Weighted Gene Co-expression Network Analysis (WGCNA) methods, results, and discussion associated with Manuscript "*Transcriptomic changes associated with maternal care in the brain of mouthbrooding cichlid Astatotilapia burtoni reflect adaptation to self-induced metabolic stress*"

Methods

Gene co-expression networks were built using WGCNA v1.69 in R-Studio. VST count tables, exported using DESeq2, were used as WGCNA input. From the initial 25572 expressed genes, 469 were removed due to low expression for a total of 25103 input gene counts. Sample clustering did not reveal outliers, so all brain transcriptomes were used in subsequent analysis.

To build a signed network, or a network that only accounts for positive correlations between genes, scale-free analysis was conducted to determine an optimal soft-thresholding power. A power of 12 was chosen because it was the lowest power with a Scale Free Topology Model Fit above 0.90 and a Mean Connectivity below 100. The deep split parameter was set to 0 and all other parameters were left at default values. A signed network was preferred over an unsigned network for simpler interpretation of module eigengene values used in downstream statistics.

Following network construction, module-trait relationships were examined using specific time-points and a binary brooding vs release state for traits. In addition, module eigengenes were used in follow-up LRT analysis to determine which modules showed significant variation between any pair of timepoints with the lmtest R package v0.9-38, and post-hoc pairwise comparisons between timepoints were conducted using the contrasts function from the emmeans R package v1.5.3. GO enrichment analysis was performed using BiNGO, as described in the main article methods section, for each network module to determine their potential functions.

Results

The signed co-expression network built with WGCNA contained 49 modules (Fig S1), with only 7 genes falling in the gray (ME0) module, indicating they fail to group with other genes. A module-trait relationship heatmap highlights 8 modules with expression that significantly correlates with a specific time-point or general brooding state (Fig S2). Modules 20, 43, and 47 correlate with B02, modules 9, 10, 29, and 36 correlate with R02, and module 34 correlates with the binary release vs. brooding maternal states.

LRT results are more stringent than Fisher's asymptotic p-values listed in the correlation heatmap. No modules showed significant differences between timepoints at a significance threshold of Pr(>Chisq) < 0.05. Three modules (9, 34, and 47), however, approach significance with Pr(>Chisq) < 0.1, all of which were identified as correlated with specific time-points in module-trait relationships (Figs S3, S4, S5).

Pairwise time-point contrasts for ME9 reveal that R02 samples have higher expression than B02 at $P < 0.1$ (Fig S3) and significantly higher expression than B14 and R14 at $P < 0.05$. This module is enriched for GO terms including 'neuropeptide hormone activity' and 'feeding behavior' with

an FDR < 0.25. These enriched terms include a number of DE and DE-trending genes, such as *pituitary adenylate cyclase-activating polypeptide*, *pro-opiomelanocortin B-like*, *cholecystokinin-like*, and more. ME34 is significantly higher in R02 than B14 and approaches significance in R02 relative to B02. In addition, R14 approaches significance in contrast with lower expression B14 (Fig S4). Several top enriched terms for this module are 'hormone activity' and 'response to hormone stimulus', which contain notable DE and DE-trending genes including *androgen receptor*, *isotocin-neurophysin IT1*, *gonadotropin subunit beta-1-like*, *glycoprotein hormone beta-5-like*, *glycoprotein hormones alpha polypeptide*, and *prolactin receptor-like*. Pairwise contrasts for ME47 show that expression in B02 is significantly higher than in R02 and approaches significance (P < 0.1) in comparison with B14 and R14 (Fig S5). This module does not yield any enriched GO terms, but the DE and DE-trending genes within this module are slightly enriched for terms like 'anatomical structure development' and 'developmental process', which both contain 9/11 genes in this subset.

Discussion

Because several significant modules are enriched with behavior and hormone related terms, this might suggest some candidates that regulate feeding behavior. However, the direction of expression changes and what is known about their orexigenic or anorexogenic properties of the DE genes they contain instead suggest that these modules are responding to diet after the resumption of feeding post-release of fry. In ME34, for example, 9/13 genes annotated with the enriched GO term 'hormone activity' are also annotated under enriched terms such as 'response to endogenous stimulus' and/or 'response to carbohydrate'. Behavioral candidates like neurotensin, galanin, and other co-expressed genes discussed in the main MS text fall within ME 17. Overall, the expression pattern of this module shows steadily decreasing expression from B02 to R14 even though no time-points significantly differ from each other.

Modules enriched with GO terms related to oxygen transportation, respiration, and ATP production show expected eigengene expression patterns related to member DE globin genes, but the modules themselves show no significant differences between time-points via pairwise contrasts. DE globin genes are divided between modules 2, 4, and 5, genes that respond to starvation, such as *pdk2*, *pdk4* are found in modules 1 and 6, respectively, and *hypoxia-inducible factor 1-alpha-like* is found in module 2.

The observation that DE and DE trending genes are split across a wide variety of modules, most of which show no significant differences across time points using eigengenes to represent expression, suggests that major behavioral and metabolic changes experienced in maternal *A. burtoni* likely result from large expression changes in a small number of genes or more minor expression changes in larger gene networks, or a combination of such trends. Alternatively, unsigned networks that include genes with positive and negative correlations of expression may better define the complex interactive gene networks in the brain, although downstream statistics to examine an unsigned module's relevance across maternal time-points would not directly correlate to associated parental behaviors or physiological changes.

Tables

Table 1. LRT and post-hoc pairwise contrasts between time-points for WGCNA module eigengenes that trend towards differential expression via LRT with Pr(>Chisq) < 0.1.

	LRT results		Pairwise contrast results	B02 - B14	B02 - R02	B02 - R14	B14 - R02	B14 - R14	R02 - R14
ME9	*LogLik*	6.449	*estimate*	0.0701	-0.2203	0.0534	-0.2904	-0.0167	0.2737
	df	3	*df*	18	18	18	18	18	18
	Chisq	7.328	*t.ratio*	0.58	-1.746	0.442	-2.403	-0.145	2.265
	Pr(>Chisq)	0.062	*p.value*	0.5691	0.0979	0.6639	0.0273	0.8864	0.0361
ME34	*LogLik*	6.370	*estimate*	0.0359	-0.2228	-0.1852	-0.2587	-0.2211	0.0376
	df	3	*df*	18	18	18	18	18	18
	Chisq	7.170	*t.ratio*	0.296	-1.759	-1.528	-2.133	-1.912	0.31
	Pr(>Chisq)	0.067	*p.value*	0.7708	0.0955	0.144	0.0469	0.0719	0.7602
ME47	*LogLik*	6.480	*estimate*	0.2257	0.32	0.2395	0.0943	0.0138	-0.0805
	df	3	*df*	18	18	18	18	18	18
	Chisq	7.389	*t.ratio*	1.87	2.539	1.985	0.782	0.12	-0.667
	Pr(>Chisq)	0.060	*p.value*	0.0778	0.0206	0.0626	0.4446	0.9059	0.513

Table 2. GO enrichment results summary for three WGCNA modules that trend towards differential expression via LRT with Pr(>Chisq) < 0.1. These results include the top ten significantly enriched GO terms for the full sets of genes within each module, and subsets of genes that trend towards differential expression via DESeq2 results with an unadjusted p < 0.05. The total number of genes passing WGCNA filters and assigned to modules was 25,103, although only 22,924 genes had associated GO information.

Module	Gene set	GO-ID	p-value	corr p-value	annot genes in module	total genes in module	annot genes in bg set	total genes in bg set	Description
ME9	*All genes*	19226	1.73E-06	6.02E-03	76	647	1549	22924	transmission of nerve impulse
		21879	2.54E-06	6.02E-03	11	647	67	22924	forebrain neuron differentiation
		1829	2.87E-06	6.02E-03	8	647	33	22924	trophectodermal cell differentiation
		7154	8.44E-06	1.33E-02	101	647	2333	22924	cell communication
		7268	1.55E-05	1.96E-02	64	647	1315	22924	synaptic transmission
		5184	1.95E-05	2.05E-02	8	647	42	22924	neuropeptide hormone activity
		7631	3.09E-05	2.78E-02	18	647	210	22924	feeding behavior
		7267	5.10E-05	3.95E-02	78	647	1764	22924	cell-cell signaling
		21953	5.64E-05	3.95E-02	26	647	392	22924	central nervous system neuron differentiation
		51591	7.48E-05	4.71E-02	14	647	147	22924	response to cAMP

	DE trending	48176	1.22E-05	7.06E-03	2	78	2	22194	regulation of hepatocyte growth factor biosynthetic process
		48178	1.22E-05	7.06E-03	2	78	2	22194	negative regulation of hepatocyte growth factor biosynthetic process
		32646	1.22E-05	7.06E-03	2	78	2	22194	regulation of hepatocyte growth factor production
		51799	3.65E-05	1.58E-02	2	78	3	22194	negative regulation of hair follicle development
		46879	6.81E-05	1.59E-02	5	78	121	22194	hormone secretion
		1553	7.03E-05	1.59E-02	3	78	23	22194	luteinization
		22602	7.65E-05	1.59E-02	5	78	124	22194	ovulation cycle process
		42698	8.25E-05	1.59E-02	5	78	126	22194	ovulation cycle
		61097	8.25E-05	1.59E-02	5	78	126	22194	regulation of protein tyrosine kinase activity
		21879	9.15E-05	1.59E-02	4	78	67	22194	forebrain neuron differentiation
ME34	*All genes*	5179	4.76E-09	1.70E-05	13	164	213	22924	hormone activity
		5615	8.72E-08	1.56E-04	45	164	2786	22924	extracellular space
		16913	1.43E-06	1.71E-03	3	164	4	22924	follicle-stimulating hormone activity
		5102	2.12E-06	1.90E-03	46	164	3222	22924	receptor binding
		71495	4.07E-06	2.91E-03	22	164	1023	22924	cellular response to endogenous stimulus
		9719	1.47E-05	8.75E-03	25	164	1369	22924	response to endogenous stimulus
		5576	2.70E-05	1.32E-02	66	164	5869	22924	extracellular region
		9755	3.17E-05	1.32E-02	8	164	170	22924	hormone-mediated signaling pathway
		32870	3.53E-05	1.32E-02	17	164	763	22924	cellular response to hormone stimulus
		44421	3.68E-05	1.32E-02	62	164	5438	22924	extracellular region part
	DE trending	34446	4.11E-05	4.03E-02	4	27	161	22194	substrate adhesion-dependent cell spreading
		9719	1.58E-04	4.22E-02	8	27	1369	22194	response to endogenous stimulus
		71495	1.71E-04	4.22E-02	7	27	1023	22194	cellular response to endogenous stimulus
		42221	1.82E-04	4.22E-02	13	27	3775	22194	response to chemical stimulus
		9725	2.54E-04	4.22E-02	7	27	1091	22194	response to hormone stimulus
		32870	2.58E-04	4.22E-02	6	27	763	22194	cellular response to hormone stimulus
		70887	3.84E-04	5.37E-02	10	27	2455	22194	cellular response to chemical stimulus
		7625	5.67E-04	6.72E-02	2	27	29	22194	grooming behavior
		1568	6.23E-04	6.72E-02	7	27	1265	22194	blood vessel development
		1944	7.06E-04	6.72E-02	7	27	1292	22194	vasculature development
ME47	*All genes*	NA	NA	NA	NA	NA	NA	NA	NA
	DE trending	21503	2.48E-03	2.21E-01	1	11	5	22194	neural fold bending
		22	2.97E-03	2.21E-01	1	11	6	22194	mitotic spindle elongation
		51231	3.46E-03	2.21E-01	1	11	7	22194	spindle elongation

		48246	3.96E-03	2.21E-01	1	11	8	22194	macrophage chemotaxis
		3100	4.45E-03	2.21E-01	1	11	9	22194	regulation of systemic arterial blood pressure by endothelin
		90327	5.93E-03	2.21E-01	1	11	12	22194	negative regulation of locomotion involved in locomotory behavior
		90325	6.43E-03	2.21E-01	1	11	13	22194	regulation of locomotion involved in locomotory behavior
		48856	6.51E-03	2.21E-01	9	11	8987	22194	anatomical structure development
		14826	6.92E-03	2.21E-01	1	11	14	22194	vein smooth muscle contraction
		1842	6.92E-03	2.21E-01	1	11	14	22194	neural fold formation

Figures

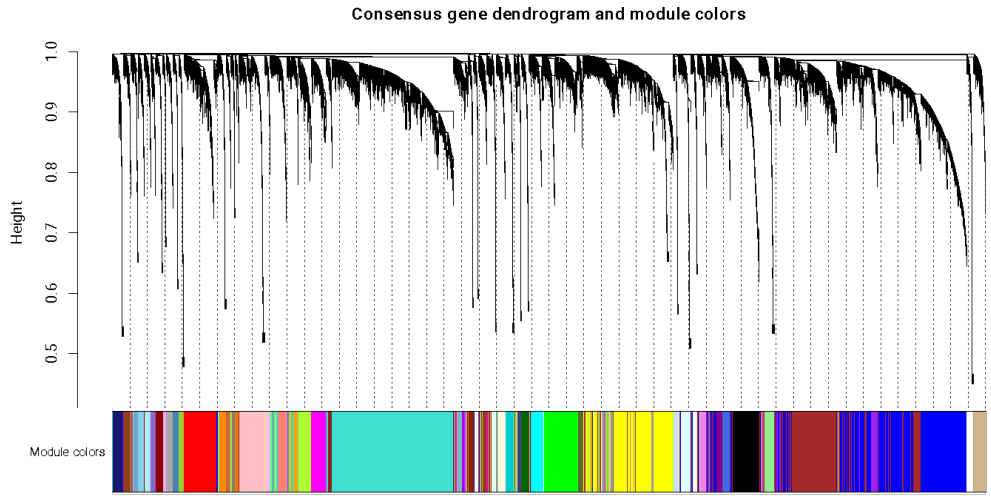

Figure 1. Gene dendrogram produced by WGCNA showing signed co-expression patterns and module membership.

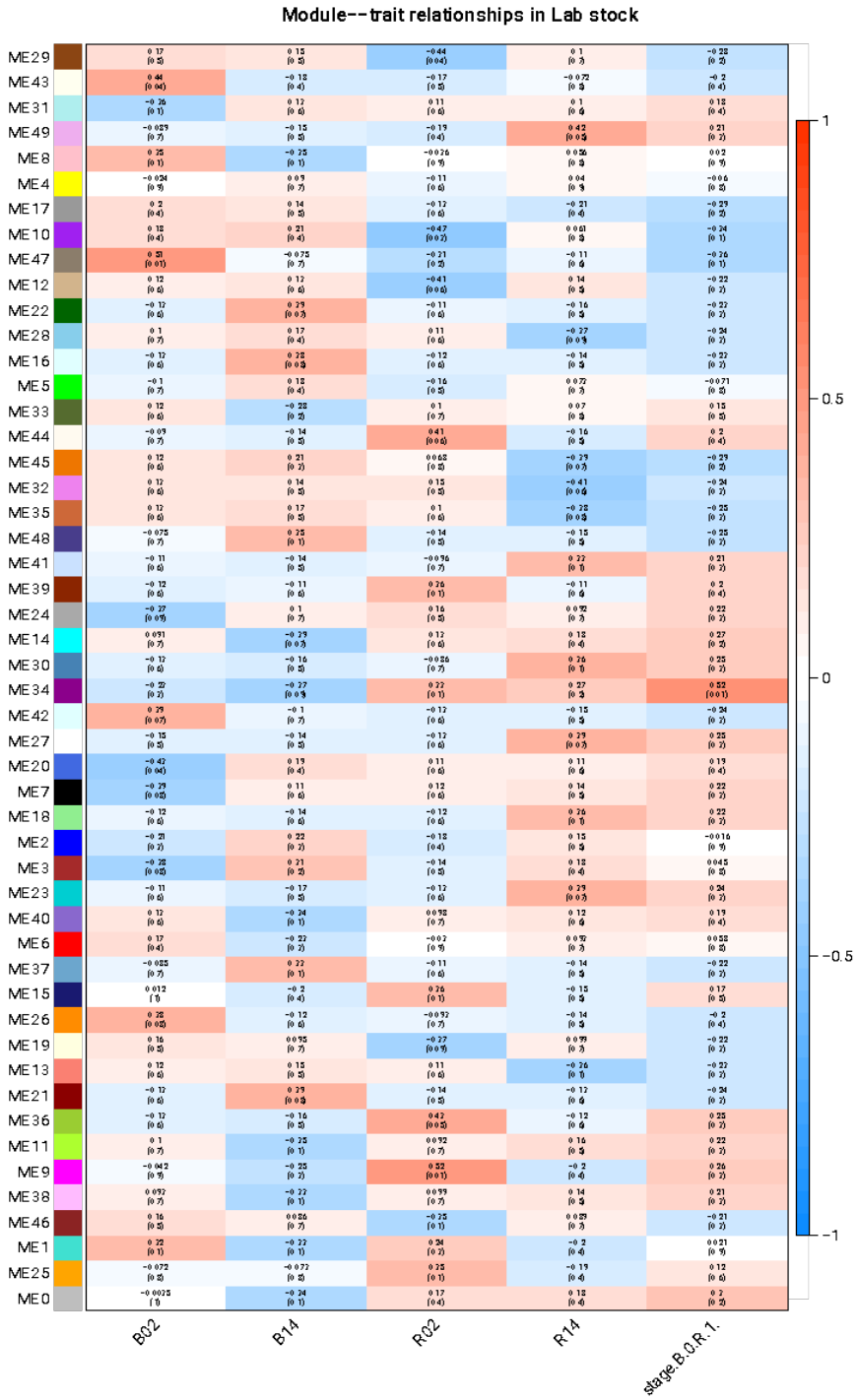

Figure 2. Module-trait relationship heatmap showing correlations between eigengenes and individual time-points or time-points grouped by associated brooding or release behavior. The first number within each cell indicated the direction and strength of correlation and the second number in parentheses are p-values indicating significance of correlation. For the column representing time-points grouped by stage, positive correlations indicate expression is higher in release stages than brooding stages, and vice versa.

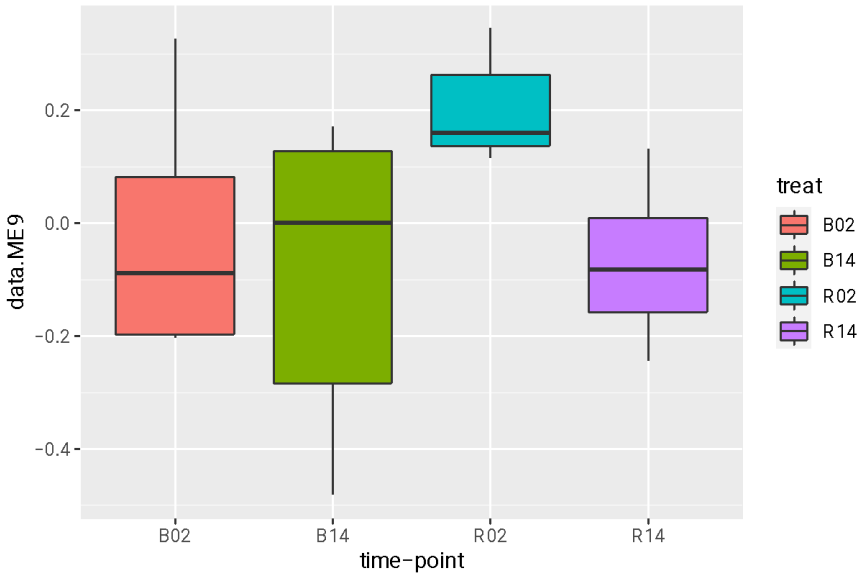

Figure 3. Boxplot of ME9 eigengene values grouped by time-point for all samples.

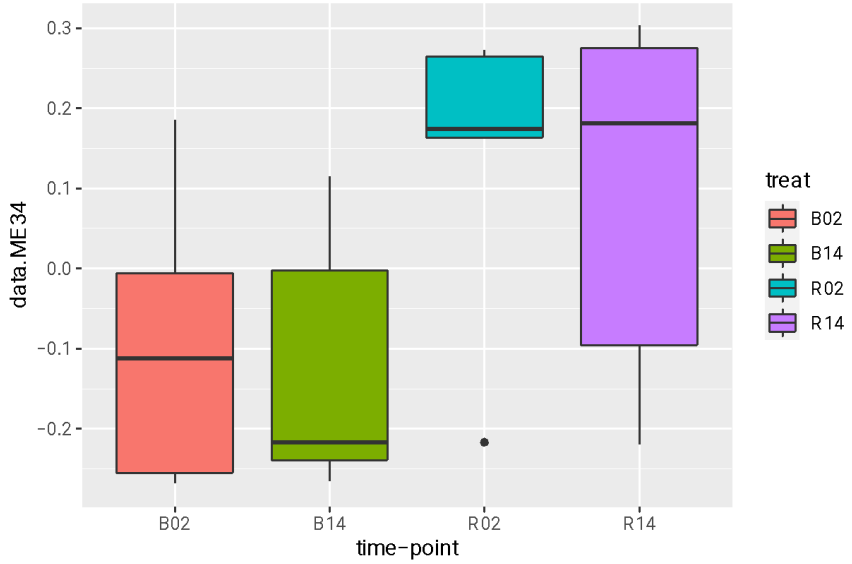

Figure 4. Boxplot of ME34 eigengene values grouped by time-point for all samples.

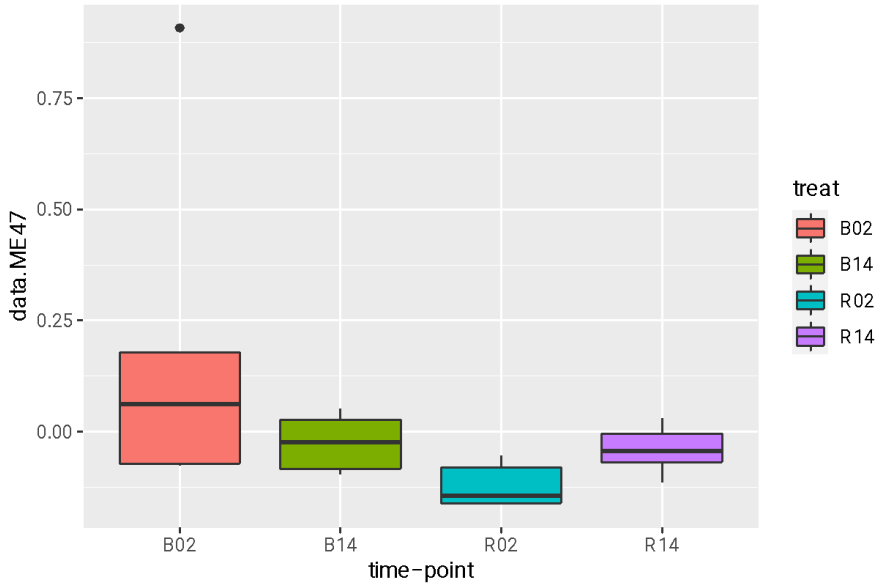

Figure 5. Boxplot of ME47 eigengene values grouped by time-point for all samples.