

# A Brief Primer on Transcriptomics Studies and Presenting the Results

Aaron Isaacs, Ph.D.

**"There are three kinds of lies: lies,  
damned lies, and statistics."**

- Samuel Clemens (a.k.a Mark Twain) who attributed it to Benjamin Disraeli

**“Lies, Damned Lies, and Medical Science”**

- Title of a 2010 article in The Atlantic magazine

# A transcriptomics study

- Differential expression (PCs, volcano plots)
- Database lookups
- Enrichment (GO terms and/or pathways)
- Networks
- Etc.

# Databases

- Ensembl/UCSC/NCBI
- OMIM
- GO/KEGG/Etc.
- STRING, Protein Atlas
- Gene Cards
- Etc.



# Catch-Me



Discovery

- 32 paired left/right samples
- No need to worry about confounding! (For the most part...)



Replication

- 98 left and 69 right samples
- Adjustments for age, sex, batch, rhythm, and heart failure



Interaction

- 98 left and 101 right samples
- Same adjustments



Results

- Transcriptome-wide FDR  $\leq 0.05$  in discovery and replication, with concordant effect direction
- Bonferroni-corrected  $P \leq 0.05$  for interaction (likelihood ratio test)



Maastricht University



# Sample overview

	Paired	Left	Right	P-value
n	32	98	69	
Age (SD)	62.2 (12.2)	64.4 (10.7)	67.2 (10.9)	0.10
Male (n, %)	25 (78.1%)	64 (65.3%)	52 (75.4%)	0.18
Heart failure (n, %)	15 (46.9%)	37 (37.8%)	15 (21.7%)	0.03
Sinus (n, %)	11 (34.4%)	38 (38.8%)	39 (56.5%)	0.08
Paroxysmal (n, %)	10 (31.3%)	32 (32.7%)	15 (21.7%)	
Persistent (n, %)	11 (34.4%)	28 (28.6%)	15 (21.7%)	

Values presented are mean (standard deviation) for continuous distributions and n (%) for categorical distributions. Differences were tested using Student's t-test or Fisher's exact test.

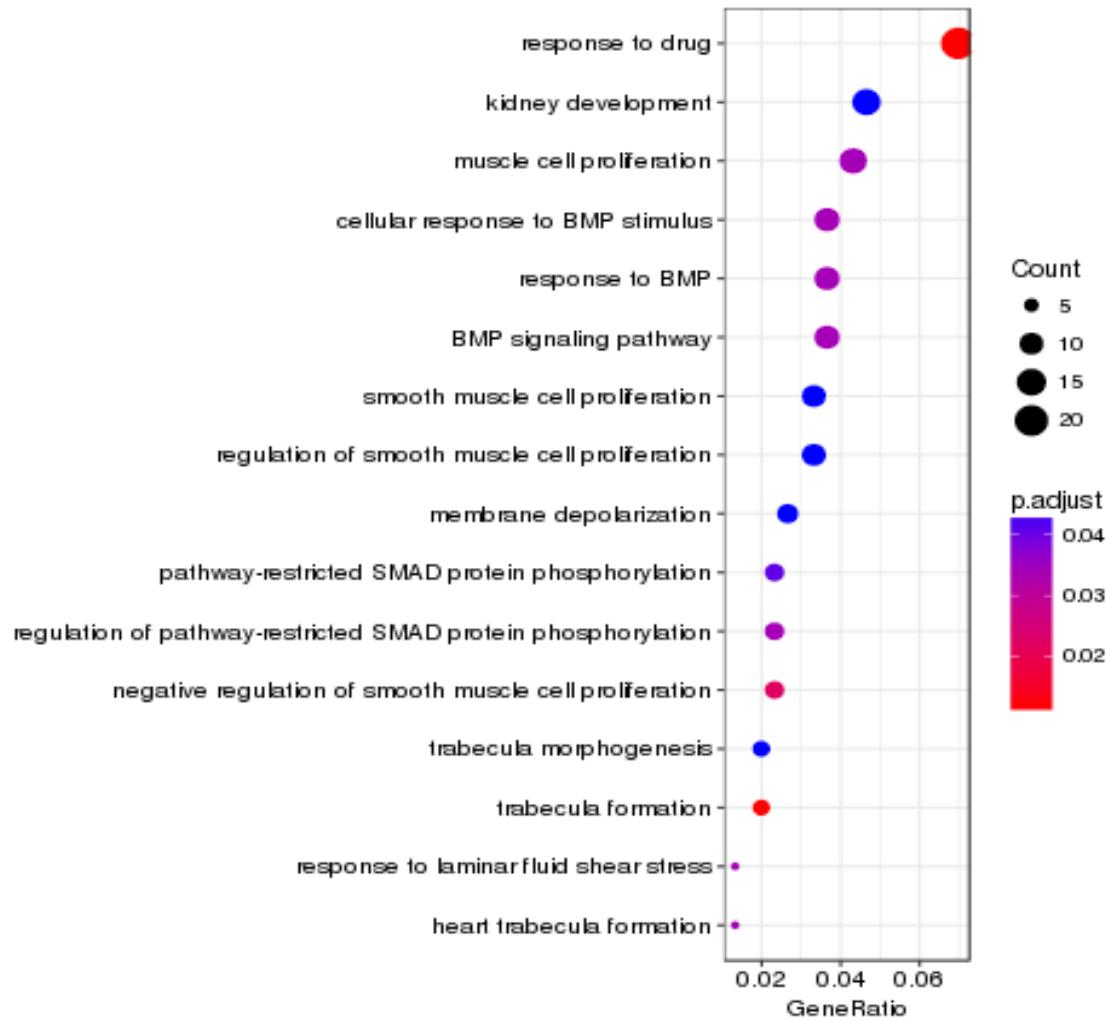
# Left/right differential expression

- 1187 features at transcriptome-wide significance in paired samples
- 1163 (98.7%) of those had concordant effect direction in unpaired samples
- 714 (61.4%) of those at transcriptome-wide significance

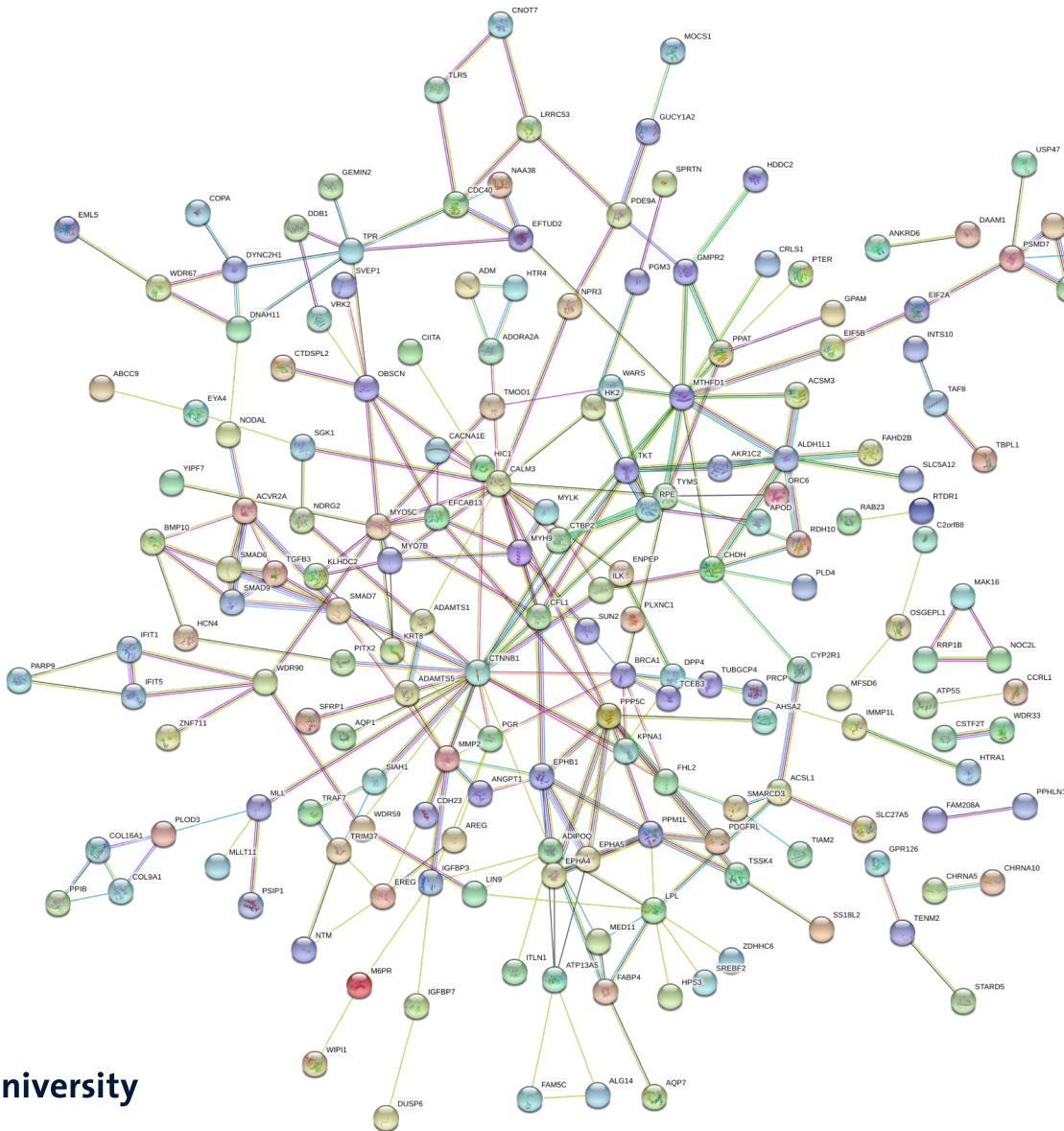
# 714 Significant transcripts

- 350 (49%) coding
- 256 (36%) antisense
- 36 (5%) lncRNA
- 36 (5%) pseudogenes
- 47% of the coding had a non-coding in close proximity

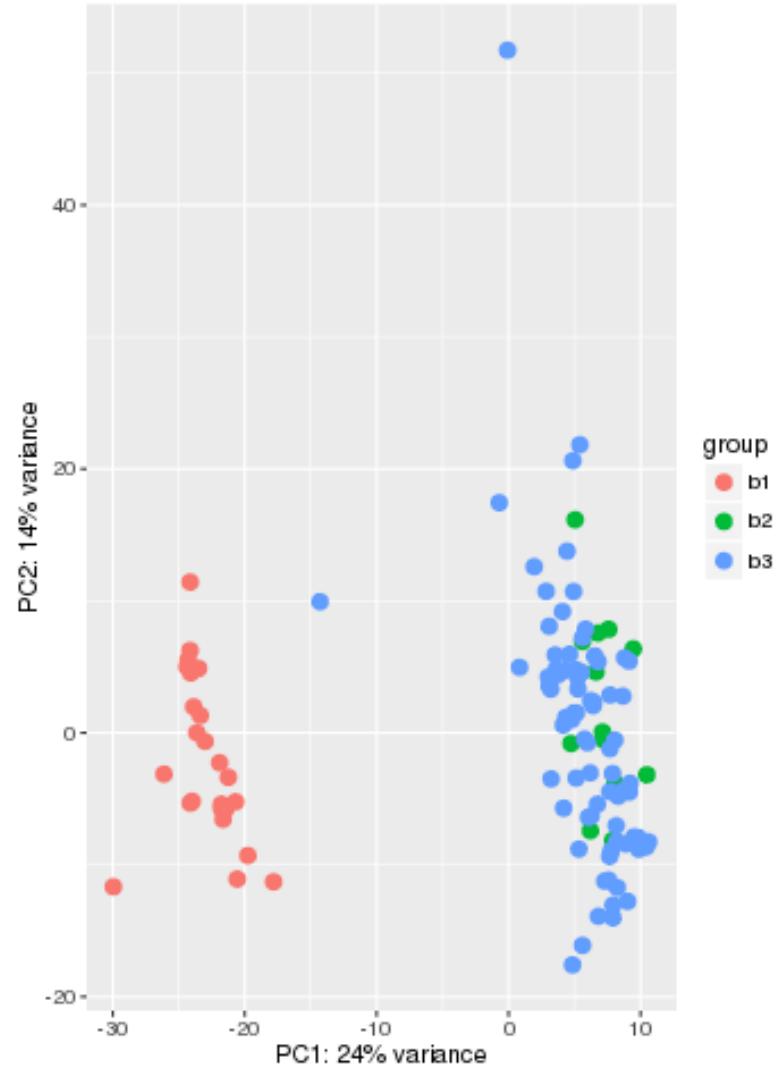
# Enriched GO terms



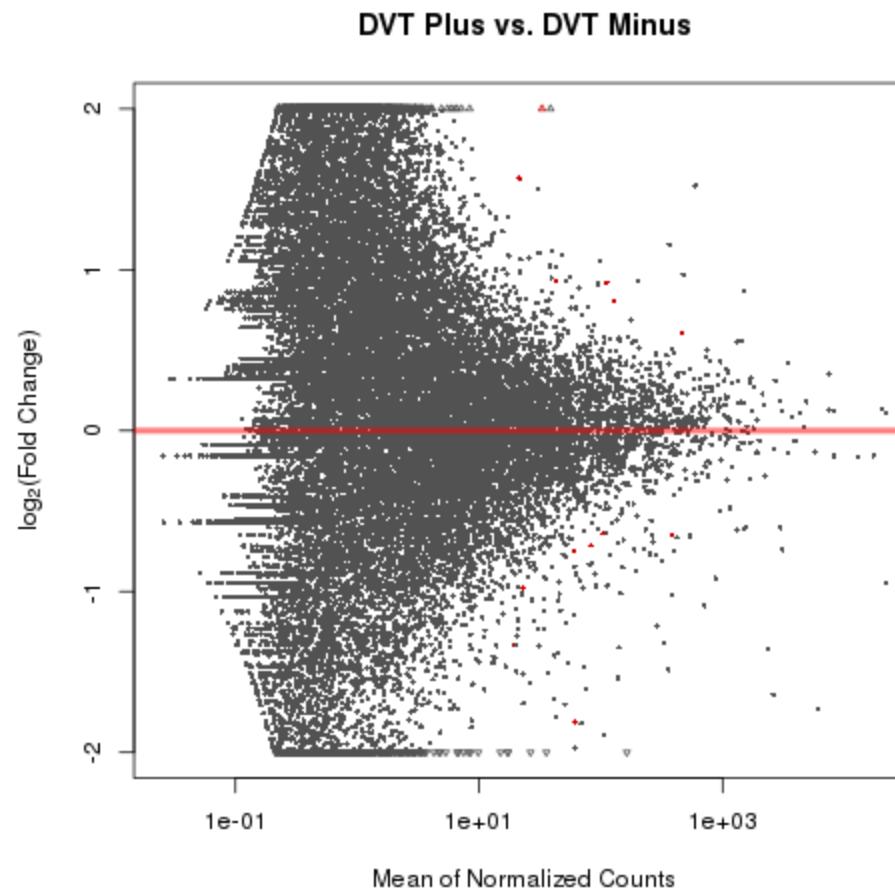
# STRING (protein interaction network)



# Clustering by PCA



# Volcano plots



# Preparing a presentation

# In general...

- Avoid overly busy backgrounds
- Use contrasting background and text colors
- Make sure the fonts are big enough (legibility)
- Follow manuscript order  
(intro/methods/results/conclusions)
- Use a spell checker

## In general... (2)

- Gist of topic should be comprehensible from slides alone
- Finer details should be in oral part
  - If the slides contain every detail, viewers will just read the slides!
  - Corollary: Putting the fine points only in the speech compels viewer attention!

# Text

- Avoid using more than 4 or 5 bullet points (lines of text) per slide
- Be concise – avoid long runs of text
- Avoid “jargonism”
- Be mindful of time

# Figures and tables

- Clear and interpretable
- Sufficient font/image size
- Visually appealing
- Avoid “color blind” color schemes

# Preparing a manuscript

# In general...

- Be concise
- Avoid repetition
- Define abbreviations (at first use)
- Proper attribution
- Use a spell-checker

# Manuscript organization

- Title page (authors and affiliations)
- Abstract
- Introduction
- Materials and methods

# Manuscript organization (2)

- Results
- Discussion
- References
- Tables and figures

# Title page

- Short, succinct title
- Ordered list of authors
- Author affiliations
- Number of tables and figures, word count
- Corresponding author information

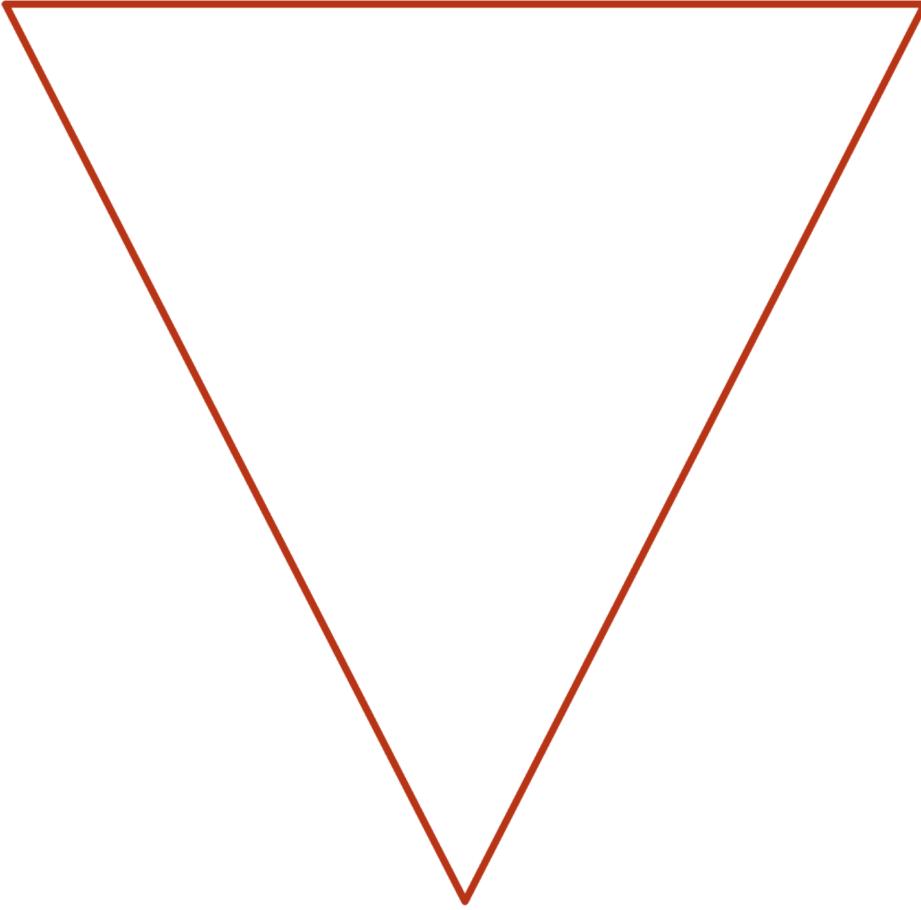
# Abstract

- 250 – 300 words (usually)
- Structured or unstructured
- Brief description of study
  - Goals
  - Methods
  - Main findings
  - Conclusion

# Introduction

- (Brief) overview of the disease
- (Brief) summary of evidence leading to experimental approach
- (Maybe) brief description of methodological utility
- Concise statement of the research question

# Introduction (2)

- 
- Heart disease is bad.
  - A major cause of heart problems is atherosclerosis.
  - Lipids levels are correlated with atherosclerosis.
  - Lipid levels have a strong genetic component.
  - GWAS is a good method for identifying genes.
  - The GWAS approach was implemented to identify novel lipid genes.

# Material and methods

- Sampling of study participants
- Collection of participant data
- Description of experimental methodology
- Description of statistical methodology

# Material and methods (2)

- Statement of medical ethics approval
- Statement of (written) informed consent
- Conformation to Declaration of Helsinki

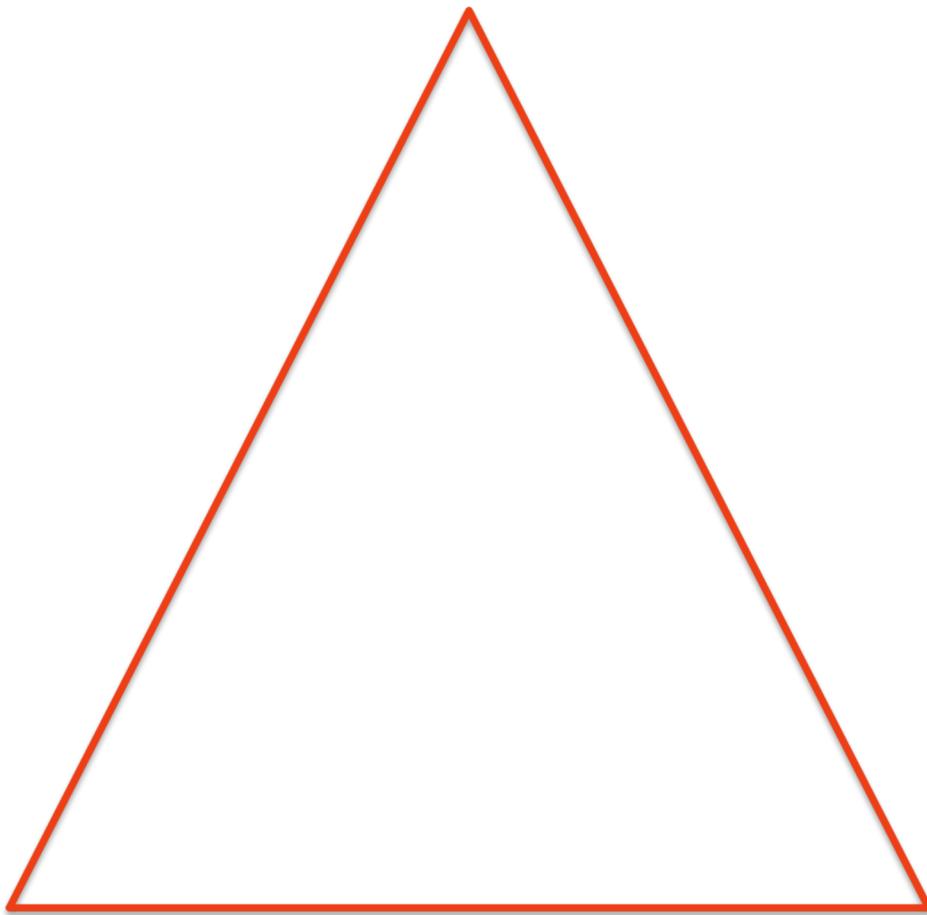
# Results

- Succinct presentation of findings
- Avoid including methodological details
- Refrain from contextualizing
- Reference to tables and/or figures

# Discussion

- Recapitulate main findings
- Contextualize findings
- Strengths and limitations of study
- (Hypothetical) implications of study results

# Discussion (2)



- These are the major findings.
- The biological underpinning of the primary findings.
- These are the secondary findings and their implications.
- The current study has strengths and weaknesses.
- These are the implications for diagnosis/patient stratification/treatment.
- This study suggests these broad conclusions.

# References

- Make sure to use proper attribution (= no plagiarism, credit where credit is due)
- Use a reference manager to allow easy formatting, re-ordering (e.g. EndNote)
- Usually in order of citation in text

# Tables

- Clear and concise
- Notation and abbreviations consistent with main text
- In order of citation in text
- Each table should have a title and a legend

# Figures

- Visualization of data not easily presented as text or table
- Eye-catching
- In order of citation in text
- Each figure should have a title and a legend