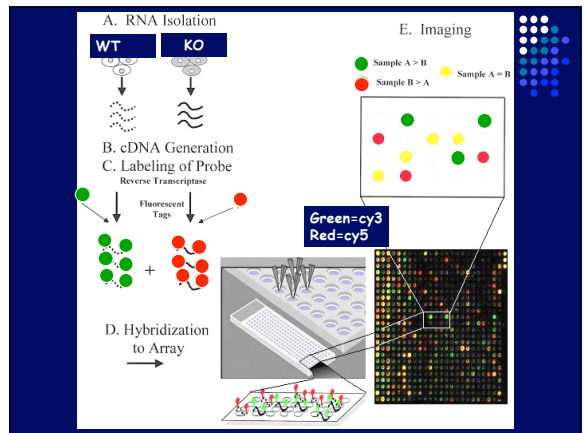
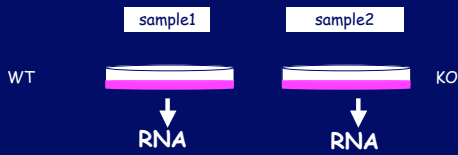


Data Analysis of DNA Microarrays



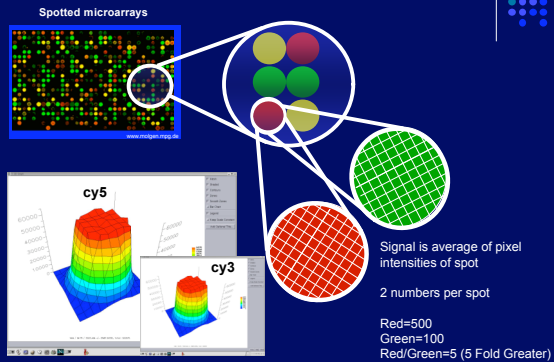
Can we detect siRNA-induced knockdown of gene expression using DNA microarrays?

Starting with two biological samples

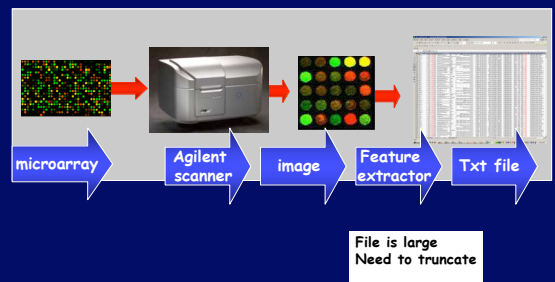


Microarray Measurements

Signal: Spotted arrays

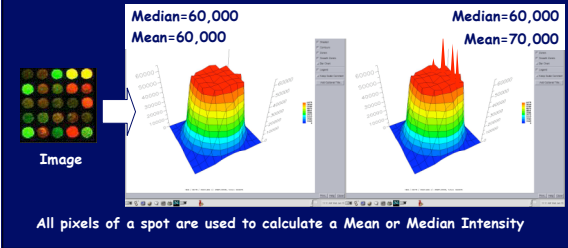


Processing microarrays: Scanning and Image analysis

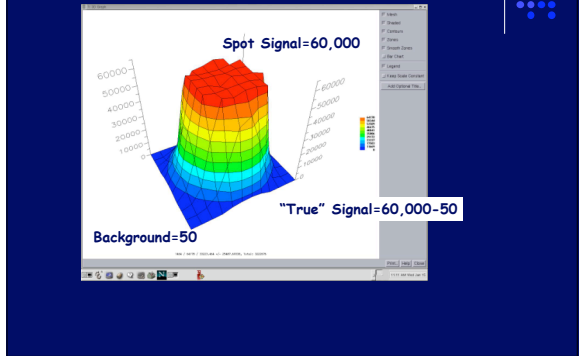


Spot Intensity: Mean or Median?

- Which is more affected by extremes?
- Which is better estimate of spot intensity?



Subtracting Background



TYPE	best	text	text	text	integer	float	float	text	text	integer	integer	integer
1	FEPARAN	Protocol	Scan	Scan	Scan	Num	Micro	Scan	Orig	Grid	Date	Grid
2	DATA	GE2-w4_31	#####	Agilent	Te	2	5	c151915-c014006_D_#####		1	532	
3	TYPE	float	float	float	integer	float	float	integer	integer	float	float	float
4	STATS	gMean	gStdev	gMean	gStdev	gMean	gStdev	gMean	gStdev	gMean	gStdev	gMean
5	DATA	38.455	39	7.04102	1000	65261	38.817	39	24.6422	1000	65196	1.2247
6	TYPE	integer	integer	integer	text	integer	text	integer	text	integer	text	float
7	FEATURE	FeatureID	Row	Col	SubType	SubType	Start	Sequence	ProbedUD	ControlType	ProbName	GeneName
8	DATA	1	1	1	0	0	0	1	GE_Bright	GE_Single	GE_Bright	Comer
9	DATA	2	1	2	66	Structural	0	1	1	Dark	Comer	Dark
10	DATA	3	1	3	66	Structural	0	1	1	Dark	Comer	Dark
11	DATA	4	1	4	66	Structural	0	1	1	Dark	Comer	Dark
12	DATA	5	1	5	66	Structural	0	1	1	Dark	Comer	Dark
13	DATA	6	1	6	66	Structural	0	1	1	Dark	Comer	Dark
14	DATA	7	1	7	66	Structural	0	1	1	Dark	Comer	Dark
15	DATA	8	1	8	66	Structural	0	1	1	Dark	Comer	Dark
16	DATA	9	1	9	66	Structural	0	1	1	Dark	Comer	Dark
17	DATA	10	1	10	66	Structural	0	1	1	Dark	Comer	Dark
18	DATA	11	1	11	66	Structural	0	1	1	Dark	Comer	Dark
19	DATA	12	1	12	66	Structural	0	1	1	Dark	Comer	Dark
20	DATA	13	1	13	66	Structural	0	1	1	Dark	Comer	Dark
21	DATA	14	1	14	66	Structural	0	1	1	Dark	Comer	Dark
22	DATA	15	1	15	66	Structural	0	1	1	Dark	Comer	Dark
23	DATA	16	1	16	66	Structural	0	1	1	Dark	Comer	Dark
24	DATA	17	1	17	66	Structural	0	1	1	Dark	Comer	Dark
25	DATA	18	1	18	66	Structural	0	1	1	Dark	Comer	Dark
26	DATA	19	1	19	66	Structural	0	1	1	Dark	Comer	Dark
27	DATA	20	1	20	66	Structural	0	1	1	Dark	Comer	Dark
28	DATA	21	1	21	66	Structural	0	1	1	Dark	Comer	Dark
29	DATA	22	1	22	66	Structural	0	1	1	Dark	Comer	Dark
30	DATA	23	1	23	66	Structural	0	1	1	Dark	Comer	Dark
31	DATA	24	1	24	66	Structural	0	1	1	Dark	Comer	Dark
32	DATA	24	1	24	66	Structural	0	1	1	Dark	Comer	Dark
33	DATA	24	1	24	66	Structural	0	1	1	Dark	Comer	Dark

Reduce number of rows
Remove rows 1-21..but save
Row 10!!

GeneName	Description	gMeanSignal	gStdevSignal	gMedianSignal	gStdevMedianSignal	gMeanSignal	gStdevSignal	gMedianSignal	gStdevMedianSignal	gMeanSignal	gStdevSignal	gMedianSignal	gStdevMedianSignal
1	Cori1	NM_009912	Mus musculus c	5.91E+01	7.95E+01	59	80	5.99E+01	7.94E+01	58	80	5.99E+01	7.94E+01
2	Hspa	NM_009725	Mus musculus n	7.96E+01	8.49E+01	66	78.5	7.79E+01	7.72E+01	68	78.5	7.79E+01	7.72E+01
3	Agp7	NM_007473	Mus musculus a	5.99E+01	7.49E+01	59	74	5.79E+01	7.89E+01	58	74	5.79E+01	7.89E+01
4	ApoA4l2	AK044412	Mus musculus a	9.71E+01	8.33E+01	68	83	8.84E+01	7.79E+01	58	83	8.84E+01	7.79E+01
5	Hcn1	NM_001042489	Mus musculus h	5.99E+01	7.55E+01	56	74	5.94E+01	7.77E+01	58	74	5.94E+01	7.77E+01
6	Gpr33	NM_008109	Mus musculus c	5.72E+01	7.43E+01	57	74	5.89E+01	7.89E+01	58	74	5.89E+01	7.89E+01
7	LOC434369	BC033388	Mus musculus a	5.99E+01	7.99E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
8	LOC434369	BC033388	Mus musculus c	8.39E+01	9.51E+01	79	82.5	8.89E+01	7.89E+01	58	82.5	8.89E+01	7.89E+01
9	LOC434369	BC033388	Mus musculus a	5.97E+01	7.99E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
10	Pdss2	NM_013782	Mus musculus p	1.09E+02	8.42E+01	95	83	8.89E+01	7.79E+01	58	83	8.89E+01	7.79E+01
11	LOC434369	BC033388	Mus musculus f	5.79E+01	7.95E+01	57	79	5.99E+01	7.89E+01	58	79	5.99E+01	7.89E+01
12	Hsp1	AK141948	Mus musculus i	5.99E+01	7.82E+01	60.5	77	5.99E+01	7.84E+01	58	77	5.99E+01	7.84E+01
13	Emo1	NM_198093	Mus musculus e	5.72E+01	7.80E+01	57	79	5.94E+01	7.87E+01	58	79	5.94E+01	7.87E+01
14	Cnec1	NM_144193	Mus musculus c	5.91E+01	7.49E+01	56	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
15	Pgpt1	NM_144123	Mus musculus p	5.99E+01	7.89E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
16	Scp9	NM_144123	Mus musculus p	5.99E+01	7.89E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
17	Timen41	NM_144123	Mus musculus p	5.99E+01	7.89E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
18	ApoB190	NM_144123	Mus musculus p	5.99E+01	7.89E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
19	Sytn1	NM_144123	Mus musculus p	5.99E+01	7.89E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
20	Ank1	NM_144123	Mus musculus p	5.99E+01	7.89E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
21	ANKRD1	NM_144123	Mus musculus p	5.99E+01	7.89E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01
22	Hsp70	NM_008194	Mus musculus h	5.92E+01	7.94E+01	58	78.5	6.13E+01	8.19E+01	60.5	78.5	6.13E+01	8.19E+01
23	BC036923	NM_170395	Mus musculus c	6.17E+01	8.39E+01	61	85	6.13E+01	7.97E+01	60	85	6.13E+01	7.97E+01
24	Phn1	NM_151452	Mus musculus b	5.79E+01	7.57E+01	58	75	5.99E+01	7.77E+01	58	75	5.99E+01	7.77E+01
25	LOC434369	BC033388	Mus musculus f	5.99E+01	7.47E+01	57.5	74.5	5.99E+01	7.99E+01	58	74.5	5.99E+01	7.99E+01
26	Jak	NM_017070	Mus musculus j	5.94E+01	8.13E+01	59	79	5.94E+01	7.89E+01	58	79	5.94E+01	7.89E+01
27	Hsp65	NM_00105247	Mus musculus f	5.91E+01	7.29E+01	59	73	5.93E+01	7.89E+01	58	73	5.93E+01	7.89E+01
28	Thec1	AK042548	Mus musculus t	5.94E+01	7.95E+01	59.5	76.5	5.99E+01	7.89E+01	58	76.5	5.99E+01	7.89E+01
29	Wnk1	NM_198703	Mus musculus v	5.99E+01	7.79E+01	58	76	5.99E+01	7.59E+01	58	76	5.99E+01	7.59E+01
30	ApoD2795	AK023795	Mus musculus i	5.95E+01	7.79E+01	58	77	5.91E+01	7.89E+01	58	77	5.91E+01	7.89E+01
31	Scn1	NM_009127	Mus musculus s	5.97E+01	7.89E+01	60	79	5.91E+01	7.87E+01	58	79	5.91E+01	7.87E+01
32	KRFB	AK031402	Mus musculus s	5.79E+01	7.99E+01	58	78.5	5.92E+01	7.93E+01	58	78.5	5.92E+01	7.93E+01
33	BC031329	NM_144149	Mus musculus c	6.94E+01	8.33E+01	61	81	6.15E+01	8.39E+01	61	81	6.15E+01	8.39E+01
34	BE956575	BE956575	LM-BC02-844-h	5.87E+01	7.94E+01	58	78.5	5.99E+01	8.02E+01	60	78.5	5.99E+01	8.02E+01
35	BC093265	BC093265	BC093265-602c	6.19E+01	7.33E+01	63	72	6.54E+01	7.89E+01	58	72	6.54E+01	7.89E+01

Copy columns L, M, N, AH, AI, AJ, AK
AR, AS, AT, AU

GeneName	Description	gMeanSignal	gStdevSignal	gMedianSignal	gStdevMedianSignal	gMeanSignal	gStdevSignal	gMedianSignal	gStdevMedianSignal	gMeanSignal	gStdevSignal	gMedianSignal	gStdevMedianSignal
1	Cori1	5.91E+01	7.95E+01	59	80	5.99E+01	7.94E+01	58	80	5.99E+01	7.94E+01	58	80
2	Hspa	7.96E+01	8.49E+01	66	78.5	7.79E+01	7.72E+01	68	78.5	7.79E+01	7.72E+01	68	78.5
3	Agp7	5.99E+01	7.49E+01	59	74	5.79E+01	7.89E+01	58	74	5.79E+01	7.89E+01	58	74
4	ApoA4l2	9.71E+01	8.33E+01	68	83	8.84E+01	7.79E+01	58	83	8.84E+01	7.79E+01	58	83
5	Hcn1	5.99E+01	7.55E+01	56	74	5.94E+01	7.77E+01	58	74	5.94E+01	7.77E+01	58	74
6	Gpr33	5.72E+01	7.43E+01	57	74	5.89E+01	7.89E+01	58	74	5.89E+01	7.89E+01	58	74
7	LOC434369	5.99E+01	7.99E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01	58	74
8	LOC434369	8.39E+01	9.51E+01	79	82.5	8.89E+01	7.89E+01	58	82.5	8.89E+01	7.89E+01	58	82.5
9	LOC434369	5.97E+01	7.99E+01	57	74	5.99E+01	7.89E+01	58	74	5.99E+01	7.89E+01	58	74
10	Pdss2	1.09E+02	8.42E+01	95	83	8.89E+01	7.79E+01	58	83	8.89E+01	7.79E+01	58	83
11	LOC434369	5.79E+01	7.95E+01	57	79	5.99E+01	7.89E+01	58	79	5.99E+01	7.89E+01	58	79
12	Hsp1	5.99E+01	7.82E+01	60.5	77	5.99E+01	7.84E+01	58	77	5.99E+01	7.84E+01	58	77
13	Emo1	5.72E+01	7.80E+01	57	79	5.94E+01	7.87E+01	58	79	5.94E+01	7.87E+01	58	79
14	Cn												

cy3 and cy5: Commonly used dyes

cy5
~700 emission

cy3 cy5

Light sensitivity: cy5 more easily degraded

cy3
~500 emission

Let's begin the normalization process:

Green background red background corrected

B	C	D	E	F	G	H	I	J	K	L	M	N
561A_06_P61.ZRG17	YNR039C	430	288.5	65	83	365	205.5					
2744A_06_P26.ZRG8	YER033C	105.5	213	74	83	31.5	130					
2744A_06_P26.ZRG8	YER033C	183	338.5	66	84	117	254.5					
4872A_06_P32.ZRT1	YGL255W	173.5	236	64	82	109.5	154					
3450A_06_P49.ZRT2	YLR130C	1618	2062.5	71	83	1747	2779.5					
2594A_06_P45.ZRT3	YKL175W	252.5	291.5	72	84	160.5	207.5					
2594A_06_P45.ZRT3	YKL175W	1206	1538	69	83	1136	1455					
5969A_06_P12.ZTA1	YBR046C	1375	2191	61	80	1314	2021					
5969A_06_P12.ZTA1	YBR046C	2970	4300.5	68	83	2902	4297.5					
4654A_06_P35.ZU01	YGR285C	2815	2136.5	71	83	2744	2053.5					
2452A_06_P60.ZWF1	YNL241C	5174	13544	65	83	14500	13461					

Formula: $\text{AVERAGE}(2, J10275)$
 $\text{AVERAGE}(\text{number1}, \text{number3}, \dots)$

Create mean signal to estimate dye bias

green red

B	C	D	E	F	G	H	I	J	K	L	M	N
4872A_06_P32.ZRT1	YGL255W	173.5	236	64	82	109.5	154					
3450A_06_P49.ZRT2	YLR130C	1618	2062.5	71	83	1747	2779.5					
2594A_06_P45.ZRT3	YKL175W	252.5	291.5	72	84	160.5	207.5					
2594A_06_P45.ZRT3	YKL175W	1206	1538	69	83	1136	1455					
5969A_06_P12.ZTA1	YBR046C	1375	2191	61	80	1314	2021					
5969A_06_P12.ZTA1	YBR046C	2970	4300.5	68	83	2902	4297.5					
4654A_06_P35.ZU01	YGR285C	2815	2136.5	71	83	2744	2053.5					
2452A_06_P60.ZWF1	YNL241C	5174	13544	65	83	14500	13461					

Formula: $\text{AVERAGE}(2, J10277)$
 $\text{AVERAGE}(\text{number1}, \text{number3}, \dots)$

1.7 FC difference

Create mean signal to estimate dye bias

Formula: $\text{AVERAGE}(2, K17)$

B	C	D	E	F	G	H	I	J	K	L	M	N
3095A_06_P45	1-Oct YKL4C	568.5	301	73	84	435.5	217	137.6473				
3095A_06_P45	1-Oct YKL4C	325	263.5	60	80	265	183.5	107.3412				
2965A_06_P54.AAC1	YMR066C	1941	2177	64	82	1877	2095	1232.353				
6119A_06_P14.AAC3	YBR085W	141	111.5	65	82	76	29.5	12.26294				
6119A_06_P14.AAC3	YBR085W	95.5	102.5	64	81	31.5	21.5	12.64706				
1626A_06_P43.AAD10	YJR159W	223	135	87.5	84	135.5	51	30				
1626A_06_P43.AAD10	YJR159W	1760	604.5	72	83	1626	521.5	306.547				
2361A_06_P61.AAD14	YNL331C	235	117	68	83	167	34	20				
2093A_06_P63.AAD15	YOL165C	103	102	65	83	38	19	11.17647				
5989A_06_P17.AAD3	YCR107W	242	216	74	84	168	132	77.64706				
1459A_06_P19.AAD4	YDL243C	113	94	69	84	44	10	5.882353				
1459A_06_P19.AAD4	YDL243C	282	395	72	83	210	312	183.5294				
5116A_06_P28.AAD6	YFL066C	215.5	163	62	81	153.5	82	48.25229				
5116A_06_P28.AAD6	YFL066C	314	270	67	81	247	189	111.1765				
2545A_06_P59.AAH1	YIL141W	233.5	164	91	85	142.5	79	45.47059				
2545A_06_P59.AAH1	YIL141W	138	110	64	81	74	29	11.68882				
6299A_06_P10.AAP1	QO080	587	253	77	85	510	168	98.82353				
6299A_06_P10.AAP1	QO080	529	250	66	85	464	161	88.29353				
4558A_06_P36.AAP1	YHR047C	626	2123									
6191A_06_P12.AAR2	YBL674C	275	139									
3821A_06_P45.AA11	YKL180W	636	629									
3821A_06_P45.AA11	YKL180W	280.5	430									
3537A_06_P48.AA12	YLR077C	6453	6905									
4971A_06_P30.ABC1	YGL119W	400.5	459.5									
4971A_06_P30.ABC1	YGL119W	220	184.5	63	81	157	103.5	60.88235				
6019A_06_P14.ABD1	YBR296C	362.5	515	75	83	287.5	452	289.4118				
6019A_06_P14.ABD1	YBR296C	473	530	68	82	405	849	498.6235				
787A_06_P45.ABF1	YKL121W	1307.5	176	71	83	1236.5	83	54.78588				
787A_06_P45.ABF1	YKL121W	4447	4509	76	87	4071	322	189.4118				
2546A_06_P54.ABF2	YMR072W	7888.5	12876	71	85	7817.5	12791	7524.118				
2546A_06_P54.ABF2	YMR072W	5054	7319.5	67	83	4987	7236.5	4256.765				
1881A_06_P43.ABM1	YJR188W	180	331	66	83	132	248	145.8234				
1881A_06_P43.ABM1	YJR188W	225.5	521.5	68	84	157.5	437.5	257.3529				
5884A_06_P17.ABP1	YCR088W	1150.5	9076	71	83	1079.5	8993	5290				

Normalize your lower abundant channel (increase) by factor to have mean expression across the array equal

Calculating Differences in Gene Expression

Spotted microarrays

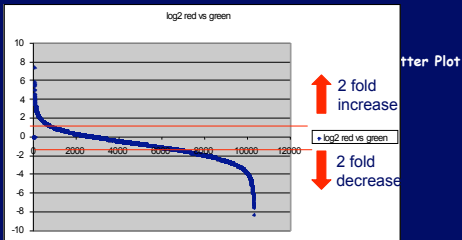
2 numbers per spot

Red=500
Green=100
Red/Green=5 (5 Fold Greater in Red)

Red=100
Green=500
Red/Green=0.2 (5 Fold Less in Red)

$\text{Log}_2(5) \approx +2$
 $\text{Log}_2(0.2) \approx -2$

Create scatter plot of log₂ ratios (green versus red)



Distribution of log₂ ratios

- What are we expecting????
- What color would all of these spots be??

Trends in Data

- How many changes do you see?
- What could these changes mean?
- How can we find out more about these genes and their functions?
- Which biological processes are up-regulated, down-regulated, no change?

Good luck!!

