Quality and Use of Genome-wide Assays for methylation and RNA Analysis

Illumina Seminar Series, Munich,

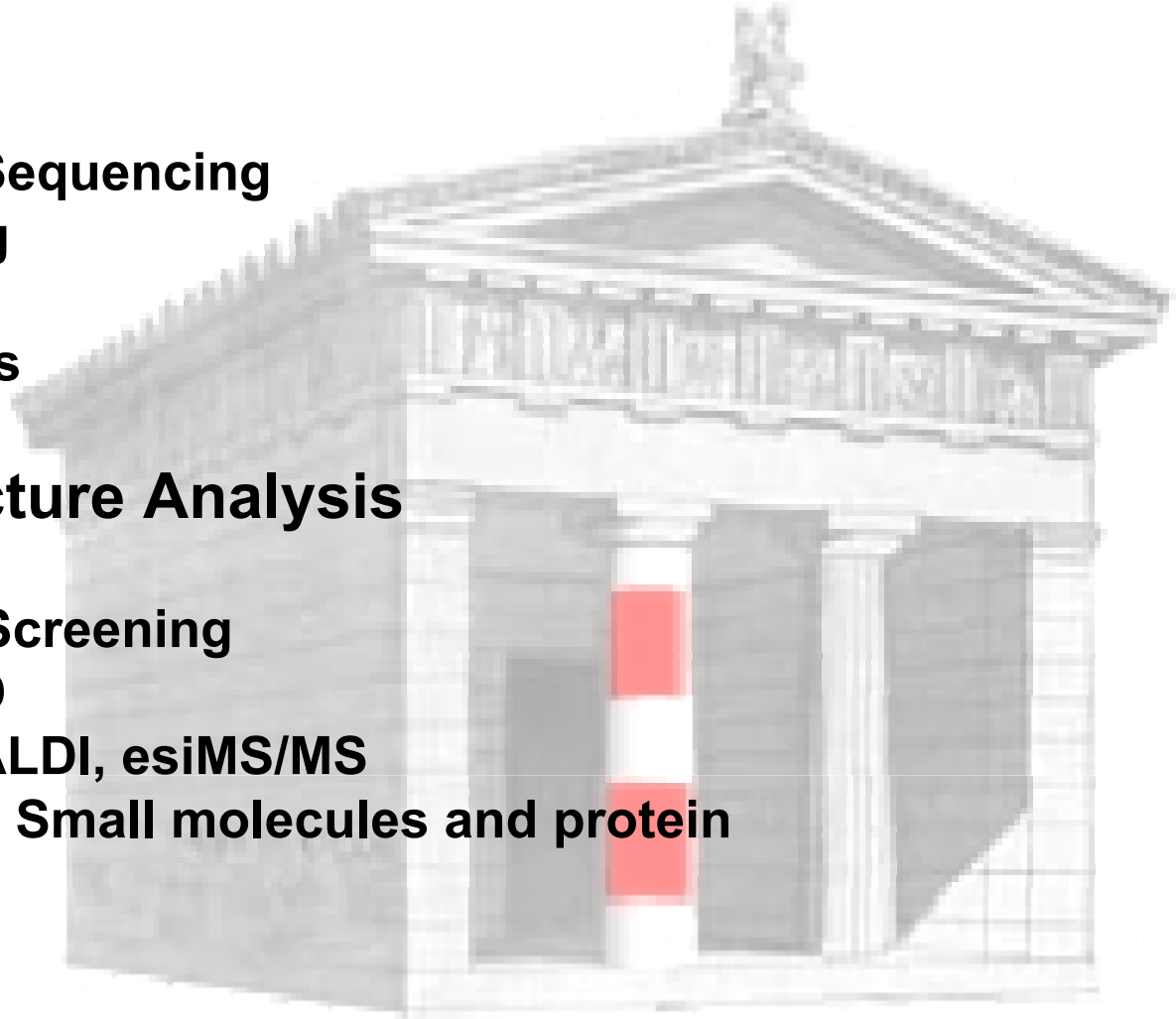
06.07.2009

Bernhard Korn, Genomcis & Proteomics Core Facility, German Cancer Research Center (DKFZ), Heidelberg, Germany

dkfz. GERMAN CANCER RESEARCH CENTER IN THE HELMHOLTZ ASSOCIATION

Scientific Services at DKFZ

- **Genomics**
  - Sanger Sequencing
  - “Next Generation” Sequencing
  - Expression Profiling
  - Genotyping
  - Methylation Analysis
  - Clone Repository
- **Proteomics & Structure Analysis**
  - Peptide synthesis
  - Protein Interaction Screening
  - Protein Analysis, 2D
  - Protein Analysis MALDI, esiMS/MS
  - Mass Spectrometry: Small molecules and protein modifications
  - NMR
  - Molecular Modeling



scientiam adiuvamus

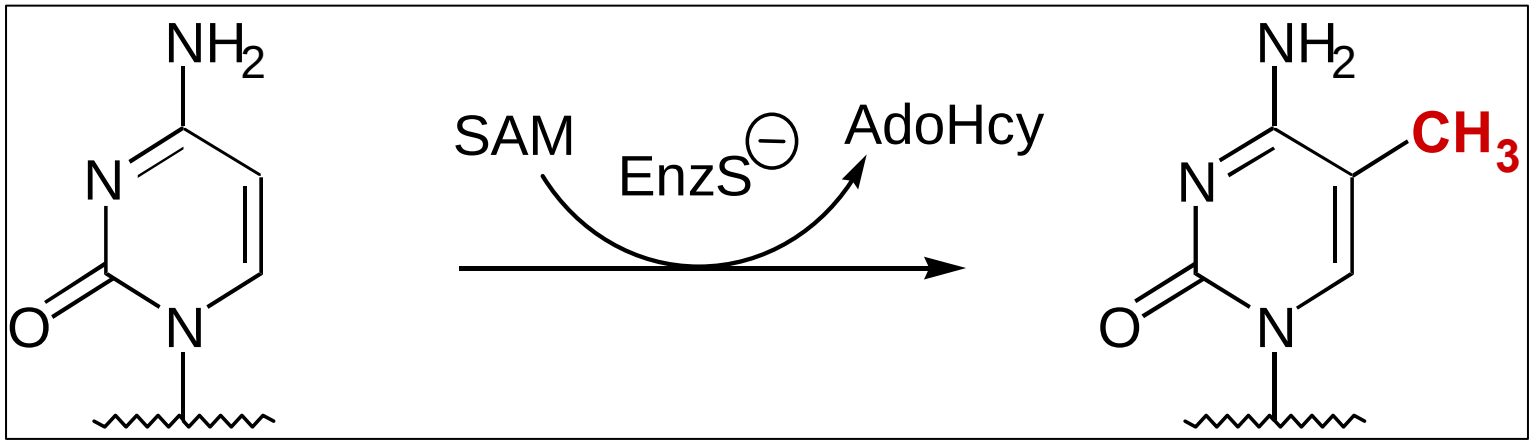
- Genome-wide methylation analysis
  - Evaluation of technology
  - Preliminary data on lymphomas

- Expression profiling using ‚deraded RNA‘
  - WG-DASL technology
  - Potential

What is epigenetics?

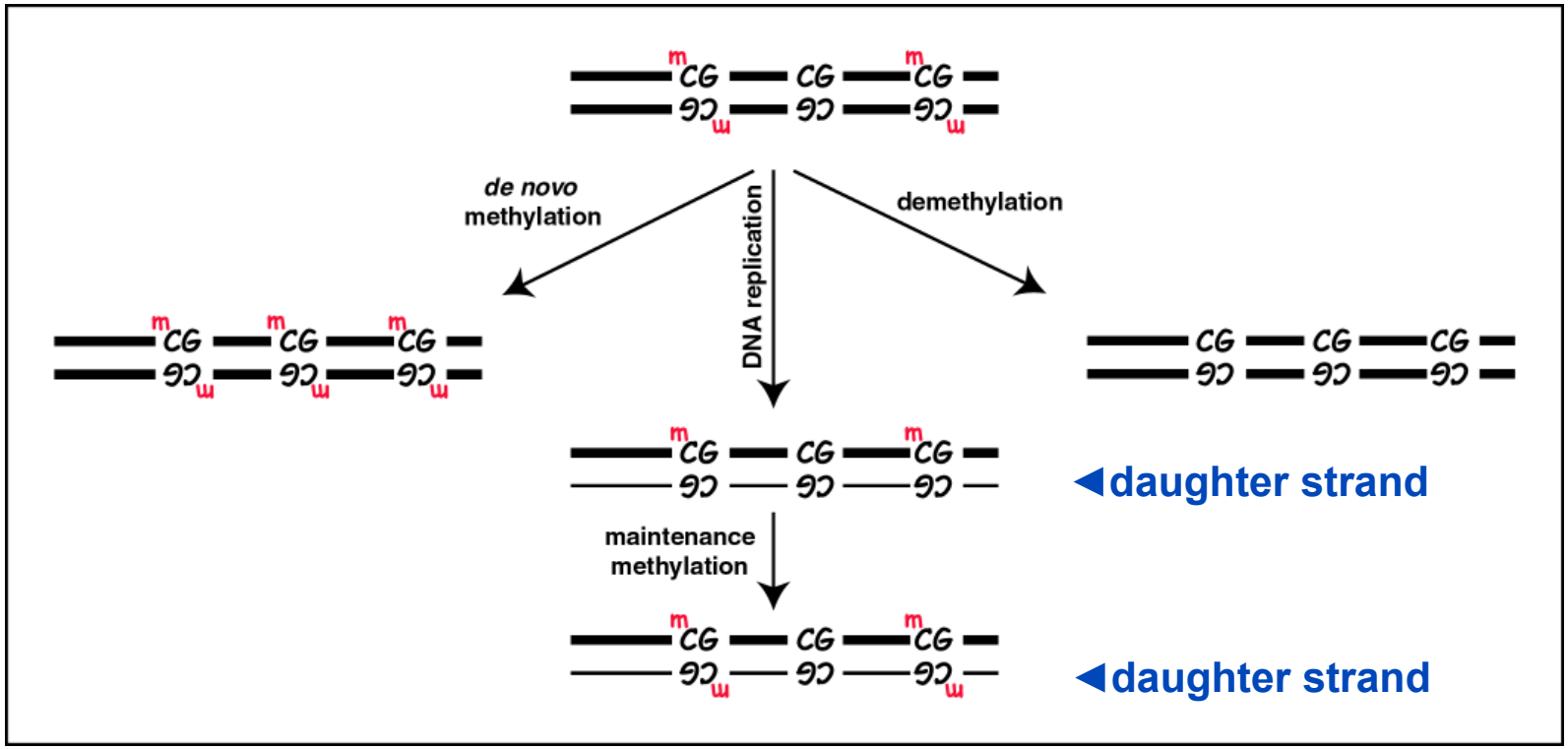
- Study of heritable changes in gene function that do NOT involve changes to the nucleotide sequence of DNA
- When a cell undergoes mitosis or meiosis, the epigenetic information is stably transmitted to the subsequent generation
- Epigenetic controls add an ‘extra layer’ of transcriptional control
- Epigenetic changes include:
  - DNA methylation
  - Histone modifications
  - RNA interference

DNA Methylation



- Methyl group introduced in the 5’ position of cytosine
- In mammals, occurs in CpG dinucleotides
- Catalyzed by DNA methyltransferases (DNMT)

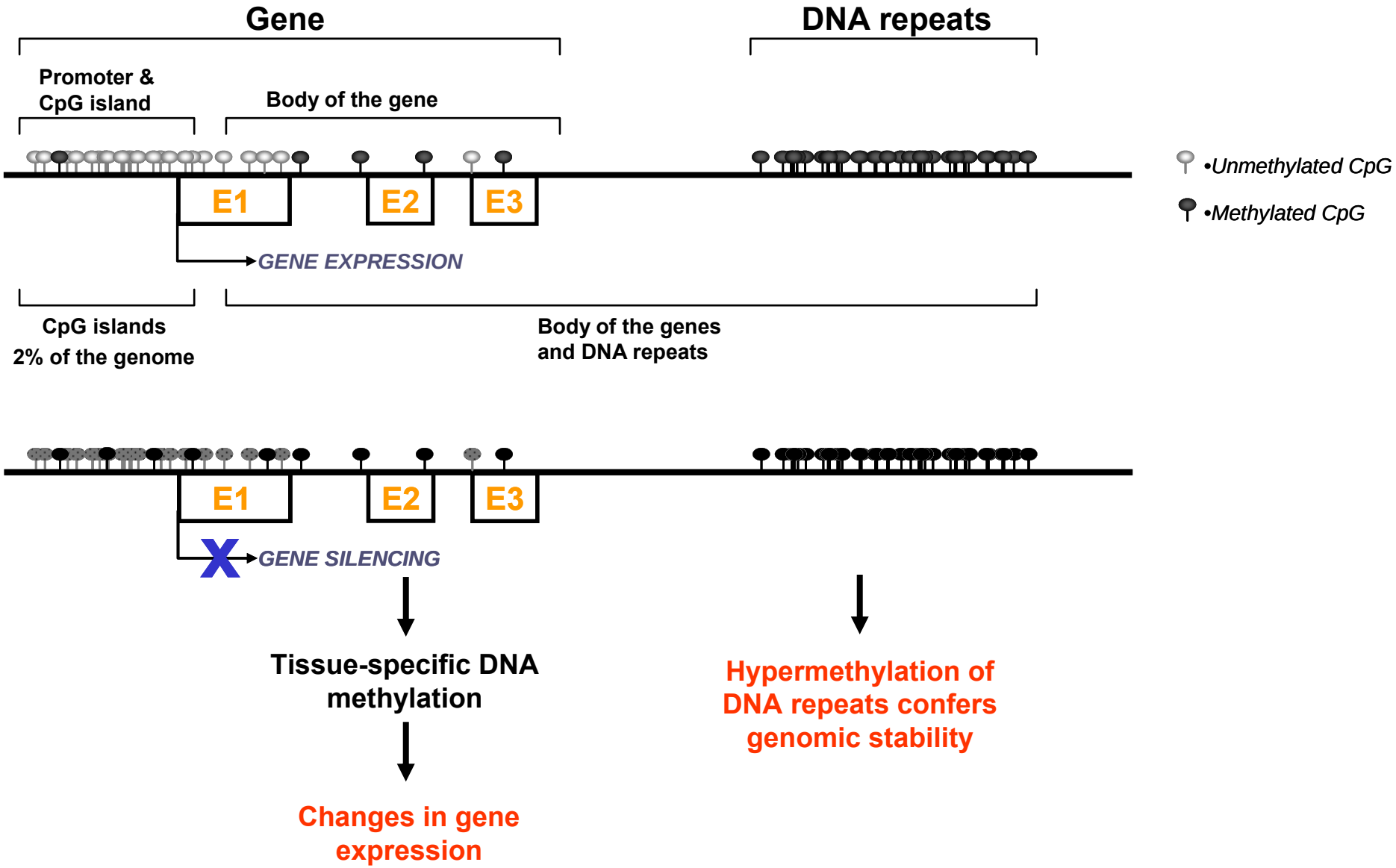
DNA methyltransferases (DNMTs)



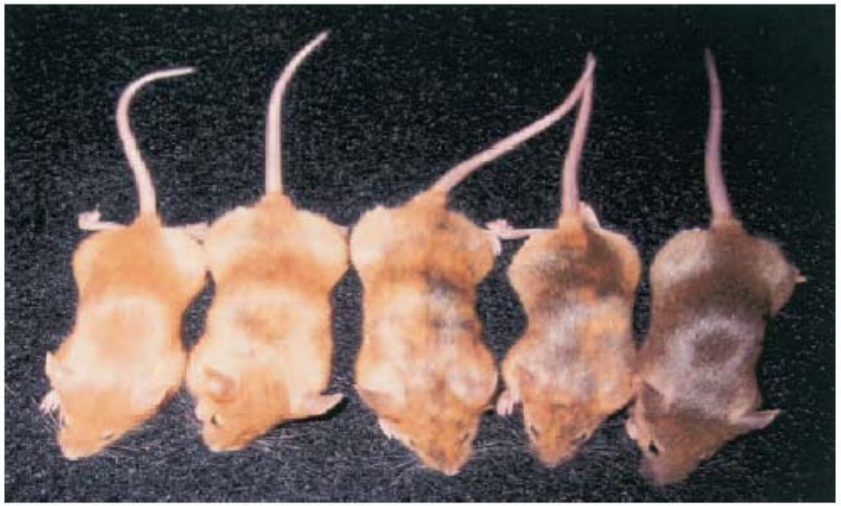
- DNMT1 → maintenance methyltransferases
- DNMT3a, DNMT3b → *de novo* methyltransferases

Bernhard Korn
Genomics & Proteomics
Core Facility

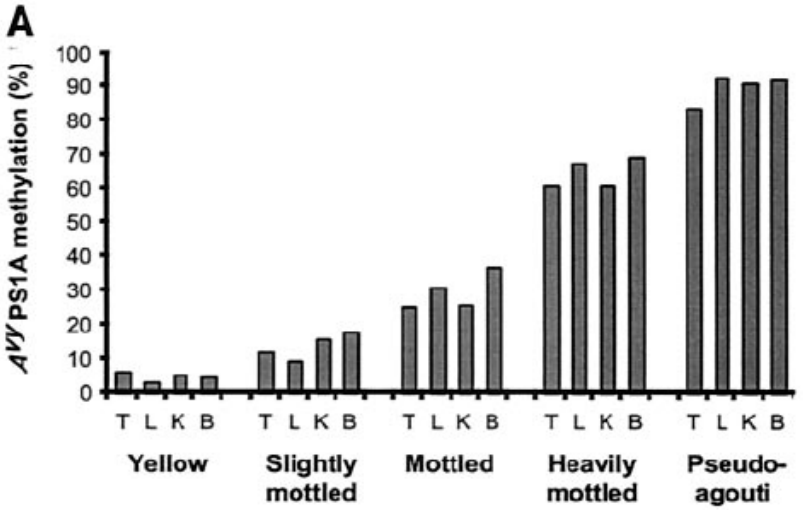
Where does DNA methylation occur in the genome?



Coat Color defined by A^{vy} Methylation



Geneticallly identical mice



EPIGENETICS AND EPISTASIS
Nutrition 20:63–68, 2004

Early Nutrition, Epigenetic Changes at Transposons and Imprinted Genes, and Enhanced Susceptibility to Adult Chronic Diseases

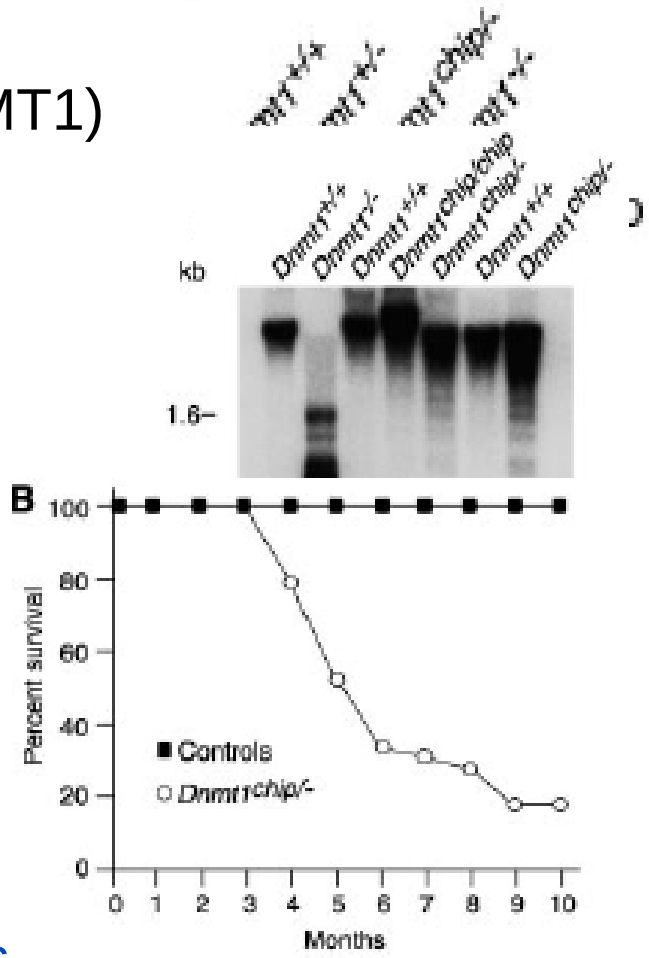
Robert A. Waterland, PhD, and Randy L. Jirtle, PhD
Department of Radiation Oncology, Duke University Medical Center, Durham, North Carolina, USA

Genomic Hypomethylation and Cancer

DNA methyltransferase 1 (DNMT1) down to 10%

Reduced DNA methylation

Poor survival rate



Mice develop T cell lymphomas
many have instable T cell receptor ß locus

Gaudet (2003), Science, 300, 489-92

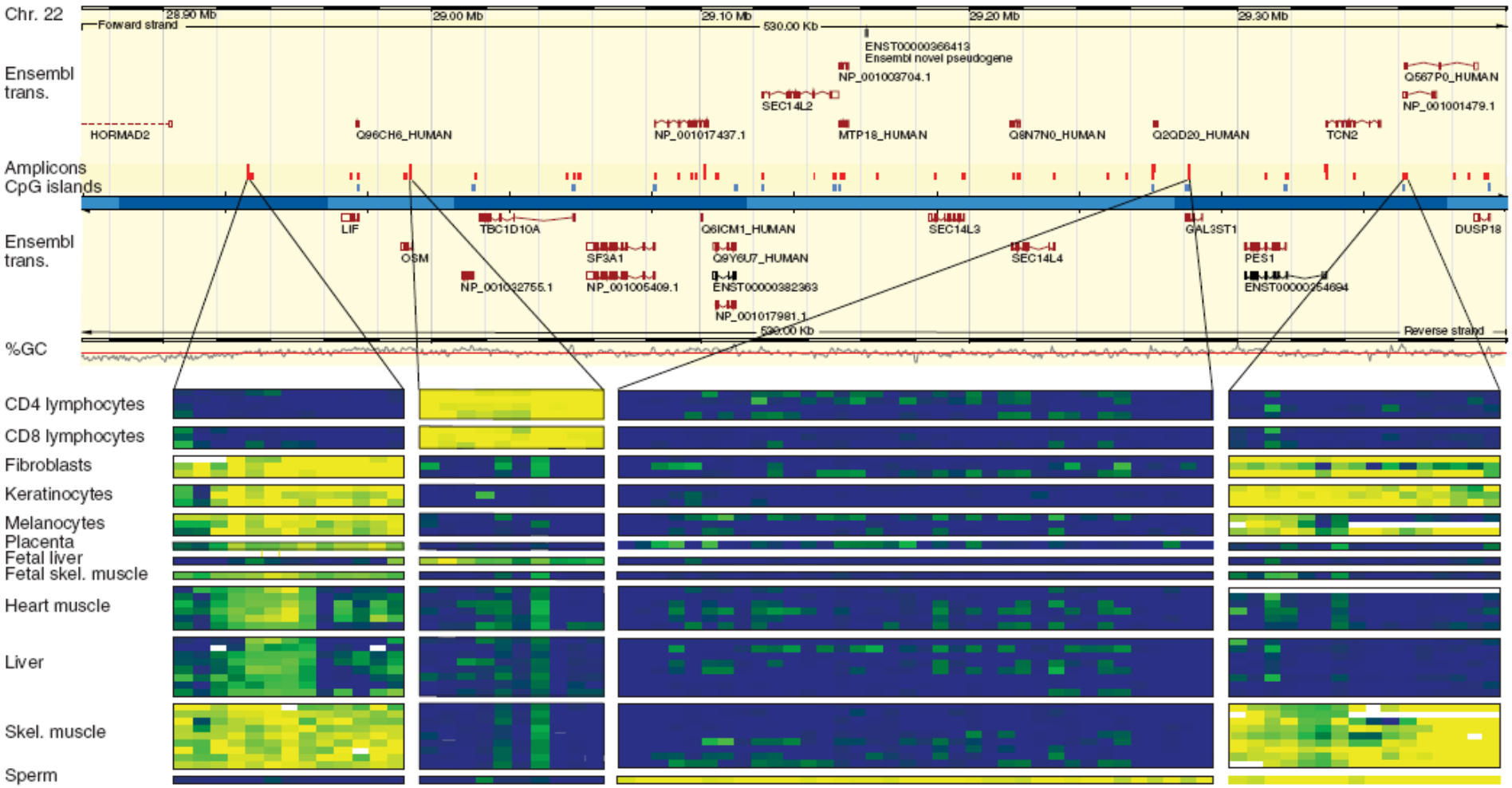
Genomic Hypomethylation and Cancer

- Which genes are involved?
- How is their promoter status?
- The methylation of which CpGs (pattern of CpGs) does have predictive information?

Chromosome-wide Methylation View

Eckhardt *et al.* Nat Genet 2006: 1.9 million CpGs in 12 different tissues

→ **Methylation within a promoter is homogeneous**



DNA Methylation Analysis

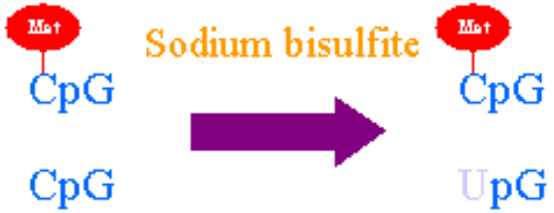
→ Methylation within a promoter is homogeneous

- Consequences
- Test individual CpGs!
  - → use genotyping assay
- Test only few CpGs per promoter?
  - Select CpGs carefully (stability of assay, functional relevance)
  - Conclude overall methylation status of each promoter
- Infinium assay for methylation analysis

Infinium HumanMethylation27 Beadchip

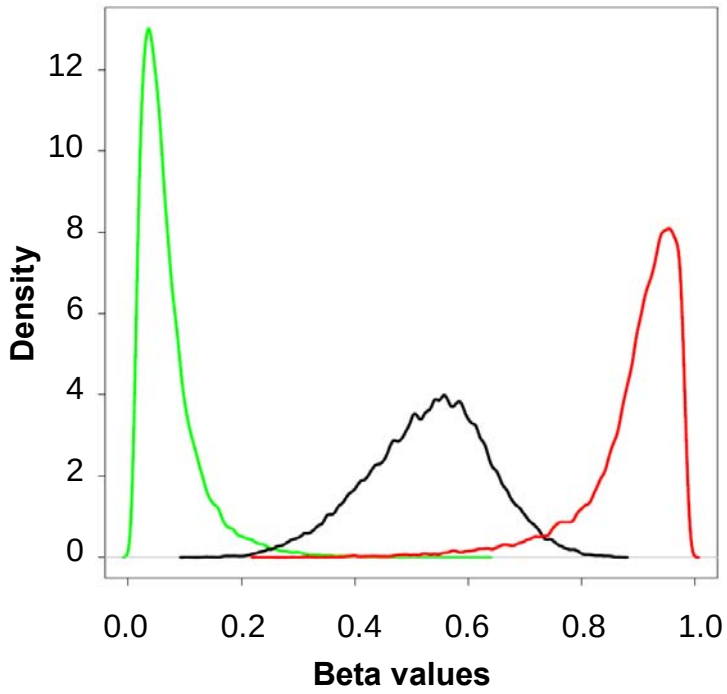
- 27.578 CpG sites (incl. 254 miRNA CpGs)
- Representing >14.000 promoters / genes
- 1-3 CpGs / promoter
- 12 arrays / BeadChip

- 1 µg genomic DNA
  - ~200 ng bisulfite treated DNA
  - Infinium assay

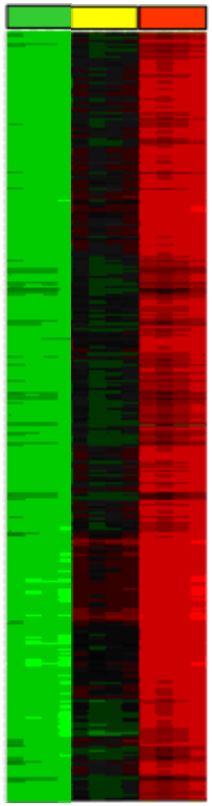


- Two-color fluorescent scanning
- Methylation measure: ß-value (0-1)

Do ß-Values reflect Methylation Levels?



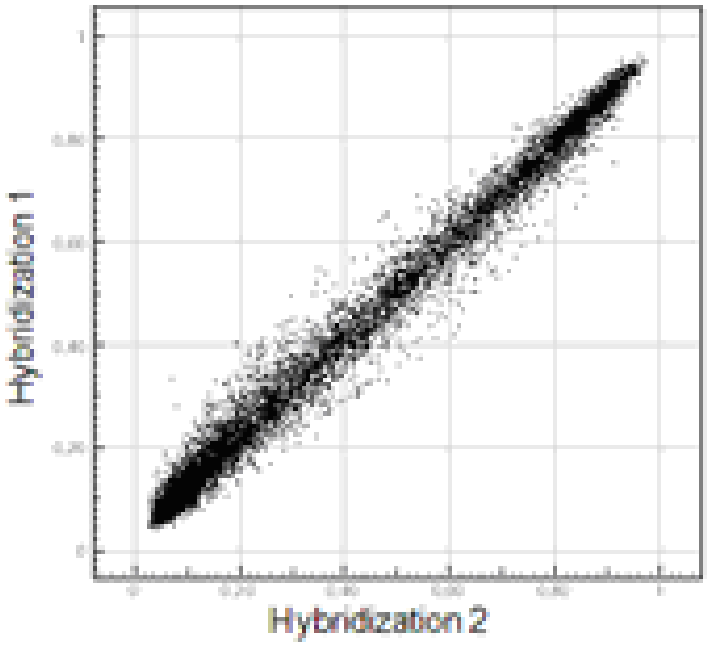
Density plot
mean ß: 0.070, 0.530, 0.894



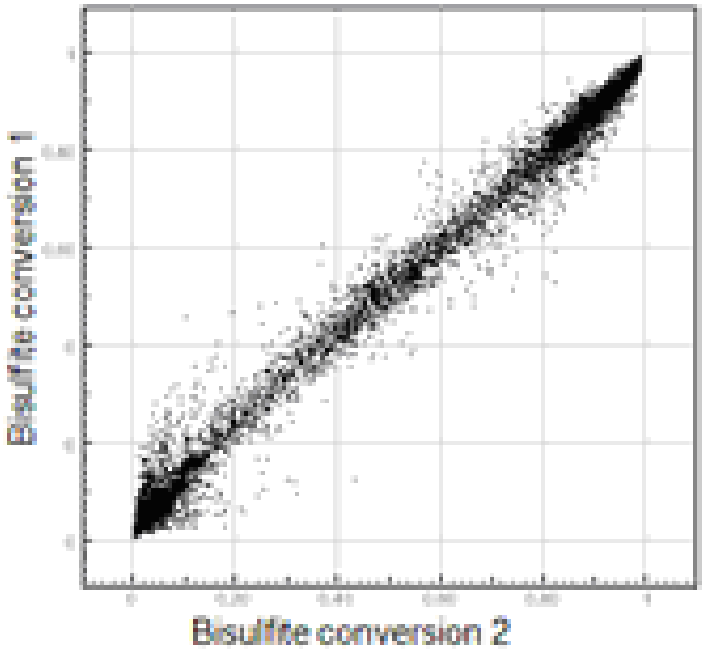
Cluster analysis of 26,492 CpGs

→ Assays are capable to discriminate methylation status
→ Quantitation is possible

Reproducibility and Robustness



Technical replicates, Hybs:
r^2=0.977 +/-0.011

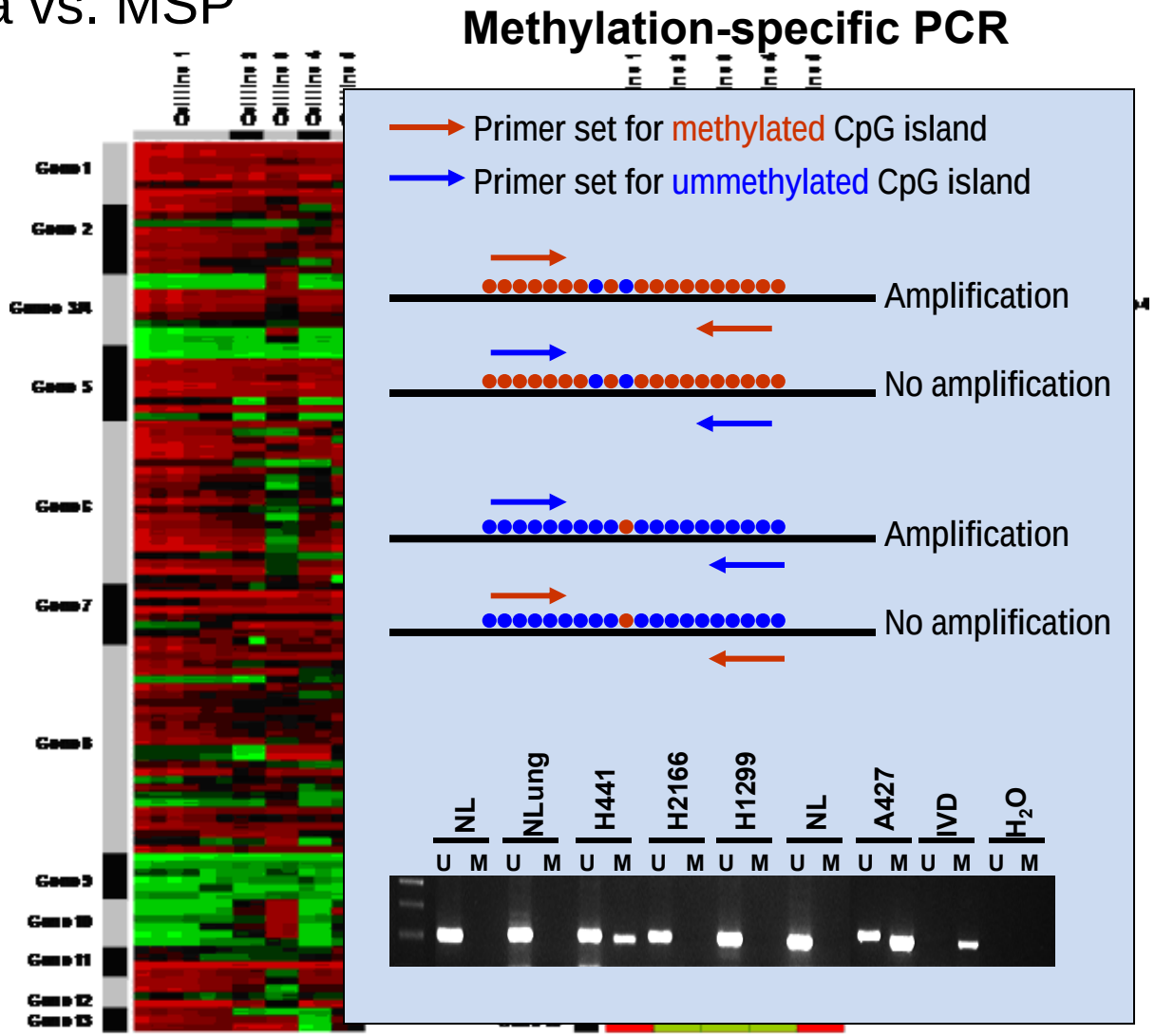


Technical replicates, bisulfite treatments:
r^2=0.955 – 0.992

→ Reproducibility is good
→ Assay outcome is robust (chemicals!)

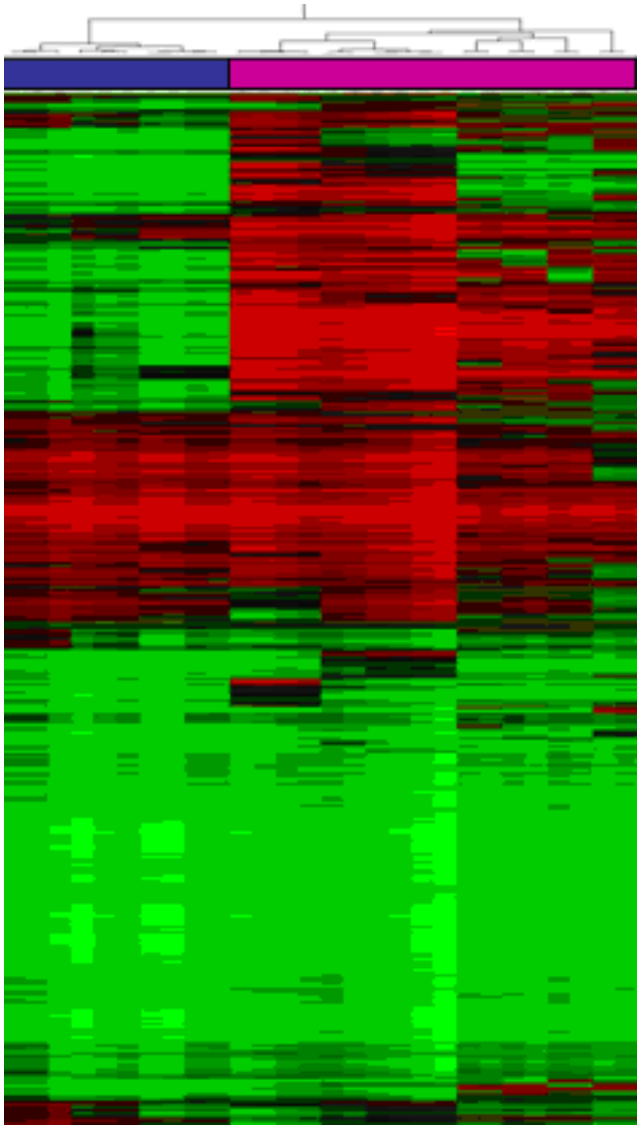
Validation of Results

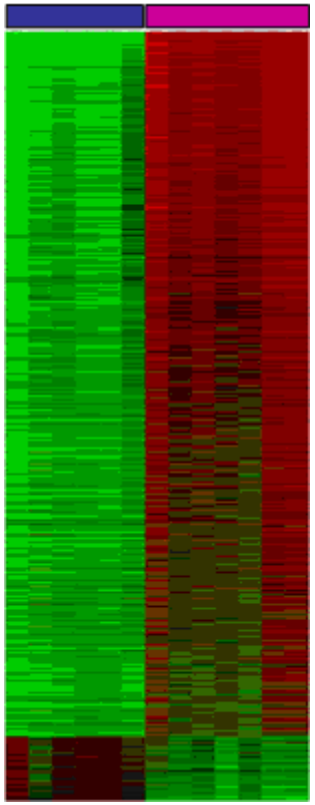
Infinium data vs. MSP



→ Strong correlation between Infinium and MSP (and bisulfite sequencing)

Meaningful Data for Cancer Research

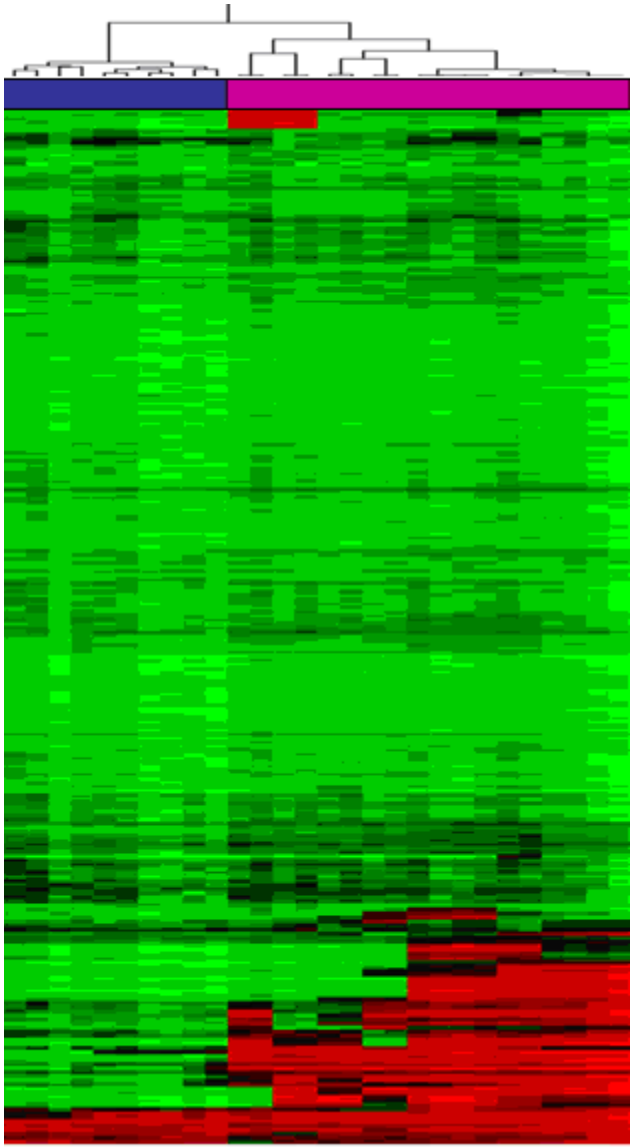




- 3,916 CpGs are hypermethylated
- 127 CpGs are hypomethylated

in lymphoma cell lines compared to normal controls; $p<0.001$

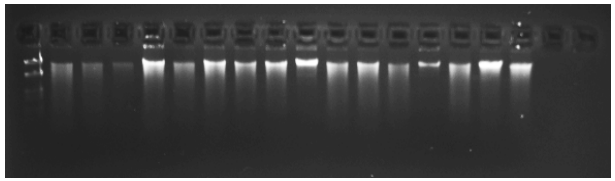
Are miRNAs regulated by Methylation?

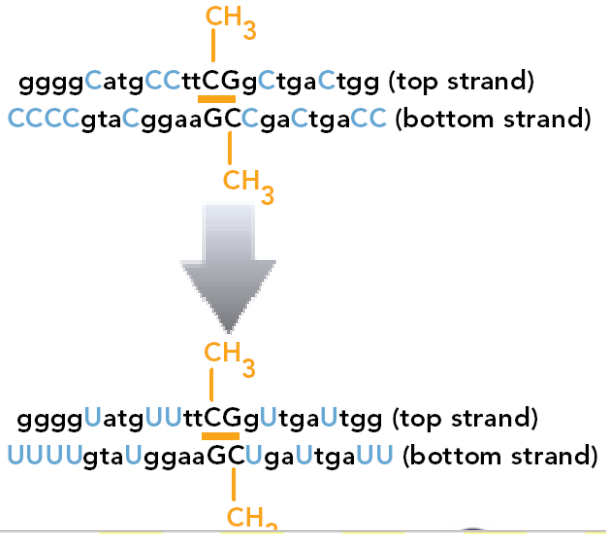


- 22/254 miRNA associated CpGs are hypermethylated
- No miRNA assoc. CpGs are hypomethylated

in lymphoma cell lines compared to normal controls; $p<0.001$

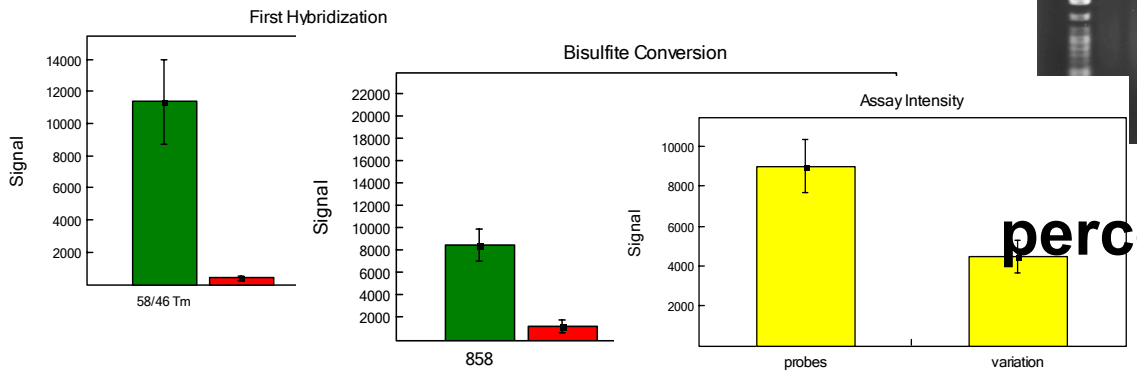
Methylation Analysis Pipeline

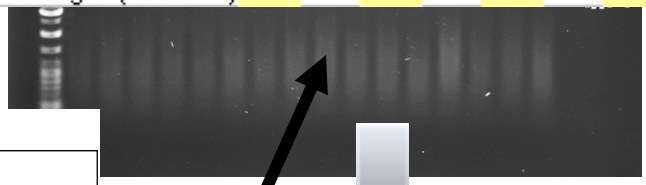




Excel-like Output with Links

SYMBOL	GeneCards-L	Entrez-Link	DIS	CP	ANNOTATION	(%	(%	(%	(%	(%	(%	(%	(%	(%
CDKN1A	GC_CDKN1A	NM_000389.2	-242	Y	Homo sapiens cyclin-dependent kinase inhibitor 1A	93	28	87	89	93	93	89	78	0
MAGEL2	GC_MAGEL2	NM_019066.2	-170	Y	Homo sapiens MAGE-like 2 (MAGEL2), mRNA.	97	98	97	98	97	91	97	98	24
MAP3K9	GC_MAP3K9	NM_033141.2	17	Y	Homo sapiens mitogen-activated protein kinase kina	85	8	62	64	88	96	69	47	3
PDGFB	GC_PDGFB	NM_002608.1	25	Y	Homo sapiens platelet-derived growth factor beta pol	89	29	78	83	91	91	83	64	0
PLXDC2	GC_PLXDC2	NM_032812.7	-914	Y	Homo sapiens plexin domain containing 2 (PLXDC2)	85	18	75	73	84	89	69	42	4





percent methylation

Golden Gate
Infinium

Summary

- DNA methylation plays crucial role in many diseases
- Genome-wide Infinium methylation assays are
  - Reproducible
  - Robust
  - Can be vaildated by other techniques
  - Allow accurate quantitation of methylation
- Useful for hypothesis-free screening and in-depth analysis

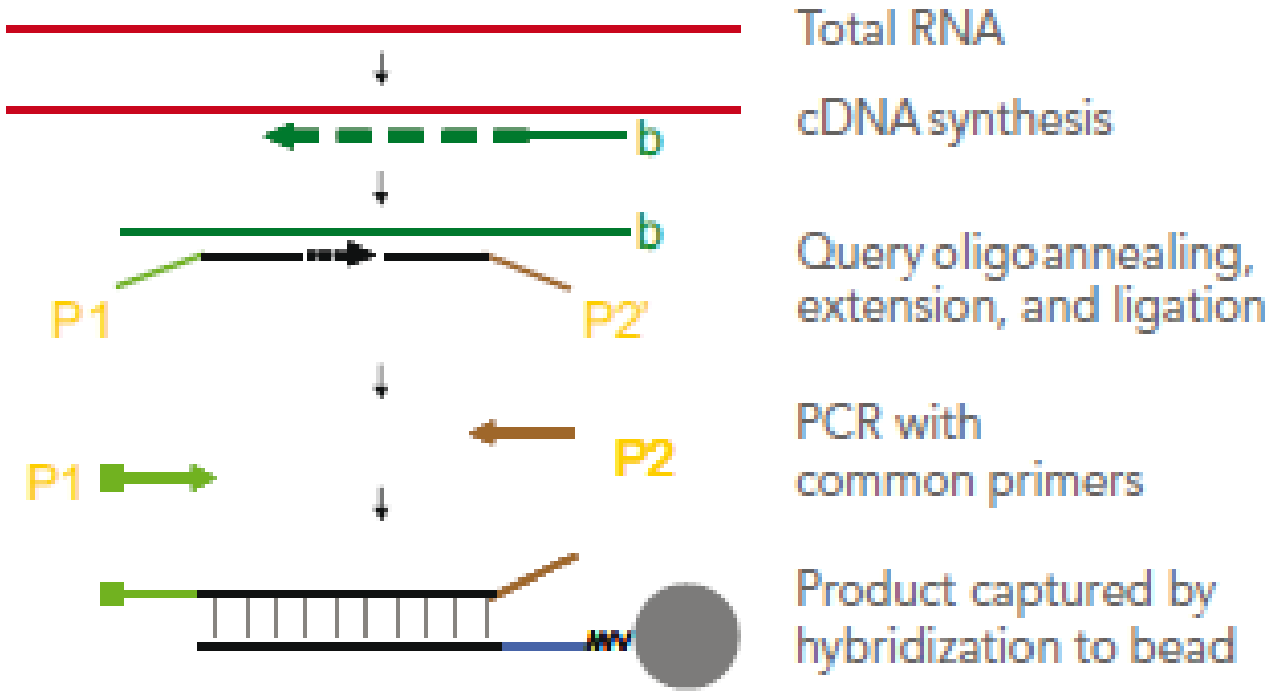
Expresssion Profiling of Formalin-fixed Parafin Embedded Samples

- Archived patient tissue
- Follow-up data
- Large patient cohorts
- Ease of storage (pathology)
- Microdissection

- **BUT**
- RNA quality
  - Degradation
  - Cross-link
- cDNA synthesis
- Labelling
- Relative probe position

DASL Technology

- Proven for GT/customized expression profiling
  - Address code
- Now possible on whole genome basis?



adapted from Illumina

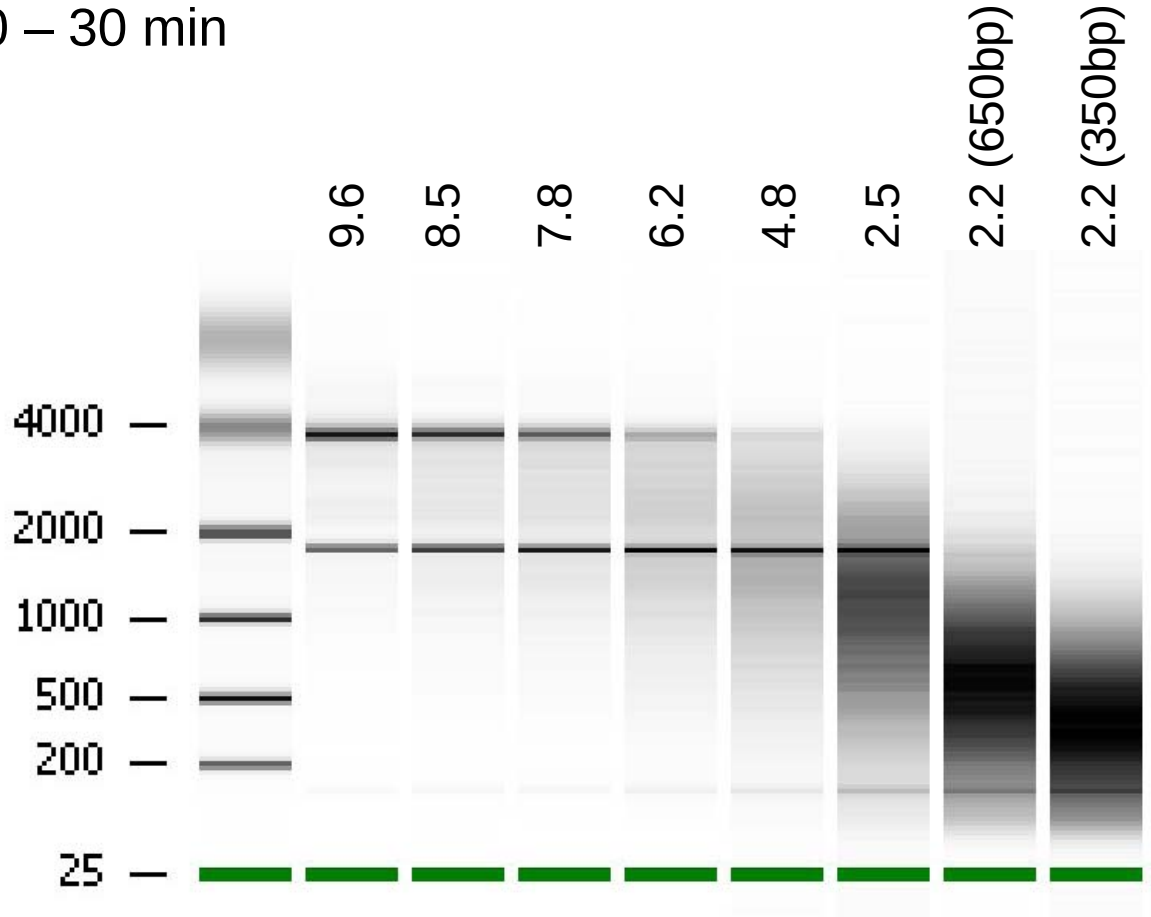
Whole Genome DASL

- Reproducibility/Robustness
  - Technical
  - RIN dependance
  - Influence of RNA amounts
  - IVT comparison

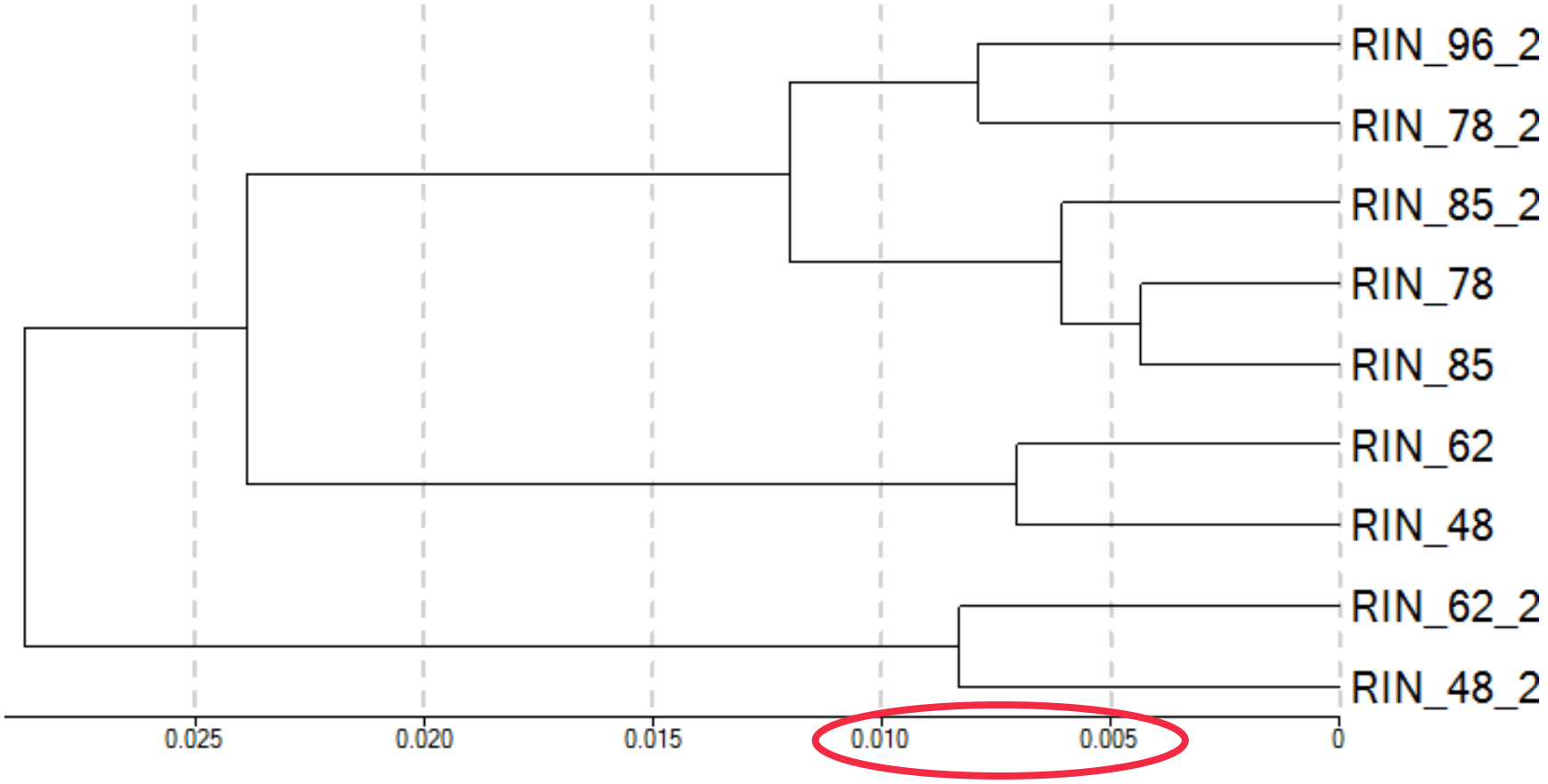
- Setup:
  - Artificial degradation series of HeLa total RNA
  - Technical replicates
  - Detecion level

Degradation Series

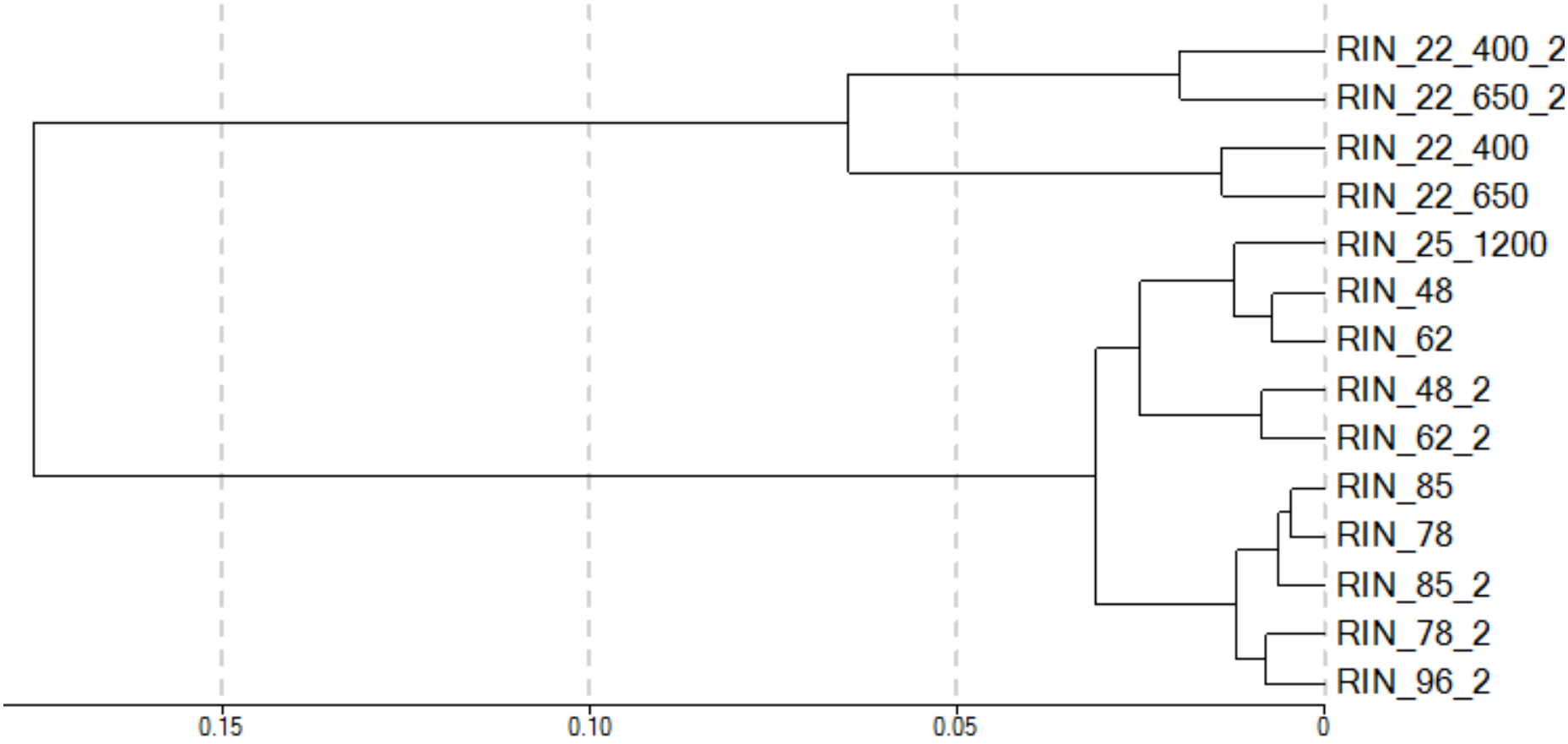
- Boiling total RNA
  - 0 – 30 min



Clustering of WG-DASL



Clustering of WG-DASL



- Correlation drops, when comparing high and low RIN

WHY?

Number of Genes expressed

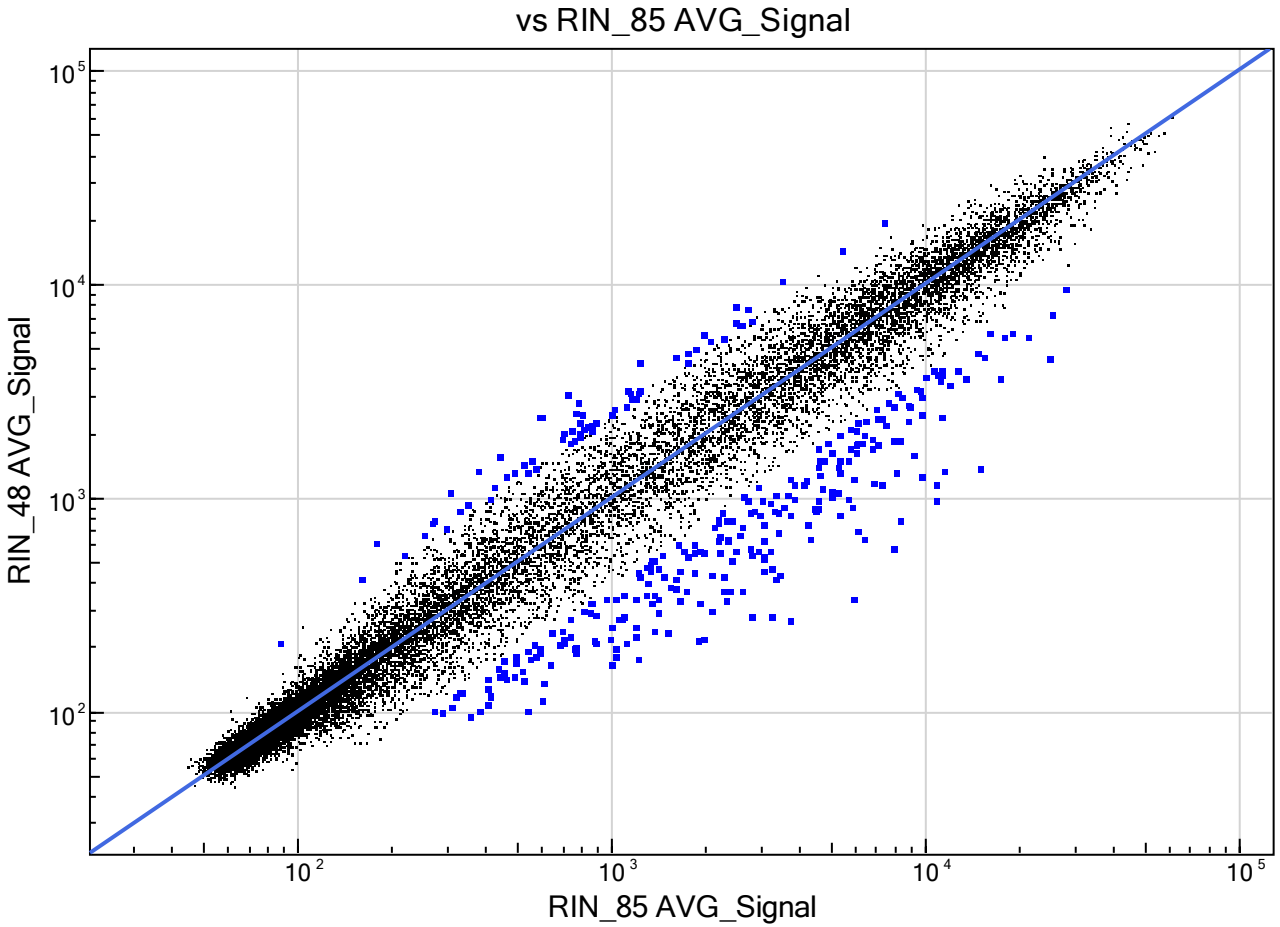
		exp p-value		
	0.001	0.005	0.01	0.05
RIN 9.6	9,101	9,224	10,369	12,015
RIN 8.5	8,466	8,525	9,421	11,158
RIN 7.8	8,488	8,494	9,778	11,545
RIN 6.2	8,111	8,258	9,055	10,559
RIN 4.8	8,238	8,345	9,256	10,749
RIN 2.5 (1200 bp)	7,893	7,921	8,913	10,112
RIN 2.2 (650 bp)	7,933	8,183	8,579	10,015
RIN 2.2 (350 bp)	7,037	7,494	7,653	9,963

Reduction of RIN → failure to detect transcripts

Data Correlation

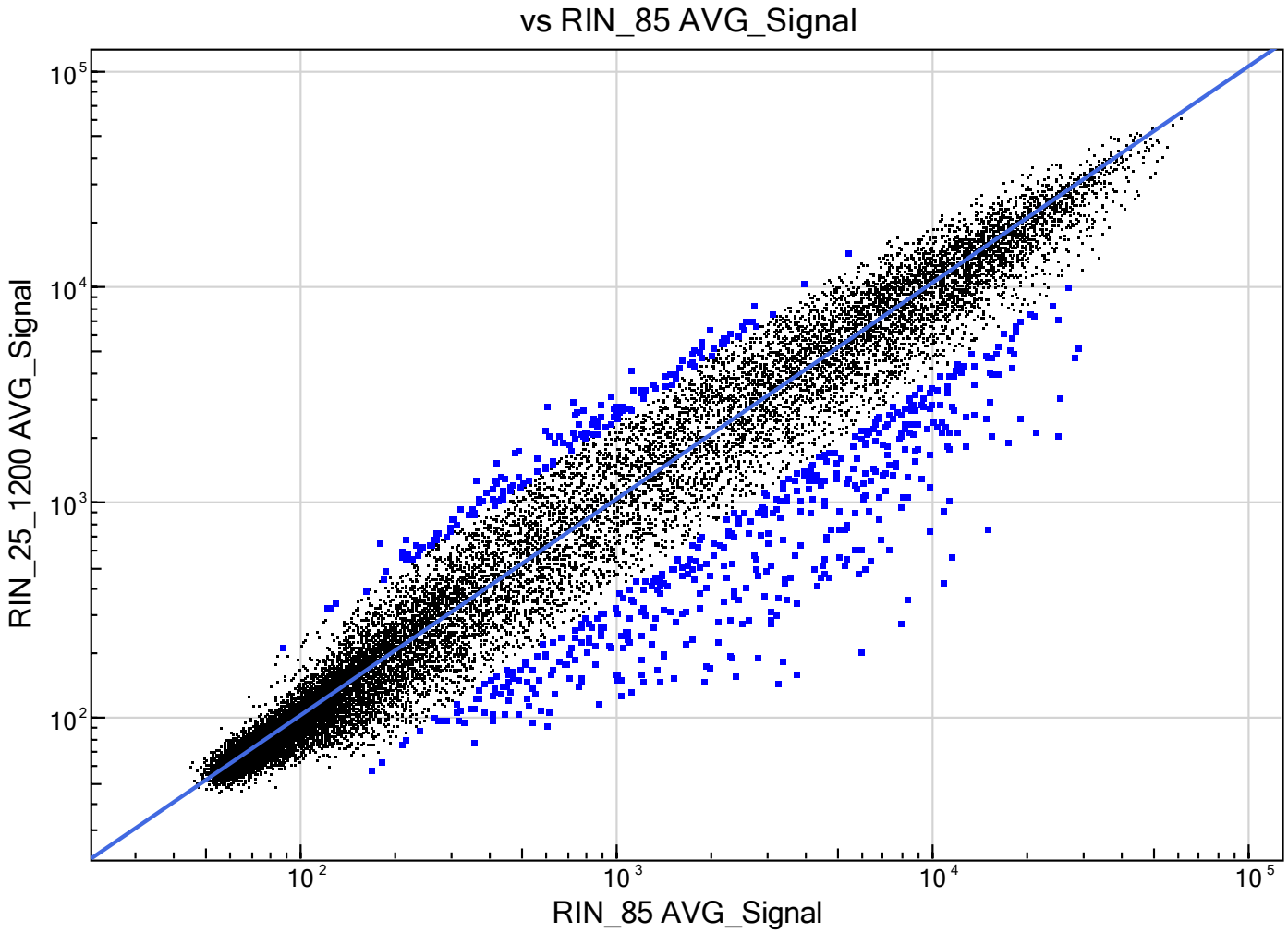
		Diff genes							
		RIN 9.6	RIN 8.5	RIN 7.8	RIN 6.2	RIN 4.8	RIN 2.5	RIN 2.2	RIN 2.2
Pearson r^2	RIN 9.6		59	107	254	366	763	2546	2675
	RIN 8.5			0	4	88	224	1834	2021
	RIN 7.8				1	64	213	1953	2205
	RIN 6.2					3	99	1545	1980
	RIN 4.8						0	1037	1631
	RIN 2.5 (1200 bp)							548	1199
	RIN 2.2 (650 bp)								11
	RIN 2.2 (350 bp)								

Scatter Plot



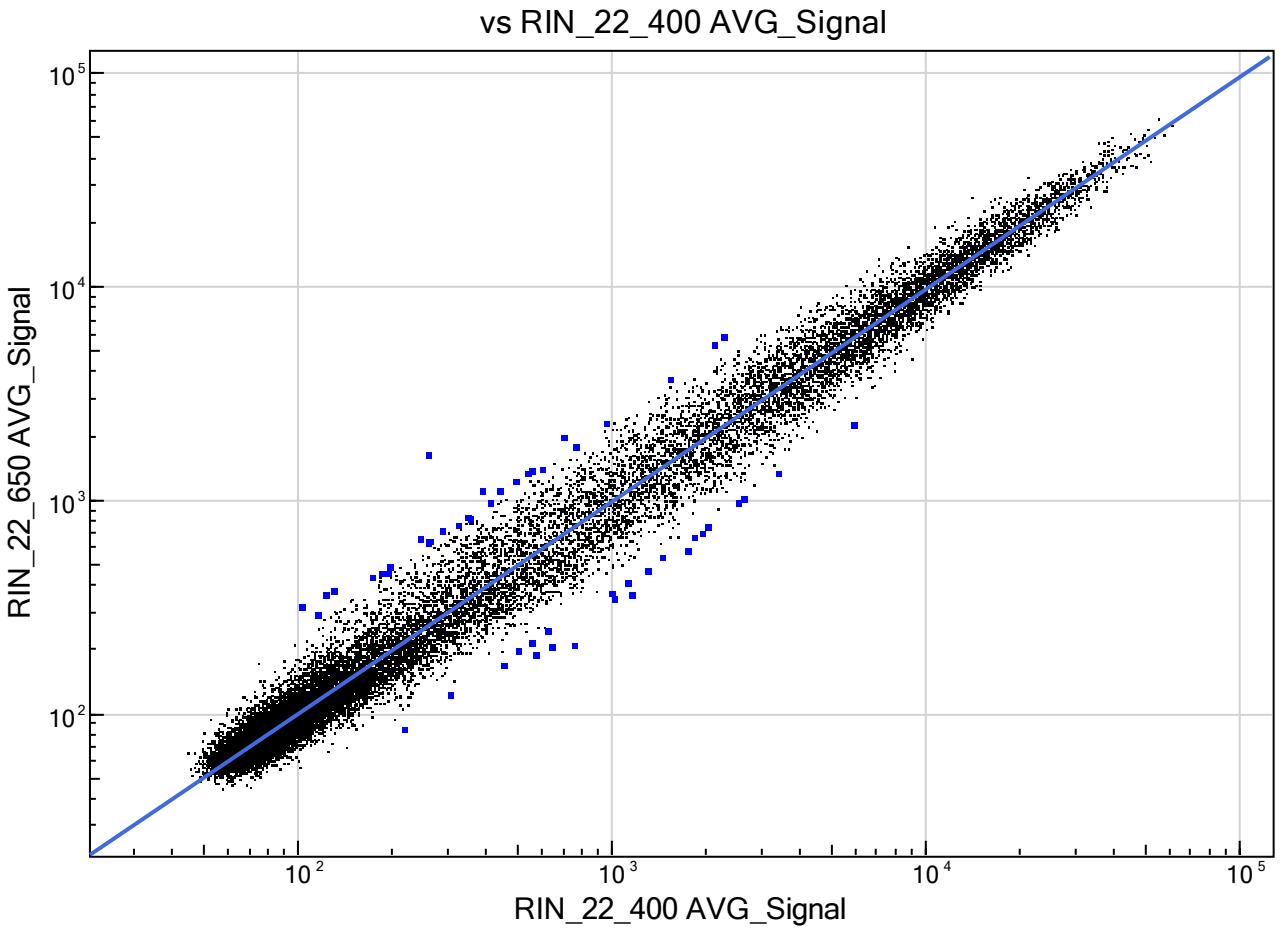
Decreasing RIN, reduced detection

Scatter Plot



Don‘t compare samples of highly different RINs

Correlation at low RIN



Excellent correlation with ‚fully' degraded RNA

Gains and Losses

	Transcripts detected	Transcripts overlap	Percentage overlap	Transcripts new	Percentage new
RIN 9.6	9,101	9101	100	0	0.0
RIN 8.5	8,466	8334	91.6	132	1.6
RIN 7.8	8,488	8362	91.9	126	1.5
RIN 6.2	8,111	7988	87.8	123	1.5
RIN 4.8	8,238	8039	88.3	199	2.4
RIN 2.5 (1200 bp)	7,893	7697	84.6	196	2.5
RIN 2.2 (650 bp)	7,933	7522	82.7	411	5.2
RIN 2.2 (350 bp)	7,037	6769	74.4	268	3.8

- High reproducibility
- Low ‚false positives‘

Genes ‚regulated': RIN 9.6 – 2.2

Symbol	ProbeID	regulation'
ACTN4	270437	-8.8
ANXA2	770730	-9.0
ATP13A2	7100326	-5.7
ATP5G2	3190711	-7.7
BDP1	7570563	-6.2
CARM1	460382	-4.8
CDV3	5720746	-29.5
CFH	3360279	-16.9
CSNK2B	3450156	-6.8
DKFZp667G2	830397	-18.7
EIF4G2	4290446	-3.2
FAM127B	1070347	-6.2
GPX4	6520128	-8.3
KBTBD4	5700017	-24.1
KRIT1	7610477	-27.4

Symbol	ProbeID	regulation'
MGC71993	5670661	-4.4
MTX1	2750551	-6.2
NDUFA3	50240	-5.6
PACS1	4540164	-20.7
PHB2	4060722	-10.1
RALGDS	2140735	-6.6
RAVER1	7560328	-6.3
REST	1090139	-30.6
RGPD5	1340541	-7.9
RNF38	1230632	-17.1
RPL29	2450167	-16.0
RPS15A	7100717	-22.0
SDCCAG1	5220541	-16.8
SPOP	6370341	-19.2
XPO6	6420189	-8.9
XRN1	110338	-17.3
ZWILCH	780114	-11.0

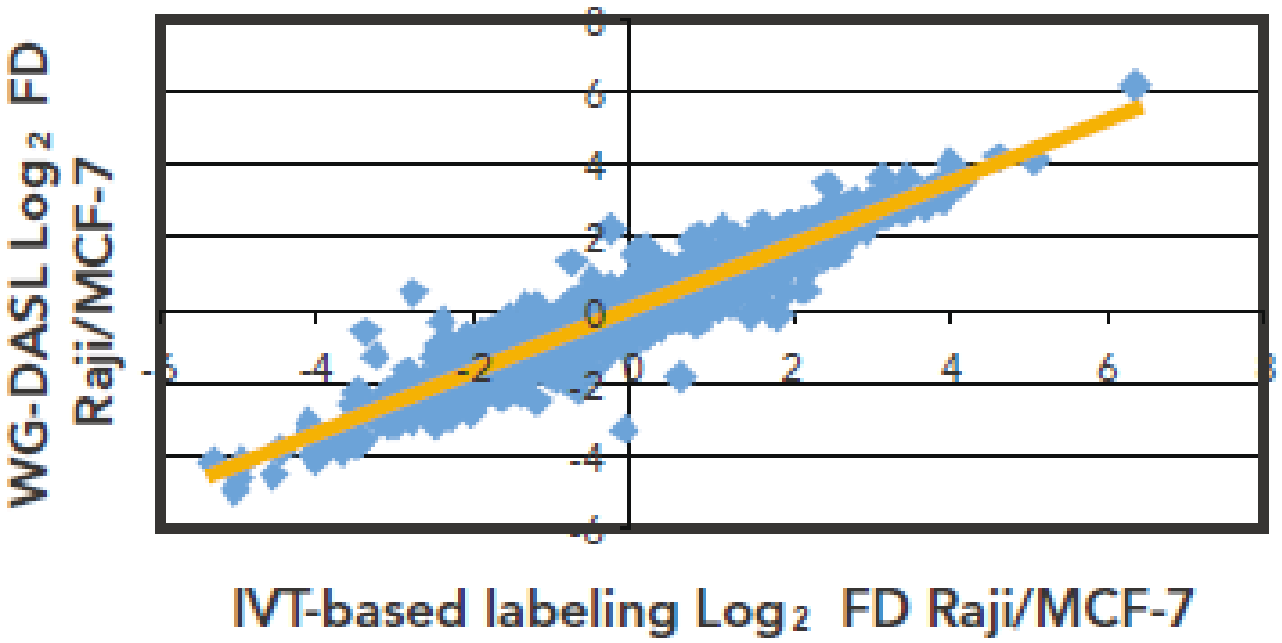
Sensitivity to RNA Amounts

	r^2
200 ng	1.0000
100 ng	0.9912
50 ng	0.9915
25 ng	0.9878
10 ng	0.9345
5 ng	0.8872
200 ng (IVT)	0.6599

- Assay is highly robust to varying amounts of RNA
- No comparison to standard IVT assay possible

Fold-changes: WG-DASL vs. IVT

$r^2 = 0.8931$



- Reproducibility on the level of expression changes!

data from John Quackenbush, Dana Faber

Summary WG-DASL

- Highly robust assay: technical replicates
  r^2 = 0.98 - 0.99 (RIN dependent)
- Only RNAs of similar RIN should be compared
  - Artifically adjust RNA quality?
  - Use more biological replicates!
- High sensitivity: significant and reproducible result down to 25 ng total RNA
- Good concordance with IVT results, based on fold-changes
- Reproducible loss of number of transcripts detected, with decreasing RIN – can we correct for it?
- Requirements
  - Highly standardized procedure
  - Calibrated equipment
- Biased RNA deradation in FFPE samples? - unsolved

Outlook

- WG-DASL Core service for
  - FFPE samples
  - Microdisected samples
  - (partially) FACS samples
- Needs and Wishes
  - WG-DASL for Sentrix-6
  - Reproducibility down to smaller amounts of RNA (<1000 cells)
  - Approach for direct labelling (avoiding RNA isolation)

Acknowledgements

- University Kiel
  - Ole Ammerpohl
  - José Martin-Subero
  - Julia Richter
  - Reiner Siebert
- German Cancer Res. Center
  - Tamara Fries
  - Roger Fischer
  - Sabine Henze
  - Bernhard Korn