Green Etxabe, Amaia; Pini, Jennifer M.; Short, Stephen; Cunha, Luis; Kille, Peter; Watson, Gordon J. 2021. **Identifying conserved polychaete molecular markers of metal exposure: comparative analyses using the Alitta virens (Annelida, Lophotrochozoa) transcriptome.**

The definitive version was published in *Comparative Biochemistry and Physiology, Part C: Toxicology & Pharmacology*, 240, 108913.
https://doi.org/10.1016/j.cbpc.2020.108913

Identifying conserved polychaete molecular markers of metal exposure: comparative analyses using the *Alitta virens* (Annelida, Lophotrochozoa) transcriptome.

Supplementary material

Table S1. Exposure conditions and animal numbers associated with RNA isolated for RNASeq and qPCR samples. The two RNAseq libraries were created using equimolar amounts of RNA to produce a control and exposed sample.

RNASeq		**qPCR**	
Experimental condition	**Animals per sample**	**Experimental condition**	**Animals per sample**
3 months control	n=16	3 months control	n=16
6 months control	n=16	6 months control	n=16
Total control RNASeq sample	n=32		
3 months low Cu	n=3	3 months low Cu	n=11
3 months high Cu	n=3	3 months high Cu	n=15
3 months low Zn	n=3	3 months low Zn	n=15
3 months high Zn	n=3	3 months high Zn	n=12
3 months high Cu-Zn	n=3	3 months high Cu-Zn	n=12
3 months low Cu-Zn	n=3	3 months low Cu-Zn	n=14
6 months low Cu	n=3	6 months low Cu	n=17
6 months low Cu	n=3	6 months low Cu	n=15
6 months low Zn	n=3	6 months low Zn	n=11
6 months low Zn	n=3	6 months low Zn	n=14
6 months low Cu-Zn	n=3	6 months low Cu-Zn	n=12
6 months low Cu-Zn	n=3	6 months low Cu-Zn	n=11
Total exposed RNASeq sample	n=36		

Table S2. Candidate genes and primer sets used for qPCR.

Gene	Primer	Sequence (5'-3')
Elong-like	AvElong-F	TATTCTCAAGCCCGGTATGG
	AvElong-R	GATCTCTCCAGGGTGGTTGA
GAPDH-like	AvGapdh-F	TCCTGACCTCAACGGAAAAC
	AvGapdh-R	AGGTGTGAGTGCAATGCAAG
Beta actin-like	AvBeta_actin-F	GCTCCATCCACCATGAAGAT
	AvBeta_actin-R	GTGAAAGATGGCAGAGCACA
Alpha tubulin-like	AvAlpha_tubulin-F	CAACTTGGTGCCCTACCCTA
	AvAlpha_tubulin-R	GAACGCTTGGTCTTGATGGT
Ubiquitin 60S-like	AvUbiquitin_60S-F	CCCGATCAACAGCGTCTTAT
	AvUbiquitin-60S-R	GTGTGTCCGCACTTCCTTTT
Ubiquitin-like	AvUnk1-F	ACGAAGTTGCAGCTGATGGA
	AvUnk1-R	CCAGAAGTGCAAAGTGCCAC
Unk1-like	AvVit-F	AAGTGAGAGCCAAGGCCATC
	AvVit-R	CGATCTGCTTGGCAATCACG
Actin-like	AvActin-F	CCGGTGCTCATGAGTGATGT
	AvActin-R	CTGCAATGCCAGTTGGTGTC
Calbind-like	AvCalbind-F	TGGCCTCCTCAGCTTCAATG
	AvCalbind-R	GTTCTTCCCTGTGGCGATCA
GST-Omega-like	AvGSTOmega-F	TGCCCCTATGCTCAGAGAGT
	AvGSTOmega-R	GGCTTGGGAAGACGTCATCA
Unk2-like	AvUnk2-F	TCGTCATGTCTGCCCTAAGC
	AvUnk2-R	CGCGTGATGATGAGAGACGA
GST-Mu-like	AvGSTMu-F	AACGCCATCTATCGCCACAT
	AvGSTMu-R	GAGAGTGACCTTGTCGCCAA

Table S3. Read, assembly, annotation and BUSCO metrics for the *A. virens* transcriptome.

Reads/assembly/annotation metrics	
Total number of reads	388.1 million
Total number of contigs	233333
Smallest contig (bp)	201
Largest contig (bp)	26608
Average contig (bp)	829
N50 for contigs >200 bp	1668
N50 for contigs >500 bp	2375
Total bases (bp)	193552380
GC content (%)	40.2
Swiss Prot BLASTX hits, E-value of $\leq 1 \times 10^{-5}$ (%)	29.9
BUSCO completeness test (954 genes in v3 odb10 Metazoa set)	
Complete BUSCOs (%)	98.6
Fragmented BUSCOs (%)	0.6
Missing BUSCOs (%)	0.8

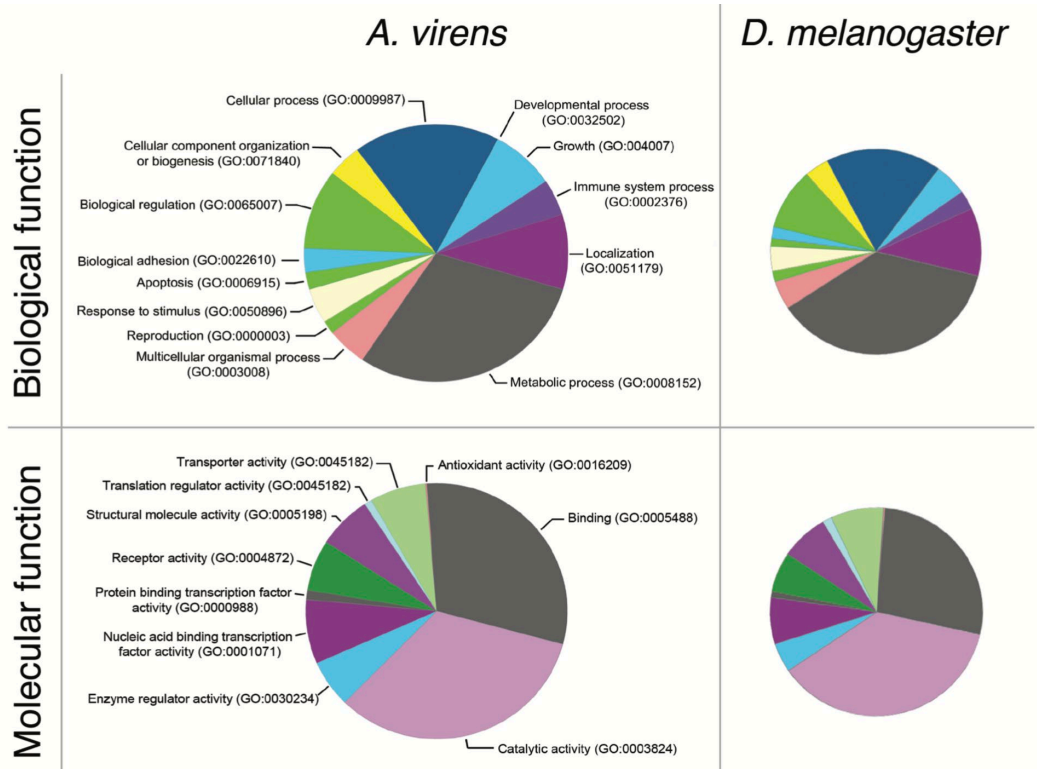


Figure S1. Panther (v. 13.1) characterisation of biological and molecular functions associated with both the *Alitta virens* transcriptome and the *Drosophila melanogaster* genome. Overall, the *A. virens* transcriptome presents equivalent levels of molecular and biological functional representation as the *D. melanogaster* genome.

Table S4. Annotation/RNASeq expression of five candidate reference genes selected for ΔCq qPCR analysis. Annotation produced using top hit following BLASTX search against UniProtKB/Swiss-Prot database.

Contig name	Putative name	Annotation	E-value	Uniprot accession	RNASeq control TMM	RNASeq exposed TMM
c74009_g1_i1	*GAPDH-like*	GAPDH	0	A3FKF7	984	771
c81218_g2_i3	*beta-Actin-like*	Actin-1	0	P0DM41	2.91	1.27
c76985_g1_i2	*Elong-like*	Elongation factor 1-alpha	0	P10126	0.5	0.54
c64848_g1_i1	*Ubiquitin 60S-like*	Ubiquitin-60S ribosomal protein	5.10E-87	P18101	692	752
c84430_g1_i1	*alpha-Tubulin-like*	Tubulin alpha-3 chain	0	P05214	2196	1929

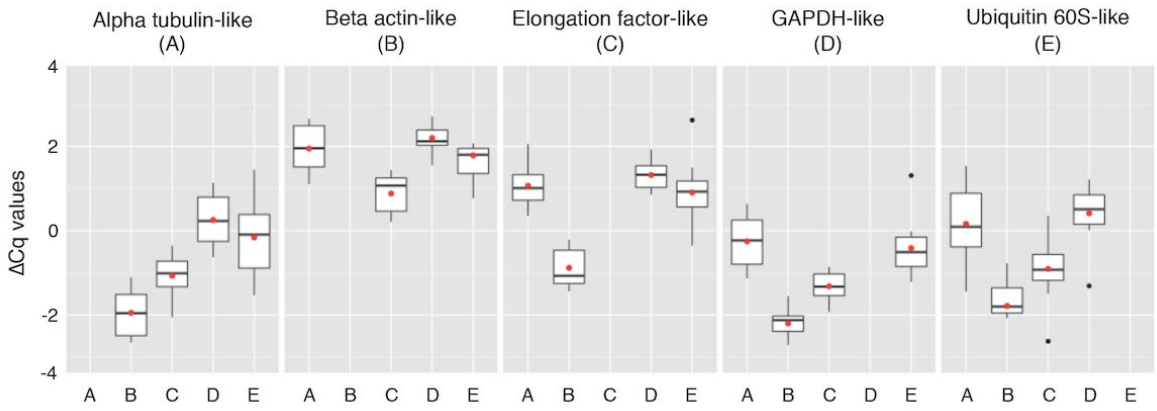


Figure S2. Reference gene selection by ΔCq approach (Silver et al., 2006). The ΔCq variability observed in the 20 gene comparisons across five candidate reference genes. A mean ΔCq value and standard deviation (std. dv.) was calculated for each gene in twelve independent cDNA samples (representing a range of control and exposed animals). A mean std. dv. was calculated to reflect the expression stability of all genes relative to the other four candidates. Comparisons presented as mean (red dots), median (lines), 25th percentile to 75th percentile (boxes) and ranges (whiskers). Black dots represent outliers. Each gene comparison is based on expression across twelve samples (representing a range of control and exposed animals). Smaller boxes and whisker range reveal smaller expression variability, therefore, greater stability for the given candidate reference gene across the comparison range.

Table S5. Reference gene selection by ΔCq approach (Silver et al., 2006). The ΔCq variability observed in the 20 gene comparisons across five candidate reference genes. Lower mean standard deviations for any given gene represents greater expression stability relative to all other genes. Each gene comparison is based on expression across twelve samples (representing a range of control and exposed animals). Genes with lower mean standard deviation values have greater expression stability.

Reference gene	Comparison gene	Mean ΔCq	Standard deviation	Mean standard deviation
alpha-tubulin-like	*beta-actin*	-1.9469	0.5893	0.6442
	Elongation	-1.0661	0.5111	
	Ubiquitin	-0.1601	0.8854	
	GAPDH	0.2533	0.5912	
beta-actin-like	*alpha-tubulin*	1.9470	0.5893	0.5448
	Elongation	0.8808	0.4557	
	Ubiquitin	2.2002	0.3171	
	GAPDH	1.7868	0.8170	
Elong-like	*alpha-tubulin*	1.0661	0.5110	0.5117
	beta-actin	-0.8808	0.4557	
	Ubiquitin	1.3194	0.3353	
	GAPDH	0.9060	0.7447	
GAPDH-like	*alpha-tubulin*	-0.2533	0.5912	0.4755

	beta-actin	-2.2002	0.3171	
	Elongation	-1.3194	0.3353	
	Ubiquitin	-0.4134	0.6585	
Ubiquitin-like	*alpha-tubulin*	0.1601	0.8854	0.7764
	beta-actin	-1.7868	0.8170	
	Elongation	-0.9060	0.7447	
	GAPDH	0.4134	0.6585	

Table S6. Annotation/RNASeq expression of seven genes with apparent differential expression selected for qPCR analysis. Annotation produced using top hit following BLASTX search against UniProtKB/Swiss-Prot database. *As no reads were present in the control sample, this represents a minimum putative fold-change (estimated using pseudo-counts). Genes shaded green and blue represent successful and failed validations respectively, following qPCR analysis.

Contig name	Putative name	Annotation	E-value	Uniprot accession	RNASeq control TMM	RNASeq exposed TMM	Fold change	Log_{10} fold change	Associated P-value
c121332_g1_i1	*Unk1*	None	n/a	n/a	117.824	0.02	5891	-3.77	3.52e-25
c65732_g3_i1	*Vit-like*	Vitellogenin	2.54e-28	Q90243	18.764	0.399	47	-1.67	1.48e-11
c68928_g1_i1	*Actin-like*	Actin	1.54e-81	Q2U7A3	0.01	106.438	10643	4.03	7.79e-30
c84530_g1_i1	*Calbind1-like*	CALCOCO1	2.29e-07	O18737	0	89.895	8990*	4.35	2.72e-32
c59077_g1_i1	*GST-Omega-like*	Glutathione S-transferase Omega-2	2.32e-41	Q6AXV9	0	10.264	1026*	2.71	5.36e-16
c77290_g4_i2	*Unk2*	None	n/a	n/a	0.029	148.999	5138	3.71	8.16e-32
c82836_g1_i1	*GST-Mu-like*	Glutathione S-transferase Mu 5	2.15e-61	P48774	0.068	1.977	29	1.46	3.67e-09

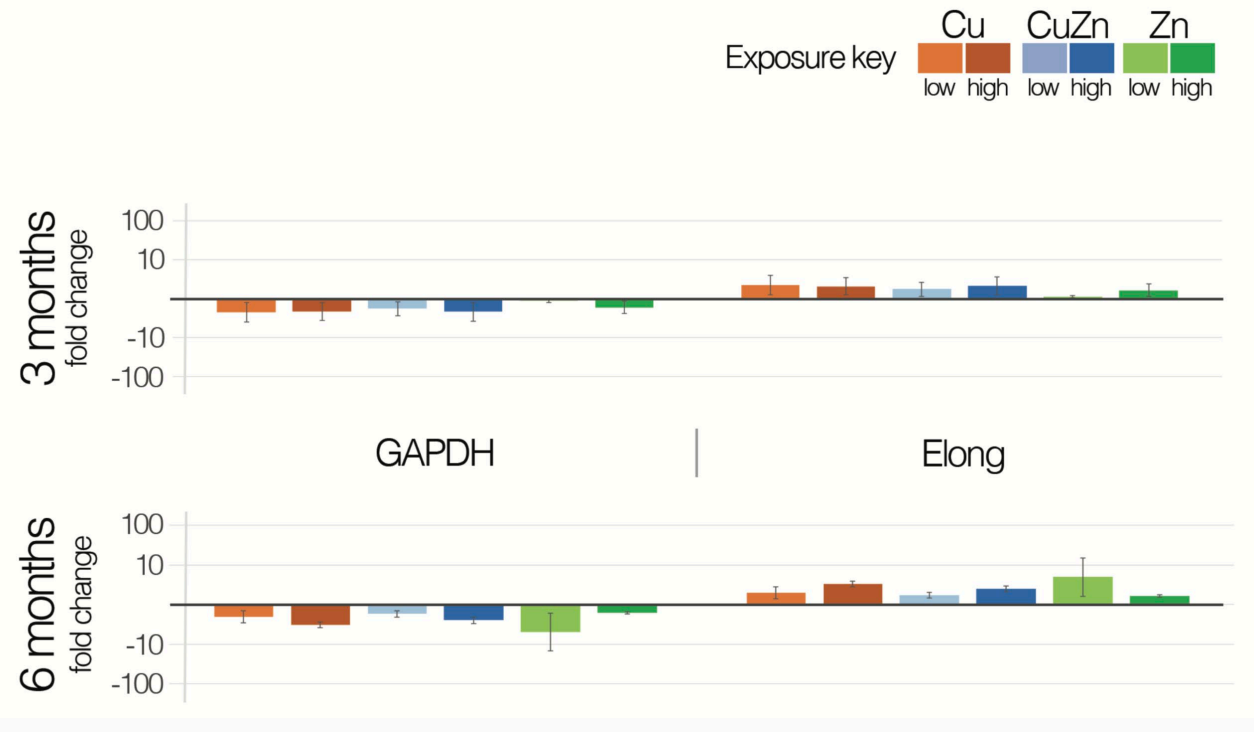


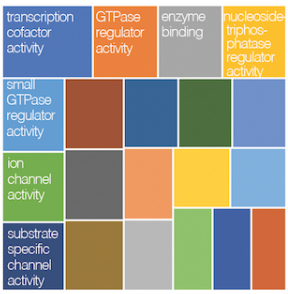
Figure S3. Comparative ΔΔCq qPCR expression analysis of the two selected *A. virens* reference genes (*GAPDH-like* and *Elong-like*). Each column represents the average expression (3 technical repeats) of each sample relative to controls, normalised to the other reference gene (*GAPDH-like* or *Elong-like*). Samples were performed in triplicate and error bars represent the standard deviation. Expression was determined at two concentrations (low and high) of Cu, Zn and Cu-Zn for three and six months (see Table S4 for details on gene annotations and RNASeq expression).

Table S7. The 'traditional' metal-responsive genes either present inconsistent expression across polychaetes or are not induced by metal exposure (1- McQuillan et al., 2014, 2- Rhee et al., 2012, 3- Rhee et al., 2011, 4- Breton et al., 2019, 5- Neave et al., 2012, 6- Rhee et al., 2007b 7- Rhee et al., 2007a, 8- Won et al., 2011). *P. nuntia* - *Perinereis nuntia*, *H. diversicolor* - *Hediste diversicolor, A. virens* - *Alitta virens.*

Classical 'metal-responsive' genes	Function	Notes about gene expression in metal-exposed *A. virens* and other polychaetes
Superoxide dismutase (SOD)	Destroys radicals which are toxic to biological systems.	No expression of '*SOD1*' or '*SOD2*' in metal-exposed *H. diversicolor*[1]. A 'CuZnSOD' and 'MnSOD' are upregulated in metal exposed *P. nuntia*[2] and Cu exposed *N. succinea*[3]. Multiple *SOD* genes present in *A. virens*. An *A. virens* orthologue to the *P. nuntia CuZnSOD* gene (represented by contig c46622_g1_i1) is upregulated (~12x) in Cu-Zn exposed animals but falls outside qPCR validated range.
Metallothioneins (Mts)	Important for Cu and Zn homeostasis	Upregulation in response to Cu/Zn in polychaetes is inconsistent[1,4] and no upregulation seen in *A. virens*.
Atox1	Binds excess intracellular Cu and transports to secretory pathway	Inconsistent across polychaetes [1,5] and not upregulated in *A. virens*. See main text and Table 2 for details.
Phytochelatin Synthase (PCS)	Synthases Phytochelatin, important for heavy metal detoxification	*PCS* genes are present in *A.virens* but not upregulated.
Copper-transporting ATPase 1 (ATP7A)	Cu-transporting P-type ATPase.	Upregulation in response to Cu/Zn in polychaetes is inconsistent[1,4] and no upregulation seen in *A. virens*.
High affinity copper uptake protein 1 (CTR1)	Required for high affinity copper transport into the cell.	Upregulation in response to Cu/Zn in polychaetes is inconsistent[1,4]. Non-significant upregulation (~2x) observed in *A. virens*.
Copper chaperone for superoxide dismutase (CCS)	A Cu chaperone protein	No significant change in metal-exposed *H. diversicolor*[1] and no change seen in *A. virens*
Glutathione S-transferase-Mu (GSTM)	Catalyze conjugation of reduced form of glutathione to xenobiotic substrates	Inconsistent upregulation in polychaetes [1,2,4] and no upregulation in *A. virens*.
Glutathione S-transferase-Theta (GSTT)	Catalyze conjugation of reduced form of glutathione to xenobiotic substrates	Upregulation in response to Cu/Zn in polychaetes is inconsistent[1,4,6] and no upregulation seen in *A. virens*
Glutathione S-transferase-Omega (GSTO)	Catalyze conjugation of reduced form of glutathione to xenobiotic substrates	*GST-Omega-like* genes upregulated in metal-exposed polychaetes[1,2,7,8], but somewhat inconsistent[1,2,4] and specific orthologue induced seems to vary between species. See main text and Table 2 for details.
Catalase (CAT)	Catalyzes decomposition of hydrogen peroxide to water and oxygen	No significant change in metal-exposed *H. diversicolor*[1] and no change seen in *A. virens*
Glutathione peroxidase (GPX)	Catalyzes reduction of hydrogen peroxide to water and oxygen reduction of peroxide radicals to alcohols and oxygen.	A gene termed '*GPX1*'is upregulated in metal exposed *P. nuntia*[2] but it is not upregulated in metal-exposed *H. diversicolor*[1]. Although the closest *A. virens* orthologues to the *P. nuntia GPX1* gene is not upregulated, an *A. virens GPX-like* gene (represented by the c67989_g1 contigs) is upregulated. As for the *GST-Omega* genes, different species appear to upregulate different GPX genes in response to metal exposure.
Glutamate cysteine ligase (GCL)	1st enzyme of the cellular glutathione biosynthesis pathway	No significant change in metal-exposed *H. diversicolor*[1] and no change seen in *A. virens*
Glutathione synthetase (GSS)	2nd enzyme in the glutathione biosynthesis pathway	No significant change in metal-exposed *H. diversicolor*[1] and no change seen in *A. virens*

Figure S4. Enriched (A) 'Molecular Function', (B) 'Cell Component' and (C) 'Biological Process' GO terms associated with putatively upregulated genes determined using DAVID (v. 6.7).

A



- transcription cofactor activity
- GTPase regulator activity
- enzyme binding
- nucleoside-triphosphatase regulator activity
- small GTPase regulator activity
- ion channel activity
- substrate specific channel activity
- channel activity
- passive transmembrane transport activity
- transcription cofactor activity
- protein homodimerisation activity
- cation channel activity
- actin binding
- metal ion transmembrane transporter activity
- transcription factor binding
- gated channel binding
- identical protein binding
- transcription activator activity
- cytoskeletal protein binding
- magnesium ion binding

B

- cell-substrate junction
- focal adhesion
- cell-substrate adherens junction
- adherens junction
- anchoring junction
- basolateral plasma membrane
- cell junction
- postsynaptic membrane
- synapse part
- microsome
- vesicular fraction
- integral to plasma membrane
- intrinsic to plasma membrane
- synapse
- plasma membrane part
- membrane fraction
- membrane-bound vesicle
- insoluble fraction

C

- maintenance of protein location in cell
- maintenance of location in cell
- maintenance of protein location
- compound eye development
- maintenance of location
- innate immune response
- eye development
- defense response
- induction of programmed cell death
- regulation of small GTPase mediated signal transduction
- di-, tri-valent inorganic cation transport
- transmembrane receptor protein tyrosine kinase signaling pathway
- immune response
- transmission of nerve impulse
- neurological system process
- sensory organ development
- synaptic transmission
- cellular protein localization
- cellular macromolecule localization
- positive regulation of programmed cell death
- positive regulation of cell death
- cell adhesion
- biological adhesion
- protein kinase cascade
- cell death
- enzyme linked receptor protein signaling pathway
- death
- apoptosis
- intracellular protein transport
- programmed cell death
- cytoskeleton organization
- cell surface receptor linked signal transduction
- positive regulation of catalytic activity
- metal ion transport
- chromatin modification
- cation transport
- protein complex assembly
- protein complex biogenesis
- cell-cell signaling
- positive regulation of molecular function
- ion transport
- protein localisation

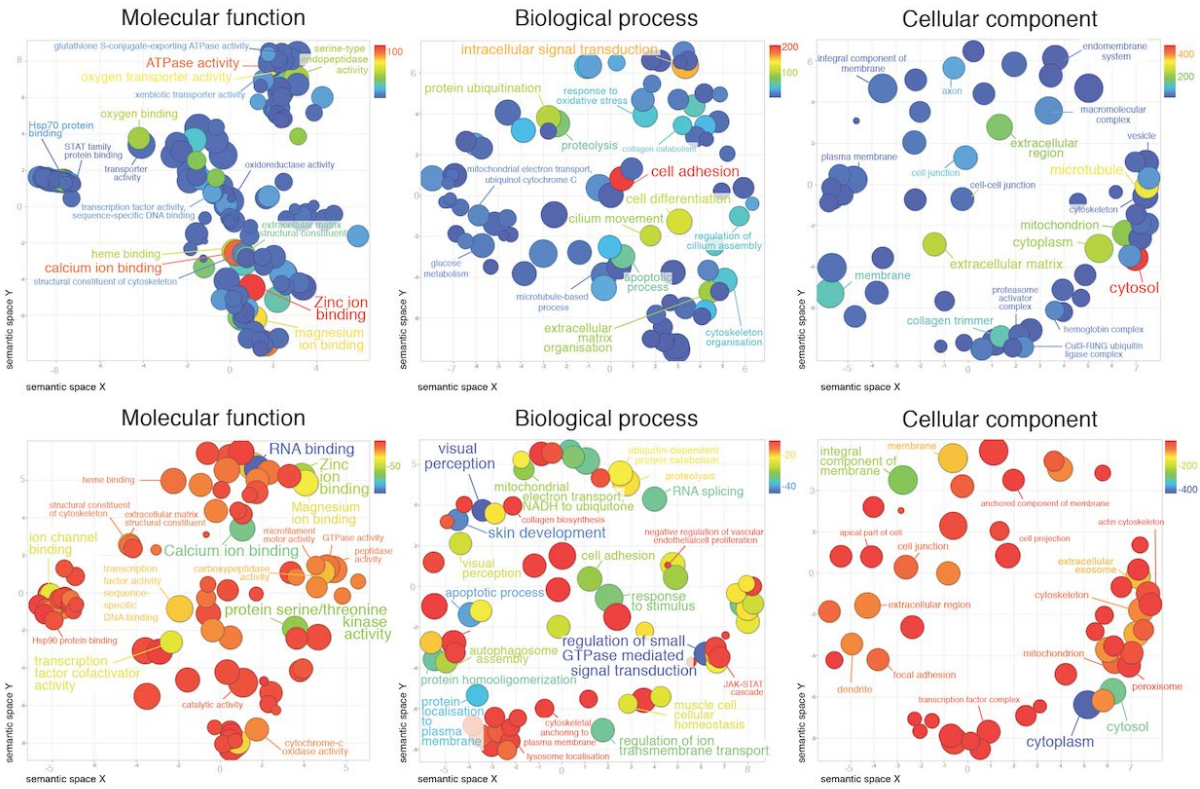


Figure S6. Summarised ReviGO analysis of ‘Molecular Function’, ‘Biological Process’ and ‘Cell Component’ Gene Ontology (GO) terms associated with putatively up (top panel) and downregulated genes (lower panel). GO terms associated with relevant genes were scored by summing fold-change levels linked to contributing contigs. The scored GO term list was then subsequently analysed using ReviGO.

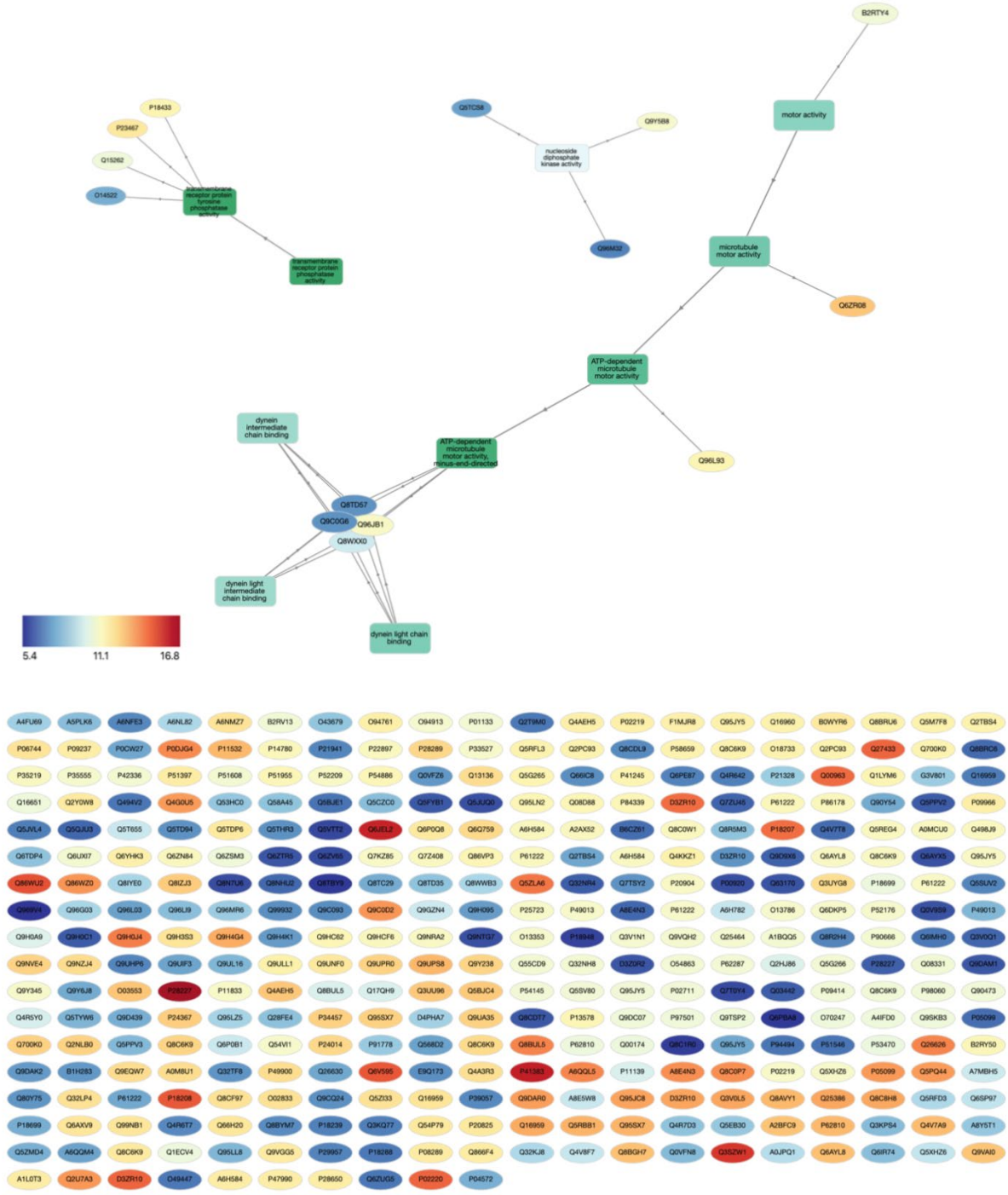


Figure S7. GOnet analysis of 'Molecular Function' Gene Ontology (GO) terms associated with putatively upregulated genes and their various relationships. Coloured UniProt accession numbers represent scored fold-change levels linked to contributing contigs. Scale represents $\log_2$ fold change.

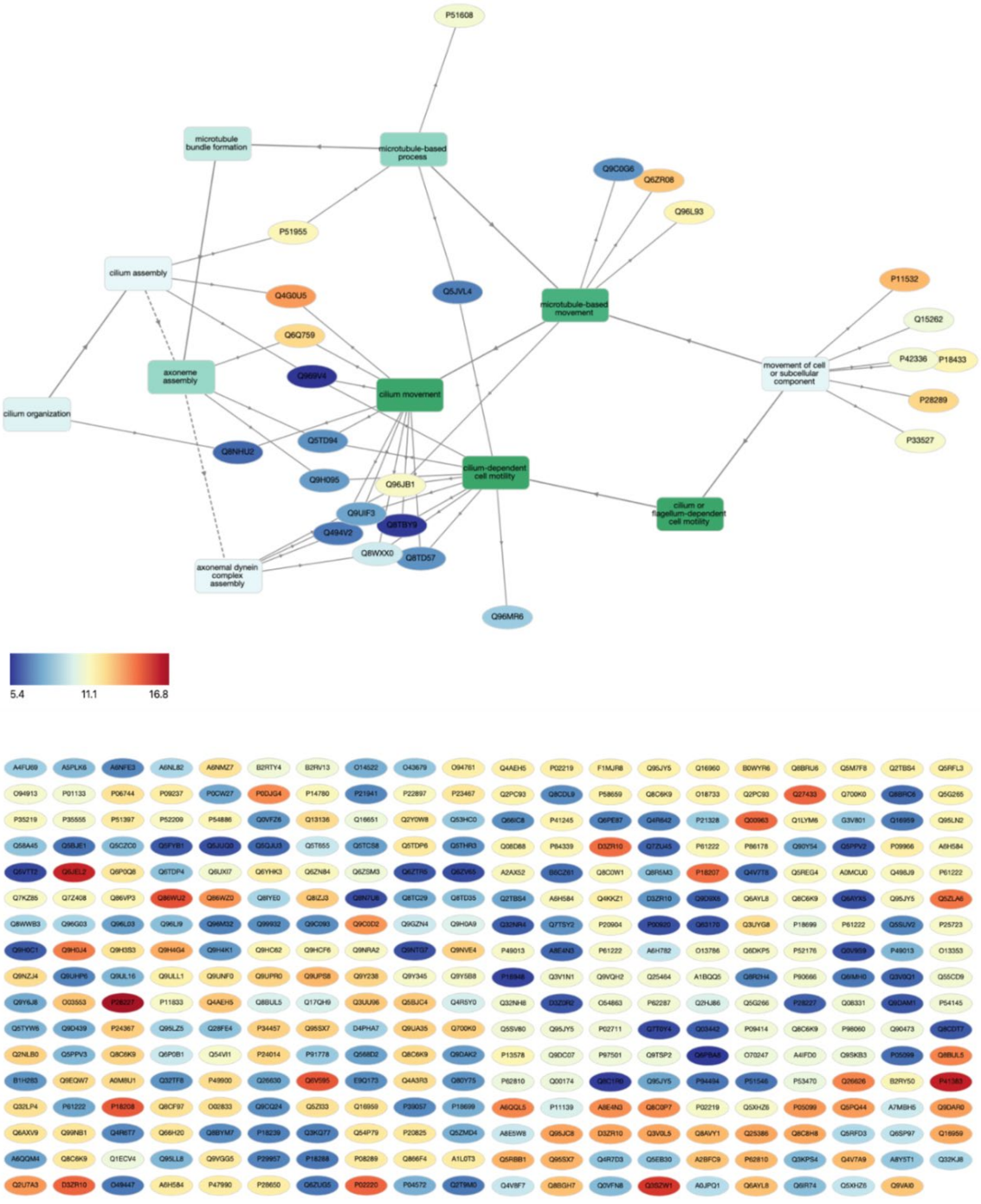


Figure S8. GOnet analysis of 'Biological Process' Gene Ontology (GO) terms associated with putatively upregulated genes and their various relationships. Coloured UniProt accession numbers represent scored fold-change levels linked to contributing contigs. Scale represents $\log_2$ fold change.

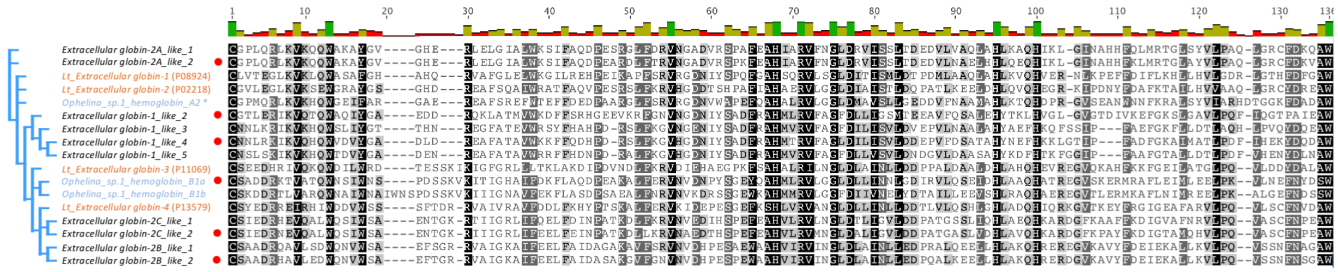


Figure S9. Alignment of haemoglobin subunits from the polychaetes *Alitta virens*- black, *Ophelina* (Neave et al., 2012)- blue and *Lumbricus terrestris* (*Lt*)- orange, with graphical identity scores for each residue. Red dots represent paralogues upregulated following metal exposure.

Table S8. Annotation and RNASeq expression of *A. virens* haemoglobins subunits and linker chains. Annotation produced using top hit following BLASTX search against UniProtKB/Swiss-Prot database. * As no reads were present in the control sample, this represents a minimum putative fold-change (estimated using pseudo-counts). Green shaded colour represents genes presenting upregulation in metal exposed animals, all fall within p-value range validated by the qPCR screen.

Contig name	Putative name	Annotation	E-value	Uniprot accession	RNASeq control TMM	RNASeq exposed TMM	Fold change	Log10 fold change	Associated P-value
c58545_g2_i1	Extracellular globin-2A_like_1	Extracellular globin-2A	1.10E-80	P09966	8.67	13.27	2	0.18	0.3
c104830_g1_i1	Extracellular globin-2A_like_2	Extracellular globin-2A	9.75E-82	P09966	0.02	59.41	2971	3.47	1.53E-23
c104416_g1_i1	Extracellular globin-1_like_1	Extracellular globin-1	8.92E-22	P02219	0.332	0.922	3	0.44	0.7
c68421_g1_i2	Extracellular globin-1_like_2	Extracellular globin-1	2.00E-30	P02219	0	9.065	907*	2.96	6.98E-15
c68421_g1_i3	Extracellular globin-1_like_3	Extracellular globin-1	3.25E-51	P02219	6.97	12.55	2	0.26	0.2
c55755_g1_i1	Extracellular globin-1_like_4	Extracellular globin-1	3.53E-43	P02219	0.03	62.22	2074	3.32	2.44E-21
c68421_g1_i1	Extracellular globin-1_like_5	Extracellular globin-1	5.95E-48	P02219	4.55	4.92	1	0.03	0.8
c52952_g1_i1	Extracellular globin-2C_like_1	Extracellular globin-2C	4.55E-68	P02220	10.24	15.21	1	0.17	0.3
c154766_g1_i1	Extracellular globin-2C_like_2	Extracellular globin-2C	4.97E-72	P02220	0	52.93	5293*	3.72	4.80E-24
c56032_g1_i1	Extracellular globin-2B_like_1	Extracellular globin-2B	2.80E-86	P13578	10.87	17.13	2	0.20	0.25
c56032_g2_i1	Extracellular globin-2B_like_2	Extracellular globin-2B	3.37E-83	P13578	0.02	57.57	2879	3.46	1.65E-22
c120169_g1_i1	Giant extracellular hemoglobin linker 1 chain_like_b	Giant extracellular hemoglobin linker 1 chain	3.42E-93	P18207	0	33.35	3335*	3.52	1.65E-23
c63201_g2_i1	Giant extracellular hemoglobin linker 1 chain_like_a	Giant extracellular hemoglobin linker 1 chain	3.05E-97	P18207	5.99	10.56	2	0.25	0.2
c36018_g1_i1	Giant extracellular hemoglobin linker 2 chain_like_a	Giant extracellular hemoglobin linker 2 chain	1.14E-129	P18208	0	44.63	4463*	3.65	9.36E-25
c62841_g1_i1	Giant extracellular hemoglobin linker 2 chain_like_b	Giant extracellular hemoglobin linker 2 chain	7.28E-69	P18208	11.08	15.78	1	0.15	0.4