S. (2003). Generation of hair cells by stepwise differentiation of embryonic stem cells. <https://doi.org/10.1073/pnas.2334503100>
- 24 Moshous D, Martin E, Carpentier W, Lim A, Callebaut I, Canioni D, Hauck F, Majewski J, Schwartzentruber J, Nitschke P, Sirvent N, Frange P, Picard C, Blanche S, Revy P, Fischer A, Latour S, Jabado N, de Villartay J-P. (2013). Whole-exome sequencing identifies Coronin-1A deficiency in 3 siblings with immunodeficiency and EBV-associated B-cell lymphoproliferation. <https://doi.org/10.1016/j.jaci.2013.01.042>
- 25 Shiow LR, Roadcap DW, Paris K, Watson SR, Grigorova IL, Lebet T, An J, Xu Y, Jenne CN, Föger N, Sorensen RU, Goodnow CC, Bear JE, Puck JM, Cyster JG.

- (2008). The actin regulator coronin 1A is mutant in a thymic egress-deficient mouse strain and in a patient with severe combined immunodeficiency. <https://doi.org/10.1038/ni.1662>
- 26 Dayel MJ, King N. (2014). Prey Capture and Phagocytosis in the Choanoflagellate *Salpingoeca rosetta*. <https://doi.org/10.1371/journal.pone.0095577>
- 27 Al Dhaibani M, El-Hattab A, Ismayl O, Suleiman J. (2018). B3GALNT2-Related Dystroglycanopathy: Expansion of the Phenotype with Novel Mutation Associated with Muscle-Eye-Brain Disease, Walker-Warburg Syndrome, Epileptic Encephalopathy-West Syndrome, and Sensorineural Hearing Loss. <https://doi.org/10.1055/s-0038-1651519>
- 28 Stevens E, Carss KJ, Cirak S, Foley AR, Torelli S, Willer T, Tambunan DE, Yau S, Brodd L, Sewry CA, Feng L, Haliloglu G, Orhan D, Dobyns WB, Enns GM, Manning M, Krause A, Salih MA, Walsh CA, Hurles M, Campbell KP, Manzini MC, Stemple D, Lin Y-Y, Muntoni F. (2013). Mutations in B3GALNT2 Cause Congenital Muscular Dystrophy and Hypoglycosylation of α -Dystroglycan. <https://doi.org/10.1016/j.ajhg.2013.01.016>
- 29 Ducro BJ, Schurink A, Bastiaansen JWM, Boegheim IJM, van Steenbeek FG, Vos-Loohuis M, Nijman IJ, Monroe GR, Hellinga I, Dibbits BW, Back W, Leegwater PAJ. (2015). A nonsense mutation in B3GALNT2 is concordant with hydrocephalus in Friesian horses. <https://doi.org/10.1186/s12864-015-1936-z>
- 30 Wetzel LA, Levin TC, Hulett RE, Chan D, King GA, Aldayafleh R, Booth DS, Sigg MA, King N. (2018). Predicted glycosyltransferases promote development and prevent spurious cell clumping in the choanoflagellate *S. rosetta*. <https://doi.org/10.7554/elife.41482>
- 31 Plotly Technologies Inc. (2015). Collaborative data science. <https://plot.ly>
- 32 arcadia-pycolor. (2025). <https://github.com/Arcadia-Science/arcadia-pycolor>
-