Better Analysis, Better Acceptance - Take up the Challenge of Publishing in Bioinformatics

Malay Bhattacharyya

SRF, MIU, ISI, Kolkata, India

First Question

- What is the Domain?
  - Bioinformatics [data → information]

    or
  - Computational Biology [case → model]

Update Yourself

- Keep Notice of the Recent Publications
  - For data
    - Nature [Best Choice]
    - Science
    - Cell
  - For methods
    - Molecular Systems Biology
    - Genome Research
    - Bioinformatics [Best Choice]

Tools for Graph Visualization

- Small Graphs [Lucrative]
  - InterView (Free download: http://www.bioinf.manchester.ac.uk/interview)
- Large Graphs [Boring, High Resolution]
  - Pajek (Free download: http://vlado.fmf.uni-lj.si/pub/networks/pajek)

Tools for Drawing Figures

- Graphs/Plots
  - MATLAB (Free download: http://www.mathworks.com/products/matlab/tryit.html)
- Diagram/Flowchart
  - Incscape (Free download: http://inkscape.org/download) [Not so User-friendly]

Converting Figures to EPS Format

- With XnView
  - Open ‘figureX.bmp’
  - Save as ‘figureX.ps’
  - Rename ‘figureX.ps’ to ‘figureX.eps’
- With MATLAB [Best Resolution]
  - Use the command

    imshow(imread(‘figureX.bmp’));
  - Save as ‘figureX.eps’

Tools for Pattern Recognition

- MultiExperiment Viewer (MeV) (Free download: http://www.tm4.org/mev)
  - – KMC, HCL, SAM, SVM, KNN, PCA, etc.
- Expander (Free download: http://acgt.cs.tau.ac.il/expander)
  - – Expression analysis, differential expression analysis, group analysis, etc.

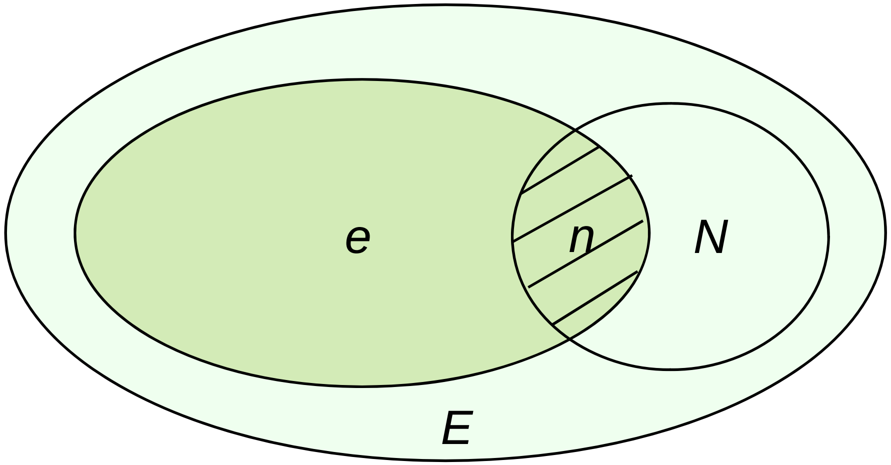
Tools for Biclustering

- Expander (Free download: http://acgt.cs.tau.ac.il/expander)
- BicAT (Free download: http://www.tik.ee.ethz.ch/sop/bicat)
  - BiMax [Best Choice, but Works only on Binary Data]
  - CC, ISA, xMotifs and OPSM

Reviewers' Choice

- Favored terms (of analysis)
  - Distribution
  - Tests of Statistical Significance
  - *p* value [Most Decisive Demand]
  - Genome-wide
  - Current issues (coding potential, conserveness, tissue-specificity, etc.)

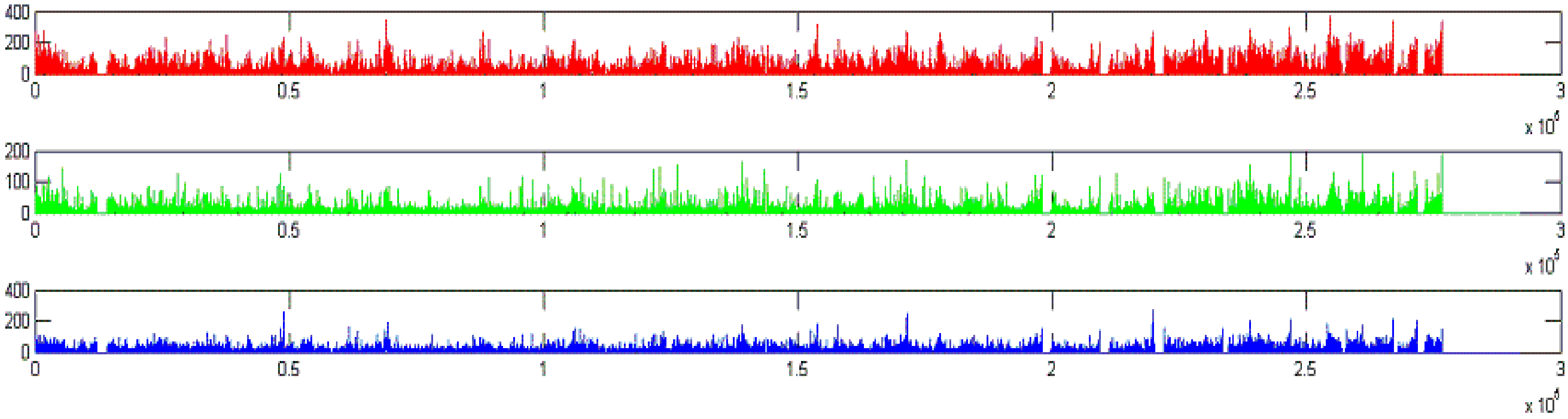
Computing Randomized *p*-value



$$p\text{-value} = \sum_{i=n}^{\min(e,N)} \frac{\binom{e}{i}\binom{E-e}{N-i}}{\binom{E}{N}}$$

- Non-parametric tests

Genome-wide Analysis



Databases to Use

- Genes and proteins – NCBI (http://www.ncbi.nlm.nih.gov)
- MicroRNAs – miRBase (http://www.mirbase.org)
- Other RNAs – RNAdb (http://research.imb.uq.edu.au/RNAdb)
- Other components – UCSC (http://genome.ucsc.edu)
- GO analysis – Funcassociate 2.0 (http://llama.mshri.on.ca/funcassociate)
- Assembly conversion – Galaxy (http://main.g2.bx.psu.edu)
- ID conversion – DAVID (http://david.abcc.ncifcrf.gov/conversion.jsp)
- Gene enrichment analysis – FatiGo (http://babelomics.bioinfo.cipf.es/functional.html)

Where to Publish

- Open-access [Beggar Request, Time Waste]
  - Molecular Systems Biology (IF 2009: 12.125)
  - PLoS Computational Biology (IF 2009: 5.759)
  - PLoS ONE (IF 2009: 4.351)
  - BMC Systems Biology (IF 2009: 4.064)
  - BMC Bioinformatics (IF 2009: 3.428)
- Subscription [No Money, No Tension]
  - Nucleic acids Research (IF 2009: 7.479)
  - RNA (IF 2009: 5.198)
  - Bioinformatics (IF 2009: 4.926) [Best Choice for Us]
  - Molecular Biosystems (IF 2009: 4.015)
  - IEEE/ACM Transactions on Computational Biology and Bioinformatics (IF 2009: 2.246)
  - Computational Biology and Chemistry (IF 2009: 1.370)

Suggestions

- Find topics in *statu nascendi* [search]
- One *ome* is not enough [integrate]
- Results should *triumph* over methods [use tools]
- Focus on hypothesizes [add new]
- Think and *share* [collaborate]

- Otherwise be a computational biologist

[Final Tip: Be Honest in Research]

To obtain long-term citations