**Computationally Predicted Gene Regulatory Networks in Molluscan Biomineralisation Identify Extracellular Matrix Production and Ion Transportation Pathways**

Victoria A. Sleight\*, Philipp Antczak, Francesco Falciani, Melody S. Clark

**Abstract**

**Motivation:** The molecular processes regulating molluscan shell production remain relatively uncharacterised, despite the clear evolutionary and societal importance of biomineralisation.

**Results:** Here we built the first computationally predicted gene regulatory network (GRN) for biomineralisation from the Antarctic clam (*Laternula elliptica*) mantle using gene expression data produced over an age-categorised shell damage-repair time-course. We used previously published *in vivo in situ* hybridisation expression data to ground truth gene interactions predicted by the GRN and show that candidate biomineralisation genes from different shell layers, and hence microstructures, were connected in unique modules. We characterised two biomineralisation modules of the GRN and hypothesise that biomineralisation module 1 is responsible for making the extracellular proteins required for producing, repairing or remodelling the nacreous shell layer, whereas the role of biomineralization module 2 is to orchestrate the transport of both ions and proteins to the shell secretion site, which are required during normal shell growth, and repair. Our findings demonstrate that unbiased computational methods are particularly valuable for studying fundamental biological processes and gene interactions in non-model species where rich sources of gene expression data exist, but annotation rates are poor and the ability to carry out true functional tests are still lacking.

**Availability and Implementation:** The raw RNA-Seq data is freely available for download from NCBI SRA (Accession: PRJNA398984), the assembled transcripts, reconstructed GRN, modules and detailed annotations are all available as supplementary files.

**\*Contact:** vas45@cam.ac.uk

**Supplementary Information:** Supplementary data are available at Bioinformatics online**.**

**Introduction**

The evolution of biomineralisation in the Late Precambrian corresponded with a huge expansion in morphological diversity. Representatives of all Kingdoms form biominerals, they are a substantial feature of life as we know it. One phylum which owes its success to biomineralised shell is the Mollusca (Vermeij, 2005). With over 85,000 extant species, molluscs are the second most speciose animal phyla. They are essential components in worldwide ecosystem functioning, an important source of protein for growing human demand and show biomimicry potential for the development of strong, low-energy, materials from vastly abundant soluble calcium carbonate (Finnemore, et al., 2012). In light of the clear social and economic importance of biominerals and biomineralising organisms, ocean acidification is a cause of concern (Gazeau, et al., 2013), and environmental scientists are trying to predict the fate of calcifiers under future ocean acidification scenarios. There is therefore, a clear requirement to understand fundamental mechanisms of molluscan biomineralisation, yet to date, such mechanisms are poorly characterised.

Molluscan shells form through a controlled biological process producing a composite biomaterial containing 95-99% calcium carbonate (CaCO3) and 1-5% organic matrix. In order to build their shells, molluscs transport minerals, proteins, glycoproteins, lipids and carbohydrates across the mantle to the extrapallial space, where mineral components are laid down as organised crystals onto an organic matrix. Over the last decade, mantle transcriptomes and shell proteomes have been described for many mollusc species (Clark, et al., 2010; Jackson, et al., 2006; Marie, et al., 2013; Marie, et al., 2010), in addition to RNA-Seq shell damage-repair experiments (Huning, et al., 2016; Sleight, et al., 2015). A handful of candidate biomineralization genes have identified and further characterised with a combination of qPCR, *in situ* hybridisationsand protein immune-detection (Fang, et al., 2012; Marie, et al., 2012; Sleight, et al., 2016). Annotation rates vary considerably between mollusc species from around 32-34% in the Antarctic clam and the blue mussel (*Mytilus edulis)* to 87% in the Pacific oyster (Sleight, et al., 2015; Yarra, et al., 2016), which often constrains research efforts to previously characterised, annotated sequences. Much of the molluscan shell molecular tool kit remains poorly described and many of these “unknown” genes will play important roles in the biomineralization process. There is currently no data available on the gene regulatory network (GRN) controlling biomineralisation and tools are urgently required to extend analyses beyond a small set of candidate genes and proteins.

Gene regulatory networks (GRN) can be inferred from observational data, including transcriptomics; can identify regulators of complex biological processes and, using a guilty-by-association criteria, can used to formulate hypothesis on the function of non-annotated genes. There is currently no data available on the gene regulatory network (GRN) controlling biomineralisation and such tools could progress our understanding of this complex process. To address this challenge and gain insight into molluscan biomineralisation we used a global gene expression profiling dataset and an information theoretic approach to reverse engineer gene regulatory networks and infer underlying transcriptional networks. Our analyses identified strikingly biologically meaningful networks, several gene interactions within these modules were consistent with previously published *in vivo in situ* hybridisation data. Using database and enrichment analysis methods, we characterised two biomineralisation modules of the GRN, generated new hypotheses on the regulation of molluscan biomineralisation and highlight priority candidates for future functional testing.

**Methods**

Experimental design

In order to induce biomineralisation, time-course shell damage-repair experiments were conducted on three age categories of Antarctic clam *Laternula elliptica,* (juvenile = 1-2 years, adolescent = 3-5 years, adult = > 6 years)using a previously described protocol for shell damage and mantle tissue collection (Sleight, et al., 2015). In order to use data to predict gene interactions, each experiment included perturbation (shell-damage) versus control, over a time course (experimental design summarised in Figure 1).



**Fig. 1. Experimental design of the perturbation time-course. A)** Schematic representation of experimental design, 78 individual animals sampled over six time points post-perturbation, three age categories and two treatments (damaged shell and control). B) Example of shell damage treatment to adult animal C) Example of repair over time (left to right: 1 month, 2 months, 3 months, modified from Sleight et al., (2015))

RNA-Seq and gene abundance estimation

Stranded cDNA libraries (n = 78) were prepared using the NEXTflex™ Rapid Illumina Directional RNA-Seq Library Prep Kit and sequencing was carried out over 5 lanes on a Hi-Seq 2000 generating 125 base paired-end reads.

All libraries were used to *de novo* assemble a representative mantle transcriptome (available in Supplementary File 1). Briefly, adaptor sequences and ribosomal RNA reads were removed, and resulting reads were further cleaned for quality (Phred score 30) and minimum read length (80 bp) using the ea-utils tool (v1.1.2) fastq-mcf. The cleaned reads were normalised using Trinity’s (v2.2.0) *in silico* Read Normalisation tool (Haas, et al., 2013), with default parameters. Normalised reads were *de novo* assembled using Trinity (v2.2.0) with default parameters (Grabherr, et al., 2011). The longest isoform of each gene was extracted from the transcriptome for annotation as per An *et al.* (2014) using the Trinity utility script get\_longest\_isoform\_seq\_per (assembly statistics and assembled transcripts available for download in Supplementary File 1). The longest isoforms of each gene were compared to a local NCBI non-redundant (nr) database (updated 01 June 2016) using Basic Local Alignment Search Tool (blastx, cut-off <1e-10) to search for sequence similarity and putative gene annotation (Altschul, et al., 1990).

Transcript abundance was estimated by alignment-based quantification using Trinity (v2.2.0) utilities (Grabherr, et al., 2011; Haas, et al., 2013). Transcripts were aligned to the *de novo* transcriptome using bowtie with default parameters and transcript abundance estimates were calculated using RNA-Seq by Expectation-Maximization (RSEM). A matrix of Trimmed Mean of M-values [TMM] normalised Fragments Per Kilobase Of Exon Per Million Fragments Mapped [FPKM] values was loaded into TM4 MultiExperiment Viewer (Howe, et al., 2011) and clustered based on expression profiles for each gene using SOTA (Herrero, et al., 2001).

ARACNe prediction of GRN

To construct a predicted GRN from the SOTA expression cluster profiles, ARACNe was implemented using “ARACNe for GNU/Linux” with a p-value cut-off of 1e-7 and DPI set to 0.1 (Margolin, et al., 2006). The ARACNe network output (Supplementary file 2) was loaded in Cytoscape v3.4.0 for visualisation and exploration (Shannon, et al., 2003). To identify highly interconnected sub-networks, GLay (clusterMaker) was applied (Supplementary information Fig.S1, (Morris, et al., 2011)). Gene annotations (Blastx nr, as above) were mapped onto nodes in the regulatory gene network.

*In vivo* ground truth of biomineralisation modules

We have previously identified and characterised biomineralisation gene candidates in the Antarctic clam using a combination of semi-qPCR, *in situ* hybridisation (ISH) and proteomics (Sleight, et al., 2016). Six of these candidate genes whose expression is specific to the calcifying outer epithelium of the mollusc mantle can be split into two groups. One set of translated proteins are also present in the nacreous shell proteome (*tyrA*, *chitin-binding domain, mytilin, pif*), whilst a second set do not have a translated protein detected in the nacreous shell proteome (*tyrB* and *Contig01043-unannotated transcript,* summarised in Fig. 2). These six candidate biomineralisation genes were used as molecular markers of biomineralization, they were firstly all confirmed to be present in the current data set, and were then investigated in the ARACNe predicted GRN to test for predictions of co-expression and interaction via first and second neighbour connections.

Differential expression analysis

The gene-level abundance estimates (raw counts) for each of the libraries were constructed into a matrix for downstream expression analyses (using the Trinity abundance\_estimates\_to\_matrix.pl script). Differentially expressed genes were identified using the Bioconductor (v3.4) edgeR package in R (v3.1.1) with a false discovery rate (FDR) of 5% (Robinson et al. 2010, McCarthy et al. 2012) and genes which showed significant temporal changes in response to damage, termed here, time-dependant damage-response genes, were identified. Genes that qualified as a time-dependent damage response gene in any of the three age categories were mapped onto the network. To investigate if the biomineralisation modules and priority candidates were enriched in time-dependent damage-response genes, all genes were ranked by FDR and the distributions of significant genes were assessed for different module conditions in each age category.

Enrichment analysis and identification of priority candidates using biomineralisation markers

The two modules that contained our previously defined biomineralisation markers were investigated. All of the trinity transcripts within each module were extracted and re-annotated using three databases: NCBI nr, Uniref90 and Swissprot human (Supplementary Files 3 & 4). Annotations were then analysed using StringDB and tested for enrichment against a whole genome background. In order to perform an *in silico* screen for priority candidate genes within the two biomineralization modules two criteria were used: firstly, genes which had GO terms relating to functional categories with potential relevance to biomineralization were identified (categories = receptors and signal transduction, ion transport, shell matrix proteins, biomineralisation enzymes and protein folding/chaperoning) and secondly, using literature searches, each gene was assessed to identify if it had previously been reported to play a role in the regulation of biomineralisation in other systems.

**Results and Discussion**

First computationally predicted mantle-specific GRN resource for molluscs

We present the first computationally predicted GRN for molluscan biomineralisation constructed from age-categorised damage-repair mantle gene expression data (Supplementary File 2, Fig. 2, Supplementary information Fig.S1). Owing to the use of mantle tissue-specific gene expression profiles the presented gene network is, to a large extent, constrained by the biomineralisation process. The treatment of shell-damage and the multifunctional nature of the mantle however, makes it challenging to experimentally differentiate between fundamental biomineral production and other processors in the shell and mantle, namely the immune response. The GRN analyses attempts to simplify the interactions of pleiotropic genes and highlights their role in biomineralisation, these genes are often obscured in standard differential expression analyses (such as those including environmental stressors) and indeed many of the genes identified in our network analysis were not significant in the differential expression analysis also performed, highlighting the strength of the network method (Fig. 2, Supplementary information Fig.S2). Currently, there are few functional tools available to study any member of the molluscan phyla. Only a handful of papers have documented RNAi (Suzuki, et al., 2009; Zhao, et al., 2014; Zhao, et al., 2016), and genome editing via CRISPR/Cas9 has been reported for a single gene in only two species of the Mollusca, *Crepidula fornicata* and *Lymnaea stagnalis* (Perry and Henry, 2015; Abe and Kuroda, 2019). Given the lack of tools available to study fundamental biological questions in this phyla, computational tools are particularly valuable and hence we believe the data presented here represents a significant resource for the community, which will underpin both hypothesis generation and unbiased discovery methods.

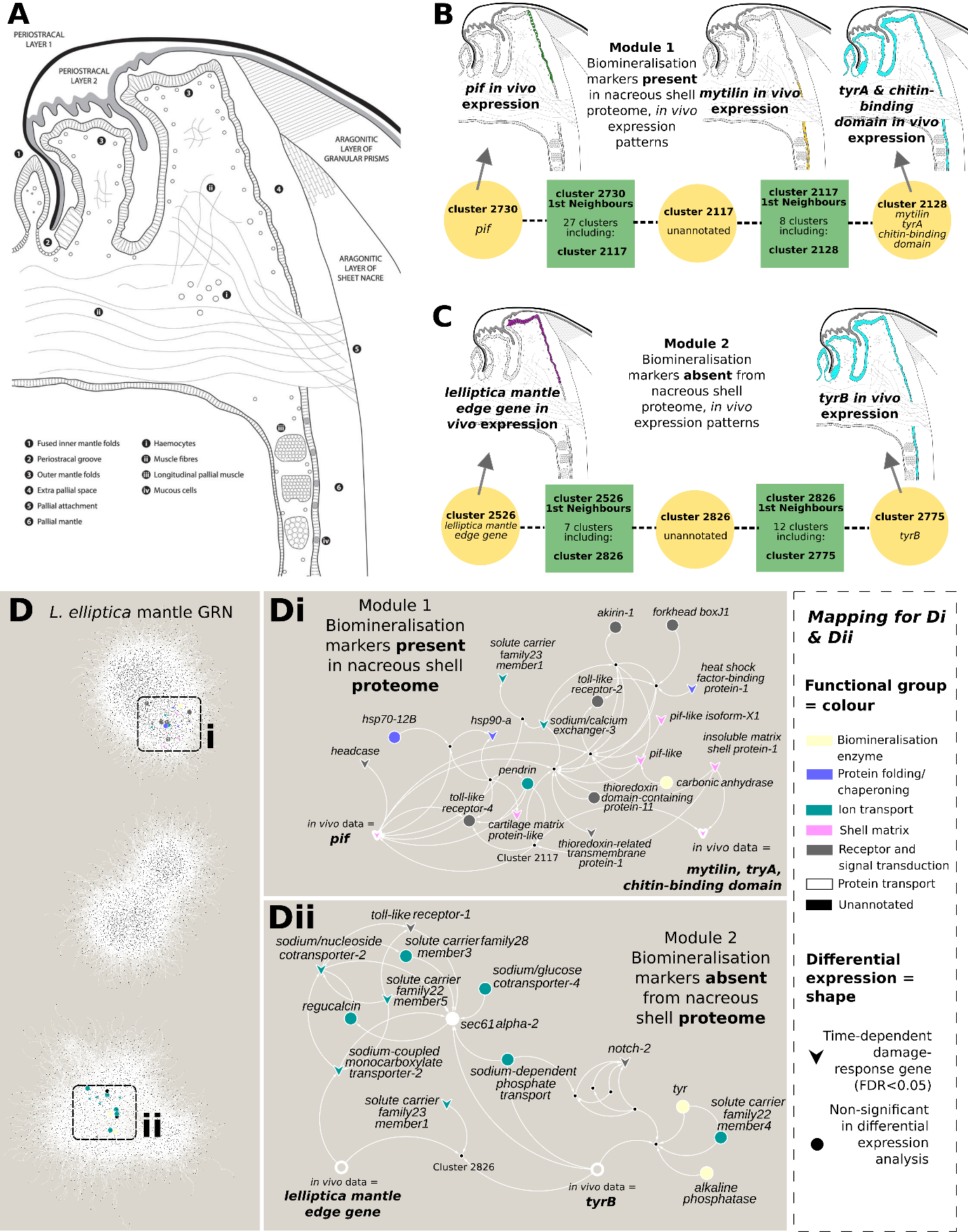
The limitations of predicting complex biological interactions from gene expression data alone are well known (Banf and Rhee, 2017), especially in the absence of physical regulatory models such as DNA-binding motifs and chromatin accessibility maps, as is the case here. It is therefore important to highlight that this resource is proposed as an exploratory tool to provide unbiased insights that complement other avenues of investigation. Below we show examples of how our biomineralisation GRN can be used in combination with *in vivo* data to provide insight into regulatory mechanisms, generate new hypotheses and highlight future priority candidates.

Co-expression genes are clustered by SOTA

SOTA was used to cluster 199,321 Trinity genes into 18,862 expression clusters based on shared expression profiles. We demonstrate that expression clusters generated by SOTA include genes which are co-expressed at the cellular level *in vivo.* For example, expression cluster 2128 includes three of our candidate biomineralisation marker genes (*mytilin, tyrA* and *chitin-binding*), which we have previously shown to have overlapping expression in the calcifying epithelium of the mollusc mantle (Fig. 2).

ARACNe predicts interaction between *in vivo* ground truthed biomineralisation markers generating unique modules correlated to shell microstructure

Expression cluster profiles from SOTA were used as an input for the network inference algorithm ANACNe, which is an information theory-based approach discriminating between direct and indirect gene-to-gene relationships. Here, we term the resulting relevance network, a computationally predicted GRN. We discovered that our six biomineralisation marker genes fall into two modules of the GRN (Fig. 2). The candidates that are all present in the nacreous shell proteome are connected via a mutual first neighbour (unannotated expression cluster 2117, module 1). Likewise, in a separate module, the candidates which are absent from the nacreous proteome, but are still co-expressed at the cellular level in the mantle epithelium, are also connected by an unannotated mutual first neighbour (unannotated expression cluster 2826, module 2). These two unique biomineralisation modules suggest that genes involved in the production of the nacreous shell layer are in a separate regulatory network to those of other shell layers, and hence microstructures. Previous immunolocalizations of shell matrix proteins corresponding to specific shell layers and mantle regions have also suggested that different secretory repertoires control the biomineralisation processes of each layer (Marie, et al., 2012), and in addition, that the mantle organ is modular in nature both at the anatomical and molecular level (Jackson, et al., 2006; Sleight, et al., 2016). The computationally predicted GRN data here provide further evidence for both of these hypotheses, in addition to highlighting, for the first time, the possible regulatory networks involved.



**Fig. 2.** **Biomineralisation modules of the GRN ground-truthed with previously published *in situ* hybridisation *in vivo* expression data.** A) Schematic of *L. elliptica* mantle anatomy and shell (reproduced with permission from Sleight, et al., (2016)). B) Module 1 is involved in the secretion of nacreous shell layer. C) Module 2 is likely involved in the transportation of ions and proteins required for both nacreous and prismatic shell layers. D) Priority gene candidates from modules 1 (i) and 2 (ii) and their closest connections, colour and shape mapping functional group and differential expression status, respectively. More information on all of the transcripts involved in these modules, their annotations, as well as the sequence data for the transcripts, is available in Supplementary Files 3&4 for module 1&2 respectively.

Hypothesised functional roles of the biomineralisation modules: extracellular matrix production and transportation

The two biomineralisation GRN modules contained both relatively well-characterized candidate biomineralisation genes, in addition to numerous unannotated transcripts. True unbiased genetic screens to test the function of large numbers of genes, for example all of the 1190 genes present in biomineralisation modules 1 and 2, are currently impossible in non-model organisms. Instead, annotations were explored using database-based tools, such as StringDB and enrichment analysis. Finally genes were *in silico* screened for priority candidates that we hypothesise are crucial for biomineralisation (Table 1).

**Table 1.** Hypothesised functional role and new priority annotated candidates from each of the biomineralisation modules in the computationally predicted mantle GRN.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Module** | **GRN Cluster ID** | **Trinity ID** | **Putative annotation** | **Functional group** |
| **Module 1**  **Functional hypothesis =** *responsible for transcribing and translating the extracellular proteins required for biomineralisation of nacreous shell layer (see* *Supplementary information Table S1 and Fig.S3)* | Cluster-9690.txt | TRINITY\_DN255896\_c3\_g2 | Toll-like receptor 4 | Receptor and signal transduction |
| Cluster-9690.txt | TRINITY\_DN245288\_c2\_g1 | Toll-like receptor 2 |
| Cluster-2849.txt | TRINITY\_DN246350\_c2\_g1 | Toll-like receptor 4 |
| Cluster-2128.txt | TRINITY\_DN248655\_c2\_g5 | Neurogenic locus notch homolog protein 2 |
| Cluster-10028.txt | **TRINITY\_DN251622\_c0\_g1** | Headcase protein homolog |
| Cluster-10469.txt | TRINITY\_DN251675\_c0\_g1 | Forkhead box protein J1 |
| Cluster-2128.txt | TRINITY\_DN253939\_c2\_g1 | Fibroblast growth factor receptor 3 |
| Cluster-18226.txt | TRINITY\_DN242791\_c2\_g1 | Thioredoxin domain-containing protein 11 |
| Cluster-9587.txt | **TRINITY\_DN256751\_c1\_g1** | Thioredoxin-related transmembrane protein 1 |
| Cluster-15239.txt | TRINITY\_DN246576\_c0\_g1 | Akirin-1 |
| Cluster-2346.txt | **TRINITY\_DN259824\_c2\_g1** | Solute carrier family 23 member 1 | Ion transport |
| Cluster-10050.txt | **TRINITY\_DN256986\_c3\_g4** | Sodium/calcium exchanger 3 |
| Cluster-10335.txt | TRINITY\_DN246390\_c1\_g1 | Pendrin |
| Cluster-13458.txt | **TRINITY\_DN251420\_c3\_g1** | cartilage matrix protein-like | Shell matrix |
| Cluster-12543.txt | **TRINITY\_DN242528\_c0\_g1** | Insoluble matrix shell protein 1 |
| Cluster-9662.txt | **TRINITY\_DN250104\_c5\_g8** | PIF-like isoform X1 |
| Cluster-9665.txt | **TRINITY\_DN244470\_c1\_g2** | PIF-like |
| Cluster-2079.txt | TRINITY\_DN256420\_c0\_g2 | Carbonic anhydrase 2-like isoform X3 | Biomineralisation enzyme |
| Cluster-10478.txt | TRINITY\_DN239234\_c0\_g1 | Heat shock 70 kDa protein 12B | Protein folding/chaperoning |
| Cluster-9701.txt | **TRINITY\_DN249833\_c1\_g2** | Heat shock protein HSP 90-alpha |
| Cluster-16489.txt | **TRINITY\_DN255921\_c0\_g1** | Heat shock factor-binding protein 1 |
| **Module 2**  **Functional hypothesis =** *responsible for the transport of both ions and proteins to biomineralisation site (see* *Supplementary information Table S2 and Fig.S4)* | Cluster-3198.txt | TRINITY\_DN259845\_c3\_g3 | Alkaline phosphatase, tissue-nonspecific | Biomineralisation enzyme |
| Cluster-2772.txt | TRINITY\_DN256081\_c2\_g5 | Tyrosinase |
| Cluster-71.txt | **TRINITY\_DN253708\_c3\_g3** | Neurogenic locus notch homolog protein 2 | Receptor |
| Cluster-2764.txt | **TRINITY\_DN248447\_c0\_g2** | Toll-like receptor 1 |
| Cluster-6852.txt | TRINITY\_DN255895\_c0\_g3 | Protein transport protein Sec61 alpha-2 | Protein transport |
| Cluster-687.txt | TRINITY\_DN244879\_c4\_g1 | Regucalcin | Ion transport |
| Cluster-6827.txt | TRINITY\_DN237157\_c0\_g1 | Sodium/glucose cotransporter 4 |
| Cluster-2631.txt | **TRINITY\_DN231473\_c0\_g2** | Sodium/nucleoside cotransporter 2 |
| Cluster-2568.txt | TRINITY\_DN239292\_c0\_g1 | Sodium-coupled monocarboxylate transporter 2 |
| Cluster-606.txt | TRINITY\_DN247229\_c2\_g1 | Sodium-dependent phosphate transport protein |
| Cluster-596.txt | TRINITY\_DN253103\_c2\_g3 | Solute carrier family 22 member 4 |
| Cluster-10275.txt | **TRINITY\_DN255933\_c0\_g1** | Solute carrier family 22 member 5 |
| Cluster-2690.txt | **TRINITY\_DN245321\_c4\_g2** | Solute carrier family 23 member 1 |
| Cluster-2652.txt | TRINITY\_DN250916\_c0\_g3 | Solute carrier family 28 member 3 |

The first biomineralisation module (module 1) contained four of our biomineralisation markers (*pif, mytilin, tryA* and *chitin-binding-domain*) all of which encode proteins that are present in the nacreous shell proteome, these genes were connected by a mutual second neighbour (Fig. 2). Module 1 comprised of 36 GRN clusters made-up of 1023 trinity genes (Supplementary File 3), 467 of the trinity genes could be assigned putative annotation after sequence similarity searches against NCBI non-redundant, uniref90 and swissprot human databases. Enrichment analysis revealed module 1 was significantly enriched in KEGG pathways, molecular functions (GO), biological processes (GO) and cellular components (GO) related to protein production and turnover (e.g. transcription, translation, ribosomes, Supplementary information, Fig.S3 & Table S1), and we therefore hypothesise this module is responsible for making the extracellular proteins required for repairing and remodelling the damaged nacreous shell layer. Although enrichment analysis suggests the primary function of module 1 is the production of extracellular matrix components, we also found genes in this module related to ion transport. One explanation for the ion transporters in this module could be that they are specific to the nacreous shell layer and creating the correct conditions for nacreous tablet formation, where as module 2 has a more general role in transport of all ions and proteins for both of the calcified shell layers and the periostracum. From module 1 we have highlighted twenty one priority gene candidates for future characterisation (Table 1).

The second biomineralisation module (module 2) contained two of our biomineralisation markers (*tryB* and *contig01043* the “*unannotated transcript”*) which have specific expression in the mantle epithelial/shell secreting cells but are absent from the nacreous shell proteome. Similar to module 1, these two genes were connected by a mutual second neighbour (Fig. 2.). Module 2 comprised of 23 GRN clusters made-up of 167 trinity genes (Supplementary File 4), 63 of the trinity genes could be assigned putative annotation after sequence similarity searches. Further exploration of these annotations revealed the module was functionally enriched in ion and protein transport genes, as well as containing enzymes that catalyse mineralisation and possible upstream receptors and ion exchange regulators (Supplementary information Fig.S4 & Table S2). Taken together with the lack of translated products present in the nacreous shell proteome, we hypothesise that the primary functional role of module 2 is to orchestrate the transport of both ions and proteins to the shell secretion site, which are required during normal shell growth, and repair. From module 2 we have highlighted fourteen biomineralisation priority gene candidates for future characterisation (Table 1).

*In silico* screen for priority candidates for future genome editing in the Mollusca

Both of the biomineralisation modules were *in silico* screened for five functional categories of genes: receptors and signal transduction, ion transport, shell matrix proteins, biomineralisation enzymes and protein folding/chaperoning, in addition to literature searches for genes known to be involved in biomineralisation in other systems. Using this two-step screen criteria we were able to identify genes novel to the context of biomineralisation (such as *toll-like receptors*)*,* as well as genes that have a well-established role in biomineralisation in other organisms but, have not previously reported in molluscs (such as *SLC23A1*).

*Receptors and signal transduction*

The presence of receptors in molluscan shell proteomes was recently discussed by Herlitze et al., 2018 (2018), it is speculated that signalling molecules could be incorporated into the shell to allow the acellular structure a means by which to communicate with the underlying mantle epithelium. If the shell is damaged, for example, these signalling molecules would be released and detected by the mantle, triggering the upregulation of calcification to repair the shell. In addition to being a possible means for feedback between the extracellular shell and the mantle, receptors are likely to be pivotal in the regulation of mantle modularity, normal shell growth, and repair. Genes encoding receptors and proteins involved in signal transduction were in both biomineralisation modules. Four of these were *toll-like receptors* that have been previously found in many molluscan mantle transcriptomes, where they are hypothesised to be functioning as pattern recognition receptors for the identification of pathogens as part of the innate immune system (Zheng, et al., 2005). More recently toll-like receptors have also been identified in shell proteomes, where they have been speculated to be a part of a phenoloxidase precursor to immunity pathways such as melanization, phagocytosis, capsulation, opsonisation (Arivalagan, et al., 2017).

Receptors and signal transduction genes important in development and regeneration were also identified in both of the biomineralisation modules. Two candidates were putatively annotated as *notch2*, a cell-cell communication receptor widely important for metazoan (including molluscan) development (De Oliveira, et al., 2016), and more specifically pivotal in bone development and regeneration in vertebrates (Shao, et al., 2018). *Notch2* has also been reported in the developing shell gland of a gastropod mollusc, *Crepidula fornicata (Perry, et al., 2015).* Further inspection of the *notch2* alignments here however, revealed the transcripts contained a repeated string of epidermal growth factor (EGF) motifs matching just the extracellular domain of notch, rather than the full complement of notch domains. We therefore speculate the new candidates could be novel receptors involved in cell-cell communication, which require more functional testing. Two additional genes that are also known to be involved in the development of vertebrate teeth and bones were highlighted from module 1; *forkhead box protein J1* (Jheon, et al., 2013) and *fibroblast growth factor receptor 3* (Colvin, et al., 1996; Tapaltsyan, et al., 2016). The expression of other *fox* genes, such as *foxa*, have been documented in the gastropod developing shell gland (Perry, et al., 2015), but to our knowledge there is very little data to date on the involvement of classic vertebrate biomineral development genes in the regulation of mollusc biomineralisation. Given a number of these vertebrate genes were found in both of the biomineralisation modules presented here, we speculate they could be involved in the development and regeneration of molluscan shell, highlighting the requirement for more functional testing and comparative work to elucidate a possible deeply conserved metazoan biomineralisation GRN.

*Ion transport*

In order to calcify the shell, molluscs must transport calcium ions across the mantle tissue to the extrapallial fluid and shell depositions site (Sillanpää, et al., 2018). Genes that are likely to be involved in this transport process were found in both of the biomineralisation modules. Three ion transport candidates found in module 1, all of which transport ions potentially relevant for molluscan biomineralision. Firstly, *solute carrier family 23 member 1* (*SLC23A1*) is a vitamin C transporter which mediates the uptake of Vitamin C in exchange for sodium ions. In vertebrate bone Vitamin C is an essential cofactor for collagen assembly (Aghajanian, et al., 2015), and in addition, recent studies in coral larvae show *SLC23A1* it is the single most up-regulated gene at the onset of larval calcification (Rosenthal, et al., 2016). Secondly, *sodium/calcium exchanger 3* is an antiporter membrane protein that removes calcium from cells, it has been proposed as a candidate for biomineralisation in both echinoderms (Flores and Livingston, 2017) and molluscs (Shi, et al., 2013). And thirdly, *pendrin* (*SLC26A4*) is a sodium-independent transporter of various ions, including bicarbonates required for shell biomineralisation. Mutations to the pendrin gene in mice and humans lead to a variety of deafness phenotypes, such as reduced calcification of the inner ear bones (Dror, et al., 2014).

Module 2 contained eight biomineralisation candidates putatively annotated as ion transporters. Similar to module 1, some of these ion transporters have previously been related to biomineralisation in other systems, whereas some are novel in the context on biomineralisation. In addition to the ion channel/transport genes module 2 also contained regucalcin. Regucalcin is pivotal to the regulation of calcium ion homeostasis and biomineralisation in vertebrates (Yamaguchi, 2014), which also shows high expression levels in other mollusc mantle transcriptomes (Li, et al., 2017; Shi, et al., 2013). We speculate that similar to vertebrates, regucalcin could be acting as a regulator of ion transport activity in this module.

*Protein transport*

Similar to the transport of ions, proteins in the extracellular matrix of the shell also need to be transported across the mantle to the shell deposition site, and therefore we screened the two biomineralisation modules for protein transport genes. Module 2 included one gene putatively annotated as a protein transporter, *sec61 alpha-2*, which is a component of the sec dependent transport pathway. This pathway is a general, highly conserved, protein export system, transporting newly synthesized proteins into or across the cell membrane (Linxweiler, et al., 2017; Rapoport, 2007). We hypothesise molluscs could use this pathway to transport extracellular matrix proteins across the mantle epithelial membrane, to the extrapallial fluid and biomineralisation site.

*Protein folding/chaperoning*

The role of protein folding and heat shock proteins (HSPs) in molluscan biomineralisation is relatively unexplored. Our previous work has demonstrated high levels of *hsp* gene expression, both constitutively and induced in response to shell damage and heat stress, in the mantle tissue of two bivalve species (*L. ellipitca* and *Mya truncata*) (Clark, et al., 2010; Sleight, et al., 2018; Sleight, et al., 2015; Sleight, et al., 2016). In the present study, the computationally predicted GRN allocated three hsps into module 1: *Heat shock 70 kDa protein 12B*, *HSP 90-alpha*, *Heat shock factor-binding protein 1*, leading us to hypothesise that hsps are important in the folding and chaperoning of shell matrix proteins during shell repair and remodelling. In addition to our own work on marine invertebrates, it was recently demonstrated in mice that *HSP 90-alpha* is secreted and directly influences remodelling of the extracellular matrix during wound repair (Bhatia, et al., 2018).

*Biomineralisation-specific enzymes*

Enzymes are crucial for the production of all biominerals (Weiss and Marin, 2008), they can be categorised in many different ways based on type of biochemical catalytic activity, or biological function carried out, from protein synthesis to signalling cascascades, ion transport and mineral nucleation. Here we describe the genes in modules 1 and 2 encoding enzymes that can be categorised as specific to mineralisation (according to (Weiss and Marin, 2008)).

Carbonic anhydrase catalyses the hydration of carbon dioxide and accelerates the formation of bicarbonate ions, it is thought of as a deeply conserved requirement for biomineralisation across many taxa (Le Roy, et al., 2014) and carbonic anhydrase domains are often components of multi-functional structural matrix proteins, for example nacrein (Miyamoto, et al., 1996). In module 1 we found one gene encoding a carbonic anhydrase, we hypothesise this enzyme is therefore synthesised and secreted, along with other shell matrix proteins in this module, into the extrapallial space, where it catalyses the production of bicarbonate ions required for mollusc shell mineralisation.

Similar to carbonic anhydrase, alkaline phosphatase is common to many mineralising taxa, it hydrolyses pyrophosphate and provides inorganic phosphate ions to promote mineralization (Golub and Boesze-Battaglia, 2007). Alkaline phosphatase activity plays a role in molluscan shell production, and its activity can be used a marker of biomineralising cells (Hohagen and Jackson, 2013).We found *alkaline phosphatase* in module 2, along with other regulators of ions such as, ion transporters, and therefore highlight it as an important component of biomineralisation governed by the regulatory genes in module 2.

Contrary to the two enzymes discussed above that catalyse ion production, tyrosinase enzymes catalyse the production of quinones. Quinones then bind amino acids to form protein cross-linkages leading to the formation of insoluble proteins, such as those present in the uncalcified cuticle layer covering the molluscan shell, the periostracum (Aguilera, et al., 2014). The insoluble periostracum seals the extrapallial space to isolate it from the external environment, enabling super saturation conditions inside. It has also been speculated as an initial substrate for biomineralisation, in addition to being critically important for shell damage-repair as a means to seal the damaged area before remodelling and new mineralisation can occur. Tyrosinase domains have been proposed as evolutionary maintained and conserved components required for all molluscan biomineralisation (Arivalagan, et al., 2017), and we have previously characterised two *tyrosinases* in the *L. elliptica* mantle transcriptome, which show similar spatial expression patterns in the mantle epithelium, but only one, *tyrA* has a translated protein product present in the shell proteome (Fig 2). The two previously characterised *tryrosinase* genes were used as biomineralisation markers in this GRN and it discovered that *tryA* was included in module 1, along with other shell matrix proteins in the nacreous shell proteome, whereas *tryB* was included in module 2 along with other biomineralisation markers absent from the proteome. The inclusion of the two copies into different biomineralisation modules further supports the hypothesis that this gene duplication was followed by sub-functionalisation. The presence of an additional transcript with sequence similarity to tyrosinase in module 2 indicates there are more than two copies in *L. elliptica*, which is perhaps not surprising as many mollusc species have large gene expansions of this family (Aguilera, et al., 2014) and we sequenced 78 individual mantle tissues significantly increasing the RNA-Seq data for this species, hence increasing our likelihood of finding novel genes.

*Shell matrix*

In addition to the four previously characterised nacreous shell matrix proteins present in module 1, we highlight four additional transcripts which are also likely to encode shell matrix proteins, but were not previously found in *L. elliptica* shell proteome or mantle transcriptome studies.

Using the GRN we have highlighted many new priority candidates which are novel in the context of molluscan biomineralisation, we stress that the role of these genes in biomineralisation is, at this stage, speculative and awaits true functional testing. In addition to the priority candidates, the two biomineralisation modules described contained a total of 660 transcripts that had no sequence similarity to anything in the four databases searched (556 transcripts in module 1 and 104 transcripts in module 2). Given we cannot use database methods to assess the likely functional role of these genes it is difficult to highlight specific candidates of interest for future work. Clearly these unannotated genes are important for biomineralisation, as our network analysis has included them in biomineralisation modules of the GRN and we therefore propose another screening criteria for unannotated genes so that they canalso be included in future functional work. Firstly the time-dependent damage-response analysis could be used to filter the list of unannotated genes, and secondly genes with a high threshold of connectivity to our biomineralisation markers and new priority candidates could e selected.

**References**

Abe, M., and Kuroda, R. The development of CRISPR for a mollusc establishes the formin Lsdia1 as the long-sought gene for snail dextral/sinistral coiling. *Development* 2019;146(9): dev175976.

Aghajanian, P.*, et al.* The Roles and Mechanisms of Actions of Vitamin C in Bone: New Developments. *J Bone Miner Res* 2015;30(11):1945-1955.

Aguilera, F., McDougall, C. and Degnan, B.M. Evolution of the tyrosinase gene family in bivalve molluscs: Independent expansion of the mantle gene repertoire. *Acta biomaterialia* 2014;10(9):3855-3865.

Altschul, S.F.*, et al.* Basic local alignment search tool. *J Mol Biol* 1990;215(3):403-410.

An, J.*, et al.* Transcriptome Profiling to Discover Putative Genes Associated with Paraquat Resistance in Goosegrass (Eleusine indica L.). *Plos One* 2014;9(6).

Arivalagan, J.*, et al.* Insights from the Shell Proteome: Biomineralization to Adaptation. *Mol Biol Evol* 2017;34(1):66-77.

Banf, M. and Rhee, S.Y. Computational inference of gene regulatory networks: Approaches, limitations and opportunities. *Biochim Biophys Acta* 2017;1860(1):41-52.

Bhatia, A.*, et al.* Extracellular and Non-Chaperone Function of Heat Shock Protein− 90α Is Required for Skin Wound Healing. *Journal of Investigative Dermatology* 2018;138(2):423-433.

Clark, M.S.*, et al.* Insights into shell deposition in the Antarctic bivalve Laternula elliptica: gene discovery in the mantle transcriptome using 454 pyrosequencing. *Bmc Genomics* 2010;11.

Colvin, J.S.*, et al.* Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3. *Nat Genet* 1996;12(4):390-397.

De Oliveira, A.L.*, et al.* Comparative transcriptomics enlarges the toolkit of known developmental genes in mollusks. *BMC Genomics* 2016;17(1):905.

Dror, A.A.*, et al.* Atrophic thyroid follicles and inner ear defects reminiscent of cochlear hypothyroidism in Slc26a4-related deafness. *Mamm Genome* 2014;25(7-8):304-316.

Fang, D.*, et al.* Novel Basic Protein, PfN23, Functions as Key Macromolecule during Nacre Formation. *Journal of Biological Chemistry* 2012;287(19):15776-15785.

Finnemore, A.*, et al.* Biomimetic layer-by-layer assembly of artificial nacre. *Nature Communications* 2012;3.

Flores, R.L. and Livingston, B.T. The skeletal proteome of the sea star Patiria miniata and evolution of biomineralization in echinoderms. *BMC Evol Biol* 2017;17(1):125.

Gazeau, F.*, et al.* Impacts of ocean acidification on marine shelled molluscs. *Marine Biology* 2013;160(8):2207-2245.

Golub, E.E. and Boesze-Battaglia, K. The role of alkaline phosphatase in mineralization. *Current Opinion in Orthopaedics* 2007;18(5):444-448.

Grabherr, M.G.*, et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 2011;29(7):644-U130.

Haas, B.J.*, et al.* De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature protocols* 2013;8(8):10.1038/nprot.2013.1084.

Herlitze, I.*, et al.* Molecular modularity and asymmetry of the molluscan mantle revealed by a gene expression atlas. *Gigascience* 2018;7(6).

Herrero, J., Valencia, A. and Dopazo, J. A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics* 2001;17(2):126-136.

Hohagen, J. and Jackson, D.J. An ancient process in a modern mollusc: early development of the shell in Lymnaea stagnalis. *Bmc Developmental Biology* 2013;13:13.

Howe, E.A.*, et al.* RNA-Seq analysis in MeV. *Bioinformatics* 2011;27(22):3209-3210.

Huning, A.K.*, et al.* A shell regeneration assay to identify biomineralization candidate genes in mytilid mussels. *Marine Genomics* 2016;27:57-67.

Jackson, D.J.*, et al.* A rapidly evolving secretome builds and patterns a sea shell. *Bmc Biology* 2006;4.

Jheon, A.H.*, et al.* From molecules to mastication: the development and evolution of teeth. *Wiley Interdiscip Rev Dev Biol* 2013;2(2):165-182.

Le Roy, N.*, et al.* The evolution of metazoan alpha-carbonic anhydrases and their roles in calcium carbonate biomineralization. *Frontiers in Zoology* 2014;11.

Li, H.*, et al.* Identification and Differential Expression of Biomineralization Genes in the Mantle of Pearl Oyster Pinctada fucata. *Mar Biotechnol (NY)* 2017;19(3):266-276.

Linxweiler, M., Schick, B. and Zimmermann, R. Let’s talk about Secs: Sec61, Sec62 and Sec63 in signal transduction, oncology and personalized medicine. *Signal Transduction And Targeted Therapy* 2017;2:17002.

Margolin, A.A.*, et al.* ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *Bmc Bioinformatics* 2006;7.

Marie, B.*, et al.* The shell-forming proteome of Lottia gigantea reveals both deep conservations and lineage-specific novelties. *FEBS J* 2013;280(1):214-232.

Marie, B.*, et al.* Different secretory repertoires control the biomineralization processes of prism and nacre deposition of the pearl oyster shell. *Proceedings of the National Academy of Sciences of the United States of America* 2012;109(51):20986-20991.

Marie, B.*, et al.* Proteomic analysis of the organic matrix of the abalone Haliotis asinina calcified shell. *Proteome Science* 2010;8.

McCarthy, D. J., Chen, Y. S. and Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation*. Nucleic Acids Research* 2012;40(10):4288-4297.

Miyamoto, H.*, et al.* A carbonic anhydrase from the nacreous layer in oyster pearls. *Proceedings of the National Academy of Sciences of the United States of America* 1996;93(18):9657-9660.

Morris, J.H.*, et al.* clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *Bmc Bioinformatics* 2011;12.

Perry, K.J. and Henry, J.Q. CRISPR/Cas9-mediated genome modification in the mollusc, Crepidula fornicata. *Genesis* 2015;53(2):237-244.

Perry, K.J.*, et al.* Deployment of regulatory genes during gastrulation and germ layer specification in a model spiralian mollusc Crepidula. *Dev Dyn* 2015;244(10):1215-1248.

Rapoport, T.A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 2007;450(7170):663-669.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26(1):139-140.

Rosenthal, J., Roberson, L. and Vazquez, N. A Possible Role for Vitamin C in Coral Calcification. In, *American Geophysical Union, Ocean Sciences Meeting 2016, abstract# AH11A-02*. 2016.

Shannon, P.*, et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003;13(11):2498-2504.

Shao, J.*, et al.* Notch expressed by osteocytes plays a critical role in mineralisation. *J Mol Med (Berl)* 2018;96(3-4):333-347.

Shi, Y.*, et al.* Characterization of the Pearl Oyster (Pinctada martensii) Mantle Transcriptome Unravels Biomineralization Genes. *Marine Biotechnology* 2013;15(2):175-187.

Sillanpää, J.K., Sundh, H. and Sundell, K.S. Calcium transfer across the outer mantle epithelium in the Pacific oyster,. *Proc Biol Sci* 2018;285(1891).

Sleight, V.A.*, et al.* An Antarctic molluscan biomineralisation tool-kit. *Sci Rep* 2016;6:36978.

Sleight, V.A.*, et al.* Cellular stress responses to chronic heat shock and shell damage in temperate Mya truncata. *Cell Stress Chaperones* 2018.

Sleight, V.A.*, et al.* Transcriptomic response to shell damage in the Antarctic clam, Laternula elliptica: time scales and spatial localisation. *Mar Genomics* 2015;20:45-55.

Sleight, V.A.*, et al.* Characterisation of the mantle transcriptome and biomineralisation genes in the blunt-gaper clam, Mya truncata. *Marine Genomics* 2016;27:47-55.

Suzuki, M.*, et al.* An Acidic Matrix Protein, Pif, Is a Key Macromolecule for Nacre Formation. *Science* 2009;325(5946):1388-1390.

Tapaltsyan, V.*, et al.* Identification of novel Fgf enhancers and their role in dental evolution. *Evol Dev* 2016;18(1):31-40.

Vermeij, G.J. Shells inside out: The architecture, evolution and function of shell envelopment in molluscs. *Evolving Form and Function: Fossils and Development* 2005:197-221.

Weiss, I., M. and Marin, F. The role of enzymes in biomineralization processes. In: A. Sigel, H.S. and Sigel, R.K.O., editors, *Biomineralization: From Nature to Application*. Wiley; 2008. p. 71-126.

Yamaguchi, M. The role of regucalcin in bone homeostasis: involvement as a novel cytokine. *Integr Biol (Camb)* 2014;6(3):258-266.

Yarra, T.*, et al.* Characterization of the mantle transcriptome in bivalves: Pecten maximus, Mytilus edulis and Crassostrea gigas. *Mar Genomics* 2016;27:9-15.

Zhao, M.*, et al.* A homeodomain transcription factor gene, PfMSX, activates expression of Pif gene in the pearl oyster Pinctada fucata. *PLoS One* 2014;9(8):e103830.

Zhao, M.*, et al.* Functional characterization and molecular mechanism exploration of three granulin epithelin precursor splice variants in biomineralization of the pearl oyster Pinctada fucata. *Mol Genet Genomics* 2016;291(1):399-409.

Zheng, L.*, et al.* Toll-like receptors in invertebrate innate immunity. *Invertebrate Survival J* 2005;2:105-113.