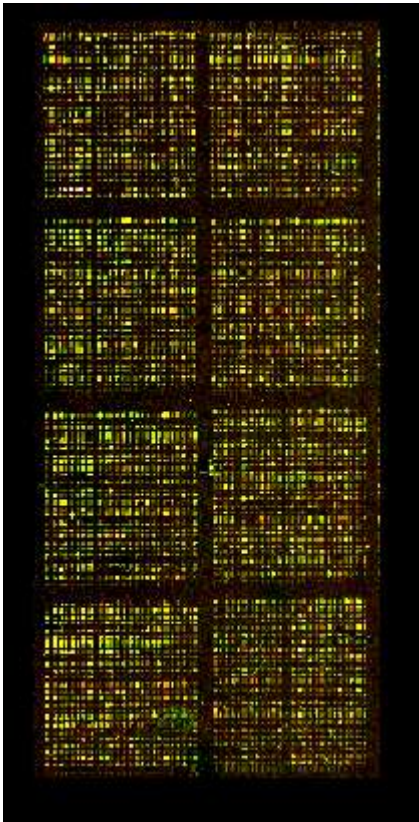
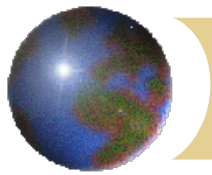


A Latent Variable Model For Estimating Gene Expression Intensities From Multiple Scanned Microarray Chips



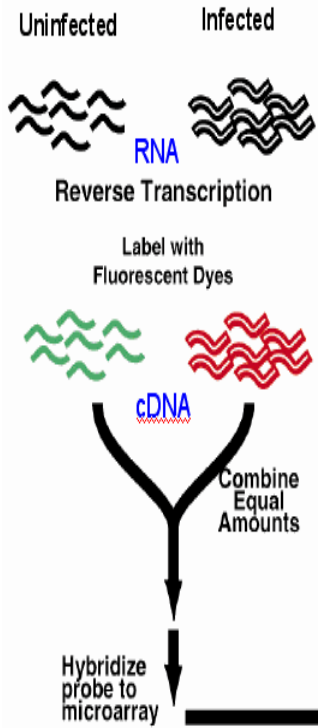
Rashi Gupta

**Department of Mathematics and Statistics
&
Institute of Biotechnology
University of Helsinki**

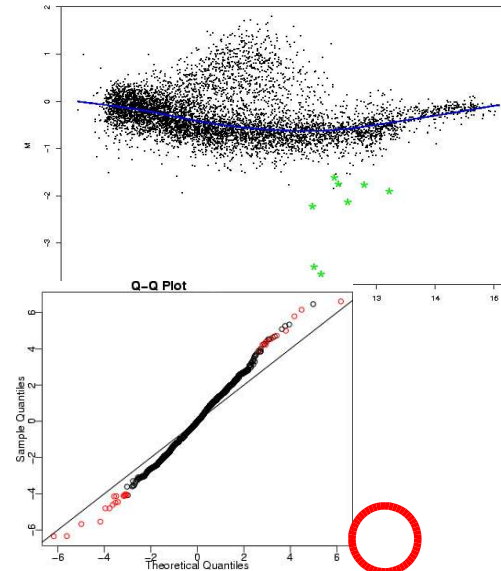
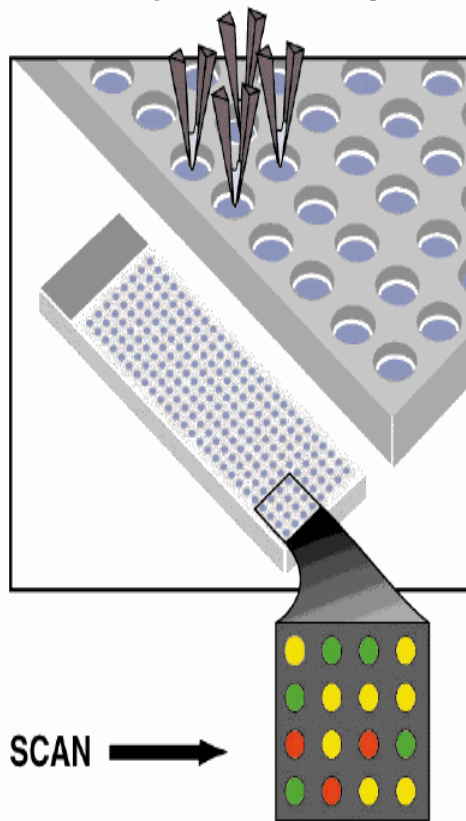


Overview - Microarray Analysis

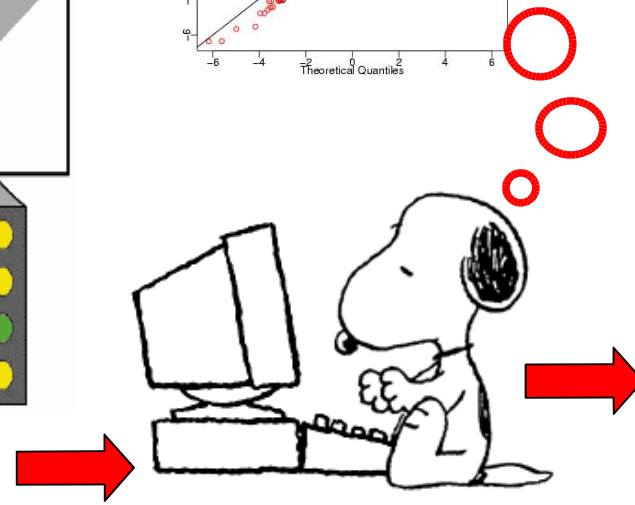
Prepare cDNA Probe



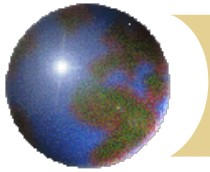
Prepare Microarray



t	ID	Name
-19.83	AA495790	ras homolog gene family
-16.83	AA598794	connective tissue growth factor
-15.22	AA488676	membrane attached signal protein 1
-14.2	AI014487	insulin-like growth factor binding protein 102
-13.62	R77252	microtubule-associated protein 7
-13.6	AA598601	insulin-like growth factor binding protein 31
-13.57	R09561	decay accelerating factor for complement (CD55)
-13.38	AA875933	EGF-containing fibulin-like extracellular matrix protein
-12.79	AA777187	cysteine-rich, angiogenic inducer
-12.63	AA598601	insulin-like growth factor binding protein 3
-12.01	AA055835	caveolin 1, caveolae protein, 22kD
-11.88	AA012944	insulin-like growth factor binding protein 102
-10.86	AA936757	heparin-binding growth factor binding protein
-10.86	AA995282	four and a half LIM domains 2
-10.35	AA677403	glycoprotein hormones, alpha polypeptide
-9.88	AA430032	pituitary tumor-transforming 1
-9.32	AI935290	cysteine and glycine-rich protein 1
-9.18	AA936757	heparin-binding growth factor binding protein2
-9.06	AA424833	bone morphogenetic protein 6
-9.02	AI985398	natriferic peptide receptor C
-8.51	AA630794	solute carrier family 3
-8.38	H29897	phospholipase C, beta 42
-8.22	W72207	cystatin A (stefin A)
-7.99	H45668	Kruppel-like factor 4 (gut)
-7.95	AA600217	activating transcription factor 4
-7.8	AA149095	dual specificity phosphatase 1
-7.68	W73874	cathepsin L
-7.61	R09561	decay accelerating factor for complement
-7.53	AW028846	trefoil factor 2 (spasmolytic protein 1)
-7.16	N66177	microphthalmia-associated transcription factor
-7.14	H03346	protease, serine, 22
-7.06	AA169469	pyruvate dehydrogenase kinase, isoenzyme 4
-6.96	AI989348	protein disulfide isomerase-related protein
-6.94	H63077	annexin A1
-6.92	AA610004	Homo sapiens putative oncogene protein
-6.84	AA59145	ZW10 (Drosophila) homolog
-6.78	AA521434	B-cell CLL/lymphoma 6
-6.77	AA400128	general transcription factor II
-6.68	T53298	insulin-like growth factor binding protein 7
-6.67	T86983	complement component 1
-6.6	AA027240	eukaryotic translation initiation factor 2
-6.57	AA482117	Ras homolog enriched in brain 2
-6.55	AA464849	thioredoxin reductase 1
-6.55	AA400893	phosphodiesterase 1A, calmodulin-dependent
-6.5	R91550	arginine-rich, mutated in early stage tumors
-6.45	AA620433	dihydropyrimidinase-like 3
-6.45	AA625628	accessory proteins BAP31/BAP29
-6.41	T77733	tubulin, gamma 1

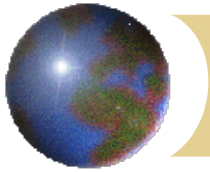


List of differentially expressed genes



Current Situation

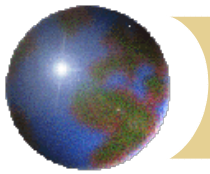
- ❖ Glass slide is scanned once using some setting of scanner so as to capture the entire range of gene expression
- ❖ Data is extracted from the digital image obtained
- ❖ This data is then used for all downstream analysis of microarray



Motivation

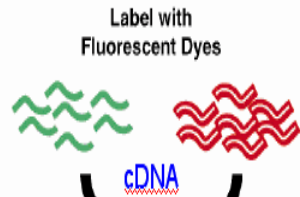
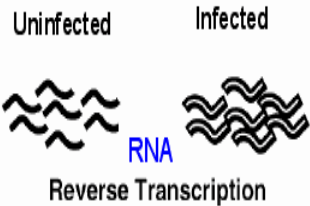
- ❖ Range of gene expression is large (0-)
- ❖ Due to the limitation of acquisition device (2^{16}) its impossible to measure beyond 65536
- ❖ All expression beyond 65536 are saturated
- ❖ Classification, finding gene interacting networks are all based on *saturated data*

Improvement possible?



Overview - Scanning

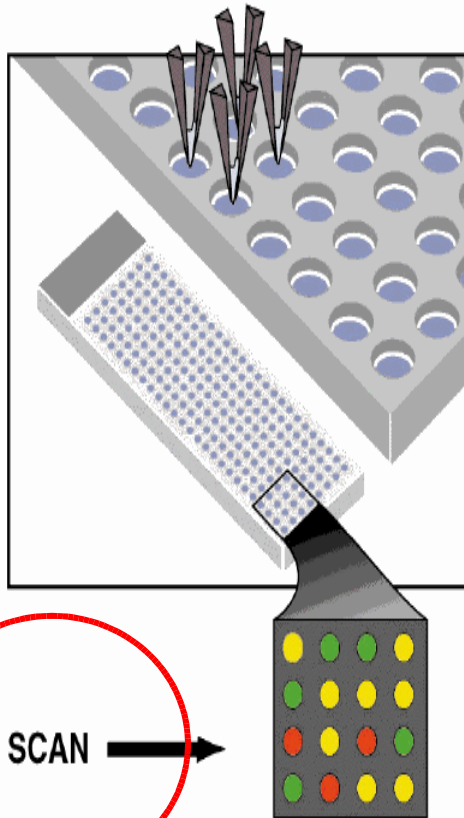
Prepare cDNA Probe



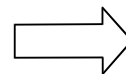
Combine Equal Amounts

Hybridize probe to microarray

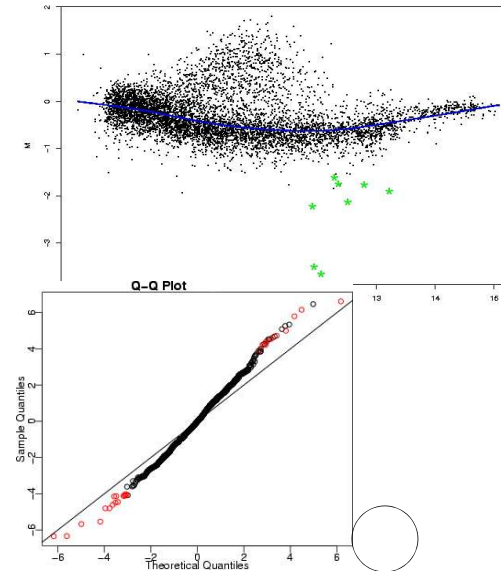
Prepare Microarray



SCAN



List of differentially expressed genes



t ID Name

- 19.83 AA495790 ras homolog gene family
- 16.83 AA598794 connective tissue growth factor
- 15.22 AA488676 membrane attached signal protein 1
- 14.2 AI014487 insulin-like growth factor binding protein 102
- 13.62 R77252 microtubule-associated protein 7
- 13.6 AA598601 insulin-like growth factor binding protein 31
- 13.57 R09561 decay accelerating factor for complement (CD55)
- 13.38 AA875933 EGF-containing fibulin-like extracellular matrix protein
- 12.79 AA777187 cysteine-rich, angiogenic inducer
- 12.63 AA598601 insulin-like growth factor binding protein 3
- 12.01 AA055835 caveolin 1, caveolae protein, 22kD
- 11.88 AA012944 insulin-like growth factor binding protein 102
- 10.86 AA936757 heparin-binding growth factor binding protein
- 10.86 AA995282 four and a half LIM domains 2
- 10.35 AA677403 glycoprotein hormones, alpha polypeptide
- 9.88 AA430032 pituitary tumor-transforming 1
- 9.32 AI935290 cysteine and glycine-rich protein 1
- 9.18 AA936757 heparin-binding growth factor binding protein 2
- 9.06 AA424833 bone morphogenetic protein 6
- 9.02 AI985398 natriuretic peptide receptor C
- 8.51 AA630794 solute carrier family 3
- 8.38 H29897 phospholipase C, beta 42
- 8.22 W72207 cystatin A (stefin A)
- 7.99 H45668 Kruppel-like factor 4 (gut)
- 7.95 AA600217 activating transcription factor 4
- 7.8 AA149095 dual specificity phosphatase 1
- 7.68 W73874 cathepsin L
- 7.61 R09561 decay accelerating factor for complement
- 7.53 AW028846 trefoil factor 2 (spasmolytic protein 1)
- 7.16 N66177 microphthalmia-associated transcription factor
- 7.14 H03346 protease, serine, 22
- 7.06 AA169469 pyruvate dehydrogenase kinase, isoenzyme 4
- 6.96 AI989348 protein disulfide isomerase-related protein annexin A1
- 6.94 H63077
- 6.92 AA610004 Homo sapiens putative oncogene protein
- 6.84 AA59145 ZW10 (Drosophila) homolog
- 6.78 AA521434 B-cell CLL/lymphoma 6
- 6.77 AA400128 general transcription factor II
- 6.68 T53298 insulin-like growth factor binding protein 7
- 6.67 T86983 complement component 1
- 6.6 AA027240 eukaryotic translation initiation factor 2
- 6.57 AA482117 Ras homolog enriched in brain 2
- 6.55 AA464849 thioredoxin reductase 1
- 6.55 AA400893 phosphodiesterase 1A, calmodulin-dependent
- 6.5 R91550 arginine-rich, mutated in early stage tumors
- 6.45 AA620433 dihydropyrimidinase-like 3
- 6.45 AA625628 accessory proteins BAP31/BAP29
- 6.41 T77733 tubulin, gamma 1

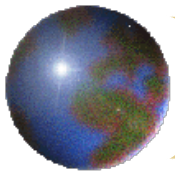
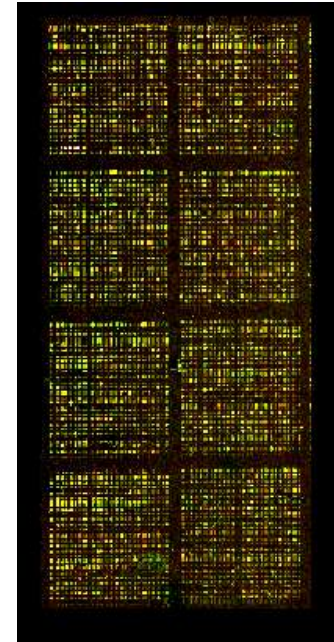
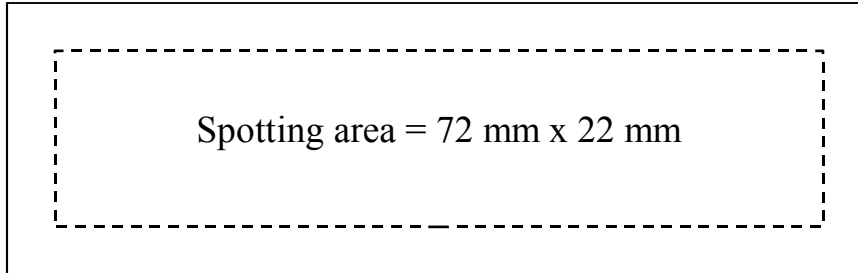


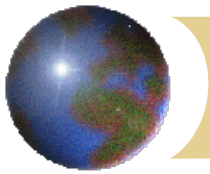
Image acquisition as a result of scanning



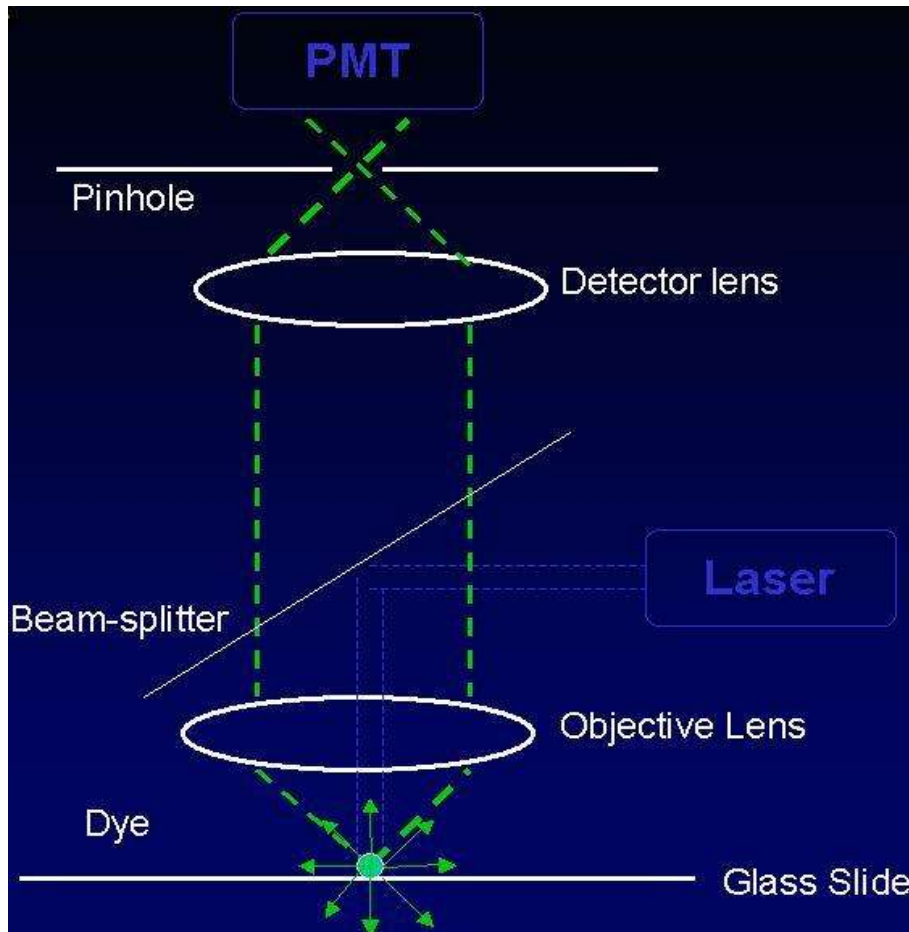
Slide dimension = 76 mm x 25 mm x 1 mm

Standard glass slide

- Image Acquisition is done using laser scanner
- The scanner performs an area scan of a slide and produces a digital map
- Parameters of scanning: *PMT (Photomultiplier Tube), Laser Power*
- PMT and Laser power are adjusted (**by approximation**) so as to capture the whole range of expressions on the chip
- ❖ *Dynamic range of genes is difficult to capture under a single setting of scanner*

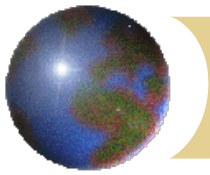


How scanner works



How are the crucial parameters of scanning chosen?

Yang, J. Speed, T. et al. *Image Analysis on cDNA Microarray Data Demo of Spot.*

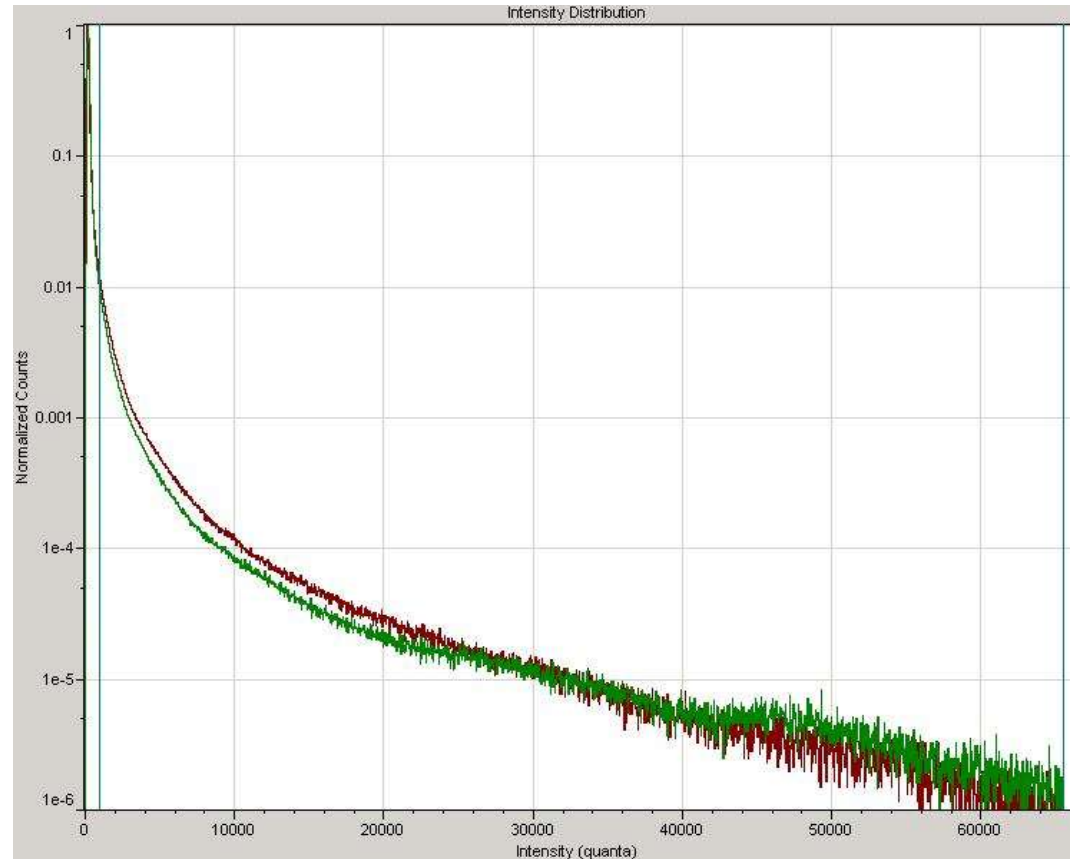


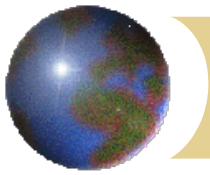
How to choose the scanner parameters

Assumption is that most gene transcripts will not change

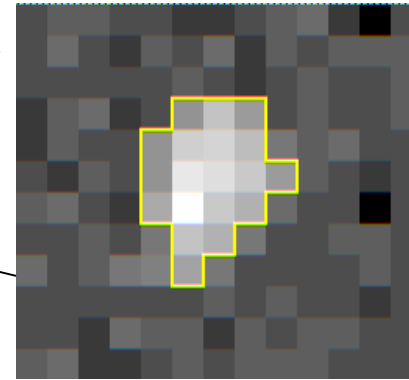
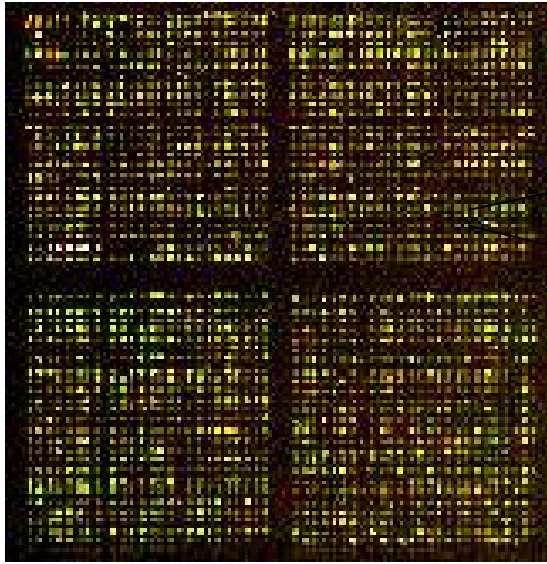
PMT gains and laser power may be adjusted so:

- Pixel intensity frequencies overlap.
- Maximum intensity with minimal saturation.



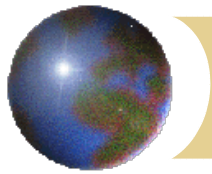


Saturation Aspect



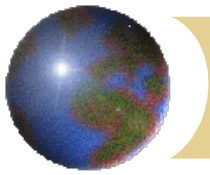
Gene / Spot

- ❖ A spot comprises of pixels
- ❖ Pixels get saturated due to limitation of acquisition device
- ❖ Averaging over pixels is commonly applied to get spot intensity



Saturation Aspect- contd

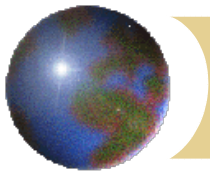
- ❖ Spots with saturated pixels ***may or may not*** be used in the analysis
- ❖ If saturated spots are ***eliminated*** we loose many genes from analysis
- ❖ If saturated spots are ***included*** then they affect all the downstream analysis of microarray



***Aim: To Handle Saturation, Capture
the Dynamic Range and Estimate the
Expression of Genes***

Phase 1- A system under which repeated scans are made to capture the expression of all genes

Phase 2- A Bayesian (Probability) model that integrates and estimates the underlying expression of genes from data extracted from different scans



Acknowledgements

Elja Arjas

Department of Mathematics and Statistics

Petri Auvinen

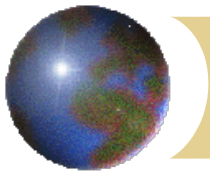
Institute of Biotechnology

Sangita Kulathinal

Department of Epidemiology and Health Promotion

Andrew Thomas

Department of Mathematics and Statistics



References

Dudley AM, Aach J, Steffen MA, Church GM: Measuring absolute expression with microarrays with a calibrated reference sample and an extended signal intensity range. *Proc Natl Acad Sci USA* 2002, 99:7554-7559

Hsiao LL, Jensen RV, Yoshida T, Clark KE, Blumenstock JE, Gullans SR: Correcting for signal saturation errors in the analysis of microarray data. *BioTechniques* 2002, 32:330-336

Schermer MJ: Confocal scanning microscopy in microarray detection. In *DNA microarrays – a practical approach Volume 1*. Edited by: Schena M. Oxford: Oxford University Press; 1999:17-42

Lyng.H, Badiee.A, Svendsrud.D.H,Hovig.E, Myklebost.O and Stokke.T: Profound influence of microarray scanner characteristics on gene expression ratios: analysis and procedure for correction, *BMC Genomics* 2004, 5:10