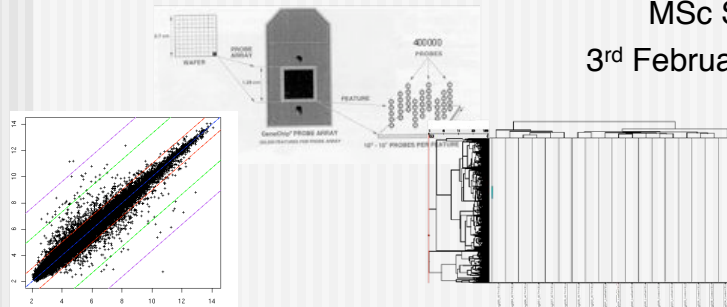
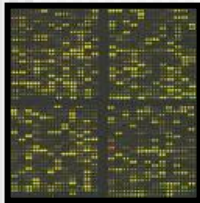


Microarray Informatics

Donald Dunbar

MSc Seminar

3rd February 2010



Aims

- To give a biologist's view of microarray experiments
- To explain some technologies involved
- To describe typical microarray experiments
- To show how to get the most from an experiment
- To show where the field is going

Introduction

- Part 1
 - Microarrays in biological research
 - A typical microarray experiment
 - Experiment design, data pre-processing
- Part 2
 - Data analysis and mining
 - Microarray standards and resources
 - Recent advances

February 3rd 2010

MSc Seminar: Donald Dunbar

Microarray Informatics

Part 1

February 3rd 2010

MSc Seminar: Donald Dunbar

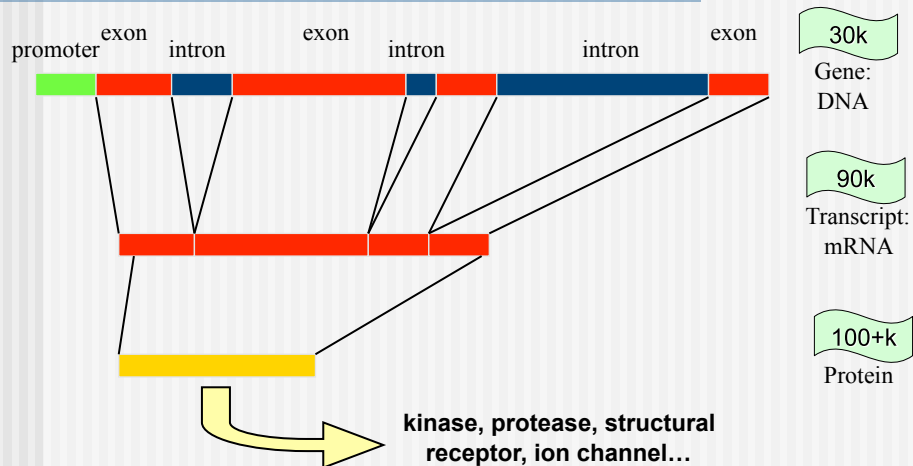
Biological research

- Using a wide range of experimental and computational methods to answer biological questions
- Genetics, physiology, molecular biology...
- Biology and informatics → bioinformatics
- Genomic revolution
- What can we measure?

February 3rd 2010

MSc Seminar: Donald Dunbar

The central dogma

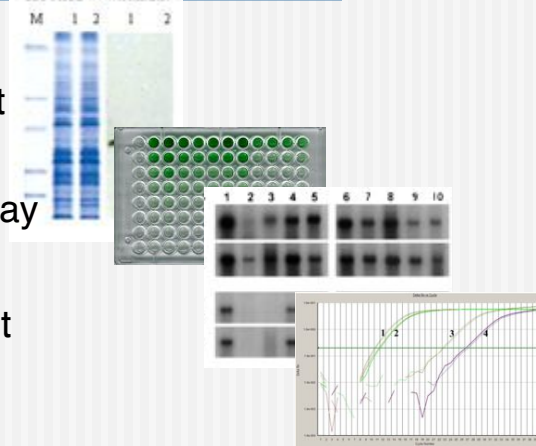


February 3rd 2010

MSc Seminar: Donald Dunbar

Measuring RNA and proteins

- Proteins
 - Western blot
 - ELISA
 - Enzyme assay
- mRNA
 - Northern blot
 - RT-PCR

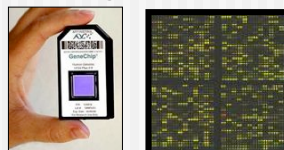


February 3rd 2010

MSc Seminar: Donald Dunbar

Measuring RNA and proteins

- Protein levels/activities would be best
 - no real high throughput method
- mRNA levels will have to do
 - genome-wide physical microarrays
 - other 'array-like' technologies
 - sequencing (see later)

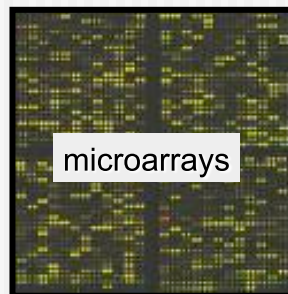


February 3rd 2010

MSc Seminar: Donald Dunbar

Measuring transcripts

- Genome level sequencing
- New miniaturisation technologies
- Better bioinformatics

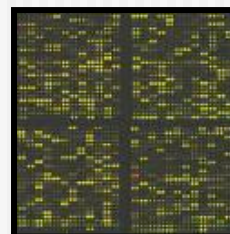


February 3rd 2010

MSc Seminar: Donald Dunbar

Microarrays: wish list

- Include all genes in the genome
- Include all splice variants
- Give reliable estimates of expression
- Easy to analyse
 - bioinformatics tools available
- Cost effective



February 3rd 2010

MSc Seminar: Donald Dunbar

Microarray technologies - 1



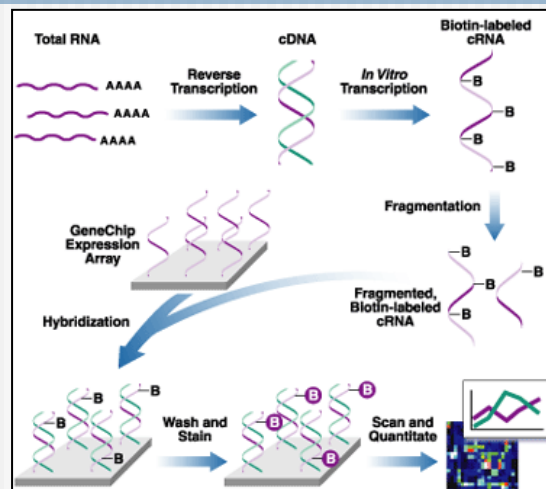
- Oligonucleotides - Affymetrix
- One chip all genes
- Chips for many species
- Several oligos per transcript
- Use of control, mismatch sequences
- One sample per chip
 - 'absolute quantification'
- Well established in research
- Expensive



February 3rd 2010

MSc Seminar: Donald Dunbar

Microarray technologies - 1



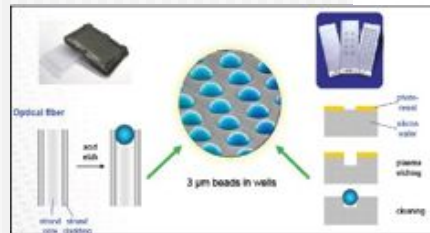
February 3rd 2010

MSc Seminar: Donald Dunbar

Microarray technologies - 2



- Illumina BeadChip
- Oligos on beads
- Hybridise in wells
- Compared to Affy
 - Higher throughput
 - Less RNA needed
 - Cheaper



February 3rd 2010

MSc Seminar: Donald Dunbar

Problems with microarrays

- The gene might not be on the chip
- Can't differentiate splice variants
- The gene might be below detection limit
- Can't differentiate RNA synthesis and degradation
- Can't tell us about post translational events
- Bioinformatics can be difficult
- Relatively expensive

February 3rd 2010

MSc Seminar: Donald Dunbar

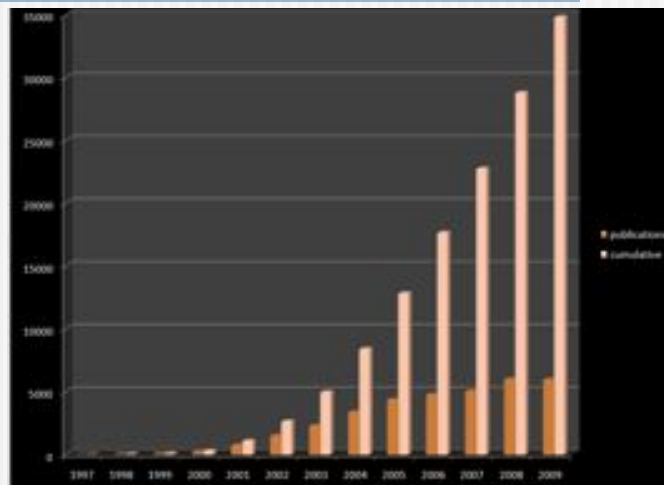
History of Microarrays

- Developed in early 1990s after larger macro-arrays (100-1000 genes)
- Microarrays were spotted on glass slides
- Labs spotted their own (Southern, Brown)
- Then companies started (Affymetrix, Agilent)
- Some early papers:
 - *Nature* 1993 364(6437): 555-6 Multiplexed biochemical assays with biological chips. Fodor SP, et al
 - *Science* 1995 Oct 20;270(5235):467-70 Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Schena M, et al

February 3rd 2010

MSc Seminar: Donald Dunbar

Microarray publications



February 3rd 2010

MSc Seminar: Donald Dunbar

Types of experiment

- Usually **control v test(s)**

Placebo

Wild-type

Healthy

Normal tissue

Time = 0

Drug treatment

Drug 2...

Knockout

Patient

Cancerous tissue

Time = 1

Time = 2...

February 3rd 2010

MSc Seminar: Donald Dunbar

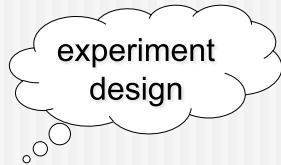
Types of experiment

- Usually **control v test(s)**
- But also **test v test(s)**
- Comparison:
 - placebo v drug treatment
 - drug 1 v drug 2
 - tissue 1 v tissue 2 v tissue 3 (pairwise)
 - time 0 v time 1, time 0 v time 2, time 0 v time 3
 - time 0 v time 1, time 1 v time 2, time 2 v time 3

February 3rd 2010

MSc Seminar: Donald Dunbar

A typical experiment



February 3rd 2010

MSc Seminar: Donald Dunbar

Experiment design: system

- What is your model?
 - animal, cell, tissue, drug, time...
- What comparison?
- What platform
 - microarray? oligo, cDNA?
- Record all information: see “standards”

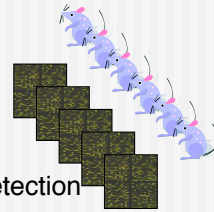


February 3rd 2010

MSc Seminar: Donald Dunbar

Experiment design: replicates

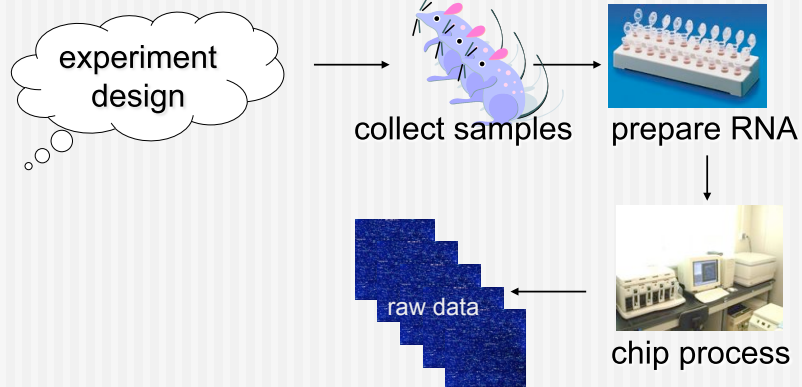
- Microarrays are noisy: need extra confidence in the measurements
- We usually don't want to know about a specific individual
 - eg not an individual mouse, but the strain
 - although sometimes we do (eg people)
- Biological replicates needed
 - independent biological samples
 - number depends on variability and required detection
- Technical replicates (same sample, different chip) usually not needed



February 3rd 2010

MSc Seminar: Donald Dunbar

A typical experiment



February 3rd 2010

MSc Seminar: Donald Dunbar

Raw data

- Affymetrix GeneChip process generates:
 - DAT image file
 - CEL raw data file
 - CDF chip definition file
- Processing then involves CEL and CDF
- Will use Bioconductor



February 3rd 2010

MSc Seminar: Donald Dunbar

Bioconductor (BioC)



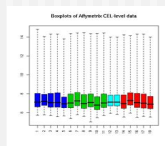
- <http://www.bioconductor.org/>
- “Bioconductor is an open source software project for the analysis and comprehension of genomic data”
- Started 2001, developed by expert volunteers
- Built on statistical programming environment “R”
- Provides a wide range of powerful statistical and graphical tools
- Use BioC for most microarray processing and analysis
- Most platforms now have BioC packages
- Tutorial: manuals.bioinformatics.ucr.edu/home/R_BioCondManual

February 3rd 2010

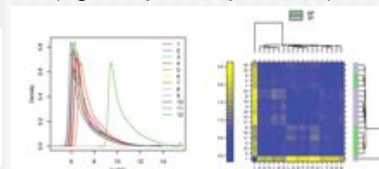
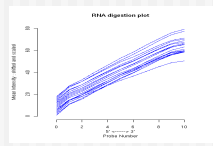
MSc Seminar: Donald Dunbar

Quality control (QC)

- Affymetrix gives data on QC
 - the microarray team will record these for you
 - scaling factor, % present, spiked probes, internal controls
- Bioconductor offers:
 - boxplots and histograms of raw and normalised data
 - RNA degradation plots
 - specialised quality control routines (eg arrayQualityMetrics)



February 3rd 2010



MSc Seminar: Donald Dunbar

Pre-processing: background

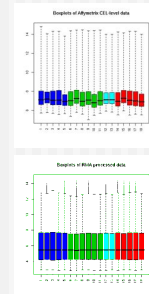
- Signal corresponds to expression...
 - plus a non-specific component (noise)
- Non specific binding of labelled target
- Need to exclude this background
- Several methods exist
 - eg Affy: PM-MM but many complications
 - eg RMA $PM=B+S$ (don't use MM)

February 3rd 2010

MSc Seminar: Donald Dunbar

Pre-processing: normalisation

- In addition to background corrections
 - chip, probe, spatial, intra and inter array variation
 - need to remove to get at real experimental differences
- Make use of statistical methods
 - combined with probe set summary: get an expression value for the gene
- But seems to be no dependency on intensity
 - additive and multiplicative errors
- Quantile normalisation often used
- Normalisation is complicated for 2-colour arrays
- Try to reduce most noise at lab stage (ie control things well statistically)

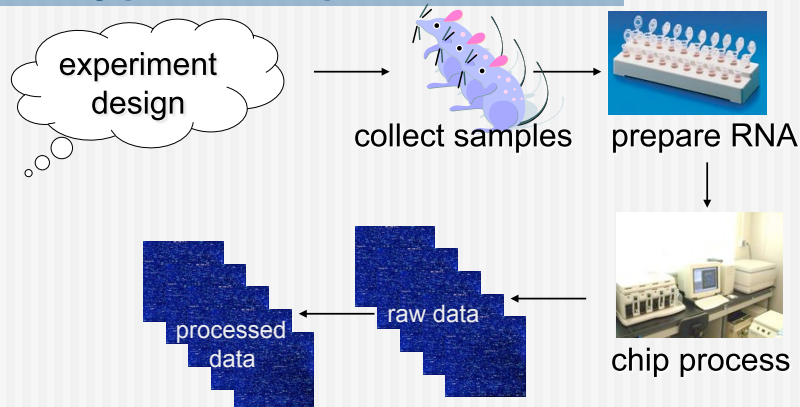


Now carry on with analysis!

February 3rd 2010

MSc Seminar: Donald Dunbar

A typical experiment



February 3rd 2010

MSc Seminar: Donald Dunbar

Part 1 Summary

- Microarrays in biological research
- Two types of microarray
- A typical microarray experiment
- Experiment design
- Data pre-processing

February 3rd 2010

MSc Seminar: Donald Dunbar

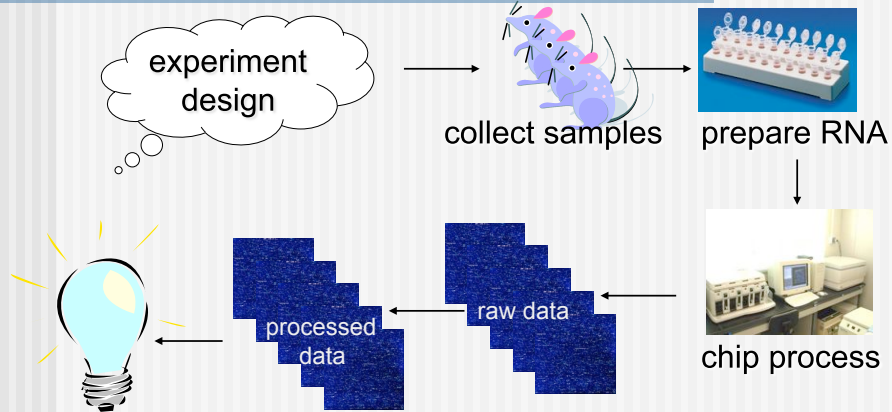
Microarray Informatics

Part 2

February 3rd 2010

MSc Seminar: Donald Dunbar

A typical experiment

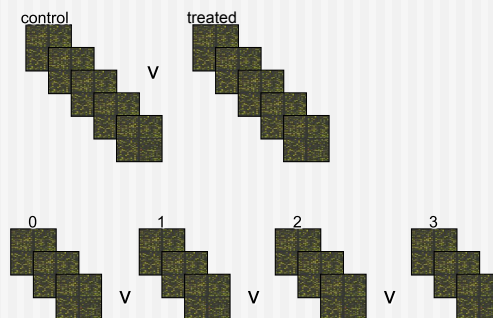


February 3rd 2010

MSc Seminar: Donald Dunbar

Data analysis

- Identifying differential expression
- Compare control and test(s)
 - t-test
 - ANOVA
 - SAM (FDR)
 - Limma
 - Rank Products
- Time series



February 3rd 2010

MSc Seminar: Donald Dunbar

Multiple testing

- Problem:
 - statistical testing of 30,000 genes
 - at $\alpha = 0.05 \rightarrow 1500$ genes
- Need to correct this
 - Multiply p-value by number of observations
 - Bonferroni, too conservative
 - False discovery
 - defines a q value: expected false positive rate
 - Less conservative, but higher chance of type I error
 - Benjamini and Hochberg
- Then regard genes as differentially expressed
- Depends on follow-up procedure!

February 3rd 2010

MSc Seminar: Donald Dunbar

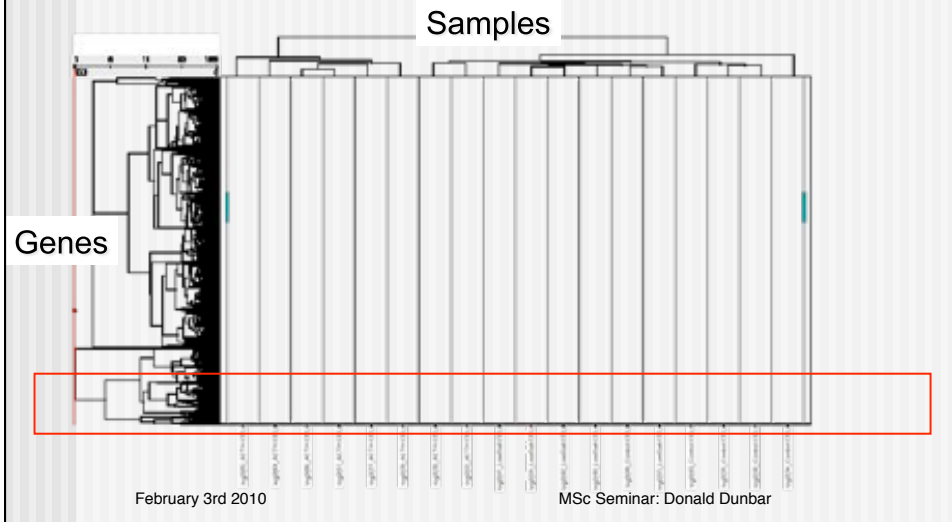
Hierarchical clustering

- Look for structure within dataset
 - similarities between genes
- Compare gene expression profiles
 - Euclidian distance
 - Correlation
 - Cosine correlation
- Calculate with distance matrix
- Combine closest, recalculate, combine closest... (or split!)
- Draw dendrogram and heatmap

February 3rd 2010

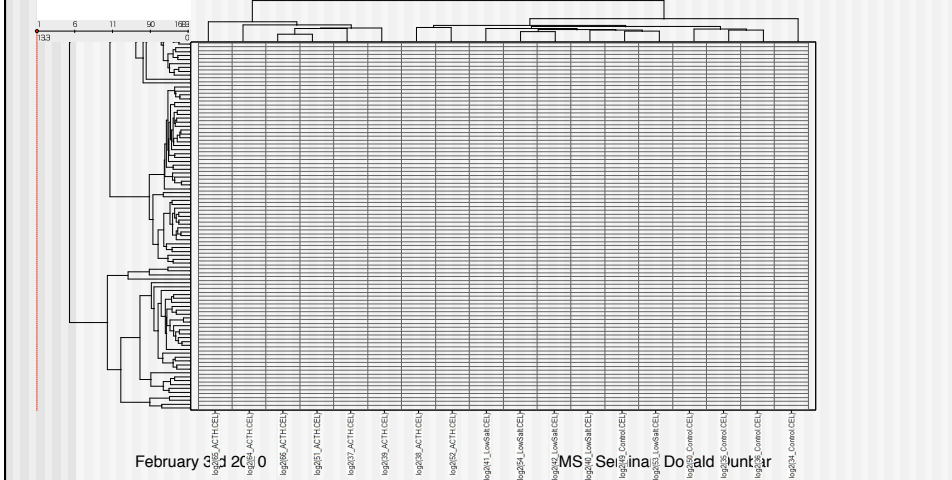
MSc Seminar: Donald Dunbar

Hierarchical clustering



Hierarchical clustering

■ Heatmaps for microarray data



Hierarchical clustering

- Predicting association of known and novel genes
- Class discovery in samples: new subtypes
- Visualising structure in data (sample outliers)
- Classifying groups of genes
- Identifying trends and rhythms in gene expression
- Caveat: you will always see clusters, even when they are not particularly meaningful (nb Ian Simpson)

February 3rd 2010

MSc Seminar: Donald Dunbar

Sample classification

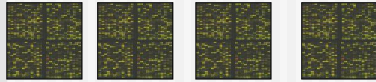
- Supervised or non-supervised
- Non-supervised
 - like hierarchical clustering of samples
- Supervised
 - have training (*known*) and test (*unknown*) datasets
 - use training sets to define robust classifier
 - apply to test set to classify new samples

February 3rd 2010

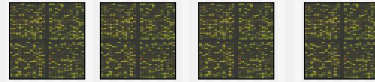
MSc Seminar: Donald Dunbar

Sample classification

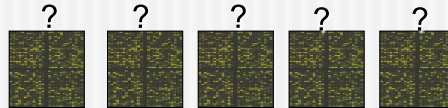
good prognosis
→ drug treatment



bad prognosis
→ surgery



Gene selection, training, cross validation →
classifier: gene x * 0.5 gene y * 0.25 gene z ...

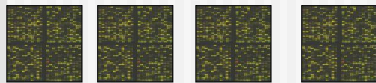


February 3rd 2010

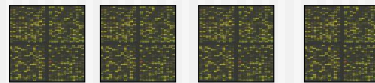
MSc Seminar: Donald Dunbar

Sample classification

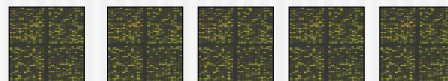
good prognosis
→ drug treatment



bad prognosis
→ surgery



Apply classifier



February 3rd 2010

MSc Seminar: Donald Dunbar

Sample classification

- Class prediction for new samples
 - cancer prognosis
 - pharmacogenomics (predict drug efficacy)
- Need to watch for overfitting
 - using too much of the data to classify
 - classifier loses specificity

February 3rd 2010

MSc Seminar: Donald Dunbar

Annotation

- Big problem for microarrays
- Genome-wide chips need genome-wide annotation
- Good bioinformatics essential
 - use several resources (Affymetrix, Ensembl)
 - keep up to date (as annotation changes)
 - genes have many attributes
 - name, symbol, gene ontology, pathway...

February 3rd 2010

MSc Seminar: Donald Dunbar

Data-mining

Microarrays are a waste
of time
...unless you do
something with the data

February 3rd 2010

MSc Seminar: Donald Dunbar

Data-mining

- Once data are statistically analysed:
 - pull out genes of interest
 - pull out pathways of interest
 - mine data based on annotation
 - what are the expression patterns of these genes
 - what are the expression patterns in this pathway
 - mine genes based on expression pattern
 - what types of genes are up-regulated ...
 - fold change, p-value, expression level, correlation
- **Should be driven by the biological question**

February 3rd 2010

MSc Seminar: Donald Dunbar

Input a query or leave blank (bad idea: lots of data!) for all data, then submit...

Annotation Queries

Affymetrix ID

Entrez Gene ID

Gene Title

Gene Symbol

Gene Ontology Term

Pathway

Chromosome

Groups

In or Out

Comments

Comments

or

Expression Queries

Genes that are "not expressed at all" are hidden. If that's fine, leave @yes if not, click @no.

	1	2	3	4	5	6	7	Filter
BAT	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
WAT	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>
Liver	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="checkbox"/>

Expression maxima and minima

BAT max/min < >

BAT max - min

WAT max/min

WAT max - min

Liver max/min

Liver max - min

Correlation with circadian gene profiles

Which gene? Tissue Rank limit

Order

Order output by and

or

[Home](#)

February 3rd 2010

MSc Seminar: Donald Dunbar

Affy ID	Entrez Gene ID	Gene title	Gene Symbol	Intensities														Group	Comments [Update]
				1A	2A	3A	4A	5A	6A	7A	1B	2B	3B	4B	5B	6B	7B		
144851_at	18026	period homolog 1 (Drosophila)	Per1 and GHOP	BAT 333	1081	1466	255	328	102	32	378	150	873	519	855	178	<input type="checkbox"/> In <input type="checkbox"/> Out	Details	
				WAT 288	786	1981	1391	1127	974	412	612	595.1	1422	1280	568	851	767	Details	
				Liver 27	120	894	1941	661	484	345	32	70	724	1040	732	729	22	Details	
146062_at	18628	period homolog 3 (Drosophila)	Per3 and GHOP	BAT 262	227	1026	1689	210	383	713	213	810	1052	1419	845	146	14	<input type="checkbox"/> In <input type="checkbox"/> Out	Details
				WAT 277	422	1153	1960	762	679	186	272	457.3	1090	1099	1054	178	281	Details	
				Liver 181	177	574	740	281	145	109	59	600	476	627	201	613	244	Details	
1434728_at	217082	hepatic leukemia factor	Hlf and GHOP	BAT 1550	2080	1322	4836	1723	2240	1254	1692	1826	2962	5122	1760	2172	1210	<input type="checkbox"/> In <input type="checkbox"/> Out	Details
				WAT 1028	1142	2039	3374	2794	1550	1317	1262	1012.6	2226	3409	2323	1958	946	Details	
				Liver 2115	1896	4029	4623	1504	3166	2083	2354	3268	4769	8126	3787	3221	2184	Details	
1450184_s_at	21088	hematopoietic embryonic factor	Tef and GHOP	BAT 680	1260	2240	2605	1056	843	829	300	1153	1074	2633	1208	616	668	<input type="checkbox"/> In <input type="checkbox"/> Out	Details
				WAT 489	690	450	1810	746	550	254	374	502	393	974	762	466	294	Details	
				Liver 191	872	1529	1671	673	472	38	310	870	1521	1509	1111	492	493	Details	
1422997_s_at	26897	chronic acyl-CoA thioesterase 1	Crot and GHOP	BAT 1493	2117	8058	10100	1774	1921	1849	1704	3137	8648	9166	2804	1893	1774	<input type="checkbox"/> In <input type="checkbox"/> Out	Details
				WAT 314	449	352	536	786	800	620	480	179.7	874	779	603	853	650	Details	
				Liver 1830	1676	2133	2410	1828	2942	1383	1132	1418	1566	1573	1918	2516	2478	Details	
1454786_at	221736	RIFKIN, cDNA (501439607) gene	501439607 and GHOP	BAT 356	301	899	1484	637	271	246	304	872	972	1146	920	618	258	<input type="checkbox"/> In <input type="checkbox"/> Out	Details
				WAT 1381	401	1113	965	123	769	404	460	849	1164	103	393	550	428	Details	
				Liver 169	132	618	191	236	126	56	67	337	626	154	310	150	117	Details	
1428306_at	74747	DNA damage-inducible transcript 4	Ddit4 and GHOP	BAT 1045	1072	1761	2483	944	956	679	835	1024	1245	2950	713	1372	1339	<input type="checkbox"/> In <input type="checkbox"/> Out	Details
				WAT 2667	2768	6696	3774	2748	2720	2816	2919	3069.6	7661	3532	2342	2850	1096	Details	
				Liver 349	678	685	873	1166	1317	1192	124	694	559	809	1261	1408	1249	Details	
1452071_at	54403	solute carrier family 4 (anion exchanger) member 3	Sclt4 and GHOP	BAT 5086	2292	9678	12544	6827	4035	3147	5317	7184	10339	12090	1221	1076	3195	<input type="checkbox"/> In <input type="checkbox"/> Out	Details
				WAT 1107	1313	1566	2149	1761	1148	1525	1139	1406.8	1294	2039	1714	1261	1542	Details	
				Liver 1375	1183	2492	2381	1094	891	1728	2532	3051	3064	1935	1449	1802	2206	Details	
		scd5lnc10F.LAO	Ccbl1 and	BAT 14	396	1218	123	146	120	17	50	316	117	100	105	116	116	<input type="checkbox"/> In <input type="checkbox"/> Out	Details

February 3rd 2010

MSc Seminar: Donald Dunbar

Detailed information: 1449851_at

Annotation Data for Per1

Affymetrix ID	1449851_at
Entrez Gene ID	18626
Gene Title	period homolog 1 (Drosophila)
Gene Symbol	Per1
Ensembl ID	ENSMUTSG00000020893
GO Biological Process information	"6355 // regulation of transcription, DNA-dependent // inferred from electronic annotation // 7623 // rhythmic behavior // inferred from electronic annotation // 7622 // circadian rhy // inferred from electronic annot"
GO Molecular Function information	4871 // signal transducer activity // inferred from electronic annotation
GO Cellular Compartment information	5634 // nucleus // inferred from direct assay
Pathway	Circadian_Exercise // Circadian_Exercise
Chromosomal location	---
Protein family	---
Protein domain	---
Interpro ID	---
Transmembrane?	---
Group	im
Comments	No comments yet

Expression Data for Per1

Intensities

View these data as a [graph](#)

(Per1 meets the general expression criteria)

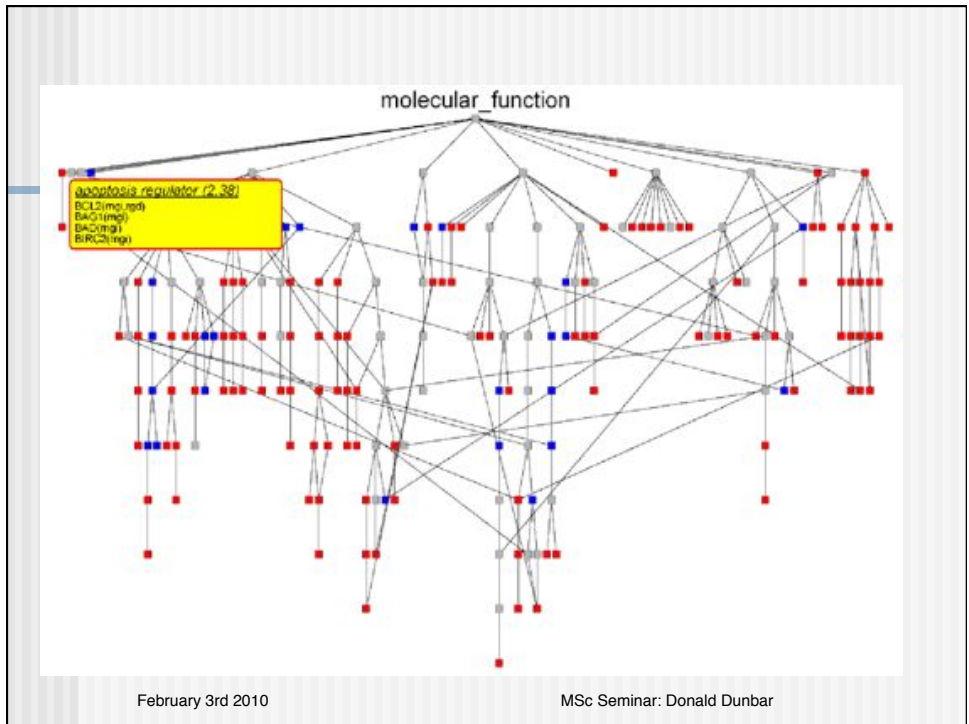
	1A	2A	3A	4A	5A	6A	7A	1B	2B	3B	4B	5B	6B	7B
BAT	413	331	1035	1466	255	328	102	32	475	556	853	539	655	178
WAT	795	786	1983	1371	1127	974	412	612	595.1	1422	1260	868	851	767
Liver	27	120	893	1041	661	444	345	32	70	723	1040	432	729	22

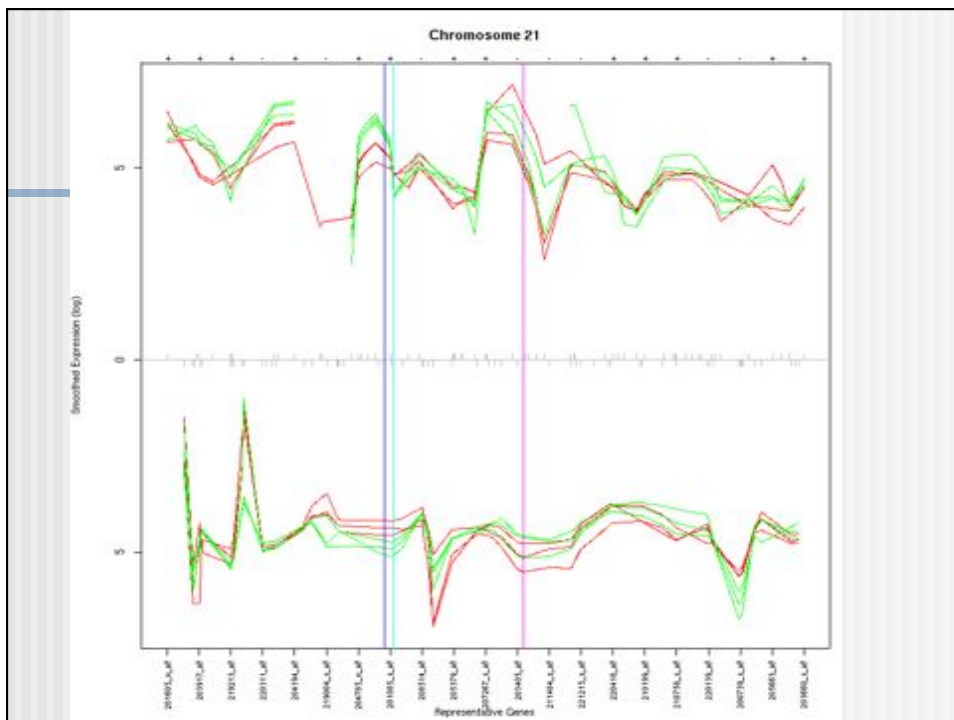
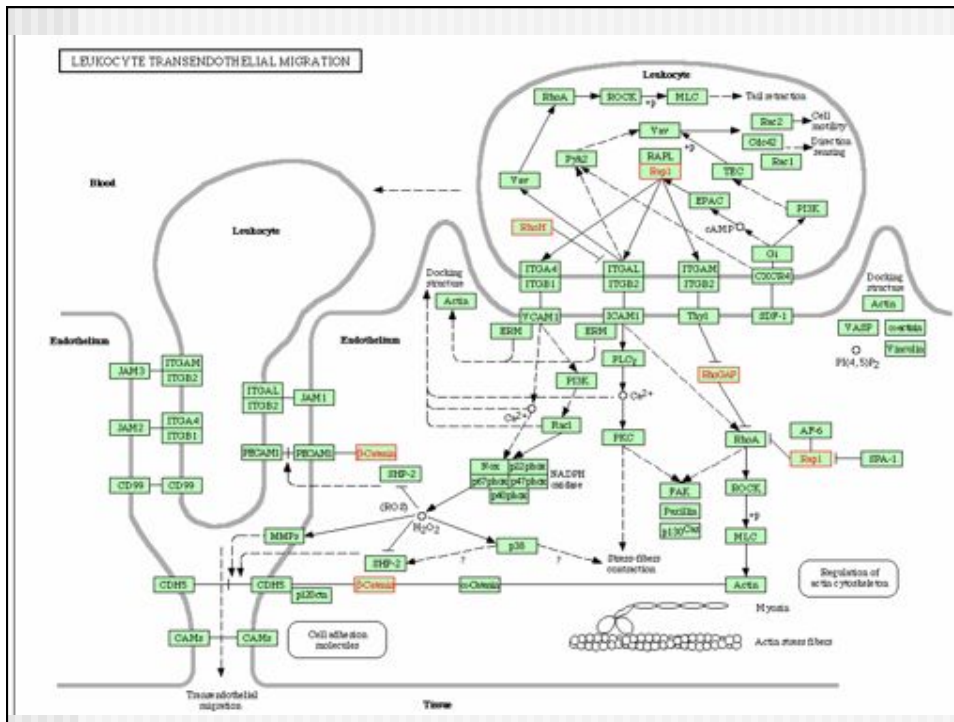
Maxima, minima, ratios, differences




	Maximum	Minimum	Max/Min	Max-Min
BAT	1466	102	14.47	1364
WAT	1983	412	4.81	1571
Liver	1041	27	38.56	1014

February 3rd 2010

MSc Seminar: Donald Dunbar





Launch New

[Cite](#)

[FAQs](#)

[Free files](#)

[Service mirrors](#)

Launch instructions

[News](#)

[Version](#)

[Mailing](#)

[Screenshots](#)

Tutorials

[Source](#)

[SOAP](#)

[Features](#)

[Acknowledgements](#)

Manual

[License](#)

[Related](#)

[FAQ](#)

Introduction

TOUCAN is a workflow for **regulatory sequence analysis on metazoan genomes**: comparative genomics, detection of significant transcription factor binding sites, and detection of cis-regulatory modules (combinations of binding sites) in sets of coexpressed/complicated genes.

It is a platform independent, standalone Java application that is tightly linked with [Ensembl](#), and was built using the [BioJava](#) package. SOAP web services are used to remotely access [multiple algorithms](#) for comparative genomics, motif detection, and module detection.

Comments, suggestions, and bug reports are welcome. Please contact the developers.

Register

If you're using TOUCAN, please register your account.

Launch TOUCAN v. 2.2.5

To run TOUCAN you need to have **two things** installed:

- Java 2 Platform, Standard Edition (J2SE), version 1.4.x. or 1.5.0
- Java Web Start which is shipped as part of J2SE. From J2SE 1.4.2 onwards it is installed together with the SDK/JRE ([more info](#))

If you have fulfilled these requirements, then you can launch TOUCAN directly using this URL: <http://www.esat.kuleuven.ac.be/~sarts/software/toucan.jsp>

Alternatively you can type this command in a terminal window:
 javaws <http://www.esat.kuleuven.ac.be/~sarts/software/toucan.jsp>

We try to encourage to use Java Web Start because this way you will always have the latest version of the software. This is important because the properties change at least once a month to follow the newest Ensembl release. If you are really unable to use Java Web Start, you can send us an email, and we can give you the JAR file of TOUCAN.

Gene set
Enriched regulatory sequences
Functional significance?

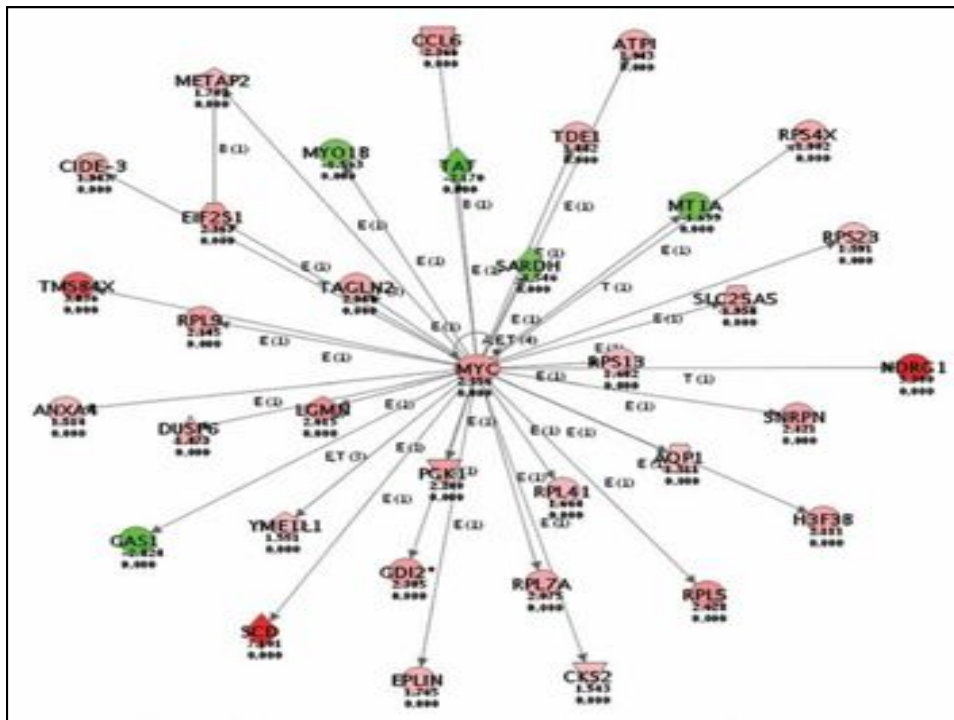
February 3rd 2010 MSc Seminar: Donald Dunbar

PubMatrix Results 13th September 2005

Genes (ACTH differentially expressed) + Terms for **ENRAGE**

PubMatrix	adrenal hypertension blood pressure	ACTH low salt low sodium	hypertrophy	hypertrophy (cardiac)	retention
(zinc finger and metalloprotease domain 8 OR Adad8)	0	0	0	0	0
(zinc finger domain containing 2 OR Abhd2)	0	0	0	0	0
(activating transcription factor 3 OR ATF3)	0	0	0	0	0
(acyl CoA oxidase long chain OR Aco3)	11	1	1	4	0
(actin OR Act)	22	18	11	1	7
(adrenomedullary lipase 1 OR Adhl1)	0	0	0	0	0
(adrenomedullary lipase 2 OR Adhl2)	0	0	0	0	0
(adrenomedullary lipase 3 OR Adhl3)	0	0	0	0	0
(adrenomedullary lipase 4 OR Adhl4)	0	0	0	0	0
(adrenomedullary lipase 5 OR Adhl5)	0	0	0	0	0
(alkaline phosphatase 1, liver OR Alpl)	22	23	15	9	28
(arginine vasopressin induced 1 OR Arvip1)	2	10	24	7	0
(arteriole permeability-increasing protein-like 2 OR Apil2)	0	0	0	0	0
(basic leucine zipper and W2 domain 1 OR Bcl1)	0	0	0	0	0
(BMP binding endothelial regulator OR MGL1820489)	0	0	0	0	0
(branched chain aminotransferase 1, cytosolic OR Bcat1)	1	0	0	0	0
(CD8 antigen-like OR Cd8)	0	0	0	0	0
(CDC28 protein kinase regulatory subunit 2 OR Cks2)	0	0	0	0	0
(CLA related cell adhesion molecule 2 OR Ccrand2)	1	0	0	0	0
(cell division cycle 2 homolog OR Cdc2)	0	0	0	0	0
(cristin, gamma 2 OR Crag2)	0	0	0	0	0
(cristin 3-like 3 OR Crl3)	0	0	0	0	0
(cyclic-dependent kinase inhibitor 1C OR Cdkic1)	11	2	0	2	0
(cystatin B OR CstB)	1	4	2	0	0
(cystostatin alpha OR CstA2)	101	11	120	18	17
(DNA (cytosine-5-) methyltransferase 3-like OR Dnm3)	0	0	0	0	0
(down regulated by C/EBP beta OR MGS14909)	0	0	0	0	0
(dystrorovin alpha OR Dna)	0	0	0	0	0
(embigin OR Emb)	1	2	2	1	0
(fatty acid desaturase 2 OR Fads2)	0	0	0	0	0
(fibronectin type 3 domain containing protein 1 OR Fbn3)	1	0	0	0	0
(galactosylase, alpha OR Gal)	18	18	18	18	18

February 3rd 2010 MSc Seminar: Donald Dunbar




Further data-mining

- Other tools available using
 - gene ontology (GO)
 - biological pathways (eg KEGG)
 - genomic localisation (Ensembl)
 - regulatory sequence data (Toucan, BioProspector)
 - literature (eg Pubmatrix, Ingenuity...)
- ... to make sense of the data
- Links at: www.bioinf.mvm.ed.ac.uk/projects/analysis_tools.html

Microarray Resources

- Microarray data repositories

- Array express (EBI, UK) 

- Gene Expression Omnibus (NCBI, USA)



- CIBEX (Japan) 

- Annotation

- NetAffx, Ensembl, TIGR, Stanford...

February 3rd 2010

MSc Seminar: Donald Dunbar

Microarray Standards

- MIAME

- Minimum annotation about a microarray experiment

- Comprehensive description of experiment

- Models experiments well, and allows replication

- chips, samples, treatments, settings, comparisons

- Required for most publications now

- MAGE-ML

- Microarray gene expression markup language

- Describes experiment (MIAME) and data

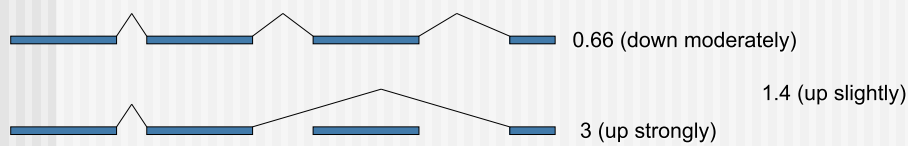
- Tools available for processing

February 3rd 2010

MSc Seminar: Donald Dunbar

Recent advances: Exon chips

- Affymetrix now have chips that allow us to measure expression of splice variants



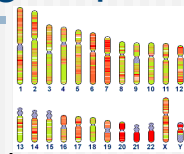
New chips will give us much more information

February 3rd 2010

MSc Seminar: Donald Dunbar

Recent advances: Genotyping chips

- All discussion on EXPRESSION chips
- Also can get chips looking at genotype
- Tell us the sequence for genome-wide markers
- Test 300,000 markers with one chip
- Look for association with disease, prognosis, trait...
- Combined with expression chips to generate
 - EXPRESSION QUANTITATIVE TRAIT LOCUS (eQTL)
 - Overlap of expression and genetic differences (cis)
 - Correlation at different locus (trans)



February 3rd 2010

MSc Seminar: Donald Dunbar

Next Generation Sequencing

- Sequence rather than hybridisation
- Gene expression, genotyping, epigenetics
- New technologies: much cheaper than before
- Gene expression, genotyping, epigenetics
- Open ended (no previous knowledge required)
- Will take over in 2 years: the end of microarrays?

February 3rd 2010

MSc Seminar: Donald Dunbar

Part 2 Summary

- Data analysis
- Data Mining
- Microarray Resources
- Microarray Standards
- Recent & future advances

February 3rd 2010

MSc Seminar: Donald Dunbar

Seminar Summary

- Part 1
 - Microarrays in biological research
 - A typical microarray experiment
- Part 2
 - Data analysis and mining
 - Recent & future advances

February 3rd 2010

MSc Seminar: Donald Dunbar

Contact

- Donald Dunbar
- QMRI Bioinformatics
- donald.dunbar@ed.ac.uk
- 0131 242 6700
- Room W3.01, QMRI, Little France
- www.bioinf.mvm.ed.ac.uk



February 3rd 2010

MSc Seminar: Donald Dunbar