Title: Shannon Entropy as a metric for conditional gene expression in *Neurospora crassa*

Authors: Abigail J. Ameri[1] and Zachary A. Lewis[1,2]

Institutional Affiliations:

[1] Department of Microbiology, University of Georgia, Athens, GA, 30602

[2] Corresponding Author: Zachary A. Lewis, University of Georgia, Athens, GA zlewis@uga.edu

ORCID

Abigail Ameri: https://orcid.org/0000-0001-5319-2630

Zachary Lewis: https://orcid.org/0000-0002-1735-8266

Short Title: A metric for variable gene expression

Key words or phrases: Shannon entropy, R, conditional gene expression, *Neurospora*

Corresponding Author: Zachary A. Lewis, University of Georgia, Athens, GA zlewis@uga.edu

Abstract

Neurospora crassa has been an important model organism for molecular biology and genetics for over 60 years. *N. crassa* has a complex life cycle, with over 28 distinct cell types and is capable of transcriptional responses to many environmental conditions including nutrient availability, temperature, and light. To quantify variation in *N. crassa* gene expression, we analyzed public expression data from 97 conditions and calculated the Shannon Entropy value for *Neurospora's* approximately 11,000 genes. Entropy values can be used to estimate the variability in expression for a single gene over a range of conditions and be used to classify individual genes as constitutive or condition-specific. Shannon entropy has previously been used measure the degree of tissue specificity of multicellular plant or animal genes. We use this metric here to measure variable gene expression in a microbe and provide this information as a resource for the *N. crassa* research community. Finally, we demonstrate the utility of this approach by using entropy values to identify genes with constitutive expression across a wide range of conditions and to identify genes that are activated exclusively during sexual development.

Introduction

Across conditions, individual genes can display expression patterns that can range from conditional to constitutive. When performing Quantitative Reverse Transcription PCR (qRT-PCR) it is crucial to identify constitutively expressed genes for experimental normalization (HUGGETT *et al.* 2005). Conversely, highly regulated, condition-specific gene promoters are often used in molecular biology to drive conditional expression of a gene under investigation (e.g., an essential gene) or to control expression of reporter genes in certain cell types or environmental conditions (e.g., a gene encoding a fluorescent protein) (GILES *et al.* 1985; HURLEY *et al.* 2012; LAMB *et al.* 2013). Moreover, identification of genes that are exclusively expressed during a condition or cell-type of interest can reveal genes that are functionally important. Such genes or promoters are often identified by examining gene expression across just a handful of experimental conditions; however, with the increase in publicly available transcriptomics data it is possible to quantify variation in gene expression across many conditions for a given organism.

In 1963, Claude Shannon laid the basis for information theory, and described the unit known as Shannon entropy (SHANNON 1997). A simplistic definition of Shannon entropy is that it describes the amount of information a variable can hold (VAJAPEYAM 2014). In our case, a variable is a gene, and the information is the collection of expression values from different conditions. If a gene is classified as having low entropy, then the expression values would be generally consistent across different conditions or possess a low amount of information. Instead, if a gene is classified as having high entropy, then the expression of this gene would be highly variable across different conditions and contain a high amount of information.

Since entropy describes information contained in a variable there are a number of uses for such a metric. Previous studies have used entropy to investigate cell and tissue specific expression of genes (SCHUG *et al.* 2005), identify potential therapeutic targets (FUHRMAN *et al.* 2000), characterize periodicity in gene expression (LANGMEAD *et al.* 2002), identify cancerous tissue samples (VAN WIERINGEN AND VAN DER VAART 2011), and make genomic comparisons (MACHADO 2012). Studies using entropy have been carried out in human cell lines (NATHANIEL D. HEINTZMAN *et al.* 2009), mouse (SCHUG *et al.* 2005), plants (ZHANG *et al.* 2006), yeast (TIMOTHY R. LEZON *et al.* 2006), bacteria, phage, and metagenomes (AKHTER *et al.* 2013) but not yet in filamentous fungi.

Neurospora crassa has a 43Mb genome encoding approximately ~11,000 genes (BORKOVICH *et al.* 2004) (add Nature paper). There is a whole genome knock out collection, and genetic, genomic, and epigenetic studies have been carried out with this organism for more than 100 years (COLOT *et al.* 2006). Indeed, *N. crassa* has been used as a model organism for epigenetics, testing fungal enzymes for biomass degradation, and circadian clock studies (DUNLAP *et al.* 2007; TIAN *et al.* 2009; ARAMAYO AND SELKER 2013). As a resource for *N. crassa* researchers, we generated an entropy value for most genes in the *N. crassa* genome using publicly available RNA-seq data, and we validated this approach using previously published lists of housekeeping or inducible genes. This resource has a number of useful applications for the *N. crassa* community.

Methods:

Public data collection:

Entropy calculations were made for all genes in the *N. crassa* genome using public RNA-seq data sets (97 conditions from a total of 173 separate sets including replicates; Table S1).

Data Analysis:

Mapping, TPM and entropy calculations:

HiSat2 (version 2.1.0) (KIM *et al.* 2019) was used to map all of the SRA accessions to the NC12 genome (NCBI assembly: GCA_000182925.2) using appropriate parameters specific for paired or single end sequence reads (with parameters –RNA-strandness RF or R) to produce bam files which were then sorted and indexed using SAMtools (version 1.3) (LI *et al.* 2009). If experiments contain replicates, the replicate bam files were merged together before obtaining counts with featureCounts from Subread (version 1.6.2) (LIAO *et al.* 2014). FeatureCounts was used with parameters -T exon to generate all counts at the gene level. Counts were imported into R where we obtained TPM using the function calculateTPM from the R package scater (MCCARTHY *et al.* 2017). This package takes in feature-level (in our case, gene-level) counts and gene lengths and outputs the TPM values for each gene. TPM values were then used to calculate the Shannon entropy using the R package BioQC (ZHANG *et al.* 2017). The function entropySpecificity was used to calculate the entropy values for all genes in the genome. To examine specific genes sets, we converted from NCU accession numbers to gene identifiers from NCBI Genome Assembly NC12 (GCA_000182925.2) and plotted the kernel density estimation with rug plots.

Data Availability Statement: All supplementary tables have been uploaded to Figshare. Table S1 contains SRA accession numbers, short descriptions, total reads, and mapped reads for each public data set used.Calculated entropy values for all *N. crassa* genes are listed in Table S2. Lists of all *N. crassa* genes used to benchmark the entropy values and generate panels in figure 2 and 3 are included in Table S3. Code used to generate the data in this manuscript is available through github. https://github.com/ajcourtney/entropy

Results and Discussion:

Shannon entropy values are useful in measuring the amount of variation in expression levels across different tissues or growth conditions**.** In order to calculate Shannon entropy values for all *Neurospora crassa* genes, we first compiled a list of available RNA-seq data sets present in the NCBI sequence read archive (SRA) (Table S1). We selected datasets that were generated with the wild type strain Oak Ridge strain background, but we used both mating types. To calculate accurate entropy values, we needed to gather many observations of gene expression across different conditions. We searched the SRA database (Leinonen *et al.* 2011) for *N. crassa* RNA-sequencing entries that were processed at different developmental stages or grown under different conditions. In total we gathered 173 accessions, which represent 97 developmental or growth conditions. We then developed a pipeline to generate entropy values for each gene (Figure 1A). Calculated entropy values are available in Table S2. We first mapped to the NC12 *N. crassa* genome using HiSat2 (Kim *et al.* 2019) to generate bam files. The bam files were then used to generate read counts for each gene in each condition using featureCounts (LIAO *et al.* 2014), which assigns reads to genomic features. Once the count file was created, we calculated normalized expression values using the Transcripts per Million (TPM) normalization method to create a matrix of normalized expression values for all genes in all conditions. We then used this expression matrix to calculate the Shannon entropy value for each gene (ZHANG *et al.* 2017). This generated entropy values for 10,300 out of 10,398 genes. The remaining 98 genes had 0 read counts in all conditions, so we were unable to calculate entropy. Our final entropy values range from 0.0506 to 6.599. 70% of the genes in the genome possess low entropy values between 0.05 and 1 (7,180/10,300) (Figure 1B). These values include the constitutively expressed genes in the genome. Entropy values above one represent only 30% of the genome (3,120/10,300), corresponding to genes with more condition-specific expression patterns.

Validation of entropy as a measure of gene expression variation in *N. crassa.*

In order to determine if entropy values are a reliable predictor of expression variability in a microbe, we examined the entropy values generated here for published gene sets expected to be enriched for constitutively expressed genes, or conversely, for gene sets expected to contain genes with highly condition-specific expression patterns. If entropy value is a reliable measure of gene expression variation across conditions, housekeeping genes should be enriched for genes with low entropy values, whereas sets of conditionally-induced genes are expected to be enriched for high entropy values. Two previous studies identified genes useful for RT-qPCR controls in *N. crassa*. One of which published a list of 38 genes classified as “housekeeping genes” based on previously generated microarray and RNA-seq datasets under three different conditions (quinic acid (QA) induction, circadian gene expression profiling, and light response) (HURLEY *et al.* 2015), and the other study identified four genes by using previous transcriptomic studies and genes used in related organisms to generate candidates that were validated by quantitative PCR under different conditions (CUSICK *et al.* 2014) (Table S1). To visualize the distribution of entropy values in this set of 42 "housekeeping" genes, we plotted a kernel density estimation (KDE) of entropy values (Figure 2A). The KDE is a smoothed version of a histogram estimated from the underlying data. As expected, the highest density of data points in the housekeeping data set is around 0.25 (low entropy) and the density falls sharply around 0.75 (Figure 2A). Two genes in this set possess entropy values above 1.6 and they encode an exo-beta-1,3-glucanase and a UDP-glucose dehydrogenase. We plotted a heatmap depicting TPM values for each gene in each condition with genes ranked by entropy values from low to high (top to bottom) (Figure 2B). Genes with higher entropy values showed significant induction of gene expression under certain conditions, whereas genes with low entropy values displayed consistent expression values across all conditions. In particular, the two genes with high entropy values showed marked induction under certain conditions. Thus, these data highlight the value of performing a comprehensive analysis of conditional gene expression when selecting constitutive control genes.

We further validated the use of entropy as a measure for constitutive gene expression by using the same approach with a published list of genes 2,624 genes involved in transcription and translation (Table S1), reasoning that genes involved in these essential processes would be expressed at similar levels in all 93 conditions we investigated. (BENZ *et al.* 2014). The distribution of entropy values for transcription and translation genes resembles the distribution of entropy values for housekeeping genes where the highest density is concentrated at the low end of entropy values (Figure 2C). Many of the genes that possess entropy values above 1.6 are either hypothetical proteins or genes associated with cellular transport or metabolism. We again examined the TPM values for each gene in this set in a heatmap ranked by entropy from low to high and again find mostly steady expression across conditions (Figure 2D).

We next asked if higher entropy values were associated with conditionally expressed genes. The highest entropy values imply that a gene must only be expressed under specific conditions and may only show expression in one or a few of the conditions in the entire RNA-seq dataset. To confirm that higher entropy values were indeed associated with condition- or tissue-specific gene expression, we created KDE plots for 513 genes induced by light (Figure 3A and Table S1) and 3,259 genes that have expression changes during sexual development (Figure 3C and Table S1) (Wu *et al.* 2014) (Wang *et al.* 2014). In both cases, there is a shift in distribution of entropy values toward higher entropy values compared to “housekeeping” or “transcription and translation” genes. We examined TPM values for each gene in each condition using a heatmap ranked by entropy values from low to high (top to bottom) and find that a majority of genes in each gene set show variable expression across conditions, as expected (Figure 3B, D). Genes that have regulation changes during perithecial (sexual) development also show a shift to the right, but with retention of more low entropy genes than in the light induced gene set (Figure 3C). Plotting the TPM values in an entropy ranked heatmap shows that approximately half of these genes are highly expressed across many conditions and half are variably expressed, corresponding to genes with lower entropy values in the density plot (Figure 3D). This implies that half of these genes are not specific to sexual or vegetative cell types even though they show transcriptional changes throughout development (WANG *et al.* 2014).

As a final confirmation that entropy can be used as a reliable metric to assess the variation or lack of variation in gene expression levels across many conditions, we plotted the expression levels of 100 genes with the highest entropy values and 100 genes with the lowest entropy values. We took the $\log_2$ TPM values for all conditions (columns) and plotted them for each gene in a heatmap that was clustered by gene (row) for both the top and bottom 100 genes. As expected, with the lowest entropy values show mostly uniform expression across all conditions (Figure 4A), and genes in the high entropy group displayed highly variable and condition-specific expression (Figure 4B). Together, these data demonstrate that entropy is an effective tool for measuring variation in gene expression levels in a filamentous fungus.

The information and code generated in the course of this study could prove useful in a number of ways. First, identifying genes that are induced in a certain condition and display a high entropy value will help identify genes that are condition-specific. In addition, examining entropy values for individual genes can be a useful approach for finding new inducible promoters to use for genetic studies. Condition-specific expressed genes are good starting targets to test for this purpose. The entropy metric determined here can also be used to confirm constitutive expression of genes chosen as controls for RT-PCR. In examining the housekeeping genes from previously published studies it is clear that not all will function as good controls under all conditions, a limitation that was discussed by Hurley and colleagues (HURLEY *et al.* 2015). We combined all of their housekeeping genes together, whereas they had them divided into housekeeping genes usable for different conditions in qRT-PCR (QA induction, light response studies, and circadian experiments). Here we can choose genes that will work across all conditions (provided the conditions were represented in the initial dataset). Our approach provides a quantitative metric that can be applied to identify condition-specific genes, as opposed to investigating individual datasets or using controls from previous studies which may not perform as expected. In addition, this methodology is scalable; the initial inclusion of more conditions will only increase the robustness of the metric produced. As more data are published, more datasets can be incorporated. This approach can be used across other fungi in addition to *N. crassa*, provided there are sufficient RNA-seq data publicly available.

Acknowledgments: We would like to thank Dr. Shannon Quinn for meaningful discussions and insight on this project, and Dr. Kristina Smith for comments on the manuscript. This work was supported by grants from the National Institutes of Health (NIH) (R01GM132644) to Z.A.L. and the National Science Foundation Graduate Research Fellowship Program Grant (DGE-1443117) to A.J.C.

References

Akhter, S., B. A. Bailey, P. Salamon, R. K. Aziz and R. A. Edwards, 2013 Applying Shannon's information theory to bacterial and phage genomes and metagenomes. Sci Rep 3**:** 1033.

Aramayo, R., and E. U. Selker, 2013 Neurospora crassa, a model system for epigenetics research. Cold Spring Harb Perspect Biol 5**:** a017921.

Benz, J. P., B. H. Chau, D. Zheng, S. Bauer, N. L. Glass *et al.*, 2014 A comparative systems analysis of polysaccharide-elicited responses in Neurospora crassa reveals carbon source-specific cellular adaptations. Mol Microbiol 91**:** 275-299.

Borkovich, K. A., L. A. Alex, O. Yarden, M. Freitag, G. E. Turner *et al.*, 2004 Lessons from the genome sequence of Neurospora crassa: tracing the path from genomic blueprint to multicellular organism. Microbiol Mol Biol Rev 68**:** 1-108.

Colot, H. V., G. Park, G. E. Turner, C. Ringelberg, C. M. Crew *et al.*, 2006 A high-throughput gene knockout procedure for Neurospora reveals functions for multiple transcription factors. Proc Natl Acad Sci U S A 103**:** 10352-10357.

Cusick, K. D., L. A. Fitzgerald, R. K. Pirlo, A. L. Cockrell, E. R. Petersen *et al.*, 2014 Selection and evaluation of reference genes for expression studies with quantitative PCR in the model fungus Neurospora crassa under different environmental conditions in continuous culture. PLoS One 9**:** e112706.

Dunlap, J. C., J. J. Loros, H. V. Colot, A. Mehra, W. J. Belden *et al.*, 2007 A circadian clock in Neurospora: how genes and proteins cooperate to produce a sustained, entrainable, and compensated biological oscillator with a period of about a day. Cold Spring Harb Symp Quant Biol 72**:** 57-68.

Fuhrman, S., M. J. Cunningham, X. Wen, G. Zweiger, J. J. Seilhamer *et al.*, 2000 The application of shannon entropy in the identification of putative drug targets. Biosystems 55**:** 5-14.

Giles, N. H., M. E. Case, J. Baum, R. Geever, L. Huiet *et al.*, 1985 Gene organization and regulation in the qa (quinic acid) gene cluster of Neurospora crassa. Microbiol Rev 49**:** 338-358.

Huggett, J., K. Dheda, S. Bustin and A. Zumla, 2005 Real-time RT-PCR normalisation; strategies and considerations. Genes Immun 6**:** 279-284.

Hurley, J. H., A. Dasgupta, P. Andrews, A. M. Crowell, C. Ringelberg *et al.*, 2015 A Tool Set for the Genome-Wide Analysis of Neurospora crassa by RT-PCR. G3 (Bethesda) 5**:** 2043-2049.

Hurley, J. M., C. H. Chen, J. J. Loros and J. C. Dunlap, 2012 Light-inducible system for tunable protein expression in Neurospora crassa. G3 (Bethesda) 2**:** 1207-1212.

Kim, D., J. M. Paggi, C. Park, C. Bennett and S. L. Salzberg, 2019 Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37**:** 907-915.

Lamb, T. M., J. Vickery and D. Bell-Pedersen, 2013 Regulation of gene expression in Neurospora crassa with a copper responsive promoter. G3 (Bethesda) 3**:** 2273-2280.

Langmead, C. J., C. R. McClung and B. R. Donald, 2002 A maximum entropy algorithm for rhythmic analysis of genome-wide expression patterns. Proc IEEE Comput Soc Bioinform Conf 1**:** 237-245.

Leinonen, R., H. Sugawara, M. Shumway and C. International Nucleotide Sequence Database, 2011 The sequence read archive. Nucleic Acids Res 39**:** D19-21.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. Bioinformatics 25**:** 2078-2079.

Liao, Y., G. K. Smyth and W. Shi, 2014 featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30**:** 923-930.

Machado, J. A. T., 2012 Shannon Entropy Analysis of the Genome Code. Mathematical Problems in Engineering 2012.

McCarthy, D. J., K. R. Campbell, A. T. L. Lun and Q. F. Wills, 2017 Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics 33**:** 1179-1186.

Nathaniel D. Heintzman, Gary C. Hon, R. David Hawkins, Pouya Kheradpour, Alexander Stark *et al.*, 2009 Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459**:** 108-112.

Schug, J., W. P. Schuller, C. Kappen, J. M. Salbaum, M. Bucan *et al.*, 2005 Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol 6**:** R33.

Shannon, C. E., 1997 The mathematical theory of communication. 1963. MD Comput 14**:** 306-317.

Tian, C., W. T. Beeson, A. T. Iavarone, J. Sun, M. A. Marletta *et al.*, 2009 Systems analysis of plant cell wall degradation by the model filamentous fungus Neurospora crassa. Proc Natl Acad Sci U S A 106**:** 22157-22162.

Timothy R. Lezon, Jayanth R. Banavar, Marek Cieplak, Amos Maritan and N. V. Fedoroff, 2006 Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. PNAS 103**:** 19033-19038.

Vajapeyam, S., 2014 Understanding Shannon's Entropy metric for Information. arXiv preprint.

van Wieringen, W. N., and A. W. van der Vaart, 2011 Statistical analysis of the cancer cell's molecular entropy using high-throughput data. Bioinformatics 27**:** 556-563.

Wang, Z., F. Lopez-Giraldez, N. Lehr, M. Farre, R. Common *et al.*, 2014 Global gene expression and focused knockout analysis reveals genes associated with fungal fruiting body development in Neurospora crassa. Eukaryot Cell 13**:** 154-169.

Wu, C., F. Yang, K. M. Smith, M. Peterson, R. Dekhang *et al.*, 2014 Genome-wide characterization of light-regulated genes in Neurospora crassa. G3 (Bethesda) 4**:** 1731-1745.

Zhang, J. D., K. Hatje, G. Sturm, C. Broger, M. Ebeling *et al.*, 2017 Detect tissue heterogeneity in gene expression data with BioQC. BMC Genomics 18**:** 277.

Zhang, X., J. Yazaki, A. Sundaresan, S. Cokus, S. W. Chan *et al.*, 2006 Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell 126**:** 1189-1201.

Figure Legends

Figure 1: Calculation of Shannon entropy for *N. crassa* genes using public RNA-seq data.

A) Schematic of our computational pipeline for calculating Shannon entropy from publically available datasets.

B) *N. crassa* genes display a broad range of entropy values. The histogram shows entropy values for all genes. The y-axis is the number of genes found in each bin. The x-axis shows the binned entropy values.

Figure 2: Constitutively expressed genes are characterized by low entropy values

A) The relative frequency of entropy values for a list of housekeeping genes is shown as a kernel density estimation (KDE) plot. The rug plot, black lines on the bottom in the KDE plot represents the individual data points that create the estimation. The y-axis is the probability density, which is the probability for each unit (gene) on the x-axis. The total area below the KDE curve integrates to one.

B) The heatmap shows the expression value for housekeeping genes across all conditions analyzed. The expression level for each gene is plotted as the $\log_2$ transformed transcript per million (TPM) value. Genes (rows) are plotted in ranked order based on the entropy value from low (top) to high (bottom). The scale on the left indicates entropy values for each gene. Each condition (column) has been assigned a category: Metabolism (gold), Development (green), or Light Response (blue). The categories are represented at the top of the heatmap in the three different colors.

C) The relative frequency of entropy values for a list of genes related to transcription and translation is shown as a kernel density estimation (KDE) plot. The rug plot, black lines on the bottom in the KDE plot represents the individual data points that create the estimation. The y-axis is the probability density, which is the probability for each unit (gene) on the x-axis.

D) Heatmap of $\log_2$ transformed TPM values from all transcription and translation related genes (rows) ranked by entropy (low to high). Entropy values are depicted by the brown to green heatmap on the left, where brown is low (top) and green is high (bottom). Each condition (column) has been assigned a category: Metabolism (gold), Development (green), or Light Response (blue). The categories are represented at the top of the heatmap in the three different colors.

Figure 3: Validating entropy values with previously published light induced genes and genes induced during sexual development

A) The relative frequency of entropy values for a list of light induced genes is shown as a kernel density estimation (KDE) plot. The rug plot, black lines on the bottom in the KDE plot represents the individual data points that create the estimation. The y-axis is the probability density, which is the probability for each unit (gene) on the x-axis. The total area below the KDE curve integrates to one.

B) The heatmap shows the expression value for light induced genes across all conditions analyzed. The expression level for each gene is plotted as the $\log_2$ transformed TPM value. Genes (rows) are plotted in ranked order based on the entropy value from low (top) to high (bottom). The scale on the left indicates entropy values for each gene. Each condition (column) has been assigned a category: Metabolism (gold), Development (green), or Light Response (blue). The categories are represented at the top of the heatmap in the three different colors.

C) The relative frequency of entropy values for a list of sexual development genes genes is shown as a kernel density estimation (KDE) plot. The rug plot, black lines on the bottom in the KDE plot represents the individual data points that create the estimation. The y-axis is the probability density, which is the probability for each unit (gene) on the x-axis. The total area below the KDE curve integrates to one.

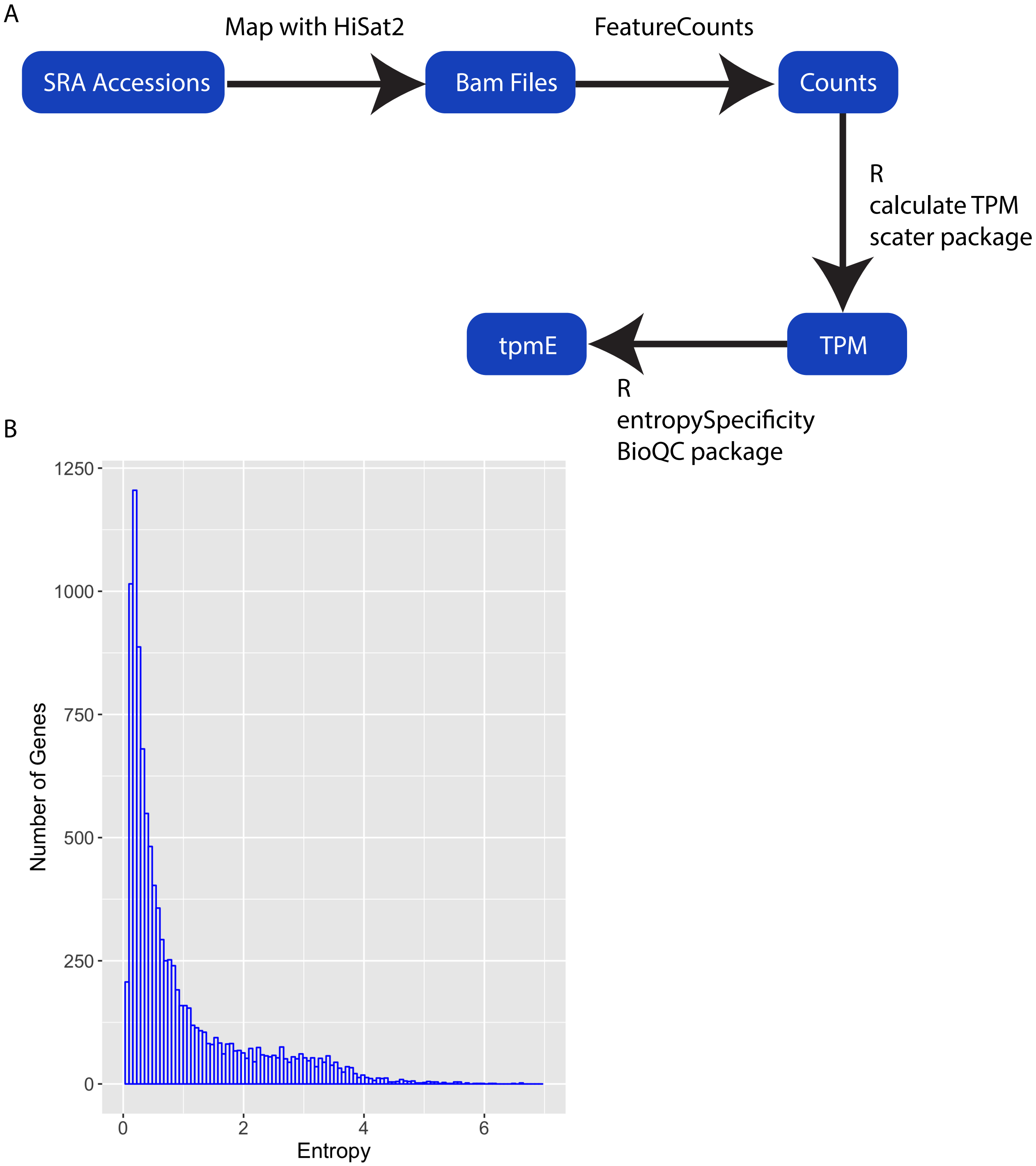
D) The heatmap shows the expression value for sexual development genes across all conditions analyzed. The expression level for each gene is plotted as the $\log_2$ transformed TPM value. Genes (rows) are plotted in ranked order based on the entropy value from low (top) to high (bottom). The scale on the left indicates entropy values for each gene. Each condition (column) has been assigned a category: Metabolism (gold), Development (green), or Light Response (blue). The categories are represented at the top of the heatmap in the three different colors.

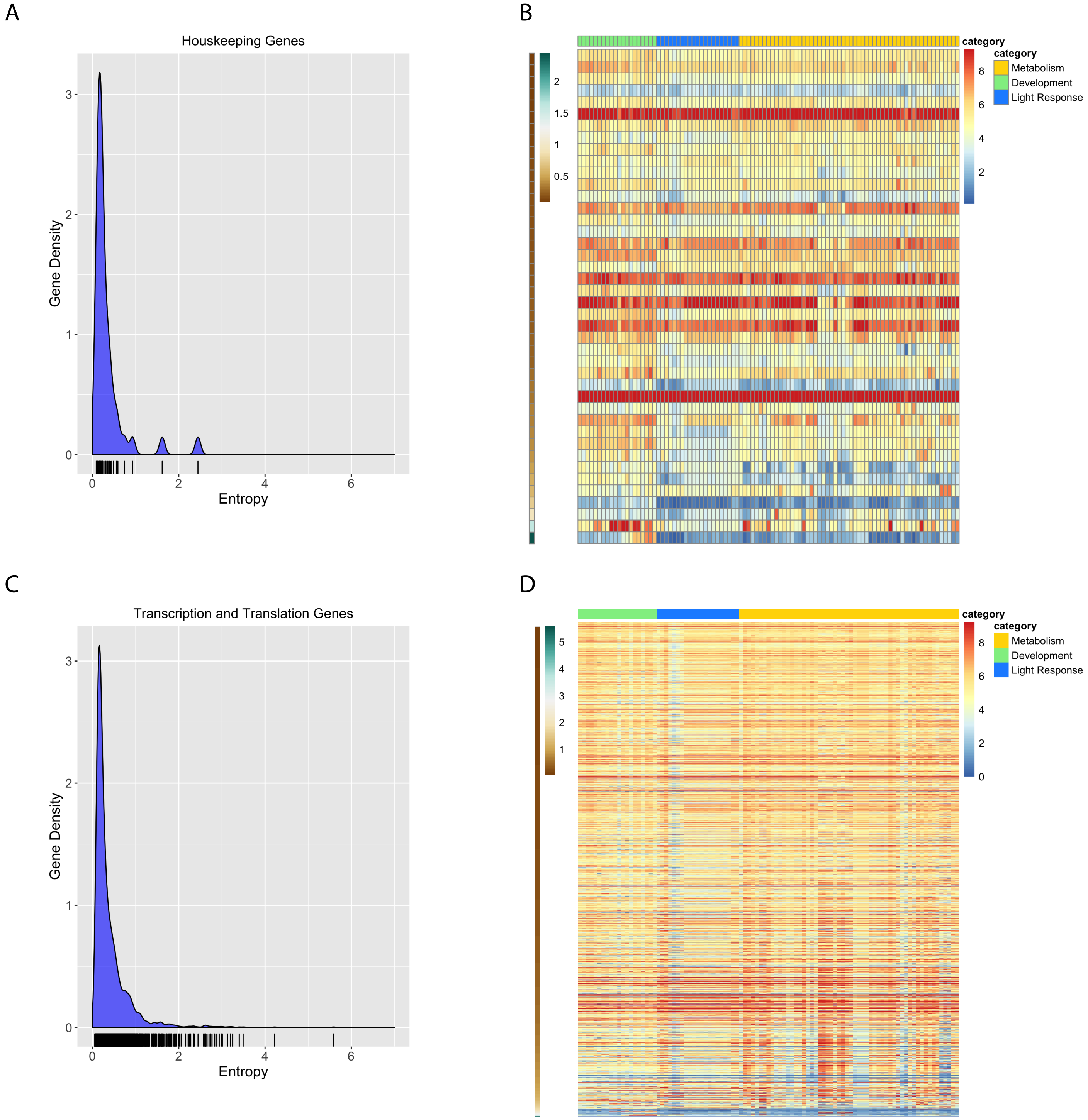
Figure 4: Log_2 TPM values for highest and lowest ranked genes

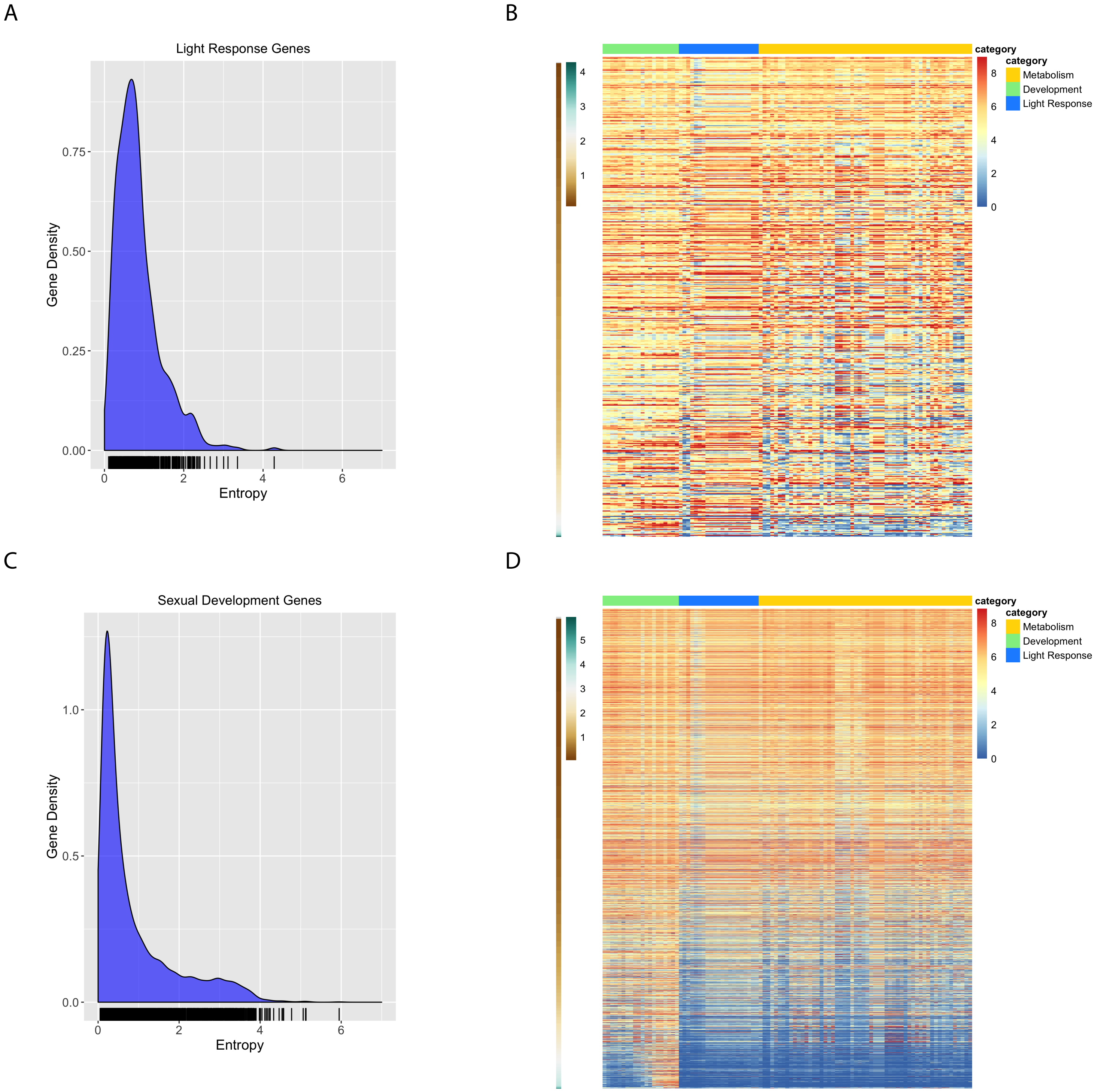
A) The heatmap shows the expression values for the 100 genes with the highest entropy values. The expression level for each gene is plotted as the $\log_2$ transformed TPM value. Each row represents a gene. Gene names are listed on the right side of the heatmap. Each condition (column) has been assigned a category: Metabolism (gold), Development (green), or Light Response (blue). The categories are represented at the top of the heatmap in the three different colors.

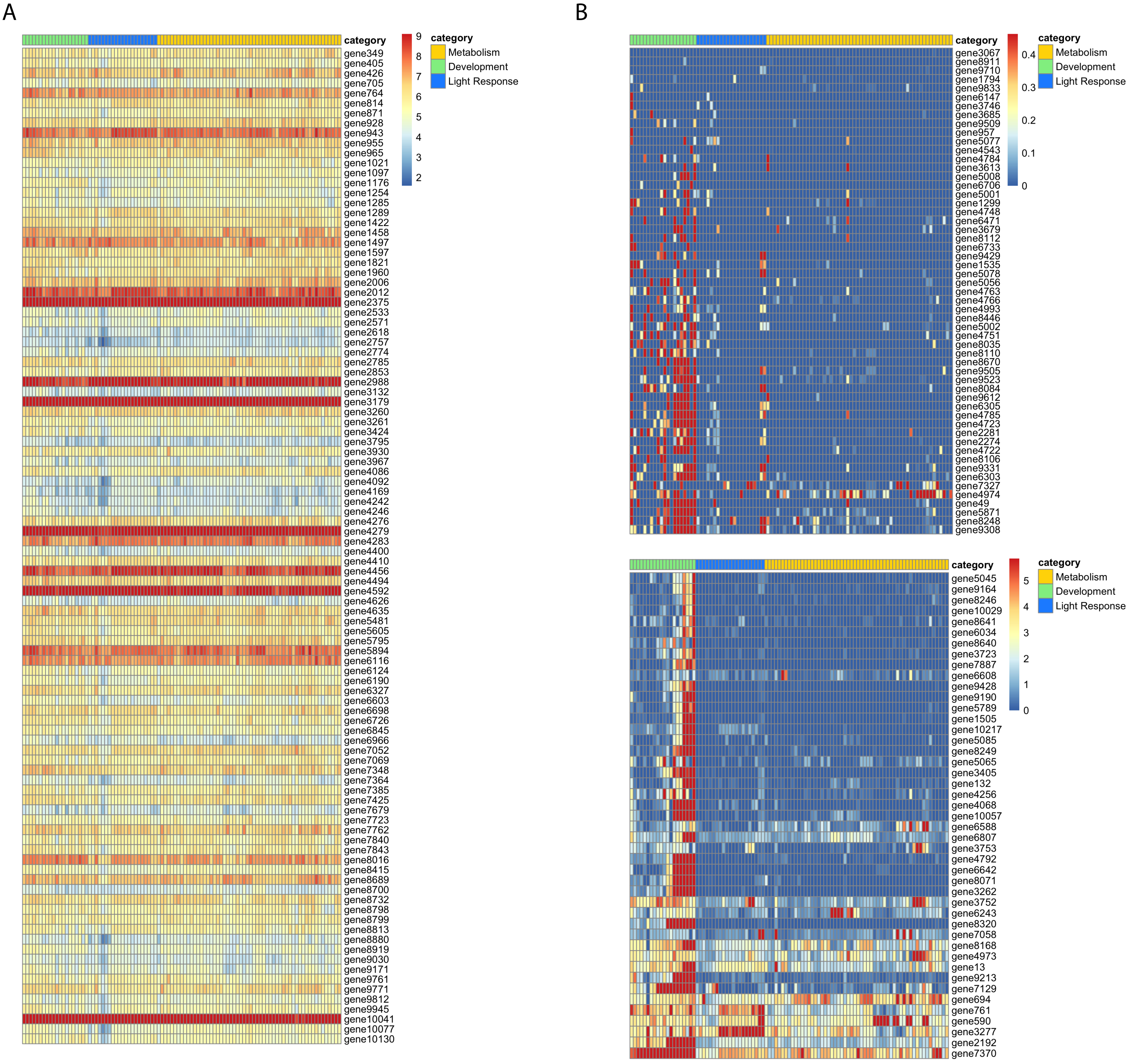
B) The heatmap shows the expression values for the 100 genes with the lowest entropy values. The expression level for each gene is plotted as the $\log_2$ transformed TPM value. Each row represents a gene. Gene names are listed on the right side of the heatmap. Each condition (column) has been assigned a category: Metabolism (gold), Development (green), or Light Response (blue). The categories are represented at the top of the heatmap in the three different colors.

Zachary Lewis




A
Houskeeping Genes
Gene Density
0
1
2
3
0
2
4
6
Entropy
B
2
1.5
1
0.5
category
category
Metabolism
Development
Light Response
8
6
4
2
C
Transcription and Translation Genes
Gene Density
0
1
2
3
0
2
4
6
Entropy
D
5
4
3
2
1
category
category
Metabolism
Development
Light Response
8
6
4
2
0


A
Light Response Genes
Gene Density
0.00
0.25
0.50
0.75
Entropy
0
2
4
6
B
4
3
2
1
category
category
8
6
4
2
0
Metabolism
Development
Light Response
C
Sexual Development Genes
Gene Density
0.0
0.5
1.0
Entropy
0
2
4
6
D
5
4
3
2
1
category
category
8
6
4
2
0
Metabolism
Development
Light Response


A
category
gene349
gene405
gene426
gene705
gene764
gene814
gene871
gene928
gene943
gene955
gene965
gene1021
gene1097
gene1176
gene1254
gene1285
gene1289
gene1422
gene1458
gene1497
gene1597
gene1821
gene1960
gene2006
gene2012
gene2375
gene2533
gene2571
gene2618
gene2757
gene2774
gene2785
gene2853
gene2988
gene3132
gene3179
gene3260
gene3261
gene3424
gene3795
gene3930
gene3967
gene4086
gene4092
gene4169
gene4242
gene4246
gene4276
gene4279
gene4283
gene4400
gene4410
gene4456
gene4494
gene4592
gene4626
gene4635
gene5481
gene5605
gene5795
gene5894
gene6116
gene6124
gene6190
gene6327
gene6603
gene6698
gene6726
gene6845
gene6966
gene7052
gene7069
gene7348
gene7364
gene7385
gene7425
gene7679
gene7723
gene7762
gene7840
gene7843
gene8016
gene8415
gene8689
gene8700
gene8732
gene8798
gene8799
gene8813
gene8880
gene8919
gene9030
gene9171
gene9761
gene9771
gene9812
gene9945
gene10041
gene10077
gene10130
9
8
7
6
5
4
3
2
category
Metabolism
Development
Light Response
B
category
gene3067
gene8911
gene9710
gene1794
gene9833
gene6147
gene3746
gene3685
gene9509
gene957
gene5077
gene4543
gene4784
gene3613
gene5008
gene6706
gene5001
gene1299
gene4748
gene6471
gene3679
gene8112
gene6733
gene9429
gene1535
gene5078
gene5056
gene4763
gene4766
gene4993
gene8446
gene5002
gene4751
gene8035
gene8110
gene8670
gene9505
gene9523
gene8084
gene9612
gene6305
gene4785
gene4723
gene2281
gene2274
gene4722
gene8106
gene9331
gene6303
gene7327
gene4974
gene49
gene5871
gene8248
gene9308
0.4
0.3
0.2
0.1
0
category
Metabolism
Development
Light Response
category
gene5045
gene9164
gene8246
gene10029
gene8641
gene6034
gene8640
gene3723
gene7887
gene6608
gene9428
gene9190
gene5789
gene1505
gene10217
gene5085
gene8249
gene5065
gene3405
gene132
gene4256
gene4068
gene10057
gene6588
gene6807
gene3753
gene4792
gene6642
gene8071
gene3262
gene3752
gene6243
gene8320
gene7058
gene8168
gene4973
gene13
gene9213
gene7129
gene694
gene761
gene590
gene3277
gene2192
gene7370
5
4
3
2
1
0
category
Metabolism
Development
Light Response