

Supplementary materials for:

Tasmanian devil (*Sarcophilus harrisi*) populations lack differential gene expression between sex and geographic location in ear tissue

Alexandra K. Fraik¹, Corey Quackenbush¹, Mark J. Margres^{1,2}, Sebastien Comte^{3,4}, David G. Hamilton³, Christopher Kozakiewicz¹, Menna Jones³, Rodrigo Hamede³, Paul A. Hohenlohe⁵, Andrew Storfer^{1,*}, Joanna L. Kelley¹

Affiliations:

¹*School of Biological Sciences, Washington State University Pullman, WA 99164, USA*

²*Department of Biological Sciences, Clemson University, Clemson, SC 29634*

³*School of Natural Sciences, Hobart, TAS 7001, Australia*

⁴*Vertebrate Pest Research Unit, NSW Department of Primary Industries, 1447 Forest Road, Orange, New South Wales 2800, Australia*

⁵*Department of Biological Sciences, University of Idaho, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, 875 Perimeter Drive, Moscow, Idaho 83844,*

	Mean Annual Temperature (°C)	Precipitation Seasonality (mm)	Isothermality	Annual Temperature Range (°C)	Enhanced Vegetation Index (EVI)	Length of Sealed Roads (km)	Elevation (m)	Surface Area of Water (m ²)
Arthur River (ARV)	12.59	34.01	49.73	12.30	52.46	9.96	48.00	280917.49
Freycinet (FRY)	12.67	14.51	54.09	17.24	74.01	61.65	57.17	4814676.77
West Pencil Pine (WPP)	7.98	31.17	47.98	17.11	61.24	245.23	703.74	77396.95

Supplementary Table 1: Abiotic environmental centroid values among populations. The centroid values for the eight abiotic

environmental variables and the biotic variable utilized in the GEA associations for each population sampled. We plotted the location of every sampled Tasmanian devil and calculated the centroid for each sampling area. We then constructed a circular buffer with a 5 km radius around each centroid to represent the sampled area for each study site, which we then used to extract the environmental data. These environmental data were obtained from www.ga.gov.au, www.worldclim.org, and www.thelist.tas.gov.au (see Supplementary Table 1 for a list of the eight abiotic environmental variables we included as qualitative descriptors of the sampling locations). We used ArcGIS version 10 (ESRI 2011) for all spatial analyses.

Sample ID	Population Sex		Number Reads Pre-Trimming	Number Reads Post-Trimming	HiSat2 Mapping Percent	Percent Coding Bases	Percent UTR Bases	Percent Intronic Bases	Percent Intergenic Bases	Percent mRNA Bases
1	ARV	Female	21,087,038	19,833,002	83.7	51.0	8.7	5.9	34.4	59.7
2	ARV	Male	21,508,474	20,618,302	85.6	49.4	9.1	4.9	36.6	58.5
3	ARV	Male	24,409,004	23,803,520	90.0	42.8	9.3	9	38.9	52.1
4	ARV	Female	30,892,554	30,086,866	89.2	46.9	9.6	5.9	37.6	56.5
5	ARV	Male	21,059,740	19,895,594	83.7	47.8	9.2	5.5	37.4	57.1
6	ARV	Female	30,273,938	29,467,814	89.1	46.4	9.6	5.9	38.1	56.0
7	FRY	Female	20,457,152	20,053,534	82.7	47.3	9.5	6.0	37.2	56.8
8	FRY	Male	20,389,108	19,612,834	81.7	47.0	10.0	5.5	37.6	57.0
9	FRY	Female	28,636,486	27,858,674	87.9	49.0	9.7	5.1	36.2	58.7
10	FRY	Female	19,282,898	18,748,094	88.4	47.1	8.8	8.7	35.5	55.9
11	FRY	Male	22,524,552	22,000,826	81.6	47.0	9.1	6.7	37.1	56.2
12	FRY	Male	13,126,976	12,821,570	78.9	47.1	10.0	5.5	37.4	57.2
13	WPP	Female	20,175,034	19,354,872	82.2	50.3	9.0	5.8	34.8	59.4
14	WPP	Male	20,659,934	20,012,910	82.4	48.5	9.3	5.7	36.5	57.8
15	WPP	Female	28,023,276	27,318,236	87.5	47.9	9.7	5.9	36.4	57.7
16	WPP	Male	30,090,240	29,205,678	88.9	46.9	9.5	6.3	37.4	56.3
17	WPP	Female	33,284,436	32,391,218	91.8	43.8	8.9	8.7	38.6	52.7
18	WPP	Female	20,433,336	19,910,918	87.1	47.8	9.8	6.3	36.2	57.6
19	WPP	Male	18,643,388	18,092,414	79.8	41.9	10.4	6.4	41.3	52.3
20	WPP	Male	24,100,656	23,427,902	85.0	47.7	9.2	5.9	37.2	56.9
Ear Sample Averages			23,461,028	22,733,820	85.0	46.8	9.5	6.2	37.4	56.3
Milk Transcriptome	Zoo	Female	2,435,1459,062	22,838,662,422	93.8	66.9	5.8	4.5	22.8	72.7

6 **Supplementary Table 2: Alignment rates from devil ear tissue and the published devil milk transcriptome.** HISAT2 alignment
7 rates and mapping statistics for the transcriptomes produced from the ear tissue as well as the published milk transcriptome
8 downloaded from NCBI (accession #PRJNA510591). The three sampled populations included Arthur River (ARV), West Pencil Pine
9 (WPP) and Freycinet (FRY).

GO Term	Genes	FDR
<i>Biological Processes</i>		
Macromolecule metabolic process (GO:0043170)	6	0.003
Cellular macromolecule metabolic process (GO:0044260)	5	0.034
Oxidation-reduction process (GO:0055114)	13	0.038
Lipid metabolic process (GO:0006629)	13	0.046
<i>Molecular Function</i>		
Binding (GO:0005488)	36	0.000
Protein binding (GO:0005515)	24	0.000
<i>Cellular Component</i>		
Cell part (GO:0044464)	48	0.004
Intracellular (GO:0005622)	40	0.007
Membrane-bounded organelle (GO:0043227)	31	0.008
Organelle part (GO:0044422)	19	0.002
Nucleus (GO:0005634)	13	0.023
Cytosol (GO:0005829)	8	0.018
Organelle lumen (GO:0043233)	8	0.018
Intracellular organelle lumen (GO:0070013)	8	0.016
Membrane-enclosed lumen (GO:0031974)	8	0.015

10

11 **Supplementary Table 3: Gene-ontology enrichment analysis of co-expressed genes associated**
12 **with FRY and ARV.** Gene-ontology enrichment analysis results for co-expressed genes in
13 Module28 which was strongly associated with the FRY and ARV populations.

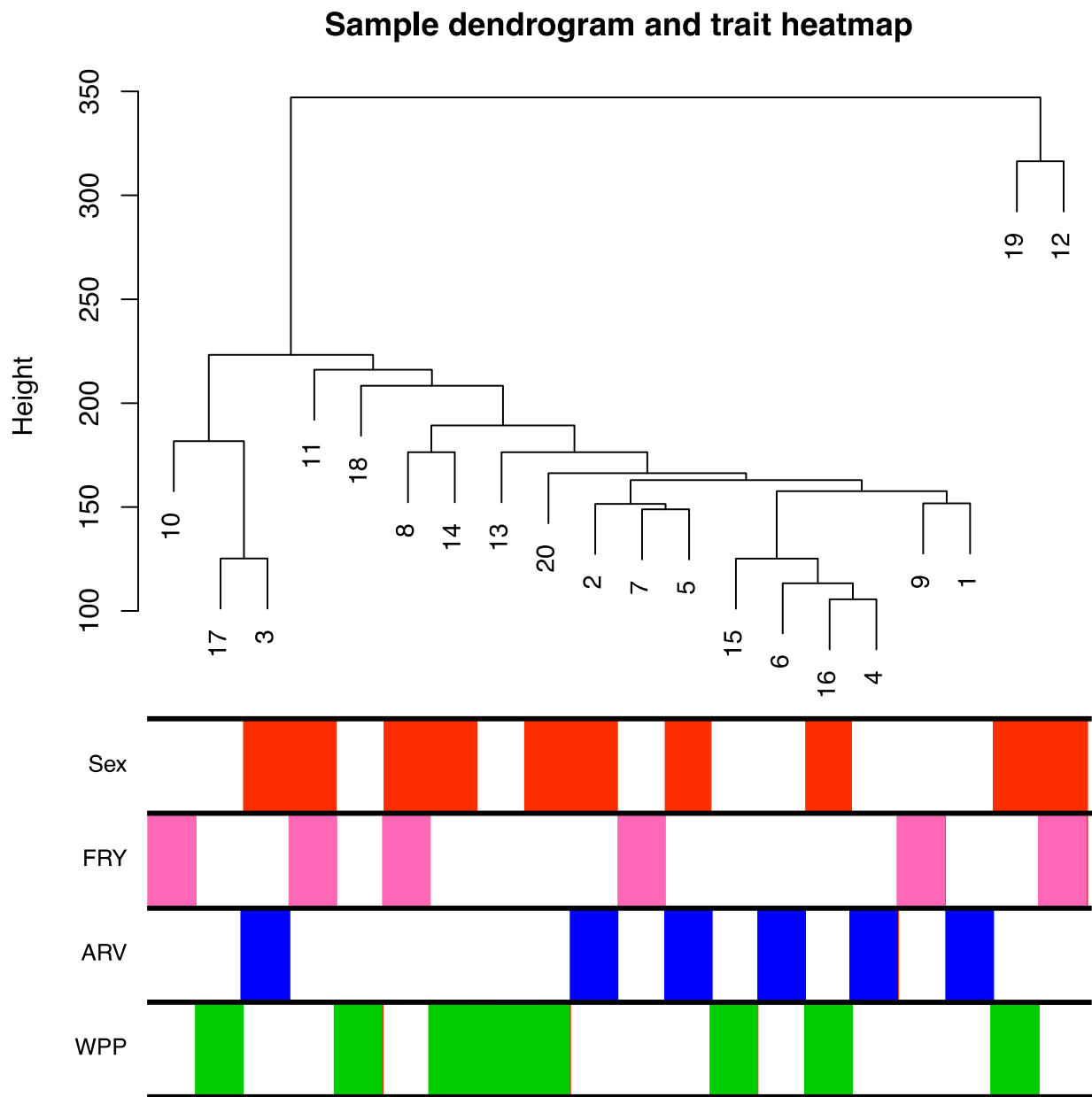
GO Term	Genes	FDR
<i>Molecular Function</i>		
Structural constituent of ribosome (GO:0003735)	7	0.001
Structural molecule activity (GO:0005198)	7	0.043
<i>Cellular Component</i>		
Regulator complex (GO:0071986)	2	0.057
RNA polymerase I complex (GO:0005736)	2	0.049
Ribosome (GO:0005840)	7	0.000
Ribosomal subunit (GO:0044391)	4	0.047
Mitochondrial protein complex (GO:0098798)	5	0.034
Mitochondrial part (GO:0044429)	7	0.045
Ribonucleoprotein complex (GO:1990904)	8	0.034
Mitochondrion (GO:0005739)	13	0.001
Cytoplasmic part (GO:0044444)	28	0.003

14

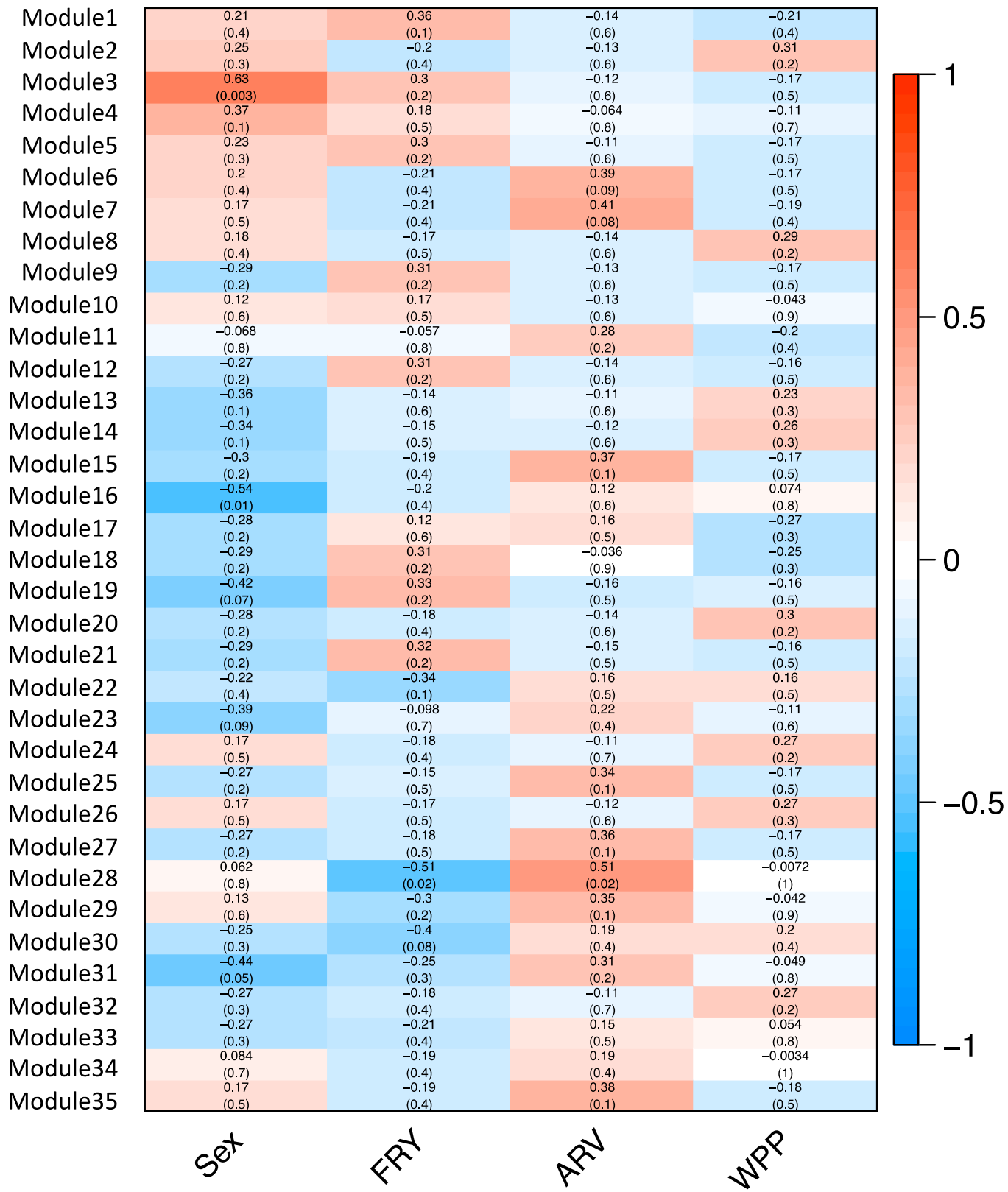
15 **Supplementary Table 4: Gene-ontology enrichment analysis of co-expressed genes**

16 **associated with sex.** Gene-ontology enrichment analysis results for co-expressed genes in

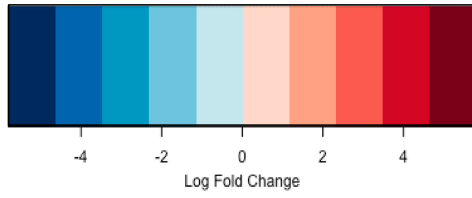
17 Module3 which was strongly associated with sex.



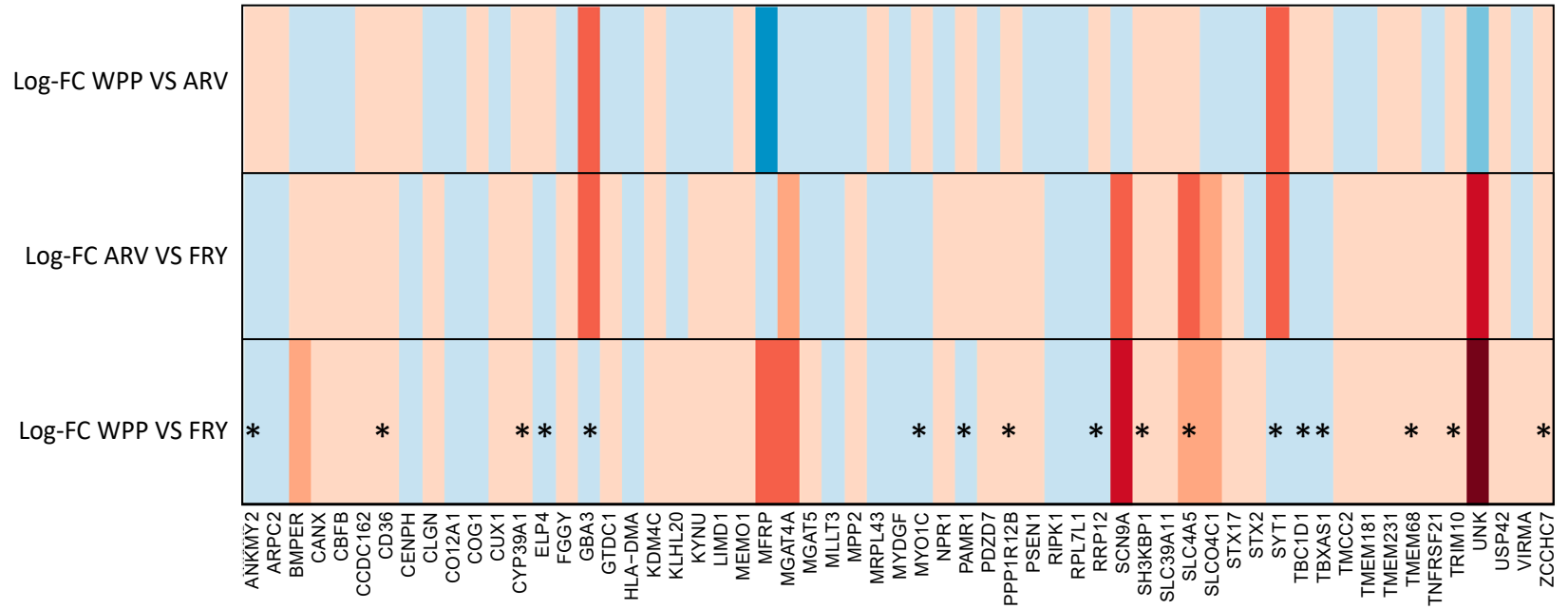
Supplementary Figure 1: Sample dendrogram and trait heatmap. Dendrograms of samples showing clustering patterns based on population and sex. Colored blocks indicate assignment of a specific sample (numbered 1-20) to the variables of interest: geographic location or sex.



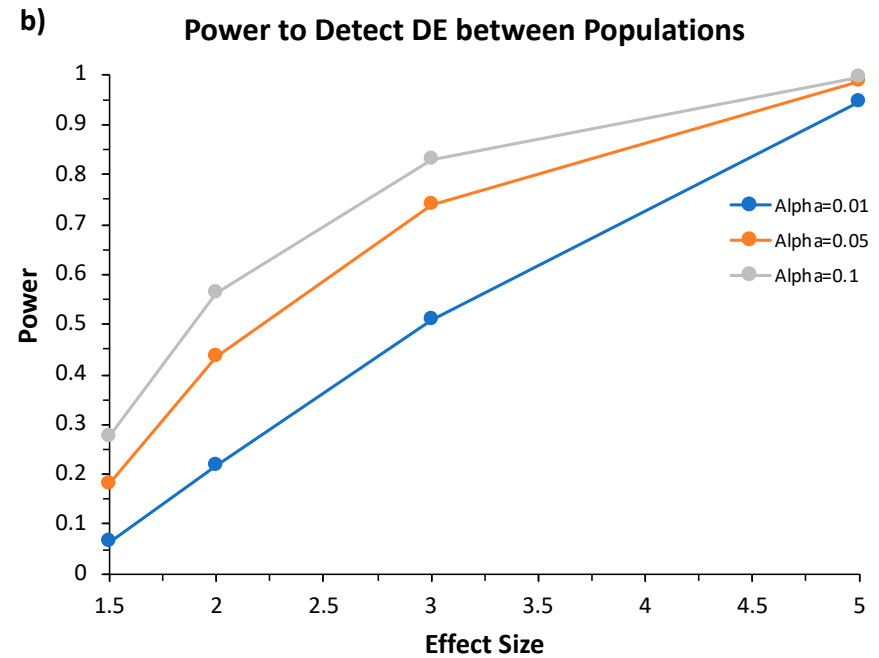
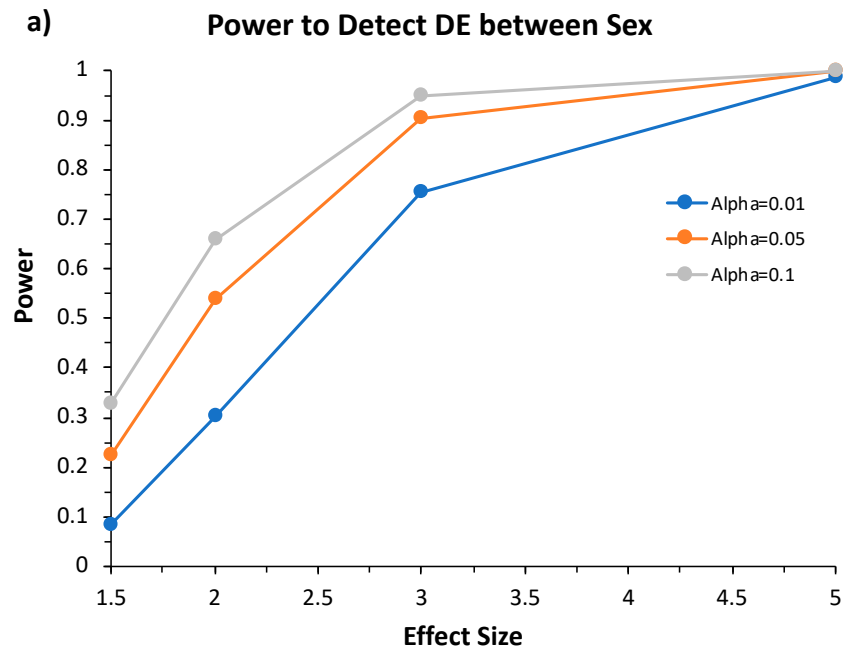
Supplementary Figure 2: Weighted gene co-expression module heat map. Weighted gene co-expression module heat map with gene modules of co-expressed genes occupying the rows and the covariates of interest representing the column values. The first column represents the comparison between modules of gene co-expression and sex. Each subsequent column value represents a population pairwise comparison (i.e., ARV is the comparison of the Arthur River population to the combination of the Freycinet and West Pencil Pine populations). The values in the block above the value in the parentheses are Pearson's correlate describing the strength of the relationship between each gene module and each covariate of interest. Values in the parentheses are the p-values indicating the significance of the relationship between gene module and covariate of interest. Cold colors indicate a negative relationship between the Pearson's correlate between the module and covariate while warm colors indicate a positive relationship.



Core Enrichment = *



2 **Supplementary Figure 3: Differential gene expression of landscape genomic candidate genes.** Heatmap showing the Log₂FC
3 (Log-transformed fold change) for each of the candidate genes for local adaptation to abiotic environment in a Tasmanian devils
4 landscape genomics (LG) study identified by Fraik et al. 2019. Asterisks indicate these genes were significantly enriched in the
5 leading edge of significant gene set enrichment analyses. Only pairwise population comparisons in gene expression between FRY &
6 ARV and WPP & FRY were significant.



7 **Supplementary Figure 4: Power analysis between sex and populations.** Power analysis was conducted to quantify whether we had
 8 sufficient power to identify significant variation in expression of genes between our biological conditions. Our experimental design
 9 had significant power (> 70%) to detect differential gene expression at an $\alpha=0.05$ between the sexes (a) and geographic sampling
 10 locations (b) at higher effect sizes ($\text{Log}_2\text{FC} > 3$).