RESEARCH

Gene co-expression network analysis facilitates identification of novel important genes for the moulting process of the salmon louse (Lepeophtheirus salmonis)

Zhaoran Zhou[1], Christiane Eichner[2], Frank Nilsen[2], Inge Jonassen[1] and Michael Dondrup[1*]

Correspondence:
nichael.dondrup@uib.no
Department of Informatics,
Jniversity of Bergen,
Thormøhlensgate 55, Bergen,
lorway
ull list of author information is
vailable at the end of the article

Abstract

Background: The salmon louse (*Lepeophtheirus salmonis*) is an obligate ectoparasitic copepod, living on Atlantic salmon and other salmonids in the marine environment. Salmon lice cause a number of environmental problems and lead to large economical losses in aquaculture every year. In order to develop novel parasite control strategies, a better understanding of the mechanisms of moulting and development of the salmon louse at the transcriptional level is required.

Method: Three weighted gene co-expression networks were constructed based on the pairwise correlations of salmon louse gene expression profiles at different life stages. Network-based approaches and gene annotation information were applied to identify genes that might be important for the moulting and development of the salmon louse. RNA interference was performed for validation. Regulatory impact factors were calculated for all the transcription factor genes by examining the changes in co-expression patterns between transcription factor genes and deferentially expressed genes in middle stages and moulting stages.

Results: Eight gene modules were predicted as important, and 10 genes from six of the eight modules have been found to show observable phenotypes in RNA interference experiments. We knocked down five hub genes from three modules and observed phenotypic consequences in all experiments. In the infection trial, no copepodids with the RAB1A-like gene knocked down were found on fish, while control samples developed to chalimus-1 larvae. The FOXO-like gene obtained highest scores in the regulatory impact factor calculation.

Conclusions: We proposed a gene co-expression network-based approach to identify genes playing an important role in the moulting and development of salmon louse. The RNA interference experiments confirmed the effectiveness of our approach and demonstrated the indispensable role of RAB1A-like gene in the development of salmon louse. In addition to salmon louse, this approach could be generalized to identify important genes associated with a phenotype of interest in other organisms.

Keywords: gene co-expression network analysis; regularized logistic regression; differential co-expression analysis; regulatory impact factor; transcriptome; salmon louse; moult; RAB1A

Background

Copepods have been suggested as the most abundant multicellular animal group, with important roles in marine ecosystems [1, 2]. The salmon louse (*Lepeophtheirus salmonis*) is an ectoparasitic copepod on salmonids, with a life cycle that has eight developmental stages separated by moulting, consisting of two nauplius stages, one copepodid stage, two chalimus stages, two preadult stages and the adult stage [3, 4]. Salmon lice are a major challenge to cage-based aquaculture of salmonids and cause large economical losses each year [5]. The emergence of salmon lice resistances against several drugs makes the situation even worse [6, 7]. Developing novel antiparasitic strategies is thus an urgent and vital issue. To achieve this, we require a thorough understanding of the molecular mechanism of life stages development of the salmon louse. Identifying key genes that influence or regulate the lifespan in the salmon louse is also of great importance for detecting novel drug targets against salmon lice.

Moulting, the production of a new cuticle and shedding of the old one, plays a crucial role in the survival and development of copepods. Moulting consists of different events, including detachment of the old cuticle, synthesis of new cuticle, shedding of the old cuticle, hardening of the new cuticle and absorption of the old cuticle. In the salmon louse, moulting occurs cyclicly during development until the adult stage [8, 9]. Steroid hormones play a crucial role in arthropod moulting by regulating a series of pathways, and one of the key hormones is 20-hydroxyecdysone (20E) [10, 11]. The binding of 20E to the ecdysone receptors (EcR) leads to cascade controlling of moulting. The polysaccharide chitin and other structural proteins are the major components of the exoskeleton, so genes involved in the chitin metabolism are supposed to be important for moulting. Particular attention had been devoted to the mechanism of copepods moulting at the molecular level [12, 13], and there have been some studies on the role of steroid hormones and ecdysone receptors in the moulting of salmon louse [14, 15]. Some studies on the impact of benzoylurea on salmon lice demonstrate the importance of chitin metabolism for both moulting and survival of salmon lice [16, 17]. However, there is still a lack of understanding of the molecular mechanism of the development of salmon louse and other copepods.

In recent years, high-throughput technologies have enabled us to study a large number of genes in parallel and thus facilitate the study of complex biological

systems [18]. Being tremendously successful, high-throughput genome sequencing produced large volumes of data and started a new era of genomics research [19]. Our group has recently performed a comprehensive transcriptome time-series analysis using RNA sequencing data from three stages of salmon lice (chalimus-1, chalimus-2 and preadult-1) and proposed a method for sorting salmon lice samples [8]. We also identified genes that potentially regulate the development and moulting of salmon louse larvae.

A research area that is particularly important for systems biology is the study of dynamic interfaces and crosslinks between different processes and components in biological systems [20]. A great deal of attention has been devoted to this area, and network-based analysis have emerged as a powerful framework for studying the large number of interactions among biological processes and components as well as the underling mechanisms. Providing an intuitive way to represent a collection of objects and connections among them, gene co-expression networks (GCNs) have been widely used to collect and depict the behaviors and interactions among components in transcriptomes [20, 21]. Signatures of hierarchical modularity have been suggested to be present in all cellular networks that have been investigated so far, ranging from metabolic to protein–protein interaction and regulatory networks [22]. In gene co-expression networks, modules are defined as groups of genes with similar expression patterns and can be identified by using clustering methods [23–25]. GCN modules have facilitated a better understanding of a number of biological phenomena [23, 26, 27], and an increasing number of studies based on GCN have been conducted to identify condition-specific gene modules and predict potential genes involved in a certain phenotype [28–31].

In this study, by re-analyzing the staged time-series data from [8], we aim at providing a framework of identifying important genes through GCN analysis and contributing to a better understanding of the molecular mechanisms of moulting in copepods. By combining GCN analysis, sample traits and annotation information from public databases, we identified modules and hub genes in order to find novel candidates with association to the moulting and development of salmon louse. For validation we performed gene knock-down by RNA interference (RNAi) for five genes.

Methods

Gene expression data and filtering

A normalized gene expression matrix was generated from the RNA-seq data provided by Eichner *et al.* [8], by extracting samples from middle instar ages and old/moulting instar ages of chalimus-1, chalimus-2 and preadult-1 larvae (Figure 1). Transcripts with low expression (not having at least 3 cpm in at least 3 samples) were discarded.

Identification of moulting-associated genes and transcription factor (TF) genes

By combining data from the published literature and LiceBase (https://licebase.org/), we collected genes which are involved in the moulting of salmon lice or known to be associated with the moulting of other arthropods with high confidence. We named these genes as "moulting-associated genes".

The GO annotation information for the salmon louse genes was obtained from LiceBase. Any salmon louse genes that were annotated under the top level of GO term associated with transcription factor ("GO:0006351", "GO:0001071", "GO:0008134", "GO:0000988" and "GO:0005667") or their children were identified as transcription factor (TF) genes.

Gene co-expression network (GCN) analysis for identifying important modules and genes associated with moulting and development of salmon louse

In this study, we define the modules and genes that might play a role in the regulation of moulting and development of salmon louse as "important modules" and "important genes" respectively, and we proposed a workflow to identify these important modules and genes based on the gene co-expression network (GCN) analysis (Figure 2). Using gene expression profiles, sample traits and gene annotation information as input, this workflow is used to predict the important modules and genes for moulting and development of salmon louse.

Gene co-expression network (GCN) construction, module identification and module eigengene calculation

Gene co-expression network (GCN) construction and power parameter estimation

The gene co-expression networks (GCNs) were constructed using the R package WGCNA [32]. A modified version of the biweight midcorrelation (bicor) [33] was

adopted to calculate the absolute correlation between pairwise genes (transcripts) (S_{ij}):

$$S_{ij} = |bicor(x_i, x_j)|, \quad (1)$$

where x_i denotes the expression profile across all samples of transcript i. The function *bicor* is implemented in the R package WGCNA.

By transforming the correlation by power function, we obtained the adjacency between pairwise transcripts (A_{ij}):

$$A_{ij} = S_{ij}^{\beta}, \quad (2)$$

where β is the power parameter, and β is determined based on whether the corresponding co-expression network exhibits scale-free characteristics and has relatively high connectivities. We chose the suitable power parameter from integers ranging from 1 to 20, by plotting the signed scale-free topology fitting index R^2 against different power parameters, and we also plotted the corresponding network mean connectivity against different power parameters. Details about how the power parameter β was estimated can be found in Additional file 1.

With the adjacency matrix A we can construct the co-expression network, where each node represents a gene, and the weight possessed by edges between nodes indicates the co-expression relationship between nodes. Although our data was from a transcriptome study and the nodes in our networks represented transcripts, we used the term "gene co-expression network" and "eigengene" for convenience in this paper.

We constructed three gene co-expression networks, based on the gene expression profiles from middle samples, old/moulting samples and all samples (samples from both middle instar ages and old/moulting instar ages).

GCN module identification and eigengene calculation

For each gene co-expression network, hierarchical clustering was performed for the nodes based on their adjacencies and a dendrogram was obtained. Using this dendrogram as input, a top-down algorithm *cutreeDynamicTree* was applied to identify gene modules. Each module was assigned a unique name as color. For each gene co-expression network, nodes that could not be assigned to any modules were moved to

a module called "grey". The grey module in each of the network was not considered in further analysis.

After identifying modules from each network, a sub-adjacency matrix can be extracted for all the gene members in each module. Then the eigengene for each module was computed as the eigenvector for the largest eigenvalue of the module gene expression matrix by the function *moduleEigengenes* in WGCNA.

Intramodular centrality measurements and intramodular hubs identification

In this study, we adopted three types of centrality measurements to measure the centralities of nodes within each module and identified intramodular hubs.

Intramodular Connectivity (kIM)

The connectivity of the ith node (k_i) in the weighted network is defined as the sum of connection weights between node i and the other nodes [34]:

$$k_i = \sum_{j \neq i} A_{ij}. \tag{3}$$

Suppose that there are Q modules detected in a network, and they are labeled by $q = 1, 2, \ldots Q$, so the connectivity of a node i within a module q is defined as intramodular connectivity ($k_i^{(q)}$or $kIM_i^{(q)}$):

$$k_i^{(q)} = kIM_i^{(q)} = \sum_{\substack{j \in M_q \\ j \neq i}} A_{ij}^{(q)}, \tag{4}$$

where M_q denotes the set of node indices that correspond to the nodes in module q, and $A^{(q)}$ is the adjacency matrix of module q. High intramodular connectivity implies that a node could be a hub within the module.

Module membership / module eigengene-based connectivity (kME)

The module membership (or module eigengene-based connectivity) is defined as the value of correlation between module eigengene and the expression profile of the genes (or transcripts) assigned to this module [35]:

$$kME_i^{(q)} = cor(x_i^{(q)}, E^{(q)}), \tag{5}$$

where $x_i^{(q)}$ specifies the expression profile in different samples of transcript i that is assigned to the module q, and $E^{(q)}$ denotes the eigengene of module q.

Since our gene co-expression networks were constructed based on the absolute correlation values between gene expression profiles, we used the absolute value of module membership to measure the centrality of each node within a module:

$$kME_i^{(q)} = \left| cor(x_i^{(q)}, E^{(q)}) \right|. \quad (6)$$

In addition, the module membership of a node for module q can be calculated for all nodes in the network:

$$kMEall_i^{(q)} = \left| cor(x_i, E^{(q)}) \right|, \quad (7)$$

and this definition can be used in the module preservation analysis. The details can be found in Additional file-1.

Intramodular weighted betweenness centrality (BC)

The betweenness centrality of a node in an unweighted network (or module) is the number of shortest paths between all other nodes in the network that pass through the node [36]. To calculate the betweenness centralities of nodes in our weighted networks, a generalization of betweenness centrality proposed by Brandes [37] was employed. The approach is implemented in the R package tnet [38].

Definition of intramodular hubs

We evaluated the centralities of nodes in each module, using intramodular connectivity, absolute module membership and intramodular weighted betweenness centrality. The nodes ranking among the highest ten percent in any of the three centrality measurements of all nodes within a module were defined as intramodular hubs. The node obtaining highest scores in all of the three centrality measurements was defined as "absolute hub".

Based on the ranks of nodes in three types of centrality measurement, we can calculate the average rank of nodes within each module. Therefore, the absolute hub should have an average rank as 1.

Module preservation analysis

The preservation of a module between the reference network and a test network can be evaluated based on the alterations in connectivity patterns and density. A

well-preservation module in two or more networks should have similar connectivity patterns and nodes in the module should remain being tightly connected. WGCNA provides a series of approaches to evaluate whether a module is preserved and reproducible in another network [39]. In this study, module preservation statistics were computed to compare the two networks constructed based on middle samples and old/moulting samples. For each module preservation statistic, permutation tests were performed to evaluate the significance of the observed value and a Z score was obtained. The Z scores for all of the module preservation statistics were integrated as a composite summary statistic $Z_{summary}$. Details about how to calculate module preservation statistics and $Z_{summary}$ can be found in Additional file 1.

The networks were unsigned, and we set the number of permutation as 200. All the correlations were calculated using the biweight midcorrelation (*bicor*). Modules with a $Z_{summary}$ smaller than 2 was regarded as non-preserved, while a $Z_{summary}$ larger than 10 indicated that a module was well preserved across different networks. Since we aimed to identify modules playing a role in the regulation of moulting, the non-preserved modules from the moulting network were of particular interest.

Regularized logistic regression using module eigengenes as independent variables

We made use of the eigengenes of modules in the global network to perform logistic regression with an elastic-net penalty (α=0.5). This task was achieved by setting the binary dependent variable as the label of middle or old/moulting (old/moulting stages were labeled as 1), and using the eigengenes of each module as independent variables.

We used the R package glmnet [40] to perform this analysis, and we adopted the λ that gives minimum mean cross-validated error.

Integrating information from external databases and enrichment analysis

Data from FlyBase (`https://flybase.org/`) and GenomeRNAi [41] were extracted and used to identify homologous observable phenotypes and lethal phenotypes enriched modules.

To detect homologous sequences in *Drosophila melanogaster*, we ran BLASTP with E-value cutoff as $1e-10$ on the corresponding protein sequences of salmon louse transcripts against protein sequences from *Drosophila*. For the BLASTP results, only the best hits were considered. After mapping the protein IDs of the

homologues from *Drosophila* to gene IDs, RNAi knock-down phenotype information were mapped to GenomeRNAi. If a salmon louse protein had more than one *Drosophila* homologue with identical maximum bitscore, all the homologues were used to search for RNAi phenotypes.

BLASTP for all salmon louse sequences against themselves was performed to find paralogues.

Enrichment analysis of modules

Based on the GO annotation file for salmon louse genes from LiceBase, GO enrichment analyses were performed for each modules identified in the middle, moulting and global network using the fisher statistic and "elim" algorithm provided by R package topGO [42].

Furthermore, with the information from the *Drosophila* homologues-based transcript-phenotype list (Additional file 2-Table S1), we conducted two enrichment analyses for each module identified in all networks. The p-values of these enrichment analyses were obtained based on hypergeometric tests, to determine whether transcripts with homologue observable phenotypes or homologue lethal phenotypes in *Drosophila* were significantly enriched within a module. Based on the suggestions from [43], we used the raw p-values of our enrichment analyses, and the cutoff of p-values was set as 0.05 for all the enrichment analyses.

Selecting important modules for further analyses

We were interested in identifying gene modules which are likely to play a role in the moulting and development of salmon louse, and we chose important modules based on three analyses: the module preservation analysis, the regularized logistic regression analysis and the *Drosophila* homologues-based enrichment analysis. According to the guilt-by-association (GBA) heuristic [44], nodes in the moulting-associated transcripts-containing modules are more likely to play a role in the moulting and development of salmon louse, and we conducted a focused search among modules containing at least one known moulting-associated transcript. Therefore, moulting-associated transcripts-containing modules satisfying any of the following criteria were chosen for further studies: 1) non-preserved modules in the moulting network ($Z_{summary} < 2$); 2) the eigengenes of modules from the global network obtained positive coefficients from the regularized logistic regression analysis (the module

with largest coefficient value should be prioritized); 3) modules that are significantly enriched by transcripts with observable and lethal RNAi phenotypes from homologues ($p - value < 0.05$) (Figure 2).

Selecting important genes as knock-down candidates from important modules

Since many researchers have proposed that hubs in a biological network tend to be more important [45–47], we chose RNAi knock-down candidates among the hubs of the important modules. For each selected module, we gave prime consideration to the absolute hub. If no absolute hub was detected, knock-down candidates were chosen from other intramodular hubs. Hubs with less paralogues and little annotation information were then given priority.

Differential gene expression (DE) analysis and regulatory impact factor (RIF) calculation

We calculated the regulatory impact factors (RIF) for all the transcripts annotated as transcription factors, based on the metric proposed in [48].

The first step was to perform differentially expressed (DE) analysis to compare the middle group and the old/moulting group, and the statistics used to test the null hypotheses were calculated based on standardized rank-sum Wilcoxon test. We computed the permutation adjusted p-values using the step-down maxT multiple testing procedures, which provide strong control of the family-wise Type I error rate (FWER). The functions are implemented in the R package multtest [49]. Transcripts with an adjusted p-value smaller than 0.05 were identified as differentially expressed between middle instar ages and old/moulting instar ages.

The first RIF value ($RIF1$) for fth TF transcript was defined as:

$$RIF1_f = \frac{1}{n_{de}} \sum_{i=1}^{n_{de}} A_i DE_i DW_{if}^2, \tag{8}$$

where n_{de} is the number of DE transcripts; A_i represents the average expression of ith DE transcript across the two groups, and DE_i is the statistics obtained from the previous DE analysis. DW_{if} is the abbreviation for differential wiring, which means the change of correlation between fth TF and the ith DE transcript across the two groups:

$$DW_{if} = ro_{if} - rm_{if}, \tag{9}$$

where ro_{if} and rm_{if} are the co-expression correlation between fth TF and the ith DE transcript in the old/moulting samples and the middle samples respectively.

The second RIF value ($RIF2$) was computed as followed:

$$RIF2_f = \frac{1}{n_{de}} \sum_{i=1}^{n_{de}} (EO_i^2 ro_{if}^2 - EM_i^2 rm_{if}^2), \tag{10}$$

where EO_i^2 and EM_i^2 denote the square of the average expression value of the ith DE transcript in the old/moulting samples and the middle samples; ro_{if}^2 and rm_{if}^2 are the square of the co-expression correlation between fth TF and the ith DE transcript in the old/moulting samples and the middle samples respectively.

RNA interference experiments

RNA interference

Double-stranded RNA (dsRNA) was produced using MEGAscript® RNAi Kit (Ambion) according to supplier's instructions using following primers added the T7 sequence (TAATACGACTCACTATAGGGAGA): EMLSAG00000001458: forward (fw)_1458: CAAGCTGTTATTGATTGGCGATTC, revers (rv)_1458: CGGCATATTTAACTGATCAGCGTA; EMLSAG00000003179: fw_3179: GCGTAAAAGTTGCGTACAATCTGA, rv_3179: GTTTATTGGGTGTGATGAATCCGA; EMLSAG00000005299: fw_5299: GTATGATGACGGACATGCTCAAGG, rv_5299: GGCCTGTTTATAGTCGGTAGCCAT; EMLSAG00000004347: fw_4347: AAACGGCGCGAGGAGGTGAATA, rv_4347: GGTGGGTTTCTTTCCTGGCTTGTT; EMLSAG00000008959: fw_8959: GCCTCCGGTTCGGATGAAGAA, rv_8959: AGGATCAGAGGGGCCACAAGTGTC.

RNA interference on nauplia and fish challenge with the emerging copepodids

RNA interference was conducted in nauplia as described in [50], but with 2 μg fragment. Infection was done in single tanks with 60 copepodids per fish on three fish for each fragment and for the control as described in [51] (10°C, full salinity). The remaining copepodids were stored on RNAlater® (Invitrogen) for later measurement

of transcript down regulation. Sampling was done after 16 days when lice from control group were in chalimus or preadult-1 stage. All lice were sampled from fish and photographs were taken under the binocular in a drop of seawater with a cover slide on top. Number of lice and size measurements on photographs were recorded. Genes to be knocked down were: EMLSAG00000001458, EMLSAG00000003179 and EMLSAG00000005299. A second trial with knock-down larvae of EMLSAG00000001458 with 100 copepodids on three fish each was done. The outflow water from the tanks was filtered and lice in the flow out were counted two hours after infection and 24 hours after infection. One fish from knock-down and control group each were terminated three days after infection and lice were fixed on Karnovsky's fixative for histological investigation. The other fish were terminated eight days after infection.

RNA interference on preadult lice

RNA interference in preadult-2 lice was done as described in [52]. In short, the fragment was injected into preadult-2 lice, which were put on fish again until most of the control lice showed its second pair of egg strings. All lice were sampled and photographed. The lice egg strings were laid into single flow through wells for hatching observation. Lice were either stored on RNAlater® for qPCR measurements or on Karnovski's reagent for histological investigation. Three different experiments were conducted, one with double-stranded RNA for EMLSAG00000001458, EMLSAG00000005299 and EMLSAG00000003179 as well as a control (35 days), one with dsRNA for EMLSAG00000004347 and a control (40 days) and one with dsRNA for EMLSAG00000008959 and a control (37 days). The cod trypsin RNA was used as a non-target control fragment [52].

RNA extraction, cDNA synthesis and qPCR measurements

Nauplia were divided into five (or four in case of control group from second infection trial) batches of 30 to 40 nauplia. RNA from nauplia was extracted by a combination of Trizol and RNAeasy micro kit as previously described [50]. RNA was frozen at -80 °C until usage. cDNA synthesis was conducted using the AffinityScript QPCR cDNA Synthesis Kit (Agilent) according to suppliers recommendations. Gene expression of the target gene was measured by quantitative real-time PCR in control and knock-down group. QPCR was carried out in duplicates using the salmon louse elongation factor 1α [53] as well as

Adenine Nucleotide Translocator 3 (ADT3) [51] (in nauplia only) as a standard. SYBR® Green PCR master mix and the following primers were used: forward (fw)_1458: CAAGCTTCTTGTGGGAAACAAATG, reverse (rv)_1458: GCTGCCATAGTCATAAAAGCTTGC; fw_3179: GGTTCGGATTCATCACACCA, rv_3179: CGTGGAGGGATAGGACAAACTTTG; fw_5299: GTTCTCGGAGAAAATAATGTTGCG, rv_5299: TGTTCAGTGATTTCCAGTGCTTCC; fw_8959: CCATCAATTTCCAAGTGGAGGA, rv_8959: CCTTCGCATATCTCTTCCTCTTCA; fw_4347: GTCTGGACAAGGAGAAAATTCTGC, rv_ 4347: CTCCGTGGTTTTGGCATTGA; fw_LsEF1α: GGTCGACAGACGTACTGGTAAATCC, rv_LsEF1α: TGCGGCCTTGGTGGTGGTTC, fw_LsADT3: CTGGAGAGGGAATTTGGCTAACGTG, rv_LsADT3: GACCCTGGACACCGTCAGACTTCA. Applied Biosystems 7500 Fast Real-Time PCR system was used for thermal cycling and quantification in 10 μl reactions (initiation 50 °C 2 min, 95 °C 2 min, then 40 cycles of 95 °C 15 seconds, 60 °C 1 min). A melting curve 60 °C to 90 °C was performed. Relative gene expression was calculated using the differences in threshold cycle (CT) between gene of interest and standard genes.

Histology

Lice were prepared as described in [54] for histological investigation. For plastic embedding, lice were washed twice in phosphate buffered saline, dehydrated in a graded ethanol series, pre-infiltrated with Technovit/ethanol (50/50) for four hours (Technovit 7100, Heraeus Kulzer Technique) and infiltrated with Technovit and hardener overnight. Two micrometre thick sections were cut with a microtome (Leica RM 2165) and stained with toluidine blue (1% in 2% borax) for one minute. The stained sections were mounted using Mountex (Histolab Products).

Results

Identification of moulting-associated and transcription factor (TF) genes

Among the transcripts in our RNA-seq data, we found 40 moulting-associated transcripts and 32 of them were retained after low expression filtering. The list of moulting-associated transcripts and the relevant papers can be found in Table 1.

There were 433 transcripts annotated as transcription factor (TF), and 231 TF were retained after low expression filtering (Additional file 2-Table S8).

Gene co-expression network (GCN) construction and module identification

To detect the genes that might be involved in the regulation of moulting and development of salmon louse, we analyzed the RNA-seq data sampled from the middle instar ages and old/moulting instar ages of chalimus-1, chalimus-2 and preadult-1 larvae (Figure 1). Of the 45 samples, 18 samples were from middle instar ages, and 27 samples were from old/moulting instar ages. After filtering transcripts with low expressions, 7108 transcripts were retained for network analysis.

Three gene co-expression networks (GCNs) were constructed for different sample groups. The first GCN was generated using all samples labeled as "middle instar age". Meanwhile, an "old/moulting instar age" GCN was created for all old/moulting samples. To further facilitate our analysis, a GCN based on all samples from middle and old/moulting instar ages was built. The three GCNs were thus denoted as middle network, moulting network and global network respectively.

We set the power parameter β as 7 to make sure that the networks satisfied scale-free topology approximately while having relatively high mean connectivities (Figure 3). The adjacency matrices of the three networks can be found in Additional file 7.

In the module identification process, 83, 60 and 78 modules were found in the middle, moulting and global network respectively, and the module sizes ranged from 32 to 333. There were 203, 444 and 506 genes assigned to the grey module of the middle, moulting and global network respectively (Additional file 2-Table S2-S4). Genes in the grey modules were not considered in the further analysis. Dendrogram and module identification results for the three networks were displayed in Additional file 3-Figure S1. Eigengenes for each module in all the three networks were also obtained.

The centralities and distribution of moulting-associated genes across modules

To preliminarily examine the essentiality of intramodular hubs, we calculated the centralities for the 32 moulting-associated transcripts. Based on our definition of intramodular hubs, there were 6, 12 and 8 moulting-associated transcripts identified as intramodular hubs in the middle, moulting and global network respectively (Additional file 2-Table S5-S7). The transcript EMLSAT00000005083 (annotated to encode hexokinase type 2, Table 1) was identified as intramodular hub in all

the three networks, and it was the absolute hub in the module "lightcyan1" of the global network.

We examined how the 32 moulting-associated transcripts were distributed across modules in the three networks. There were 25, 20 and 24 modules containing moulting-associated transcripts in the middle, moulting and global network, accounting for 30.1%, 33.3% and 30.8% in the three networks respectively. The numbers of moulting-associated transcripts in these modules ranged from 1 to 4 (Additional file 2-Table S2-S4).

Module preservation analysis

To identify genes which may play a role in the moulting of salmon lice, we detected non-preserved modules from the moulting network based on module preservation analysis. Five modules from the moulting network were found as non-preserved, and the module sizes ranged from 41 to 100 (Additional file 2-table S3). Strong correlations among genes in these non-preserved modules were only observed in the moulting network, and two non-preserved modules (yellowgreen and lavenderblush3) contained moulting-associated transcripts (EMLSAT00000008812 and EMLSAT00000012651) (Additional file 2-Table S6). Notably, the moulting-associated transcripts were also identified as intramodular hubs in these modules. The transcript EMLSAT00000008812 (annotated to encode chitinase, Table 1) was ranked eighth (based on connectivity) in the "yellowgreen" module, and the transcript EMLSAT00000012651 (annotated as EcR, Table 1) was ranked third (based on betweenness centrality) in the "lavenderblush3" module. We thus hypothesized that transcripts in these two modules could be important for salmon louse moulting, and hubs from these modules should be considered as important.

Eight modules from the middle network were identified as non-preserved, and the module sizes ranged from 54 to 109 (Additional file 2-Table S2). Three non-preserved modules (darkseagreen4, brown4 and lightcyan1) were found containing one moulting-associated transcript respectively (Additional file 2-Table S5). However, none of these moulting-associated transcripts were intramodular hubs in the middle network.

Regularized logistic regression analysis on the global co-expression network

To compare the intramodular overall gene pexression patterns between the middle sample group and old/moulting sample group, we performed elastic net regularization-based logistic regression using the eigengenes of module from the global network as independent variables. As a result, we found modules with eigengenes that were highly expressed in one sample group but lowly expressed in the other sample group. From the 78 module eigengenes, we identified 15 eigengenes with non-zero coefficient (ranging from -1.75 to 0.963), and six of the 15 corresponding modules contained one known moulting-associated transcripts (Additional file 2-Table S4). It was noteworthy that module "steelblue" possessed the largest positive coefficient and contained one moulting-associated transcript (EMLSAT00000001150) as intramodular hub, which was ranked second in the betweenness centrality measurement (Additional file 2-Table S7).

When checking the absolute value of regression coefficients, three modules (magenta, lightcyan and ivory) were found with absolute coefficient larger than 1. The moulting-associated transcripts were found in two of the three modules (lightcyan and ivory). The coefficients of all the three modules were negative, indicating that genes in these modules exhibited much higher expressions in middle samples. Notably, two modules (indianred4 and lavenderblush3) with negative regression coefficients contained moulting-associated transcripts (EMLSAT00000000733 and EMLSAT00000008543) annotated as hormone receptor 3 (Hr3) and hormone receptor 4 (Hr4) (Table 1). In the module "indianred4", EMLSAT00000000733 was ranked eighth in the connectivity measurement.

Differentially expressed transcripts between middle group and old/moulting group were found in all the modules with non-zero regression coefficients, and the proportions ranged from 36.9% to 98.4% (Additional file 2-Table S4).

Integrating information from external databases

We identified homologous genes in *Drosophila melanogaster* for the salmon louse transcripts and then searched for RNAi phenotypes for these homologues in the GenomeRNAi database. We found homologous RNAi phenotypes for 3059 salmon louse transcripts. Additional file 2-Table S1 maps salmon louse transcripts to the RNAi phenotypes of the corresponding homologues in *Drosophila*.

Enrichment analysis of modules

Based on the GO annotation file for the salmon louse transcripts, we performed GO enrichment analysis for all the modules to preliminarily elucidate the biological functions of the modules. The GO term with smallest p-value in each category (Biological Process(BP), Molecular Function(MF) and Cellular Component(CC)) were recorded (Additional file 2-Table S2-S4).

To further identify modules which are more likely to contain important genes for lice development, we conducted enrichment analyses for all the modules based on the homologues-based transcript-phenotype list. The transcripts with observable RNAi phenotypes were significantly enriched in 16, 13, and 14 modules in the middle, moulting and global network, accounting for 19.3%, 21.7%, and 17.9% in total modules respectively. Analogously, 14, 14, and 9 modules were detected as enriched by transcripts with lethal RNAi phenotypes in the middle, moulting and global network, accounting for 16.9%, 23.3% and 11.5% in total modules respectively. We found a relatively large overlap between the two enrichment analyses: 10, 11 and 7 modules (accounting for 12.0%, 18.3% and 9.0% in total modules) were identified as being significantly enriched by both observable and lethal RNAi phenotypes in the middle, moulting and global network respectively (Additional file 2-Table S2-S4).

Differential gene expression (DE) analysis and regulatory impact factor (RIF) calculation

All 45 samples were divided into middle and old/moulting groups to find DE transcripts. There were 1364 transcripts differentially expressed between the two groups. The list of DE transcripts facilitated calculation of the RIF scores for all transcripts with GO annotation as transcription factor (TF).

For the 231 TF transcripts, RIF scores were computed (Additional file 2-Table S8). It is noteworthy that EMLSAT00000003849 (annotated as forkhead box protein O(FOXO)) obtained highest RIF scores from both methods. This transcript is also an intramodular hub of a moulting-associated transcripts-containing module in both of the middle and moulting network.

Selecting important modules for further analyses

In the module preservation analysis, two modules (yellowgreen and lavenderblush3) from the moulting network were detected based on our criteria. In the regularized

logistic regression analysis, two modules (steelblue and green) from the global network passed the criteria. In the homologues-based enrichment analysis, one (mediumpurple3), two (darkolivegreen and violet) and one (turquoise) module were found satisfying the criteria from the middle, moulting and global network respectively. In summary, one, four and three modules from the middle, moulting and global network were selected for further analyses.

Examining intramodular hubs and selecting important genes as knock-down candidates from important modules

After determining important modules, we investigated the hubs of these modules to understand their roles in moulting and development of salmon louse. For each of the eight chosen modules, we examined their hub with highest average rank (Additional file 2-Table S12). The absolute hubs (EMLSAT00000003844 and EMLSAT00000001458) of the two non-preserved modules (lavenderblush3 and yellowgreen) selected from the moulting network are annotated as epithelial cell transforming 2 (ECT2) and Ras-related protein Rab-1A (RAB1A), respectively. For the four modules selected from the enrichment analysis, EMLSAT00000000929, EMLSAT00000005299, and EMLSAT00000012769 were identified as absolute hubs, annotated as high density lipoprotein-binding protein (HDLBP), ER membrane protein complex subunit 3 (EMC3) and laminin subunit beta-1 (LanB1) respectively. EMLSAT00000010555 had the highest average rank in the module (turquoise) from the global network, annotated as stress-induced-phosphoprotein 1 (STIP1). No absolute hubs were found in the two global modules selected from the regression analysis, and the nodes with highest average rank in these modules were EMLSAT00000007421 and EMLSAT00000012693, both of them were identified as differentially expressed transcripts between the middle group and old/moulting group. EMLSAT00000007421 was annotated as cuticular protein 62Bb (Cpr62Bb), and few annotation was found for the hub EMLSAT00000012693.

To validate the importance of genes in the selected modules in moulting and development of salmon louse within the limited accesses to RNAi experiments, we selected RNAi knock-down candidates from three important modules. Since the important modules were selected based on three analyses, we selected one module from each of the three analyses. Firstly, we choose the module “yellowgreen” from the

two non-preserved modules (the other one is lavenderblush3) in the moulting network. The module "yellowgreen" had larger size then the module "lavenderblush3", and the absolute hub of the module "yellowgreen" had higher score of the absolute module membership. From the two moulting modules (darkolivegreen and violet) selected from the enrichment analysis, we chose the module "violet" for further analysis, since it contained more transcripts annotated as transcription factors (TF), and these TF transcripts obtained higher scores in the regulatory impact factor analysis than those found in the module "darkolivegreen". Furthermore, with regards to the proportion of transcripts annotated as TF, the module "violet" and "yellowgreen" ranked as first and third among all modules in the moulting network. Finally, we selected the module "steelblue" in the global network for further analysis, because the eigengene of this module obtained largest coefficient in the regularized logistic regression analysis. Details of the three selected modules can be found in Table 2 and Additional file 2.

For the module "yellowgreen" and "violet" from the moulting network, we chose the absolute hub for RNAi experiment. According to the criteria discussed in the method section, we chose another one hub without paralogues from each module to knock down.

No absolute hub was found in the module "steelblue" from the global network, and the hub (EMLSAT00000007421) with highest average rank was annotated to encode cuticle protein. Among the 12 intramodular hubs found in the module "steelblue", three (EMLSAT00000012111, EMLSAT00000008158 and EMLSAT00000012113) were annotated to encode proteins with the chitin binding peritrophin-A domain (PF01607); four (EMLSAT00000007421, EMLSAT00000007422, EMLSAT00000009987, and EMLSAT00000010209) were annotated to encode cuticle proteins (PF00379); one (EMLSAT00000004870) was annotated to encode protein with the polyprenyl synthetase domain (PF00348), and the moulting-associated transcript (EMLSAT00000001150) was annotated to encode cytochrome P450 (PF00067) (Additional file 2-Table S10, Table 1). Among the three hubs with few annotation information, we chose one (EMLSAT00000004347) with least number of paralogues to knock down. The details of all the knock-down candidates were collected in Table 3.

RNA interference on nauplia and infection of salmon with the emerging copepodids

Measurement of gene expression in copepodids before infection showed down regulation of all targeted genes (t-test: $p - value < 0.05$) with varying knock-down efficiency. For gene EMLSAG00000001458, EMLSAG00000005299 and EMLSAG00000003179, efficiency was 94%, 84% and 89%, respectively. At termination after 16 days no lice were found on the fish infected with copepodids from EMLSAG00000001458-KD group (Figure 4 left panel). There was no significant difference in the number of lice between control group and EMLSAG00000005299-knock-down (KD) group as well as EMLSAG00000003179-KD group, and the development of lice from all groups found on fish was similar (Figure 4 right panel). No difference in the phenotype could be observed under the binocular or by size measurements.

Second trial with EMLSAG00000001458 knock-down

Since no lice with EMLSAG00000001458 knocked down were found at termination after 16 days on the fish, we were interested in finding out whether this could be due to reduced infection success or due to problems with development and moulting. A second infection trial for qualitative measurement was done. Knock down efficiency measured in copepodids before infection was 95%. After two hours, 30, 37 and 33 lice were found in the filtered flow through water of tanks from fish of the control group, and 32, 57 and 35 lice were found from tanks of the knock-down group. After 24 hours, 9, 9 and 4 lice were found in the flow out from control fish, and 9, 8 and 4 lice were found from knock-down fish. No lice were found in the filters after three days. At termination of the first fish at day three after infection there were 10 lice on the control fish and 14 on the knock-down fish. These were sampled for histological investigation. No differences were observed in the histological sections (Additonal file 3-Figure S3). Eight days after infection lice had developed to chalimus-1 on control fish (13 on one fish, 39 on the other), but no lice were found on one of the fish with knock-down samples and two copepodids on the other fish.

Knock-down in preadult lice

At sampling lice were in the adult stage. Down regulation was on average 77% for EMLSAG00000001458-KD group, 47% for EMLSAG00000003179-KD and 68% for EMLSAG00000004347-KD group. Lice from EMLSAG00000001458-KD group

and EMLSAG00000008959-KD group had no egg string. Length measurements for body parameters (cephalothorax and genital segment length) as well as egg strings were shown in Table 4. Egg strings of all groups with egg strings present hatched and produced viable normal looking offspring. Histological sections were done for EMLSAG00000001458-KD and EMLSAG00000004347-KD lice. Histology of different tissues is shown in Additonal file 3-Figure S2. EMLSAG00000001458 knockdown lice did not develop normal looking oocytes (Additonal file 3-Figure S2 f) and the ovaries did not contain any oogonia (Additonal file 3-Figure S2 j). The cellular structures of the subcuticular tissue of the cephalothorax were changed and only loose connection between cells was observed (Additonal file 3-Figure S2 n), while the subcuticular tissue of the genital segment seemed not to be affected in the same way. At sampling EMLSAG00000004347 knock-down lice showed a weak genital segment, which was easily squeezed and teared apart when handling the lice. In the histological sections, the cuticle and subcuticular tissue of the genital segment (Additonal file 3-Figure S2 s) did not show obvious differences to the control louse.

Examining the RNA interference experiments data from LiceBase

From LiceBase, RNAi experiments for 188 genes were collected, and 112 genes among them appeared in our three networks. 10 genes in six of the eight selected modules were found with observable RNAi phenotypes (including the RNAi experiments results from this study). One gene from the selected module "darkoliveg-reen" had been knocked down, but no phenotype was observed. No RNAi results were found for the genes in the selected module "lavenderblush3" (Additional file 2-Table S13).

Notably, one hub (EMLSAG00000009839) from one non-preserved module (skyblue) of the moulting network show reduced survival in the RNAi experiments, although this module did not contain any known moulting-associated genes. The absolute hub (EMLSAG00000005382) of the module "blue2" in the middle network show shorter eggstrings in RNAi experiments, and this module contained two known moulting-associated genes.

We also found RNAi experiment records of four genes (EMLSAG00000010968, EMLSAG00000006642, EMLSAG00000007048 and EMLSAG00000004159) which obtained highest average rank in four modules. These four modules did not satisfy

any of the criteria for being an important module for salmon louse moulting and development, and the four genes did not show any observable phenotype in RNAi experiments (Additional file 2-Table S13).

Examining the modules where the RNAi candidates were selected

In the RNAi experiments, all five selected gene candidates show observable phenotypes, and we thus examined the three modules they were from. We plotted heatmaps of scaled gene expression profiles and barplot of scaled eigengene expression for each of the three module (Figure 5).

Gene expression profiles within a module were strongly correlated. Genes in the module "yellowgreen" tended to be highly expressed in the chalimus-1 old female samples, chalimus-1 moulting samples and preadult-1 old female samples. Genes in the module "violet" were highly expressed in the chalimus-1 male samples and chalimus-1 moulting samples. For the module "steelblue" from the global network, genes were highly expressed in almost all samples from the old and moulting instar stages (except two preadult-1 old male samples), especially in the chalimus-1 moulting samples. The preadult lice with the two genes of the module "yellowgreen" knocked down failed to develop eggstrings. Further study is necessary to understand the role of genes from the module "yellowgreen" in the fecundity of female lice.

The topological graph for each of three modules (Figure 6, Additional file 4-6) shows that moulting-associated genes in the module "yellowgreen" and "steelblue" obtained relatively high average ranks, and they were tightly connected with other hubs. In the module "violet", the moulting-associated gene obtained a low average rank, but the genes annotated as transcription factor obtained high average ranks and were tightly connected with other hubs. For these modules, the proportion of differentially expressed genes was highest in the module "steelblue" identified from the global network.

The enriched GO terms in the module "yellowgreen" included GO:0008152, GO:0001071 and GO:0005667, indicating nucleic acid transcription factor activity. The enriched GO terms in the module "violet" included GO:0006355, GO:0003677 and GO:0044454, indicating transcriptional regulation. In the module "steelblue", the most enriched GO terms were GO:0042302, GO:0008061 and GO:0006030, indicating metabolic processes of chitin and cuticle.

Discussion

RNA interference (RNAi) has enormously facilitated rapid and straightforward analysis of gene function for parasites and other organisms [55–57], and the whole-genome RNAi screen have been successfully applied to detect genes with important functions for many biological processes in C.elegans and mammalian cultured cells [58–60]. Although a robust RNAi method for knocking down salmon louse genes have been established [57], genome-wide RNAi screen for salmon louse is both labour-intensive and time-consuming due to the complex life cycle of salmon louse [3, 4]. Currently, biologists choose RNAi gene candidates subjectively based on their research interests, and little work has yet been carried out to develop bioinformatics methods for objectively predicting salmon louse genes that have a crucial role in biological processes of interest and are likely to show visible phenotypes in RNAi experiments.

In this study, we systematically analyzed the RNA-seq data of salmon lice from different life stages and proposed an approach (a workflow) for identifying important genes involved in the moulting and development of salmon louse (Figure 2). This approach can be used in selecting candidate genes for RNAi experiments. The results of our RNAi experiments and the RNAi records from LiceBase indicate the effectiveness of our approach.

The module preservation analysis allowed us to identify two important genes (EMLSAG00000001458 and EMLSAG00000008959 annotated as RAB1A and DIEXF respectively) for the moulting and/or development of salmon louse, and both of the genes were from a non-preserved module (yellowgreen) in the moulting network. The non-preserved modules in the moulting network may be co-regulated and play an indispensable role in moulting and/or development of salmon louse. Further studies are required to clarify the biological meaning of the non-preserved modules in the middle network as well as the well-preserved modules between the middle and moulting network.

In the regularized logistic regression analysis, all the module eigengenes were calculated based on the scaled gene expression data using the same method, thus they were in the same unit and it is feasible to identify the most important module by comparing the regression coefficients of eigengenes. We found that the module (steelblue) whose eigengene obtained largest coefficient was enriched for GO categories

related to cuticle and chitin metabolic process. All the annotated hubs in this module are associated with chitin binding peritrophin-A domain, cuticle proteins, and cytochrome P450, which have been reported as important proteins for the moulting of arthropods [61–64]. We knocked down a hub (EMLSAG00000004347) with little annotation and observed both reduced fecundity and fragile constituents. Based on RNAi results and the annotations of other hubs in this module, we speculate that gene EMLSAG00000004347 may participate in building of the louse exoskeleton during the moulting process to adult stage. Our approach offers an effective solution in proposing and annotating novel putative genes that play a role in the moulting process of salmon louse. Although we focused on analyzing the modules containing moulting-associated genes due to the limited access to RNAi experiments, the module preservation analysis and regularized logistic regression analysis identify important modules without taking any prior knowledge into account. These methods are suitable to analyze the expression data from less well-annotated organisms.

Instead of focusing on the moulting process, the emphasis of *Drosophila* homologues-based enrichment analysis was placed on detecting important modules containing genes that are likely to have observable or lethal RNAi phenotypes. Four modules were identified from this analysis. Besides the two RNAi experiments performed in this study, RNAi records were found for genes in each of the four modules. Strong RNAi phenotypes were observed on one and three genes in the module "mediumpurple3" and module "turquoise" respectively. Therefore, homologues-based RNAi phenotype enrichment analysis can provide an additional window into the important module identification, especially for studying the less well-annotated organisms.

With the rapidly increasing availability of published protein–protein interaction (PPI) networks, researchers have found that highly connected proteins with a central role in the network's architecture are more likely to be essential [65–67]. Although the underlying reason is in dispute [68], the centrality-lethality rule [45] has been widely accepted. Since the PPI network for salmon louse is not available, we attempted to study the essentiality of hubs in gene co-expression networks. Taking the topological characteristics of weighted co-expression network into consideration, we used three different methods to identify intramodular hubs. In many

cases, the hubs identified with these three measurements were coherent and complementary, enabling us to define absolute hub for each module. This not only had the advantage of evaluating the intramodular centrality of nodes from different perspectives, but also increased the reliability of our study. 17 of the 32 moulting-associated genes were detected as intramodular hubs in the three gene co-expression networks, and the gene EMLSAG00000005083 with a BLAST hit as hexokinase was found as absolute intramodular hub in the global network and intramodular hub in the other two networks. For the two modules (yellowgreen and violet) from which we chose two hubs in each to knock down, we found that both of the two absolute hubs (EMLSAG00000001458 and EMLSAG00000005299) show stronger phenotypic consequences than the other middle-ranked hub (EMLSAG00000008959 and EMLSAG00000003179). The absolute hub (EMLSAT00000005382) of a moulting-associated transcripts-containing module has been recently identified as a potential mediator of intestinal heme absorption, and the corresponding protein has been characterized and is named as LsHSCARB (L. salmonis heme scavenger receptor class B-like) [69]. Furthermore, a recent study on gene co-expression network yielded a similar conclusion: researchers found that intramodular hubs in a network are likely to play a role in the biological processes of interest [70]. These observations provided support for the importance of intramodular hubs in gene co-expression network and module. However, we found records of salmon lice RNAi experiments on the nauplius-1, copepodid and/or preadult-2 stages for four hubs with highest averange rank in four modules. The four modules did not pass any of the criteria for module selection, and the four hubs did not show any observable RNAi phenotype. Therefore, not all the intramodular hubs are important, even if they have highest ranks. Combining the RNAi results in this study and RNAi records from the LiceBase, we argue that knocking down hubs from the modules selected based on our criteria are more likely to show observable phenotypes. Nonetheless, more work is required for an in-depth understanding of the relationship between the node centrality of gene co-expression network and the gene essentiality of salmon louse.

In addition to demonstrating the biological importance of intramodular hubs, RNAi experiments also show the indispensable role of the chosen genes in the moulting and development of salmon louse. Lice with the RAB1A-like gene (EMLSAG00000001458) knocked down show both reduced survival and reduced fecun-

dity. As a member of the Rab GTPase protein family, Ras-related protein Rab-1A has important role in many biological processes. There has been several published papers focusing on studying the role of insects Rab proteins in the secretion of prothoracicotropic hormone (PTTH) [71, 72], which is an important regulator of the ecdysteroidogenesis in insects [73]. However, few studies had been conducted to understand the function of Rab proteins in the moulting and development of crustaceans. The absolute intramodular hub of the other module chosen from the module preservation analysis is annotated as epithelial cell transforming 2 (ECT2). Both of the RAB1 and ECT2 are associated with GTPase activities [74–76]. Therefore, it is advocated that the role of Ras and Rho GTPases in the moulting and life development of salmon louse should be explored.

In the regulatory impact factor analysis, the transcript (EMLSAT00000003849) getting highest scores in both calculations exhibits a high similarity with gene FOXO (the corresponding protein is forkhead box protein O) in many other organisms. The importance of FOXO in metabolism, cellular proliferation, stress tolerance and lifespan have long been recognized by researchers [77, 78]. FOXOs are crucial regulators of cellular homeostasis that have a conserved role in modulating organismal aging and fitness [79]. More interestingly, several recent studies have demonstrated that FOXO transcription factors control the growth and moulting in insects [80–82]. The transcription factor encoded by FOXO gene in *Drosophila* (dFOXO) was reported to be involved in regulation of developmental timing through interaction with moulting hormonoe ecdysone [83]. Combining our analysis results with these published papers, we propose that it is worth investigating whether the FOXO-like transcription factor plays a crucial role in salmon louse moulting process.

Neither of the transcript (RAB1A-like) with strongest RNAi phenotype nor the transcript (FOXO-like) with highest RIF scores were detected as differentially expressed between middle samples and old/moulting samples, indicating that differential expression analysis is not the best choice when it comes to identify genes that play a key role in regulating a certain phenotype. In a typical differential expression analysis, genes are simply compared to themselves, leading to an overlook of much information. On the other hand, some genes known to be related to a mutant phenotype or a disease are not differentially expressed compared with control samples [48, 84], because many mutations or post-translational modifications alter the ac-

tivities and functions of genes without affecting their expression levels [85]. In fact, one particularly powerful advantage of network-based analysis is that it can provide an overall view of the alterations of interactions among genes in different sample groups. Furthermore, our gene co-expression network-based analysis approach circumvents the multiple testing problem that plagues conventional differential gene expression analysis.

In summary, the results of RNAi experiments provided preliminary evidence that our GCN-based analysis approach is effective for identifying genes associated with a certain phenotype. The generally accepted view that hubs in biological network tend to be essential was also testified in this study. We propose that hubs from the modules selected based on our criteria are more likely to be important for the phenotype of interest. We look forward to further investigation of the interesting genes identified in this study and generalization of our approach to the studies of other organisms.

Acknowledgements

We would like to thank Heidi Kongshaug, Lars Are Hamre and Per Gunnar Espedal for technical help in the laboratory, and we would like to thank Kjell Petersen for helpful comments and suggestions on the manuscript.

Funding

This project was funded by the Research Council of Norway, SFI-Sea Lice Research Centre, grant number 203513/O30. Further, this work was funded by the ELIXIR2 (270068) infrastructure grant from the Research Council of Norway to MD.

Abbreviations

BC: betweenness centrality; BLAST: Basic Local Alignment Search Tool; BP: biological process; CT: threshold cycle; CC: cellular component; DE: differentially expressed/differential gene expression; dsRNA: double-stranded RNA; DW: differential wiring; fw: forward; FWER: family-wise Type I error rate; GCN: gene co-expression network; GO: gene ontology; KD: knock down; kIM: intramodular connectivity; kME: module eigengene-based connectivity; MF: molecular function; RIF: regulatory impact factor; RNAi: RNA interference; RNA-seq: RNA sequencing; rv: reverse; TF: transcription factor.

Availability of data and materials

The datasets analyzed in this study and code are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

All experiments were performed according to Norwegian animal welfare regulations with the approval of the governmental Norwegian Animal Research Authority (ID7704, no 2010/245410).

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Author's contributions

IJ and FN obtained the funding for this project. MD and CE designed the study and provided the data. ZZ performed bioinformatics analyses and wrote the initial manuscript. CE performed the RNAi experiments and wrote the corresponding methods and results. ZZ, MD and CE edited and approved the submitted manuscript. All authors read and approved the final manuscript.

Author details

[1]Department of Informatics, University of Bergen, Thormøhlensgate 55, Bergen, Norway. [2]Department of Biology & Sea Lice Research Centre, University of Bergen, Thormøhlensgate 53, Bergen, Norway.

References

1. Martin, J.H.: The Possible Transport of Trace Metals Via Moulted Copepod Exoskeletons. Limnology and Oceanography **15**(5), 756–761 (1970)
2. Turner, J.T.: The importance of small planktonic copepods and their roles in pelagic marine food webs. Zool. Stud, 255–266 (2004)
3. Johnson, S.C., Albright, L.J.: The developmental stages of Lepeophtheirus salmonis (Krøyer, 1837) (Copepoda: Caligidae). Canadian Journal of Zoology **69**(4), 929–950 (1991). doi:10.1139/z91-138
4. Hamre, L.A., Eichner, C., Caipang, C.M.A., Dalvin, S.T., Bron, J.E., Nilsen, F., Boxshall, G., Skern-Mauritzen, R.: The Salmon Louse Lepeophtheirus salmonis (Copepoda: Caligidae) Life Cycle Has Only Two Chalimus Stages. PLoS ONE **8**(9) (2013). doi:10.1371/journal.pone.0073539
5. Abolofia, J., Asche, F., Wilen, J.E.: The Cost of Lice: Quantifying the Impacts of Parasitic Sea Lice on Farmed Salmon. Marine Resource Economics **32**(3), 329–349 (2017). doi:10.1086/691981
6. Denholm, I., Devine, G.J., Horsberg, T.E., Sevatdal, S., Fallang, A., Nolan, D.V., Powell, R.: Analysis and management of resistance to chemotherapeutants in salmon lice, Lepeophtheirus salmonis (Copepoda: Caligidae). Pest Management Science **58**(6), 528–536 (2002). doi:10.1002/ps.482
7. Aaen, S.M., Helgesen, K.O., Bakke, M.J., Kaur, K., Horsberg, T.E.: Drug resistance in sea lice: a threat to salmonid aquaculture. Trends in Parasitology **31**(2), 72–81 (2015). doi:10.1016/j.pt.2014.12.006
8. Eichner, C., Dondrup, M., Nilsen, F.: RNA sequencing reveals distinct gene expression patterns during the development of parasitic larval stages of the salmon louse (Lepeophtheirus salmonis). Journal of Fish Diseases **41**(6), 1005–1029 (2018). doi:10.1111/jfd.12770
9. Schram, T.A.: Supplementary descriptions of the developmental stages of Lepeophtheirus salmonis (Kroyer, 1837) (Copepoda: Caligidae). In: Pathogens of Wild and Farmed Fish: Sea Lice, pp. 30–47. Ellis Horwood Limited, ??? (1993)
10. Chang, E.S., Mykles, D.L.: Regulation of crustacean molting: A review and our perspectives. General and Comparative Endocrinology **172**(3), 323–330 (2011). doi:10.1016/j.ygcen.2011.04.003
11. Zitnan, D., Kim, Y.-J., Zitnanová, I., Roller, L., Adams, M.E.: Complex steroid-peptide-receptor cascade controls insect ecdysis. General and comparative endocrinology **153**(1-3), 88–96 (2007). doi:10.1016/j.ygcen.2007.04.002
12. Tarrant, A.M., Baumgartner, M.F., Hansen, B.H., Altin, D., Nordtug, T., Olsen, A.J.: Transcriptional profiling of reproductive development, lipid storage and molting throughout the last juvenile stage of the marine copepod Calanus finmarchicus. Frontiers in Zoology **11**(1), 91 (2014). doi:10.1186/s12983-014-0091-8
13. Johnson, C.: Ecdysteroids in the oceanic copepod Calanus pacificus: Variation during molt cycle and change associated with diapause. Marine Ecology-progress Series - MAR ECOL-PROGR SER **257**, 159–165 (2003). doi:10.3354/meps257159
14. Sandlund, L., Nilsen, F., Male, R., Grotmol, S., Kongshaug, H., Dalvin, S.: Molecular characterisation of the salmon louse, Lepeophtheirus salmonis salmonis (Krøyer, 1837), ecdysone receptor with emphasis on functional studies of female reproduction. International Journal for Parasitology **45**(2), 175–185 (2015). doi:10.1016/j.ijpara.2014.10.003
15. Eichner, C., Dalvin, S., Skern-Mauritzen, R., Malde, K., Kongshaug, H., Nilsen, F.: Characterization of a novel RXR receptor in the salmon louse (Lepeophtheirus salmonis, Copepoda) regulating growth and female reproduction. BMC Genomics **16**(1), 81 (2015). doi:10.1186/s12864-015-1277-y

16. Poley, J.D., Braden, L.M., Messmer, A.M., Igboeli, O.O., Whyte, S.K., Macdonald, A., Rodriguez, J., Gameiro, M., Rufener, L., Bouvier, J., Wadowska, D.W., Koop, B.F., Hosking, B.C., Fast, M.D.: High level efficacy of lufenuron against sea lice (Lepeophtheirus salmonis) linked to rapid impact on moulting processes. International Journal for Parasitology: Drugs and Drug Resistance **8**(2), 174–188 (2018). doi:10.1016/j.ijpddr.2018.02.007
17. Hardardottir, H.M., Male, R., Nilsen, F., Dalvin, S.: Effects of chitin synthesis inhibitor treatment on Lepeophtheirus salmonis (Copepoda, Caligidae) larvae. PLOS ONE **14**(9), 0222520 (2019)
18. Soon, W.W., Hariharan, M., Snyder, M.P.: High-throughput sequencing for biology and medicine. Molecular Systems Biology **9**(1), 640 (2013). doi:10.1038/msb.2012.61
19. Park, S.T., Kim, J.: Trends in Next-Generation Sequencing and a New Era for Whole Genome Sequencing. International neurourology journal **20**(Suppl 2), 76–83 (2016). doi:10.5213/inj.1632742.371
20. Green, S.: Philosophy of Systems and Synthetic Biology, Winter 201 edn. Metaphysics Research Lab, Stanford Universit, ??? (2019). https://plato.stanford.edu/archives/win2019/entries/systems-synthetic-biology/
21. Zheng, Z.-L., Zhao, Y.: Transcriptome comparison and gene coexpression network analysis provide a systems view of citrus response to 'CandidatusLiberibacter asiaticus' infection. BMC Genomics **14**(1), 27 (2013). doi:10.1186/1471-2164-14-27
22. Barabási, A.-L., Oltvai, Z.N.: Network biology: understanding the cell's functional organization. Nature Reviews Genetics **5**(2), 101–113 (2004). doi:10.1038/nrg1272
23. Saelens, W., Cannoodt, R., Saeys, Y.: A comprehensive evaluation of module detection methods for gene expression data. Nature Communications **9**(1) (2018). doi:10.1038/s41467-018-03424-4
24. Zhang, B., Horvath, S.: A General Framework for Weighted Gene Co-Expression Network Analysis. Statistical Applications in Genetics and Molecular Biology **4**(1) (2005). doi:10.2202/1544-6115.1128
25. Oldham, M.C., Horvath, S., Geschwind, D.H.: Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proceedings of the National Academy of Sciences **103**(47), 17973–17978 (2006). doi:10.1073/pnas.0605938103
26. Horvath, S., Zhang, B., Carlson, M., Lu, K.V., Zhu, S., Felciano, R.M., Laurance, M.F., Zhao, W., Qi, S., Chen, Z., Lee, Y., Scheck, A.C., Liau, L.M., Wu, H., Geschwind, D.H., Febbo, P.G., Kornblum, H.I., Cloughesy, T.F., Nelson, S.F., Mischel, P.S.: Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. Proceedings of the National Academy of Sciences of the United States of America **103**(46), 17402–17407 (2006). doi:10.1073/pnas.0608396103
27. Gargalovic, P.S., Imura, M., Zhang, B., Gharavi, N.M., Clark, M.J., Pagnon, J., Yang, W.-P., He, A., Truong, A., Patel, S., Nelson, S.F., Horvath, S., Berliner, J.A., Kirchgessner, T.G., Lusis, A.J.: Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. Proceedings of the National Academy of Sciences **103**(34), 12741–12746 (2006). doi:10.1073/pnas.0605457103
28. van Dam, S., Võsa, U., van der Graaf, A., Franke, L., de Magalhães, J.P.: Gene co-expression analysis for functional classification and gene-disease predictions. Briefings in bioinformatics **19**(4), 575–592 (2018). doi:10.1093/bib/bbw139
29. Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D., Friedman, N.: Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nature Genetics **34**(2), 166–176 (2003). doi:10.1038/ng1165
30. Jiang, J., Sun, X., Wu, W., Li, L., Wu, H., Zhang, L., Yu, G., Li, Y.: Construction and application of a co-expression network in Mycobacterium tuberculosis. Scientific Reports **6**(1), 28422 (2016). doi:10.1038/srep28422
31. Kommadath, A., te Pas, M.F.W., Smits, M.A.: Gene coexpression network analysis identifies genes and biological processes shared among anterior pituitary and brain areas that affect estrous behavior in dairy cows. Journal of Dairy Science **96**(4), 2583–2595 (2013). doi:10.3168/jds.2012-5814
32. Langfelder, P., Horvath, S.: WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics **9** (2008). doi:10.1186/1471-2105-9-559
33. Langfelder, P., Horvath, S.: Fast R Functions for Robust Correlations and Hierarchical Clustering. Journal of statistical software **46**(11), 11 (2012)

34. Horvath Steve: Weighted Network Analysis: Applications in Genomics and Systems Biology, pp. 4–16. Springer, ??? (2011). doi:10.1007/978-1-4419-8819-5
35. Horvath, S., Dong, J.: Geometric Interpretation of Gene Coexpression Network Analysis. PLOS Computational Biology **4**(8), 1000117 (2008)
36. Freeman, L.: A Set of Measures of Centrality Based on Betweenness. Sociometry **40**, 35–41 (1977). doi:10.2307/3033543
37. Brandes, U.: A faster algorithm for betweenness centrality. The Journal of Mathematical Sociology **25**(2), 163–177 (2001). doi:10.1080/0022250X.2001.9990249
38. Opsahl, T.: Structure and Evolution of Weighted Networks. PhD thesis, Queen Mary College University of London (2009)
39. Langfelder, P., Luo, R., Oldham, M.C., Horvath, S.: Is My Network Module Preserved and Reproducible? PLOS Computational Biology **7**(1), 1001057 (2011)
40. Friedman, J., Hastie, T., Tibshirani, R.: Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. **33**, 1–22 (2010)
41. Schmidt, E.E., Pelz, O., Buhlmann, S., Kerr, G., Horn, T., Boutros, M.: GenomeRNAi: A database for cell-based and in vivo RNAi phenotypes, 2013 update. Nucleic Acids Research **41**(D1), 1021–1026 (2013). doi:10.1093/nar/gks1170
42. Alexa, A., Rahnenfuhrer, J.: topGO: Enrichment Analysis for Gene Ontology (2018)
43. Alexa, Adrian; Rahnenfuhrer, J.: Gene set enrichment analysis with topGO (2020)
44. Wolfe, C.J., Kohane, I.S., Butte, A.J.: Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. BMC Bioinformatics **6**(1), 227 (2005). doi:10.1186/1471-2105-6-227
45. Jeong, H., Mason, S.P., Barabási, A.-L., Oltvai, Z.N.: Lethality and centrality in protein networks. Nature **411**(6833), 41–42 (2001). doi:10.1038/35075138
46. Hahn, M.W., Kern, A.D.: Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks. Molecular Biology and Evolution **22**(4), 803–806 (2004). doi:10.1093/molbev/msi072
47. Han, J.-D.J., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J.M., Cusick, M.E., Roth, F.P., Vidal, M.: Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature **430**(6995), 88–93 (2004). doi:10.1038/nature02555
48. Reverter, A., Hudson, N.J., Nagaraj, S.H., Pérez-Enciso, M., Dalrymple, B.P.: Regulatory impact factors: Unraveling the transcriptional regulation of complex traits from expression data. Bioinformatics **26**(7), 896–904 (2010). doi:10.1093/bioinformatics/btq051
49. Pollard, K.S., Dudoit, S., Van der Laan, M.J.: Multiple Testing Procedures: R multtest Package and Applications to Genomics. U.C. Berkeley Division of Biostatistics Working Paper Series. (2004)
50. Eichner, C., Nilsen, F., Grotmol, S., Dalvin, S.: A method for stable gene knock-down by RNA interference in larvae of the salmon louse (Lepeophtheirus salmonis). Experimental Parasitology **140**, 44–51 (2014). doi:10.1016/j.exppara.2014.03.014
51. Eichner, C., Øvergård, A.C., Nilsen, F., Dalvin, S.: Molecular characterization and knock-down of salmon louse (Lepeophtheirus salmonis) prostaglandin E synthase. Experimental Parasitology **159**, 79–93 (2015). doi:10.1016/j.exppara.2015.09.001
52. Dalvin, S., Frost, P., Biering, E., Hamre, L.A., Eichner, C., Krossøy, B., Nilsen, F.: Functional characterisation of the maternal yolk-associated protein (LsYAP) utilising systemic RNA interference in the salmon louse (Lepeophtheirus salmonis) (Crustacea: Copepoda). International Journal for Parasitology **39**(13), 1407–1415 (2009). doi:10.1016/j.ijpara.2009.04.004
53. Frost, P., Nilsen, F.: Validation of reference genes for transcription profiling in the salmon louse, Lepeophtheirus salmonis, by quantitative real-time PCR. Veterinary Parasitology **118**(1), 169–174 (2003). doi:10.1016/j.vetpar.2003.09.020
54. Eichner, C., Harasimczuk, E., Nilsen, F., Grotmol, S., Dalvin, S.: Molecular characterisation and functional analysis of LsChi2, a chitinase found in the salmon louse (Lepeophtheirus salmonis salmonis, Krøyer 1838). Experimental Parasitology **151-152**, 39–48 (2015). doi:10.1016/j.exppara.2015.01.011
55. Banerjee, S., Banerjee, A., Gill, S.S., Gupta, O.P., Dahuja, A., Jain, P.K., Sirohi, A.: RNA Interference: A

Novel Source of Resistance to Combat Plant Parasitic Nematodes (2017). https://www.frontiersin.org/article/10.3389/fpls.2017.00834

56. Kolev, N.G., Tschudi, C., Ullu, E.: RNA interference in protozoan parasites: achievements and challenges. Eukaryotic cell **10**(9), 1156–1163 (2011). doi:10.1128/EC.05114-11
57. Bucher, G., Scholten, J., Klingler, M.: Parental RNAi in Tribolium (Coleoptera). Current Biology **12**(3), 85–86 (2002). doi:10.1016/S0960-9822(02)00666-8
58. Dompe, N., Rivers, C.S., Li, L., Cordes, S., Schwickart, M., Punnoose, E.A., Amler, L., Seshagiri, S., Tang, J., Modrusan, Z., Davis, D.P.: A whole-genome RNAi screen identifies an 8q22 gene cluster that inhibits death receptor-mediated apoptosis. Proceedings of the National Academy of Sciences of the United States of America **108**(43), 943–951 (2011). doi:10.1073/pnas.1100132108
59. Parry, D.H., Xu, J., Ruvkun, G.: A whole-genome RNAi Screen for C. elegans miRNA pathway genes. Current biology : CB **17**(23), 2013–2022 (2007). doi:10.1016/j.cub.2007.10.058
60. Marie, C., Verkerke, H.P., Theodorescu, D., Petri, W.A.: A whole-genome RNAi screen uncovers a novel role for human potassium channels in cell killing by the parasite Entamoeba histolytica. Scientific reports **5**, 13613 (2015). doi:10.1038/srep13613
61. Hadley, N.F.: The Arthropod Cuticle. Scientific American **255**(1), 104–113 (1986)
62. Zhang, X., Chang, H., Dong, Z., Zhang, Y., Zhao, D., Ye, L., Xia, Q., Zhao, P.: Comparative Proteome Analysis Reveals that Cuticular Proteins Analogous to Peritrophin-Motif Proteins are Involved in the Regeneration of Chitin Layer in the Silk Gland of Bombyx mori at the Molting Stage. PROTEOMICS **18**(19), 1700389 (2018). doi:10.1002/pmic.201700389
63. Noh, M.Y., Muthukrishnan, S., Kramer, K.J., Arakane, Y.: A chitinase with two catalytic domains is required for organization of the cuticular extracellular matrix of a beetle. PLOS Genetics **14**(3), 1007307 (2018)
64. Aragon, S., Claudinot, S., Blais, C., Maibeche, M., Dauphin-Villemant, C.: Molting cycle-dependent expression of CYP4C15, a cytochrome P450 enzyme putatively involved in ecdysteroidogenesis in the crayfish, Orconectes limosus. Insect Biochemistry and Molecular Biology **32**(2), 153–159 (2002). doi:10.1016/S0965-1748(01)00095-9
65. Said, M.R., Begley, T.J., Oppenheim, A.V., Lauffenburger, D.A., Samson, L.D.: Global network analysis of phenotypic effects: protein networks and toxicity modulation in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America **101**(52), 18006–18011 (2004). doi:10.1073/pnas.0405996101
66. Wuchty, S.: Interaction and domain networks of yeast. PROTEOMICS **2**(12), 1715–1723 (2002). doi:10.1002/1615-9861(200212)2:12¡1715::AID-PROT1715¿3.0.CO;2-O
67. Hahn, M.W., Kern, A.D.: Comparative Genomics of Centrality and Essentiality in Three Eukaryotic Protein-Interaction Networks. Molecular Biology and Evolution **22**(4), 803–806 (2005). doi:10.1093/molbev/msi072
68. Goymer, P.: Why do we need hubs? Nature Reviews Genetics **9**(9), 651 (2008). doi:10.1038/nrg2450
69. Heggland, E.I., Eichner, C., Støve, S.I., Martinez, A., Nilsen, F., Dondrup, M.: A scavenger receptor B (CD36)-like protein is a potential mediator of intestinal heme absorption in the hematophagous ectoparasite Lepeophtheirus salmonis. Scientific Reports **9**(1), 4218 (2019). doi:10.1038/s41598-019-40590-x
70. Borin, G.P., Carazzolle, M.F., Dos Santos, R.A.C., Riaño-Pachón, D.M., Oliveira, J.V.d.C.: Gene Co-expression Network Reveals Potential New Genes Related to Sugarcane Bagasse Degradation in Trichoderma reesei RUT-30. Frontiers in bioengineering and biotechnology **6**, 151 (2018). doi:10.3389/fbioe.2018.00151
71. Uno, T., Sakamoto, K., Isoyama, Y., Hiragaki, S., Uno, Y., Kanamaru, K., Yamagata, H., Takagi, M., Mizoguchi, A., Takeda, M.: Relationship between the expression of Rab family GTPases and neuropeptide hormones in the brain of Bombyx mori. Histochemistry and Cell Biology **139**(2), 299–308 (2013). doi:10.1007/s00418-012-1021-5
72. Hiragaki, S., Uno, T., Takeda, M.: Putative regulatory mechanism of prothoracicotropic hormone (PTTH) secretion in the American cockroach, Periplaneta americana as inferred from co-localization of Rab8, PTTH, and protein kinase C in neurosecretory cells. Cell and Tissue Research **335**(3), 607–615 (2009). doi:10.1007/s00441-008-0747-9
73. Niwa, Y.S., Niwa, R.: Transcriptional regulation of insect steroid hormone biosynthesis and its role in

controlling timing of molting and metamorphosis. Development Growth and Differentiation **58**(1), 94–105 (2016). doi:10.1111/dgd.12248

74. Stenmark, H., Olkkonen, V.M.: The Rab GTPase family. Genome Biology **2**(5), 3007–1 (2001). doi:10.1186/gb-2001-2-5-reviews3007
75. Field, M.C., Horn, D., Carrington, M.B.T.-M.i.E.: Analysis of Small GTPase Function in Trypanosomes. In: Small GTPases in Disease, Part A vol. 438, pp. 57–76. Academic Press, ??? (2008). doi:10.1016/S0076-6879(07)38005-1. http://www.sciencedirect.com/science/article/pii/S0076687907380051
76. Tatsumoto, T., Xie, X., Blumenthal, R., Okamoto, I., Miki, T.: Human ECT2 is an exchange factor for Rho GTPases, phosphorylated in G2/M phases, and involved in cytokinesis. The Journal of cell biology **147**(5), 921–928 (1999). doi:10.1083/jcb.147.5.921
77. van der Horst, A., Burgering, B.M.T.: Stressing the role of FoxO proteins in lifespan and disease. Nature Reviews Molecular Cell Biology **8**(6), 440–450 (2007). doi:10.1038/nrm2190
78. Martins, R., Lithgow, G.J., Link, W.: Long live FOXO: unraveling the role of FOXO proteins in aging and longevity. Aging cell **15**(2), 196–207 (2016). doi:10.1111/acel.12427
79. Spellberg, M.J., Marr, M.T.: FOXO regulates RNA interference in Drosophila and protects from RNA virus infection. Proceedings of the National Academy of Sciences **112**(47), 14587–14592 (2015). doi:10.1073/pnas.1517124112
80. Zeng, B., Huang, Y., Xu, J., Shiotsuki, T., Bai, H., Palli, S.R., Huang, Y., Tan, A.: The FOXO transcription factor controls insect growth and development by regulating juvenile hormone degradation in the silkworm, Bombyx mori. The Journal of biological chemistry **292**(28), 11659–11669 (2017). doi:10.1074/jbc.M117.777797
81. Cai, M.-J., Zhao, W.-L., Jing, Y.-P., Song, Q., Zhang, X.-Q., Wang, J.-X., Zhao, X.-F.: 20-hydroxyecdysone activates Forkhead box O to promote proteolysis during Helicoverpa armigera molting. Development **143**(6), 1005–1015 (2016). doi:10.1242/dev.128694
82. Hossain, M.S., Liu, Y., Zhou, S., Li, K., Tian, L., Li, S.: 20-Hydroxyecdysone-induced transcriptional activity of FoxO upregulates brummer and acid lipase-1 and promotes lipolysis in Bombyx fat body. Insect biochemistry and molecular biology **43**(9), 829–838 (2013). doi:10.1016/j.ibmb.2013.06.007
83. Colombani, J., Bianchini, L., Layalle, S., Pondeville, E., Dauphin-Villemant, C., Antoniewski, C., Carré, C., Noselli, S., Léopold, P.: Antagonistic Actions of Ecdysone and Insulins Determine Final Size in Drosophila. Science **310**(5748), 667–670 (2005). doi:10.1126/science.1119432
84. Hudson, N.J., Reverter, A., Dalrymple, B.P.: A Differential Wiring Analysis of Expression Data Correctly Identifies the Gene Containing the Causal Mutation. PLoS Computational Biology **5**(5), 1000382 (2009). doi:10.1371/journal.pcbi.1000382
85. de la Fuente, A.: From 'differential expression' to 'differential networking' – identification of dysfunctional regulatory networks in diseases. Trends in Genetics **26**(7), 326–333 (2010). doi:10.1016/j.tig.2010.05.001
86. Sandlund, L., Nilsen, F., Male, R., Dalvin, S.: The ecdysone receptor (EcR) is a major regulator of tissue development and growth in the marine salmonid ectoparasite, Lepeophtheirus salmonis (Copepoda, Caligidae). Molecular and Biochemical Parasitology **208**(2), 65–73 (2016). doi:10.1016/j.molbiopara.2016.06.007
87. Chen, J., Liang, Z., Liang, Y., Pang, R., Zhang, W.: Conserved microRNAs miR-8-5p and miR-2a-3p modulate chitin biosynthesis in response to 20-hydroxyecdysone signaling in the brown planthopper, Nilaparvata lugens. Insect Biochemistry and Molecular Biology **43**(9), 839–848 (2013). doi:10.1016/j.ibmb.2013.06.002
88. Braden, L., Michaud, D., Igboeli, O.O., Dondrup, M., Hamre, L., Dalvin, S., Purcell, S.L., Kongshaug, H., Eichner, C., Nilsen, F., Fast, M.D.: Identification of critical enzymes in the salmon louse chitin synthesis pathway as revealed by RNA interference-mediated abrogation of infectivity. International Journal for Parasitology **50**(10), 873–889 (2020). doi:10.1016/j.ijpara.2020.06.007
89. Surholt, B.: Formation of glucosamine-6-phosphate in chitin synthesis during ecdysis of the migratory locust, Locusta migratoria. Insect Biochemistry **5**(5), 585–593 (1975). doi:10.1016/0020-1790(75)90040-2
90. Oikari, S., Makkonen, K., Deen, A.J., Tyni, I., Kärnä, R., Tammi, R.H., Tammi, M.I.: Hexosamine biosynthesis in keratinocytes: roles of GFAT and GNPDA enzymes in the maintenance of UDP-GlcNAc content and hyaluronan synthesis. Glycobiology **26**(7), 710–722 (2016). doi:10.1093/glycob/cww019
91. Vrba, J., Macháéek, J.: Release of dissolved extracellular β-N-acetylglucosaminidase during crustacean

moulting. Limnology and Oceanography **39**(3), 712–716 (1994). doi:10.4319/lo.1994.39.3.0712

92. Lyu, Z., Chen, J., Li, Z., Cheng, J., Wang, C., Lin, T.: Knockdown of β-N-acetylglucosaminidase gene disrupts molting process in Heortia vitessoides Moore. Archives of Insect Biochemistry and Physiology **101**(4), 21561 (2019). doi:10.1002/arch.21561
93. Poley, J.D., Sutherland, B.J.G., Jones, S.R.M., Koop, B.F., Fast, M.D.: Sex-biased gene expression and sequence conservation in Atlantic and Pacific salmon lice (Lepeophtheirus salmonis). BMC Genomics **17**(1), 483 (2016). doi:10.1186/s12864-016-2835-7
94. Chen, J., Tang, B., Chen, H., Yao, Q., Huang, X., Chen, J., Zhang, D., Zhang, W.: Different Functions of the Insect Soluble and Membrane-Bound Trehalase Genes in Chitin Biosynthesis Revealed by RNA Interference. PLOS ONE **5**(4), 10133 (2010)
95. Merzendorfer, H.: The cellular basis of chitin synthesis in fungi and insects: Common principles and differences. European Journal of Cell Biology **90**(9), 759–769 (2011). doi:10.1016/j.ejcb.2011.04.014
96. Frand, A.R., Russel, S., Ruvkun, G.: Functional Genomic Analysis of C. elegans Molting. PLOS Biology **3**(10), 312 (2005)
97. Schumann, I., Hering, L., Mayer, G., Kenny, N., Hui, J.: Halloween genes in panarthropods and the evolution of the early moulting pathway in Ecdysozoa. Royal Society Open Science (2018). doi:10.1098/rsos.180888
98. Nakagawa, Y., Henrich, V.C.: Arthropod nuclear receptors and their role in molting. The FEBS Journal **276**(21), 6128–6157 (2009). doi:10.1111/j.1742-4658.2009.07347.x
99. Lavorgna, G., Karim, F.D., Thummel, C.S., Wu, C.: Potential role for a FTZ-F1 steroid receptor superfamily member in the control of Drosophila metamorphosis. Proceedings of the National Academy of Sciences **90**(7), 3004–3008 (1993). doi:10.1073/pnas.90.7.3004
100. Piulachs, M.-D., Pagone, V., Bellés, X.: Key roles of the Broad-Complex gene in insect embryogenesis. Insect Biochemistry and Molecular Biology **40**(6), 468–475 (2010). doi:10.1016/j.ibmb.2010.04.006
101. Maldonado-Aguayo, W., Chávez-Mardones, J., Gonçalves, A.T., Gallardo-Escárate, C.: Cathepsin Gene Family Reveals Transcriptome Patterns Related to the Infective Stages of the Salmon Louse Caligus rogercresseyi. PLOS ONE **10**(4), 0123954 (2015)
102. Nagai, C., Mabashi-Asazuma, H., Nagasawa, H., Nagata, S.: Identification and Characterization of Receptors for Ion Transport Peptide (ITP) and ITP-like (ITPL) in the Silkworm Bombyx mori*. Journal of Biological Chemistry **289**(46), 32166–32177 (2014). doi:10.1074/jbc.M114.590646
103. Webster, S.G., Keller, R., Dircksen, H.: The CHH-superfamily of multifunctional peptide hormones controlling crustacean metabolism, osmoregulation, moulting, and reproduction. General and Comparative Endocrinology **175**(2), 217–233 (2012). doi:10.1016/j.ygcen.2011.11.035

Figures

Tables

Additional Files

Additional file 1 — supplementary methods.pdf

The methods used to estimate the power parameter and calculate the module preservation statistics.

Additional file 2 — analysis results.ods

The main analysis results of this study.

Table S1: Homologues-based transcript-phenotype list

Table S2-S3: Analysis results for the modules in the middle network and moulting network respectively, including module size, enriched GO terms with smallest p-values, p-values of enrichment analyses based on homologues, module preservation $Z_summary$, number of known moulting-associated genes, proportion of TF and DE genes.

Table S4: Analysis results for the modules in the global network, including module size, enriched GO terms with smallest p-values, p-values of enrichment analyses based on homologues, regularized logistic regression coefficients, number of known moulting-associated genes, proportion of TF and DE genes.

Table S5-S7: Module assignment results for the known moulting-associated transcripts in the middle, moulting and global network; the ranks of these transcripts based on three types of centrality measurements within modules; whether they were intramodular hubs or not.

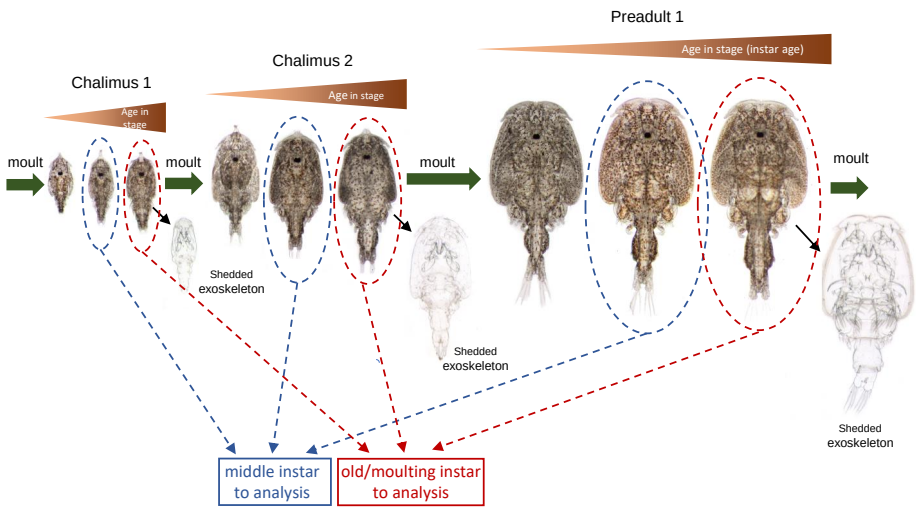

Figure 1 Photographs of representative Lepeophtheirus salmonis chalimus-1, chalimus-2 and preadult-1 larvae. Within the stage, lice were divided into groups of different instar ages: directly after moulting (young), in the middle of the stage (middle) and directly before the moult to the next stage (old/moulting). The moults are represented by a green arrow and a shedded exoskeleton. For this study data from lice of the middle and old/moulting instar age were used.

Table S8: Two types of RIF scores for the transcripts annotated as transcription factors

Table S9-S11: Centrality measurements, average ranks and annotations for the nodes in three selected modules (yellowgreen, steelblue and violet).

Table S12: Annotations of the nodes with highest average ranks from the selected modules.

Table S13: Available RNAi results for nodes from the eight selected important modules and nodes with high average ranks from modules without passing any criteria.

Additional file 3 — supplementary figures.pdf

Figure S1: Hierarchical clustering tree (dendrogram) of nodes in the middle, moulting and global network.

Figure S2: Histological sections of adult female louse tissues from control samples and samples with selected genes knock-down.

Figure S3: Histological sections of copepodids sampled three days post infection.

Additional file 4 — yellowgreen.xgmml

Visualization file for the selected important module "yellowgreen" from the moulting network.

Additional file 5 — violet.xgmml

Visualization file for the selected important module "violet" from the moulting network.

Additional file 6 — steelblueGlobal.xgmml

Visualization file for the selected important module "steelblue" from the global network.

Additional file 7 — Adjacency matrices.RData

Adjacency matrices for the middle, moulting and global network respectively.

Supplementary data has been deposited in Figshare DOI: 10.6084/m9.figshare.c.5375315

Table 1 ID, annotation and relevant publications for the moulting-associated transcripts

TranscriptID	Annotation from LiceBase	Publication
EMLSAT00000012651	Nuclear receptor subfamily 1 group I member 2	[86]
EMLSAT00000004055	Phosphoacetylglucosamine mutase	[87]
EMLSAT00000002853	Chitin synthase 8	[17]
EMLSAT00000007308	Chitin synthase 8	[88]
EMLSAT00000012864	Probable glucosamine 6-phosphate N-acetyltransferase	[89]
EMLSAT00000000683	Glutamine–fructose-6-phosphate aminotransferase [isomerizing] 2	[90]
EMLSAT00000008812	Acidic mammalian chitinase	[17]
EMLSAT00000005289	Probable chitinase 2	[17]
EMLSAT00000008235	Acidic mammalian chitinase	[17]
EMLSAT00000005464	Chitooligosaccharidolytic beta-N-acetylglucosaminidase	[91]
EMLSAT00000008487	Beta-hexosaminidase subunit alpha	[92]
EMLSAT00000000150	N-acetylglucosamine-6-phosphate deacetylase	[16]
EMLSAT00000001322	Glucosamine-6-phosphate isomerase	[16]
EMLSAT00000004643	Hexokinase type 2	[93]
EMLSAT00000005083	Hexokinase type 2	[93]
EMLSAT00000001264	Hexokinase type 2	[93]
EMLSAT00000008931	Glucose-6-phosphate isomerase	[94]
EMLSAT00000003220	Mannose-6-phosphate isomerase	[95]
EMLSAT00000007185	Phosphoglucomutase-2	[95]
EMLSAT00000011599	Phosphoglucomutase-1	[95]
EMLSAT00000001990	TMEM47	[96]
EMLSAT00000001904	Ecdysone-induced protein 78C	[97]
EMLSAT00000000733	Probable nuclear hormone receptor HR3	[98]
EMLSAT00000008543	Hormone receptor 4	[98]
EMLSAT00000008902	Nuclear hormone receptor FTZ-F1	[99]
EMLSAT00000008651	Ecdysone-induced protein 74EF isoform A	[86]
EMLSAT00000000188	Zinc finger protein 142	[100]
EMLSAT00000009170	Growth hormone secretagogue receptor type 1	[50]
EMLSAT00000008881	Cathepsin L	[101]
EMLSAT00000001278	Ion transport peptide	[102]
EMLSAT00000008115	CHH-like protein	[103]
EMLSAT00000001150	Cytochrome P450 307a1	[64]

Table 2 Information of the chosen important modules for further knock-down candidates selection

Module	Network	Size	$Z_{summary}$	Number of moulting-associated genes	Number of TF genes	$p-value$	
						RNAi phenotype	RNAi lethal phenotype
yellowgreen	moulting	81	1.92	1	7	0.84	0.54
violet	moulting	83	8.52	1	10	0.026	0.0076
steelblue	global	79	–	1	0	0.90	0.94

Table 3 Information of the chosen knock-down candidate genes

TranscriptID	Module	Network	Number of paralogues within salmon louse genome	Differentially expressed	Annotated as TF gene	Rank			
						kIM-score	BC-score	kME-score	Average
EMLSAT00000001458	yellowgreen	moulting	60	False	False	1	1	1	1
EMLSAT00000008959	yellowgreen	moulting	0	False	False	6	19	7	10.7
EMLSAT00000005299	violet	moulting	0	False	False	1	1	1	1
EMLSAT00000003179	violet	moulting	0	False	True	18	3	17	12.7
EMLSAT00000004347	steelblue	global	2	True	False	24	4	24	17.3

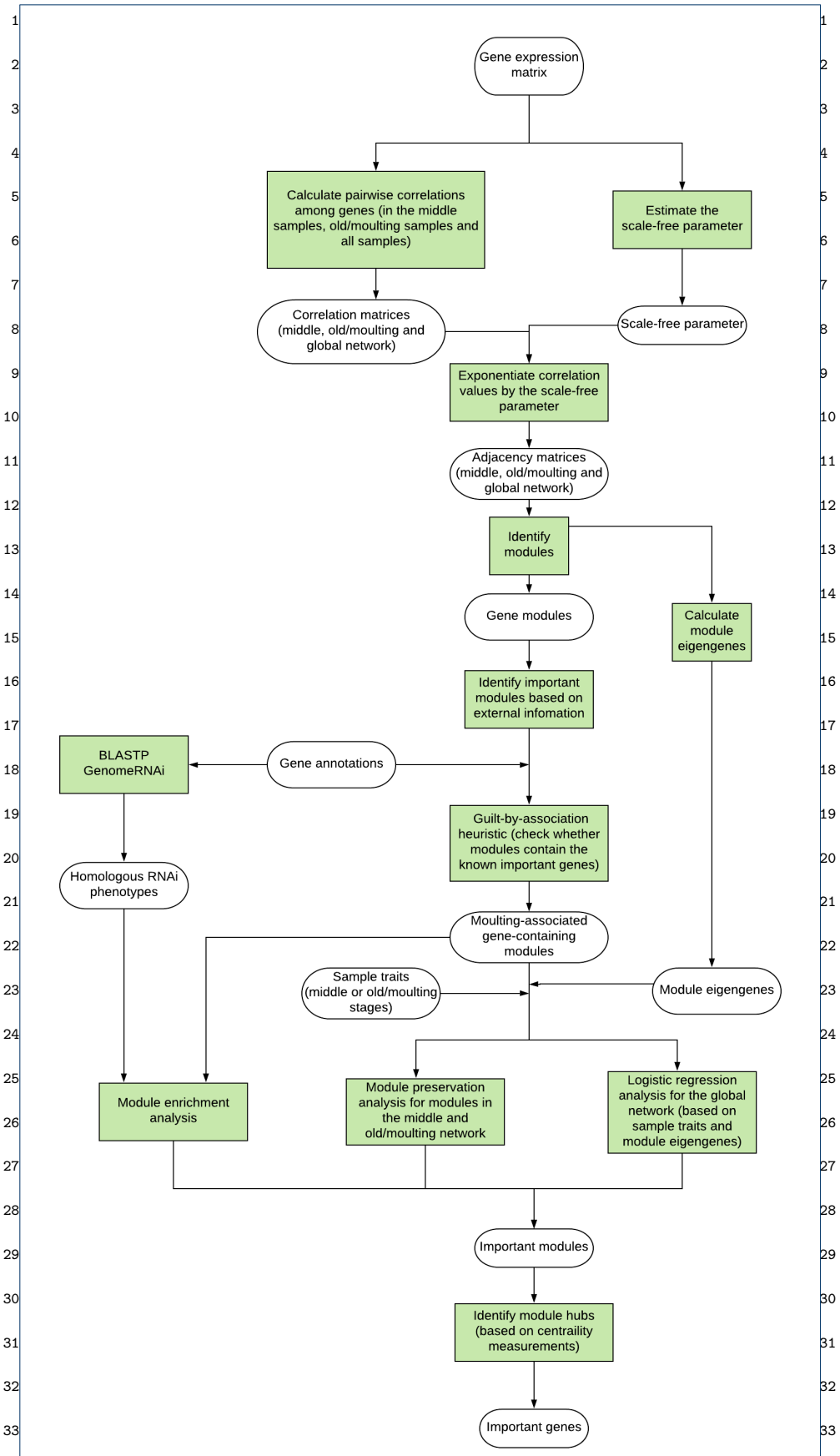

Figure 2 The workflow of gene co-expression network (GCN) analysis for identifying important modules and genes. Methods are highlighted as lightgreen.

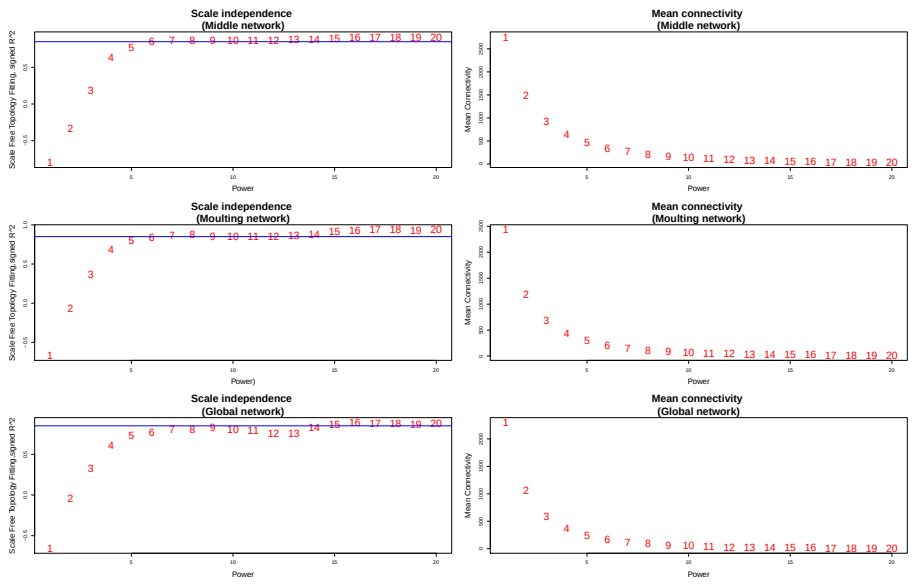

Figure 3 Network scale-free topology and mean connectivity for different powers ranging from 1 to 20. Plots on the left panel show the scale-free topology fitting index (y-axis) as a function of the power (x-axis) in the middle network, moulting network and global network respectively. A horizontal line is added at the height of 0.85 on each of the plot. Plots on the right panel show the mean connectivity (y-axis) as a function of the power (x-axis) in the middle network, moulting network and global network respectively.

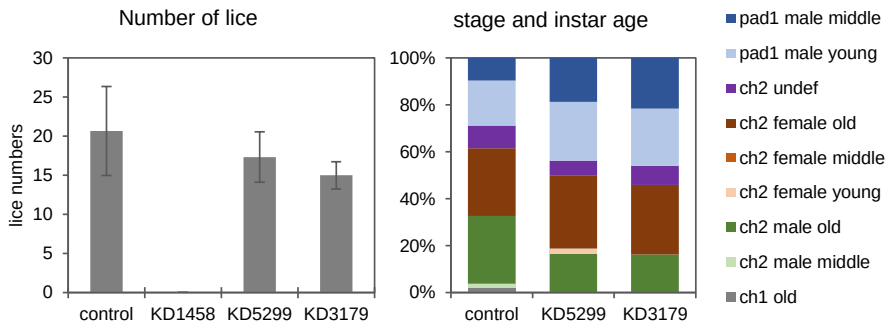

Figure 4 Number of lice (left panel) and stage and instar age of these lice (right panel) at termination 16 days post infection. N=3.

Table 4 RNAi phenotypes for the chosen knock-down candidate genes

TranscriptID	Eggstring length	Cephalothorax length	Genital segment length	Other phenotype changes
EMLSAT00000001458	no eggstring	no significant difference to control lice	significant shorter (29%) than control lice	–
EMLSAT00000008959	no eggstring	no significant difference to control lice	no significant difference to control lice	–
EMLSAT00000005299	on AVG 20% shorter	no significant difference to control lice	slightly but significant shorter (4%) than control lice	–
EMLSAT00000003179	no significant difference to control lice	slightly but significant shorter (3%) than control lice	slightly but significant shorter (7%) than control lice	–
EMLSAT00000004347	on AVG 18% shorter	slightly but significant longer (2%) than control lice	slightly but significant shorter (6%) than control lice	weak GS (easily squeezed and torn apart)

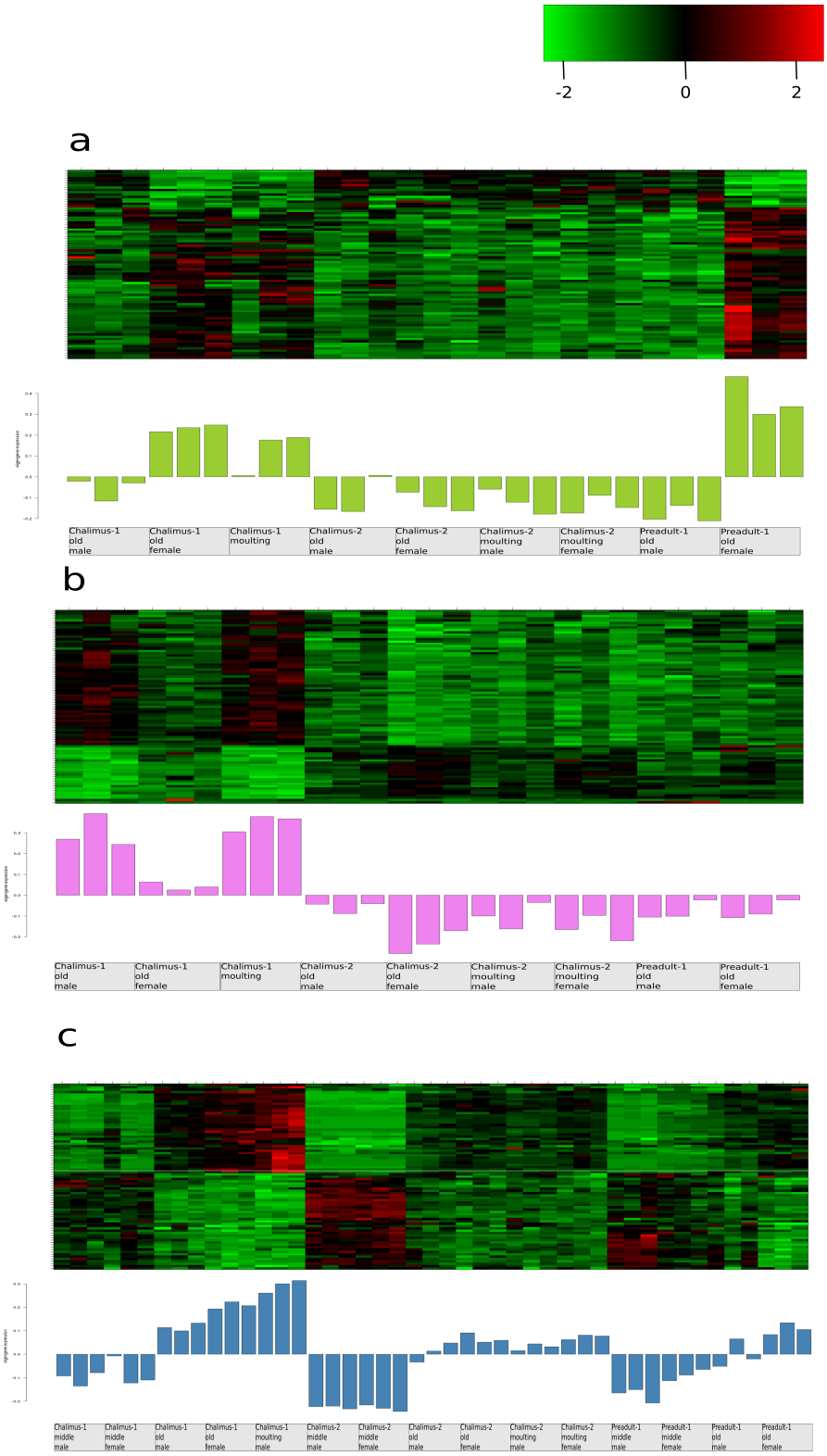

Figure 5 Heatmap of scaled gene expression profiles and barplot of scaled eigengene expression for module "yellowgreen" extracted from the moulting network (a), module "violet" extracted from the moulting network (b) and module "steelblue" extracted from the global network (c). The heatmap color-codes the entries for the gene expression matrix of each module: higher expression values are represented with reds of increasing RGB intensity, and lower expression values are represented with greens of increasing RGB intensity. The barplot underneath the heatmap shows the expression levels of module eigengene in different samples(x-axis). The barplots are colored based on the name of each module.

Figure 6 Graph representation of the module "yellowgreen" extracted from the moulting network (a), module "violet" extracted from the moulting network (b) and module "steelblue" extracted from the global network (c). To make sure the readability, only edges representing correlations with an absolute value larger than 0.84 are shown. Nodes are labeled based on their average rank in all centrality measurements, and node label font sizes are proportional to the nodes average ranks. Nodes are colored as red if the corresponding genes are annotated as transcription factors. The purple nodes represent the moulting-associated genes. Differentially expressed genes between middle group and moulting group are displayed as diamond shape nodes. Nodes with larger weighted betweenness centrality are drawn with larger sizes. Annotation information of the top 10 nodes in terms of average rank are listed.