Hannover Medical School

RCU Transcriptomics

Research Core Unit Transcriptomics

Microarray Study Designs using Agilent Expression Arrays

Version: 19.03.2012

Introduction

The right choice of the experimental design is an important prerequisite for a successful microarray study. Unlike Affymetrix, the Agilent microarray platform allows to choose between two different modes of hybridization, i.e. Single-Color or Dual-Color.

Both "color modes" have particular characteristics that should be considered before taking a final decision.

In this manual, advantages and disadvantages of both color modes are briefly discussed and examples of appropriate study designs are presented.[*] [**]

[*] Please note that all exemplarily depicted designs refer to just one loaded microarray slide, containing 4 microarrays. Thus, most of these designs are primarily suited for an initial pilot experiment rather than supporting a comprehensive study and meeting stringent statistical demands.

[**] For further information beyond the scope of this manual we recommend: Patterson et al. (2006), "Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project".

Characteristics of the Dual-Color mode

The underlying methodical concept is a co-hybridization of two differently labeled RNA samples to one microarray. One of the samples to be compared is labeled with Cy3 or Alexa555 (green channel) while the other is labeled with Cy5 or Alexa647 (red channel). Differential mRNA expression is deduced from differences in fluorescence intensities between both channels, measured at the identical position (feature / spot) of one and the same microarray.

Advantages of the Dual-Color mode

- Dual-Color results provided in our standard excel file formats are a bit more clearly arranged as compared to Single-Color results.

- Informative p-values, calculated by Agilent's Feature Extraction software, indicate the probability that the measured difference in intensities reflects a true difference in abundance of the two competing labeled RNA species. By default, these p-values are only available in the Dual-Color mode.

- Dual-Color derived ratios are more robust against moderate hybridization artifacts and thus less prone to false-positive results. This is, because the most frequently occurring hybridization artifacts tend to affect both color channels similarly. Therefore, the resulting detraction is often minimized/equalized after calculating ratios of intensities of the two channels.

- In the case of a common-reference design (Figure 3), locally restricted impairments in the hybridization performance can be recognized by an "abnormal behavior" of the intensities of the common-reference sample.

- In some cases, the dual-color approach is more reliable in detecting small intensity differences; however, there are also exceptions to this rule.

- Depending on the underlying biological study design and the experimenter's objective, the dual-color approach might be advantageous regarding the overall costs, especially if the study should primarily serve as a pilot or screening approach.

Disadvantages of the Dual-Color mode

- The ratio values are compressed, i.e. a 10-fold change in real ends in a fold change of about 6.5, indicated by the microarray results. However, this effect is largely balanced regarding both labeling directions, i.e. green/red and red/green ratios, and throughout the whole intensity (and ratio) range.

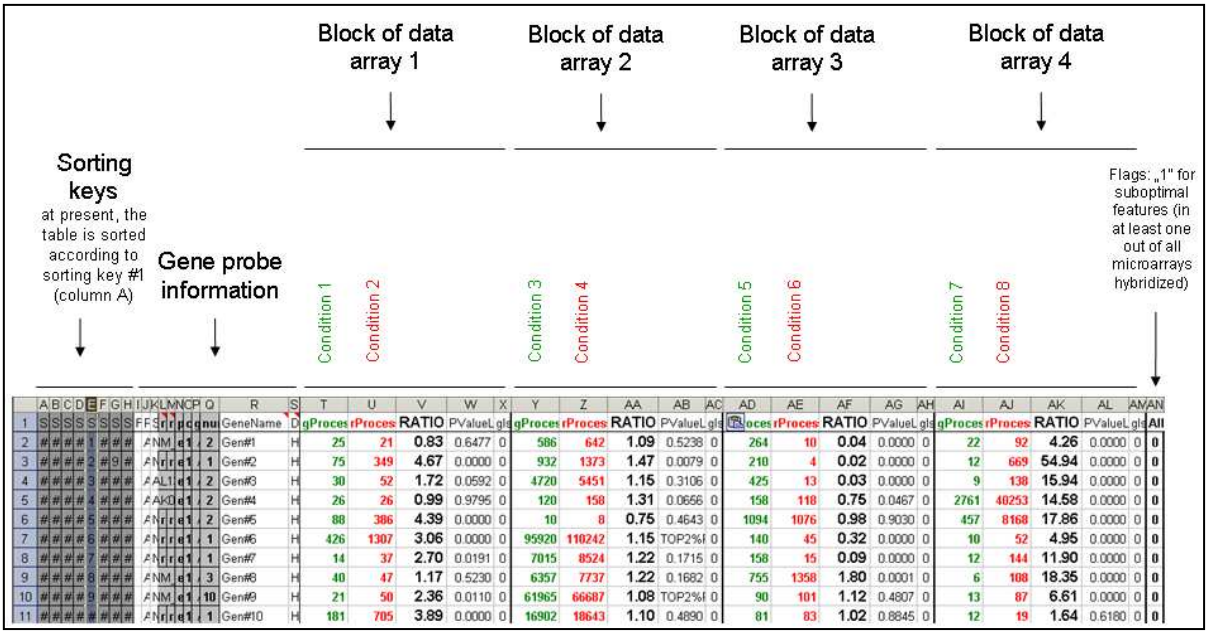
- Cy dyes slightly influence each other. Dyes have, for example, particular similarities, i.e. "red-labeled" samples are slightly more similar among each other as compared to "green-labeled" samples (and vice versa). This leads to a slight bias due to these individual dye properties. Anyhow, most of the resulting systematic impairments can be estimated by a "self-self hybridization" (see Figure 4) or can be partly equalized by a dye-swap design (see Figure 2).

- The Dual-Color design is less flexible to future extensions of the study (e.g. depending on the permanent availability of the generated batch of a fluorescently-labeled common reference sample; see Figure 3) .

Conclusion

The dual-color mode is recommended for cost-efficient screening approaches and preliminary experiments.

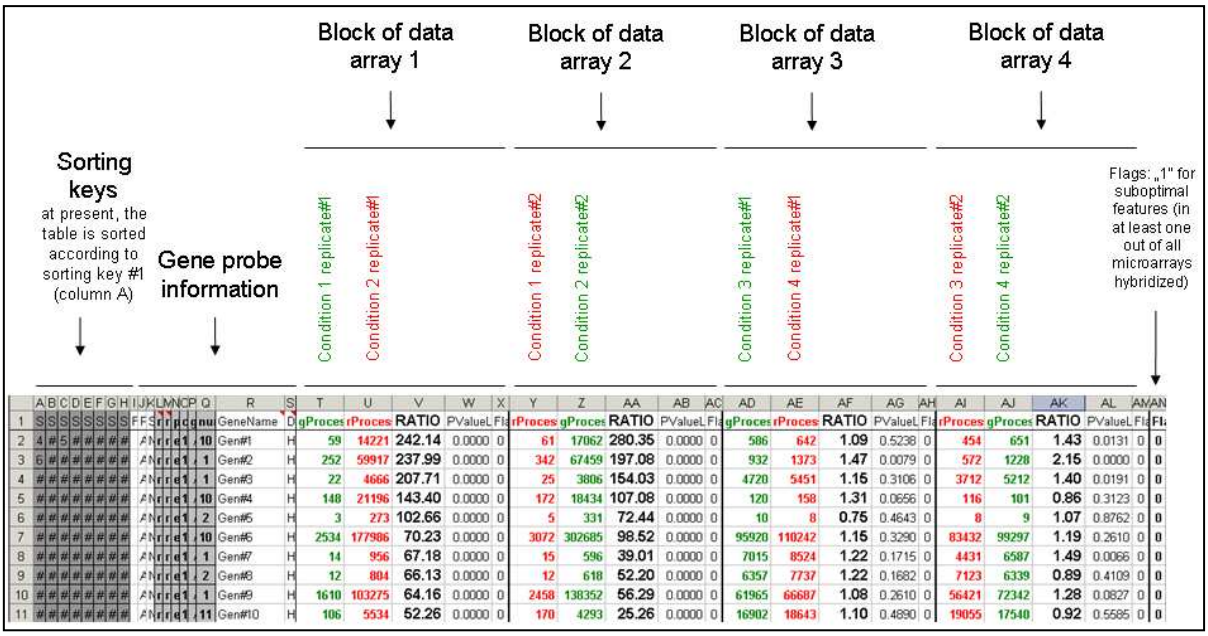
Dual-Color designs



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
1	S	S	S	S	S	S	S	S	S	F	F	S	r	r	p	c	g	n	u	GeneName	D	gProces	rProces	RATIO	PValueL	gIs	gProces	rProces	RATIO	PValueL	gIs	Proces	rProces	RATIO	PValueL	gIs	gProces	rProces	RATIO	PValueL	gIs	All
2	#	#	#	#	1	#	#	#	A	N	M	e	1	.	2			Gen#1	H	25	21	0.83	0.6477	0	586	642	1.09	0.5238	0	264	10	0.04	0.0000	0	22	92	4.26	0.0000	0	0
3	#	#	#	#	2	#	9	#	A	N	r	r	e	1	.	1		Gen#2	H	75	349	4.67	0.0000	0	932	1373	1.47	0.0079	0	210	4	0.02	0.0000	0	12	669	54.94	0.0000	0	0
4	#	#	#	#	3	#	#	#	A	A	L	1	e	1	.	2		Gen#3	H	30	52	1.72	0.0592	0	4720	5451	1.15	0.3106	0	425	13	0.03	0.0000	0	9	138	15.94	0.0000	0	0
5	#	#	#	#	4	#	#	#	A	A	K	1	e	1	.	2		Gen#4	H	26	26	0.99	0.9795	0	120	158	1.31	0.0656	0	158	118	0.75	0.0467	0	2761	40253	14.58	0.0000	0	0
6	#	#	#	#	5	#	#	#	A	N	r	r	e	1	.	2		Gen#5	H	88	386	4.39	0.0000	0	10	8	0.75	0.4643	0	1094	1076	0.98	0.9030	0	457	8168	17.86	0.0000	0	0
7	#	#	#	#	6	#	#	#	A	N	r	r	e	1	.	1		Gen#6	H	426	1307	3.06	0.0000	0	95920	110242	1.15	TOP2%I	0	140	45	0.32	0.0000	0	10	52	4.95	0.0000	0	0
8	#	#	#	#	7	#	#	#	A	N	r	r	e	1	.	1		Gen#7	H	14	37	2.70	0.0191	0	7015	8524	1.22	0.1715	0	158	15	0.09	0.0000	0	12	144	11.90	0.0000	0	0
9	#	#	#	#	8	#	#	#	A	N	M	e	1	.	3			Gen#8	H	40	47	1.17	0.5230	0	6357	7737	1.22	0.1682	0	755	1358	1.80	0.0001	0	6	108	18.35	0.0000	0	0
10	#	#	#	#	9	#	#	#	A	N	M	e	1	.	10			Gen#9	H	21	50	2.36	0.0110	0	61965	66687	1.08	TOP2%I	0	90	101	1.12	0.4807	0	13	87	6.61	0.0000	0	0
11	#	#	#	#	#	#	#	#	A	N	r	r	e	1	.	1		Gen#10	H	181	705	3.89	0.0000	0	16902	18643	1.10	0.4890	0	81	83	1.02	0.8845	0	12	19	1.64	0.6180	0	0

Figure 1: Dual-Color design#1:

8 different samples on a single slide.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
1	S	S	S	S	S	S	S	S	S	F	F	S	r	r	p	c	g	n	u	GeneName	D	gProces	rProces	RATIO	PValueL	Fla	rProces	gProces	RATIO	PValueL	Fla	gProces	rProces	RATIO	PValueL	Fla	rProces	gProces	RATIO	PValueL	Fla
2	4	#	5	#	#	#	#	#	A	N	r	r	e	1	.	10		Gen#1	H	59	14221	242.14	0.0000	0	61	17062	280.35	0.0000	0	586	642	1.09	0.5238	0	454	651	1.43	0.0131	0	0
3	6	#	#	#	#	#	#	#	A	N	r	r	e	1	.	1		Gen#2	H	252	59917	237.99	0.0000	0	342	67459	197.08	0.0000	0	932	1373	1.47	0.0079	0	572	1228	2.15	0.0000	0	0
4	#	#	#	#	#	#	#	#	A	N	r	r	e	1	.	1		Gen#3	H	22	4666	207.71	0.0000	0	25	3806	154.03	0.0000	0	4720	5451	1.15	0.3106	0	3712	5212	1.40	0.0191	0	0
5	#	#	#	#	#	#	#	#	A	N	r	r	e	1	.	10		Gen#4	H	148	21196	143.40	0.0000	0	172	18434	107.08	0.0000	0	120	158	1.31	0.0656	0	116	101	0.86	0.3123	0	0
6	#	#	#	#	#	#	#	#	A	N	r	r	e	1	.	2		Gen#5	H	3	273	102.66	0.0000	0	5	331	72.44	0.0000	0	10	8	0.75	0.4643	0	8	9	1.07	0.8762	0	0
7	#	#	#	#	#	#	#	#	A	N	r	r	e	1	.	10		Gen#6	H	2534	177986	70.23	0.0000	0	3072	302685	98.52	0.0000	0	95920	110242	1.15	0.3290	0	83432	99297	1.19	0.2610	0	0
8	#	#	#	#	#	#	#	#	A	N	r	r	e	1	.	1		Gen#7	H	14	956	67.18	0.0000	0	15	596	39.01	0.0000	0	7015	8524	1.22	0.1715	0	4431	6587	1.49	0.0066	0	0
9	#	#	#	#	#	#	#	#	A	N	r	r	e	1	.	2		Gen#8	H	12	804	66.13	0.0000	0	12	618	52.20	0.0000	0	6357	7737	1.22	0.1682	0	7123	6339	0.89	0.4109	0	0
10	#	#	#	#	#	#	#	#	A	N	r	r	e	1	.	1		Gen#9	H	1610	103275	64.16	0.0000	0	2458	138352	56.29	0.0000	0	61965	66687	1.08	0.2610	0	56421	72342	1.28	0.0827	0	0
11	#	#	#	#	#	#	#	#	A	N	r	r	e	1	.	11		Gen#10	H	106	5534	52.26	0.0000	0	170	4293	25.26	0.0000	0	16902	18643	1.10	0.4890	0	19055	17540	0.92	0.5585	0	0

Figure 2: Dual-Color design#2:

4 different samples on a slide including two replicates per sample and a dye-swap.

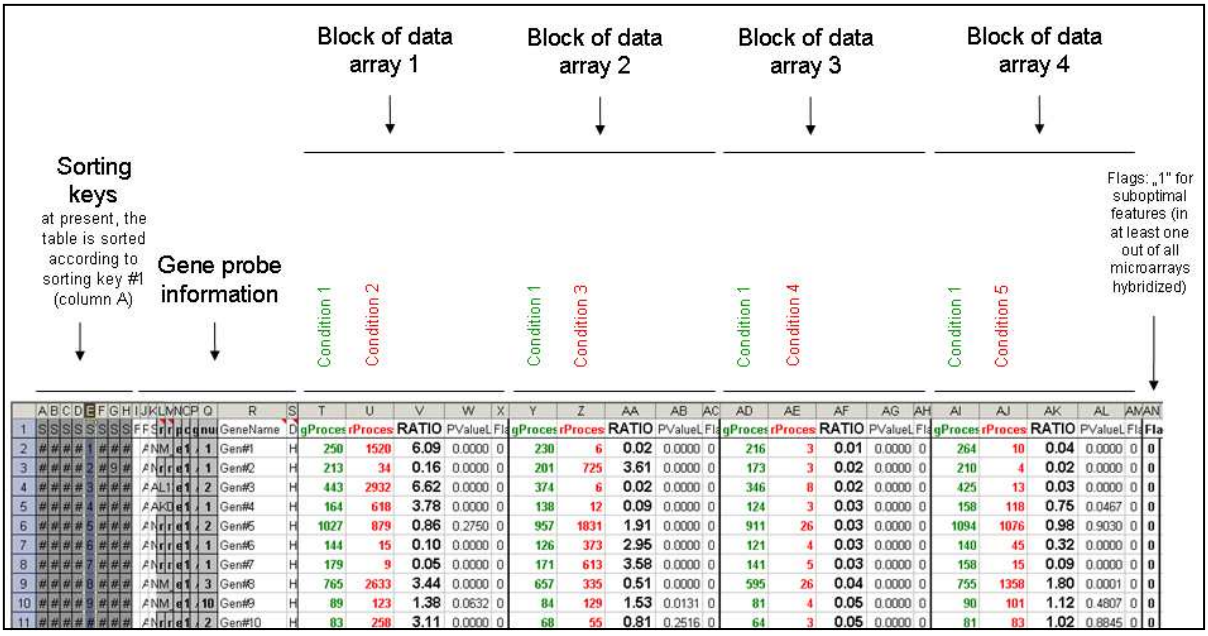


Figure 3: Dual-Color design#3:

5 (1+4) different samples on a slide including a common reference sample.

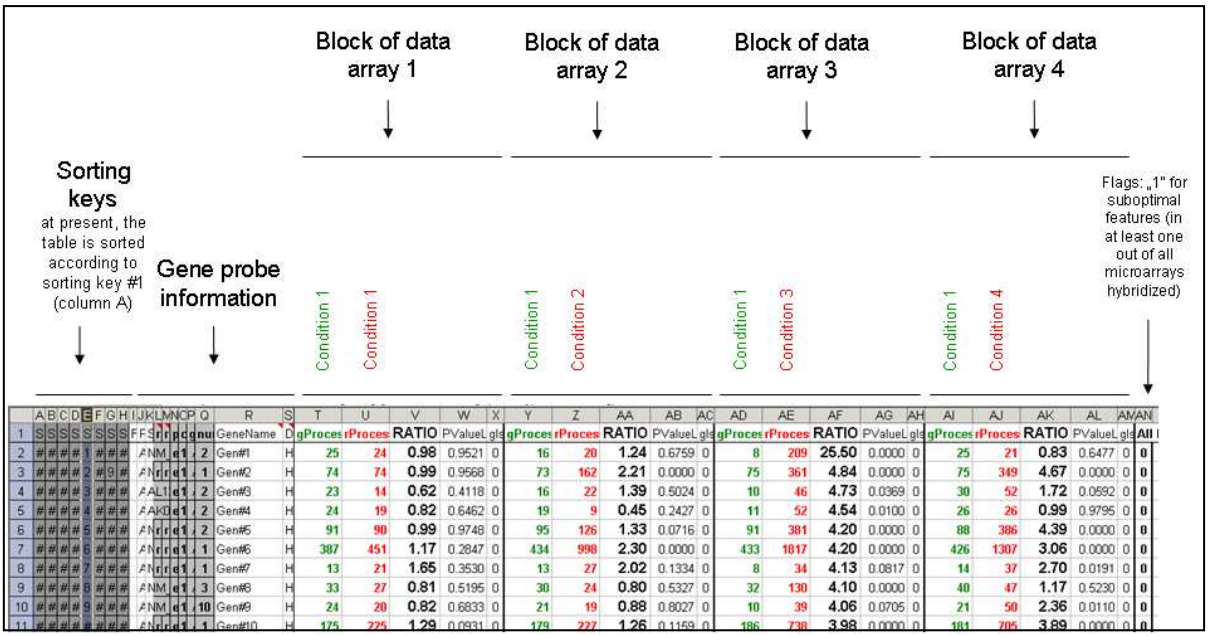


Figure 4: Dual-Color design#4:

4 (1+3) different samples on a slide including a common reference and a self-self-hybridization.

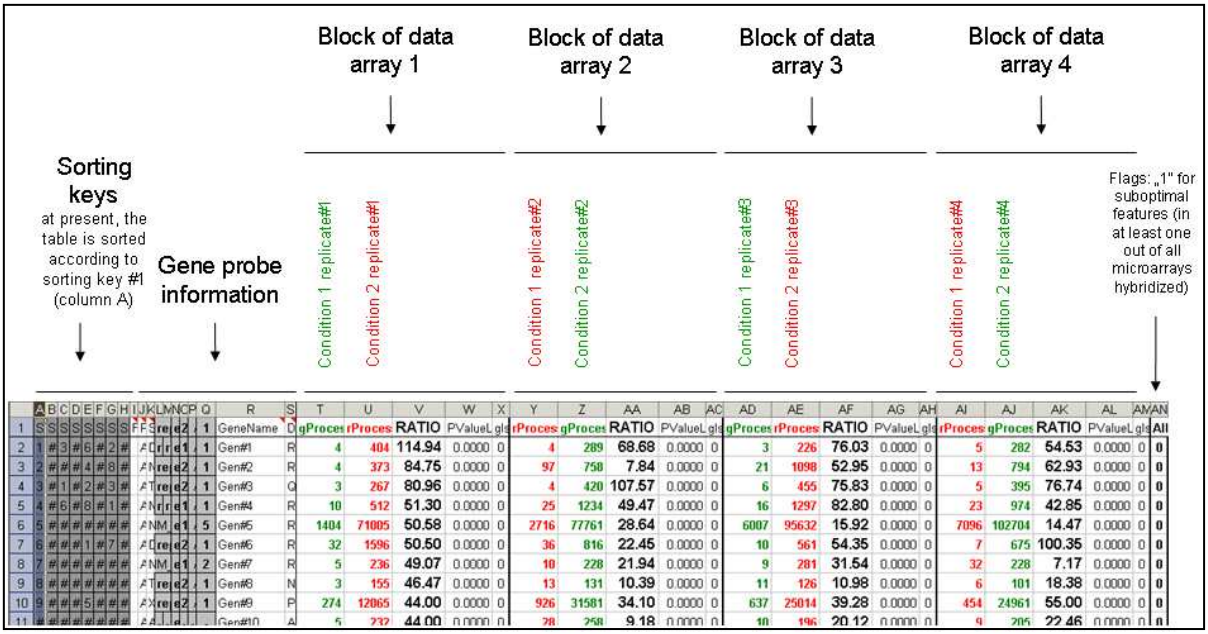


Figure 5: Dual-Color design#5:

2 (1+1) different samples on a slide with 4 replicates per sample and a balanced dye-swap design.

Characteristics of the Single-Color mode

The underlying methodical concept is a hybridization of one labeled RNA sample (labeled with Cy3 or Alexa555; green channel) to one microarray. Differential mRNA expression is deduced from differences in fluorescence intensity comparing a particular scanned region (feature/spot) of **one** microarray with the same region of **one** (or several) **other** microarray(s).

Advantages of the Single-Color mode

- The experimental design is quite simple: one sample – one microarray – one data set.

- A study can be easily extended by subsequently generated samples that shall be analyzed in the context of already existing data.

- There is no need to take dye biases into account while defining the study design**.**

- A big advantage of the single-color mode is the “equivalence” of all samples to be compared. Every single sample can be compared with every other sample under

exactly the same consistent accuracy. The deduced pairwise comparisons again can be compared with each other, i.e. “ratios of ratios”.

- Some analyses perform optimal just with single-color data such as (intensity-based) clustering and principal component analysis.

- There is no compression of ratio values as evident as with dual-color data. Single-Color designs are more appropriate to quantify absolute differences in mRNA abundance.

Disadvantages of the Single-Color mode

- Based on our experience with the Agilent microarray platform, slight impairments in hybridization performance generally occur sporadically, rather than systematically (this means: different regions and spots are adversely affected on different microarrays). Unlike in the dual-color mode, these inaccuracies are not equalized when ratios are generated. Thus, misinterpretations can result (false positively regulated candidates), especially in cases when i) only one probe per mRNA is present on the array, or ii) the number of analyzed samples (arrays) per condition is restricted to n=1. However, in the majority of cases, technical impairments should be indicated by the flag entries provided.

- The provided standard result file format might be a bit less clearly arranged than in case of a dual-color design.

- A p-value is not provided in our standard result table.

Conclusion

The single-color platform is recommended for large microarray studies and/or in cases in which comprehensive data analyses are planned.

Single-Color designs

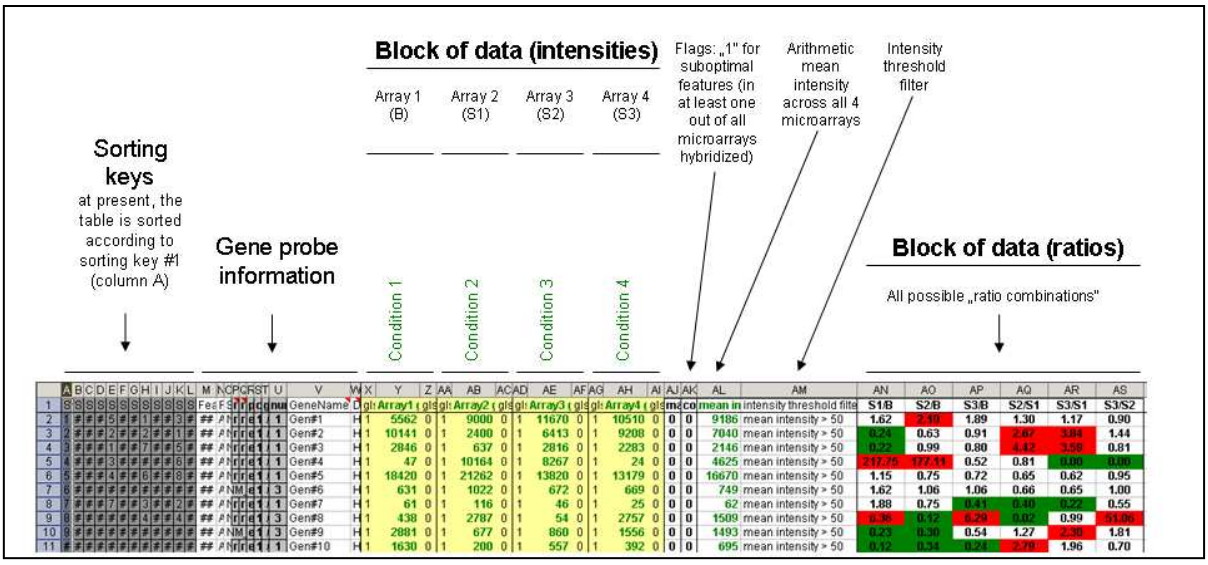


Figure 6: Single-Color design#1:

4 different samples on a slide.

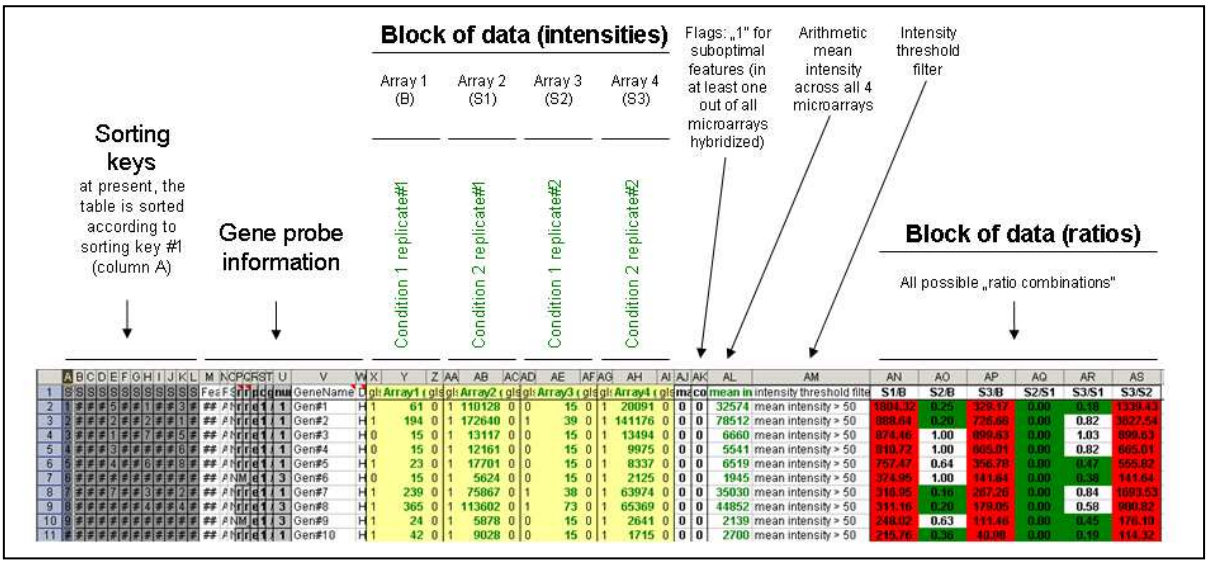


Figure 7: Single-Color design#2:

2 different samples on a slide including 2 replicates.

Kontakt:

Heike Schneider
Technische Assistentin
Gebäude I3, Ebene 1, Raum 1080
Tel: +49 (0)511-532 2830
E-Mail: Schneider.Heike@mh-hannover.de

Torsten Glomb
Diplom-Bioinformatiker
Gebäude I3, Ebene 1, Raum 1300
Tel: +49 (0)511-532 2869
E-Mail: Glomb.Torsten@mh-hannover.de

Dr. rer. nat Oliver Dittrich-Breiholz
Leiter der Zentralen Forschungseinrichtung Transcriptomics
Gebäude I3, Ebene 1, Raum 1270
Tel: +49 (0)511-532 5814
E-Mail: Dittrich.Oliver@mh-hannover.de