

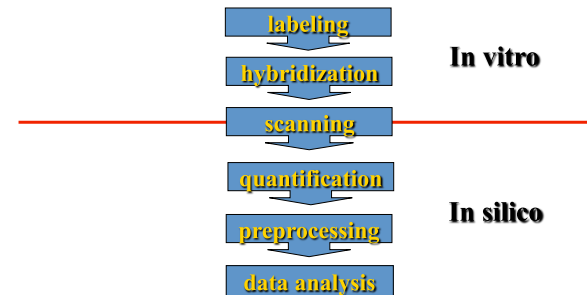
Preprocessing of Dual-channel Microarray Data

Vincent Detours
IRIBHM - Free University of Brussels

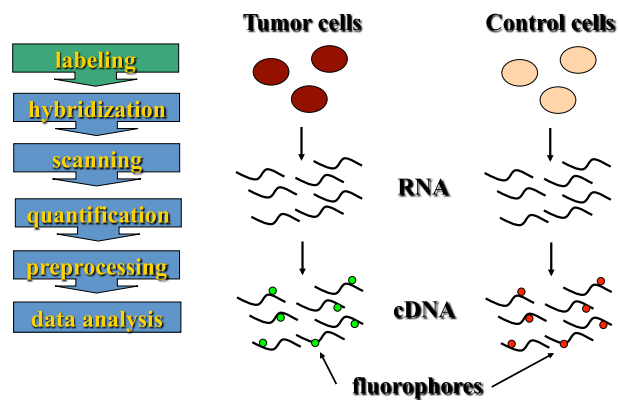
vdetours@ulb.ac.be

1

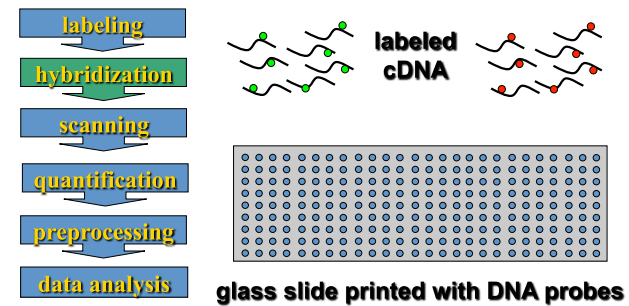
The pipeline



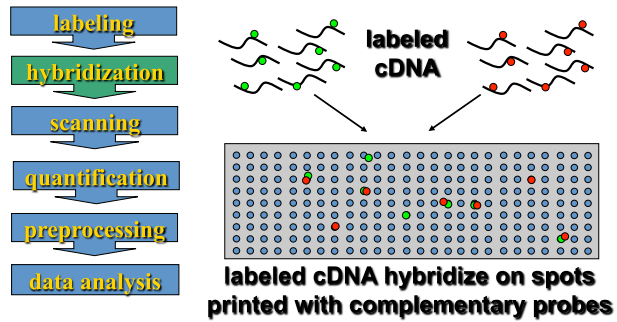
Labeling



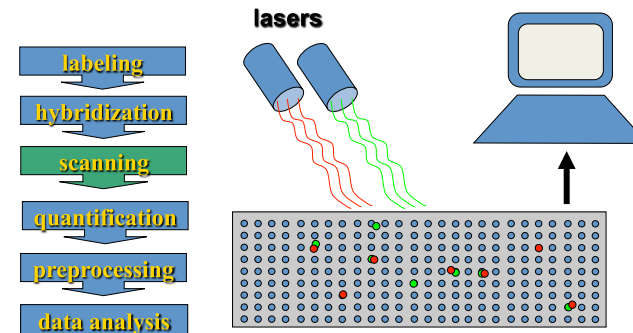
Hybridization



Hybridization

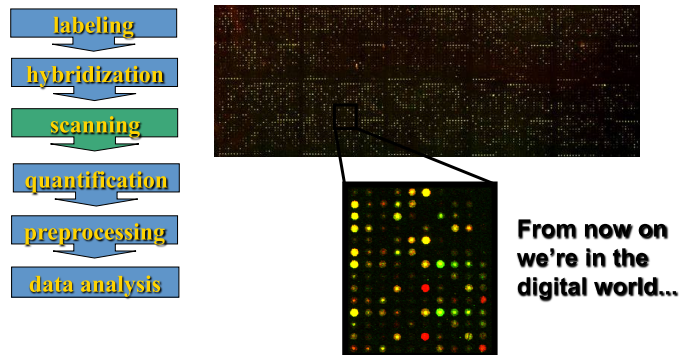


Scanning



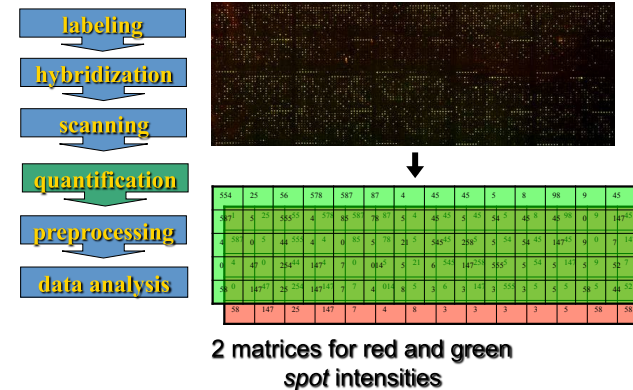
Scanning

Bitmap TIFF images

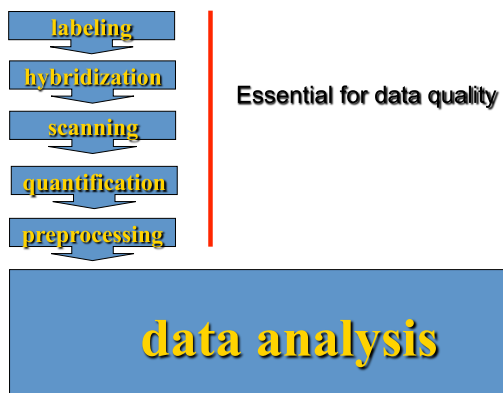
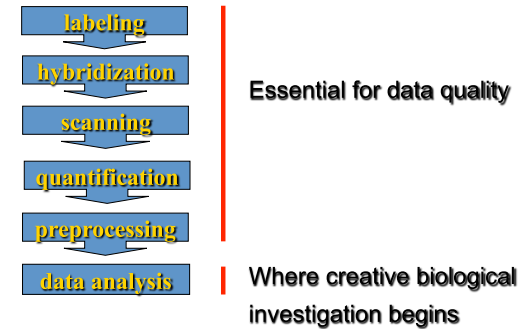
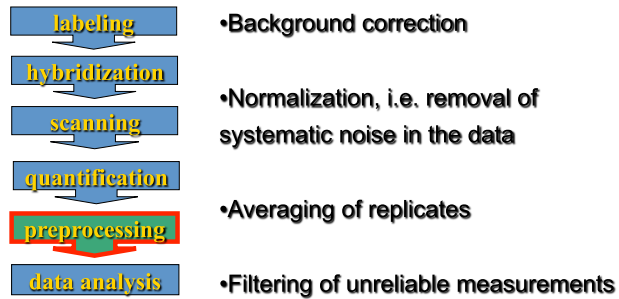


Quantification

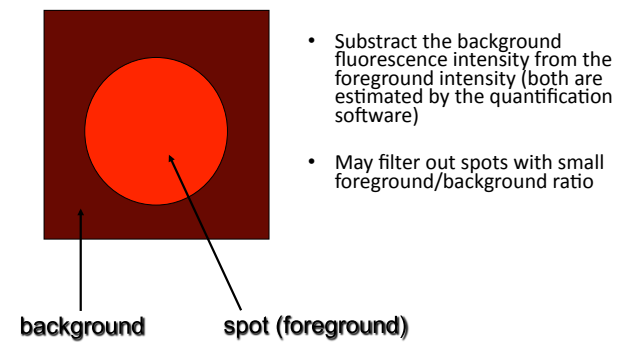
2 images with red and green *pixels'* intensities



Preprocessing

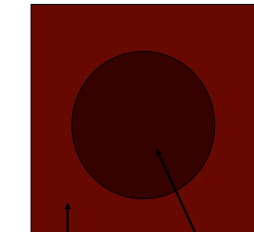


Background correction



12

A vexing problem



- Background is not exactly additive
- Some low intensity spots end up *below* background
- A conservative option is to filter out such 'negative' intensity spots

13

Normalization relies on (possibly false) hypothesis

Some assumptions about how 'good' data should look like are needed in order to detect and correct for systematic biases.

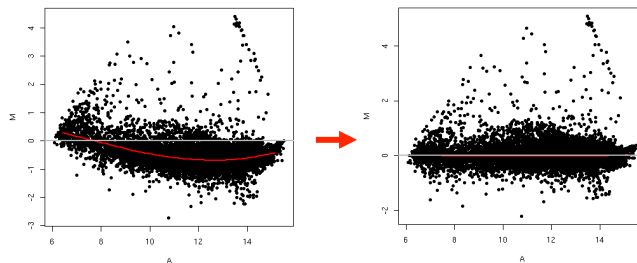
Normalization algorithms typically assume that

- Gene regulation is independent of non biological variables (e.g., the location of a spot on the slide),
- Regulation levels averaged across large numbers of genes is constant (#up = #down).

If these assumptions are not true—they clearly aren't in some biological contexts—only comparisons of expression ratios are valid.

A normalization example: intensity-based correction

Assumption: Genes with low average expression have no biological reason to be more up- or down-regulated than genes with high average expression.



Myc is a global amplifier of transcription (this has consequences for the interpretation of expression data)

Cell

Leadin
Prim

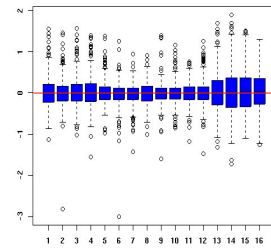
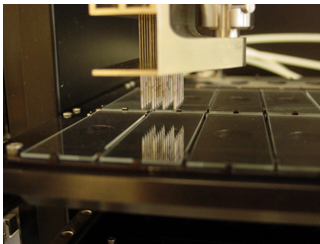
Revisiting Global Gene Expression Analysis

Jakob Lovén,^{1,2} David A. Orlando,^{1,2} Alla A. Sigova,¹ Charles Y. Lin,^{1,2} Peter B. Rahl,¹ Christopher B. Burge,³
David L. Levens,¹ Tong Ihn Lee,^{1,2} and Richard A. Young^{1,2,3,4}

16

A normalization example:
correction of print-tips biases

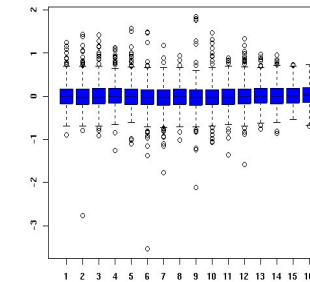
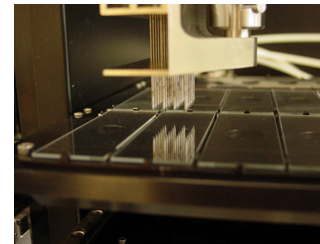
Assumption: Print tip should not influence differential expression



From Lee Hwa Yang

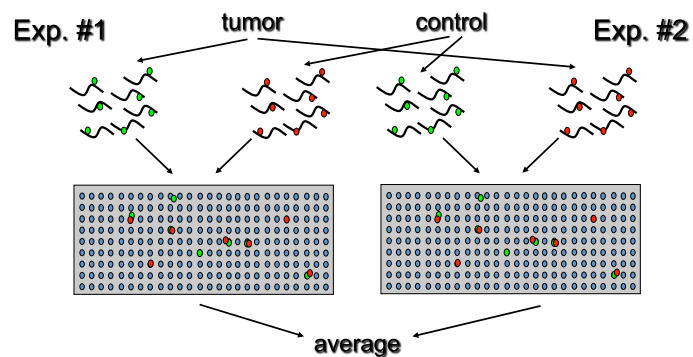
A normalization example:
correction of print-tips biases

Assumption: Print tip should not influence differential expression



From Lee Hwa Yang

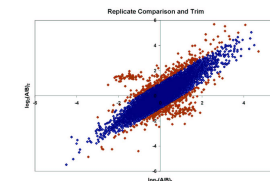
Dye incorporation biases are removed by
averaging dye swap experiments



Filtering

Filter out

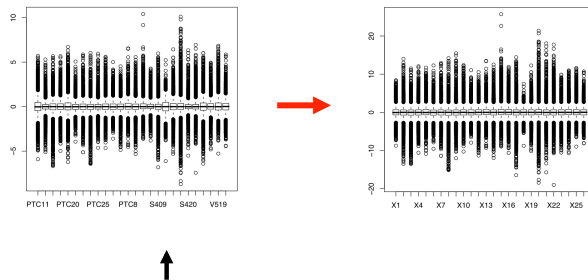
- Spots below or close to background
- Non replicated spots
- Spots inconsistent across replicates



In general: the less replicates you have the more conservative you should be

A multi array normalization example: intensity-based correction

Assumption: Distribution of expression should have the same average intensity and variance across arrays



So, what preprocessing is best?

- Hard to say in the absence of well accepted benchmark data sets!
- A reasonable pipeline is:
 - background correction
 - intensity normalization
 - space normalization
 - dye swap averaging

22

CONCLUSION

Your analysis should not depend on tweaking the preprocessing.

Cross-study, cross-platform, comparison will ultimately be the gold standard, thus

OPEN AND UNRESTRICTED ACCESS TO RAW DATA IS THE KEY.