



Microarrays

Inhaltsverzeichnis

1. Microarray

- Herstellung
- Aufbau
- Arten
- HGU133Plus2
- Grundbegriffe
- Genexpressionsanalyse

2. RMA-Normalisierung

3. Qualitätskontrolle

- Histogramm
- RNA-Degradation-Plot
- Quality-Control-Plot

Herstellung von Microarrays

- mehrere Verfahren, die sich in zwei Techniken zuordnen lassen
 1. in situ-Synthese
 2. externe Synthese

in situ-Synthese

- Probe wird auf dem Microarray hergestellt
- Nukleotidkette wird Schrittweise verlängert
- kommt insbesondere bei Oligonucleotidarrays zur Anwendung
 - > da Sequenzgenauigkeit wichtig ist und synthetisierende Kette sehr kurz ist (Aufwand dieser Methode wächst mit der Kettenlänge)

externe Synthese

- DNA-Sonden (Probes) werden extern durch PCR oder Klonen hergestellt und mittels eines Roboters zielgenau auf die richtige Zelle des Microarrays aufgebracht
- hierbei handelt es sich um längere Ketten
- dabei entstehende cDNA-Arrays werden vor allem zur Genexpressionsanalyse eingesetzt

Aufbau

- werden allgemein benutzt um zu ermitteln ob Gene exprimiert sind (kodieren für Proteine)
- hauptsächlich 2 Arten von Chips:
 1. cDNA Microarrays
 2. Oligonucleotide Microarrays (Affymetrix)
- Spots auf dem Chip müssen einzigartig für ein Gen sein (Spezifität)
- Spots müssen dieses Gen entdecken (Sensitivität)
- alle Spots sollten unter den gleichen Bedingungen funktionieren, z.B. Temperatur (Einheitlichkeit)

Aufbau von cDNA-Chips

- 10000 Gene auf einem Chip
- Oberfläche ist aus beschichtetem Glas
- hohe Sensitivität, wegen langer Probe-Sequenzen
- aber niedrige Spezifität
- Probes können nicht zwischen ähnlichen oder gleichen Subsequenzen unterscheiden
- dadurch auch nicht zwischen Gen-Familien
- mehrere cDNAs für ein Gen

Aufbau von Oligo-Chips

- 9000 Gene auf einem Chip
- Oberfläche für Oligos ist aus beschichtetem Glas oder Silizium
- Oligos sind in Arrays angeordnet (65000-500000 auf einem Chip)
- jedes Array (Spot) „erkennt“ ein Gen
- extrahierte Gene (targets) lagern sich an Oligos (probes) an -> Hybridisierung

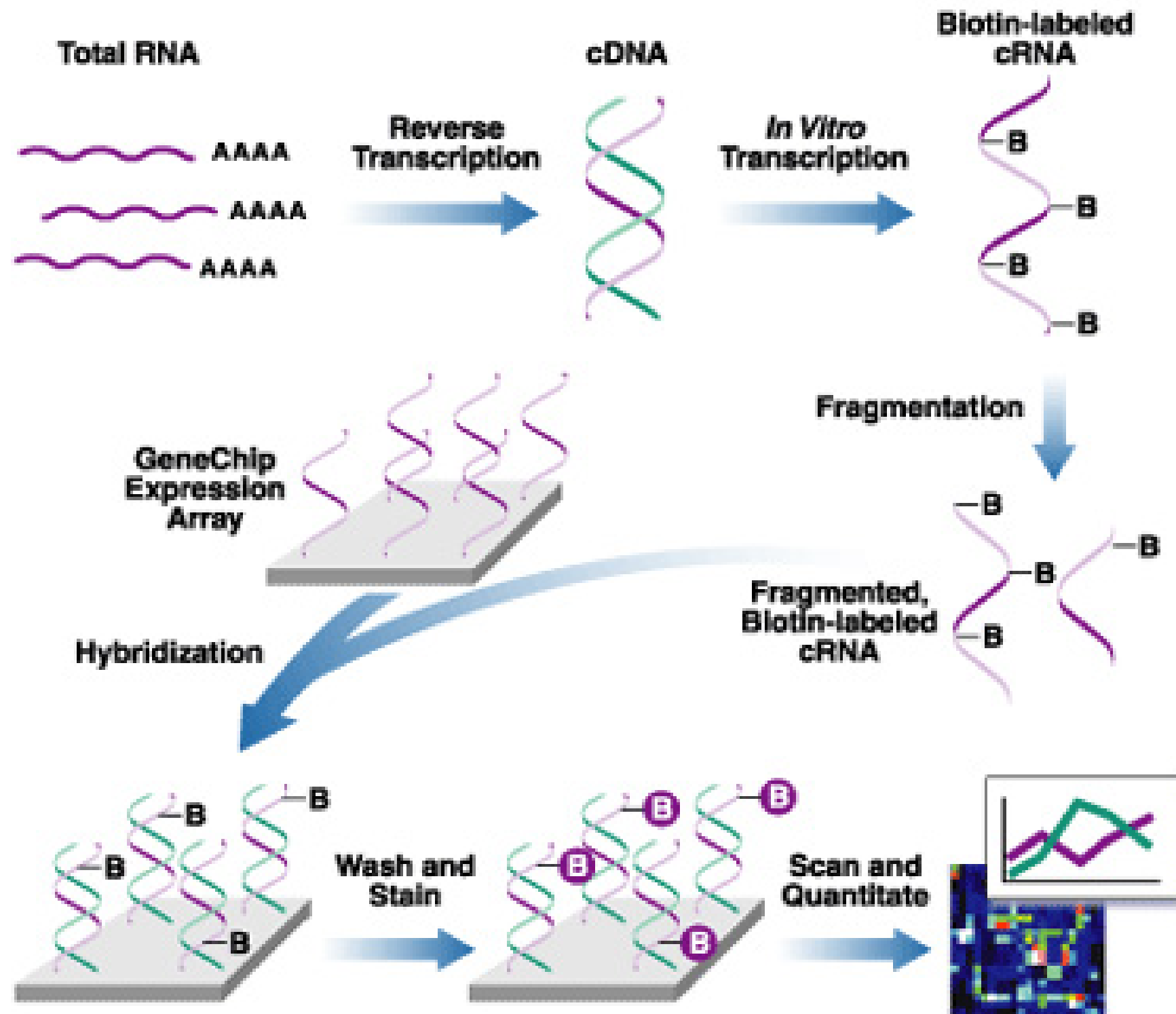
HGU133Plus2

- ermöglicht eine umfassende Analyse von genomweiten Expressionen auf einem einzelnen Array
- analysiert das Expressionslevel von über 47.000 Transkripten und Varianten, einschließlich 38.500 gut charakterisierte, menschliche Gene
- besteht aus >54.000 Probe sets und 1.300.000 verschiedenen Oligonucleotideigenschaften

Grundbegriffe

- Probe: einzelsträngiges DNA-Oligonukleotid, welches direkt auf der Oberfläche des Genchip-Arrays synthetisiert wird. Die 25-Basen Oligonukleotid wurde entworfen um komplementär zu einem spezifischen Gen-Transkript zu sein.
- Probe set: eine Reihe von Sonden, die entwickelt wurde, um ein Transkript zu erkennen. Ein Probeset besteht normalerweise aus 11-20 Probe pairs, z.B. ein 1 Probe pair set ist zusammengesetzt aus 11 PM und 11 MM. Neuere Array Designs von Affymetrix, z.B. HG-U133, enthalten Probe sets mit 11 Probe pairs. Ältere Designs haben eine mittlere Probe set Anzahl von 16 oder 20 Probe pairs.
- Probe pair: zwei Probe cells, eine PM und der dazugehörige MM. Auf dem Probe Array wird ein Probe pair mit einer PM-Zelle direkt über eine MM-Zelle angeordnet.
- Perfect Match (PM): Proben, die entworfen wurden, um komplementär zu einer Bezugssequenz zu sein
- Mismatch (MM): Proben, die entworfen wurden, um komplementär zu einer Bezugssequenz zu sein, mit Ausnahme einer homomeren Fehlpaarung an der zentralen Position (z.B. 13.Position von 25 Basenproben, A->T oder G->C). Mismatch Probes dienen als Kontrolle für eine Kreuz-Hybridisierung.

Genexpressionsanalyse Ablauf



Normalisierung Allgemeines

- Ziel: Vergleich Genexpressionslevel von verschiedenen Experimenten
- Grundidee: entferne systematische Fehler in den Daten, erhalte aber die Variation in der Genexpression, die aufgrund von biologisch relevanten Veränderungen in der Transkription auftritt

RMA (Robust Multichip Average)-Normalisierung

1. Hintergrundkorrektur
2. Quantil-Normalisierung
3. Zusammenfassung

Hintergrundkorrektur

- entfernt lokale Artefakte und „Rauschen“ der einzelnen Microarrays
- Annahmen: $PM = \text{Signal} + \text{Hintergrundsignal}$
- Korrigierte $PM = E(\text{Signal} \mid PM)$
- Hintergrundwerte sind normalverteilt und Signale sind exponentialverteilt

Quantil-Normalisierung

- entferne Arrayeffekte -> Vergleich Expressionswerte verschiedener Microarrays
- vergleicht Expressionslevel zwischen den Microarrays für verschiedene Quantile
- nutzt q-q-Plot
- robust gegenüber Ausreißern

Zusammenfassung

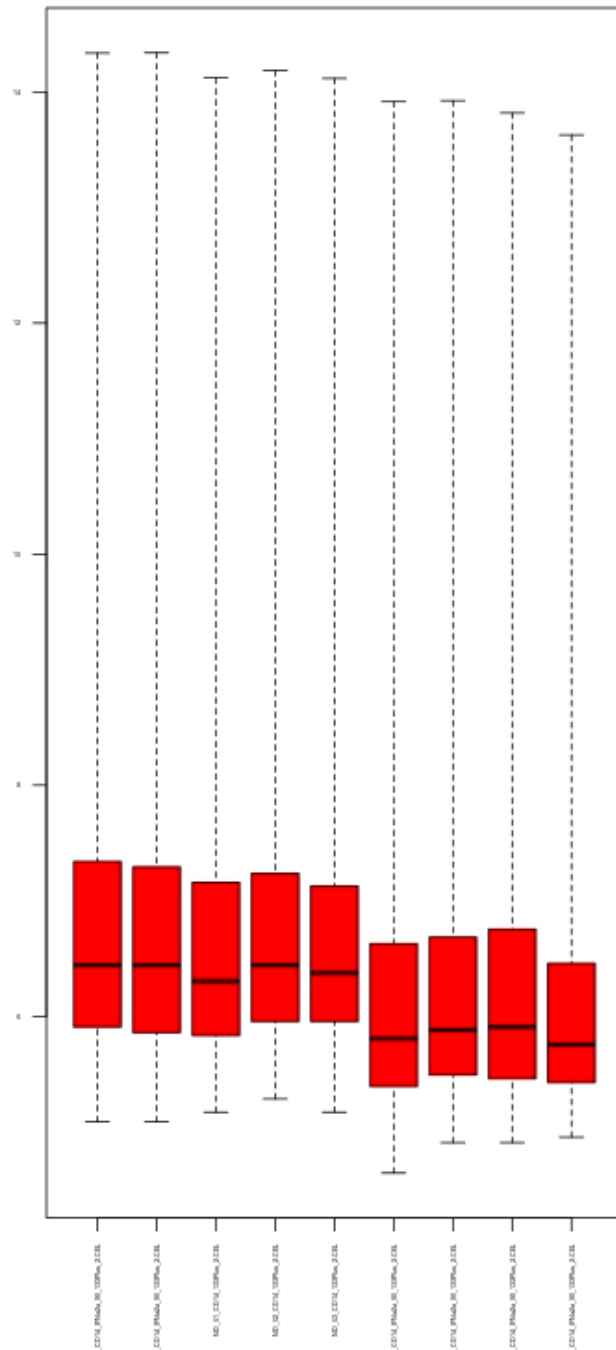
- PM Werte werden log2-transformiert und mittels linearem Modell zu einem Genexpressionswert (pro Probeset) zusammengefasst

		mRNA Samples					
Gene		sample1	sample2	sample3	sample4	sample5	...
	1	0.46	0.30	0.80	1.51	0.90	...
	2	-0.10	0.49	0.24	0.06	0.46	...
	3	0.15	0.74	0.04	0.10	0.20	...
	4	-0.45	-1.03	-0.79	-0.56	-0.32	...
	5	-0.06	1.06	1.35	1.09	-1.09	...

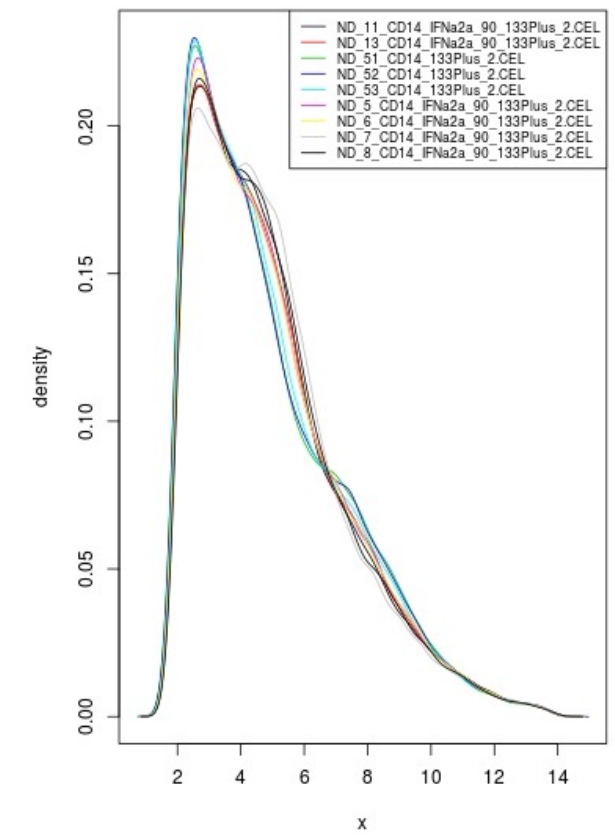
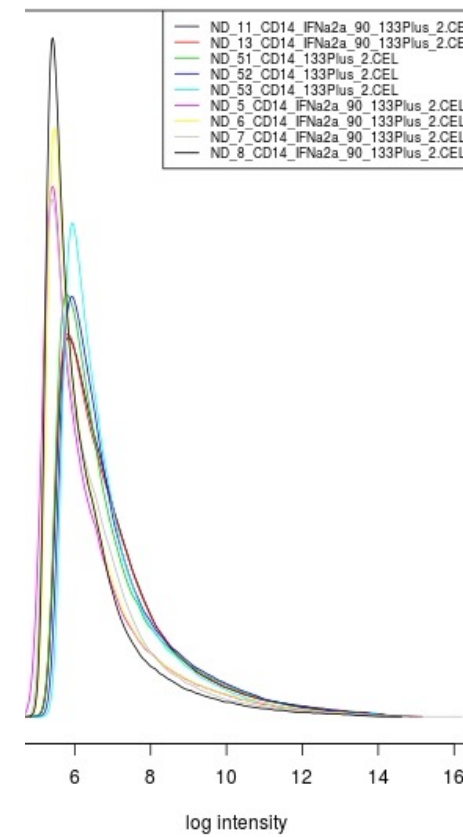
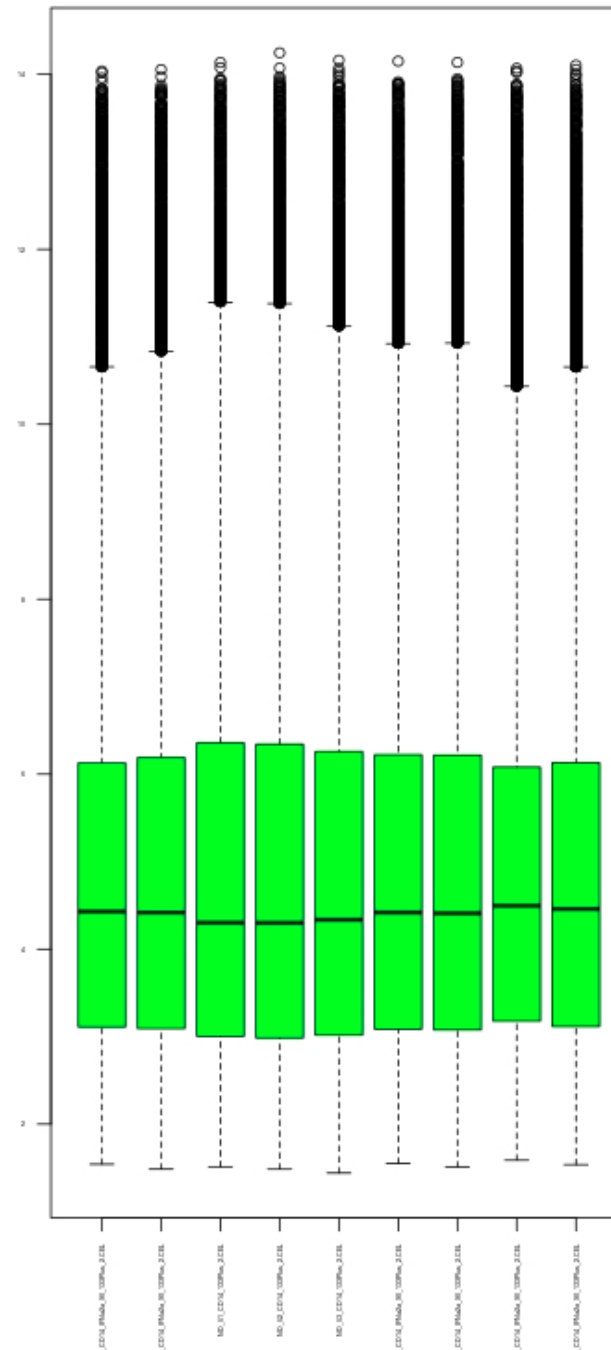
Genexpression Level for Gen i in mRNA sample j

Überprüfung RMA-Normalisierung

raw data



RMA normalisierte Daten

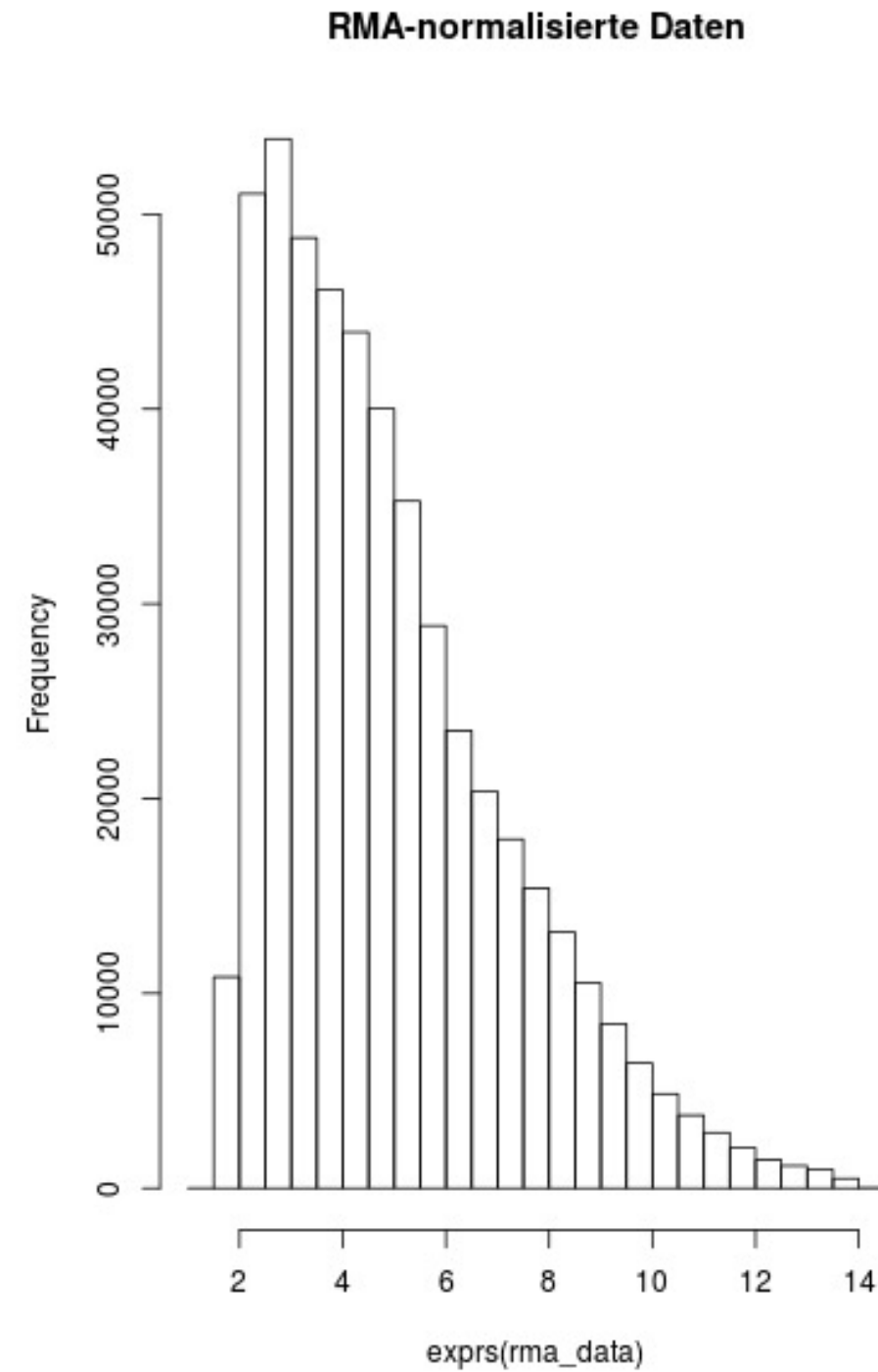
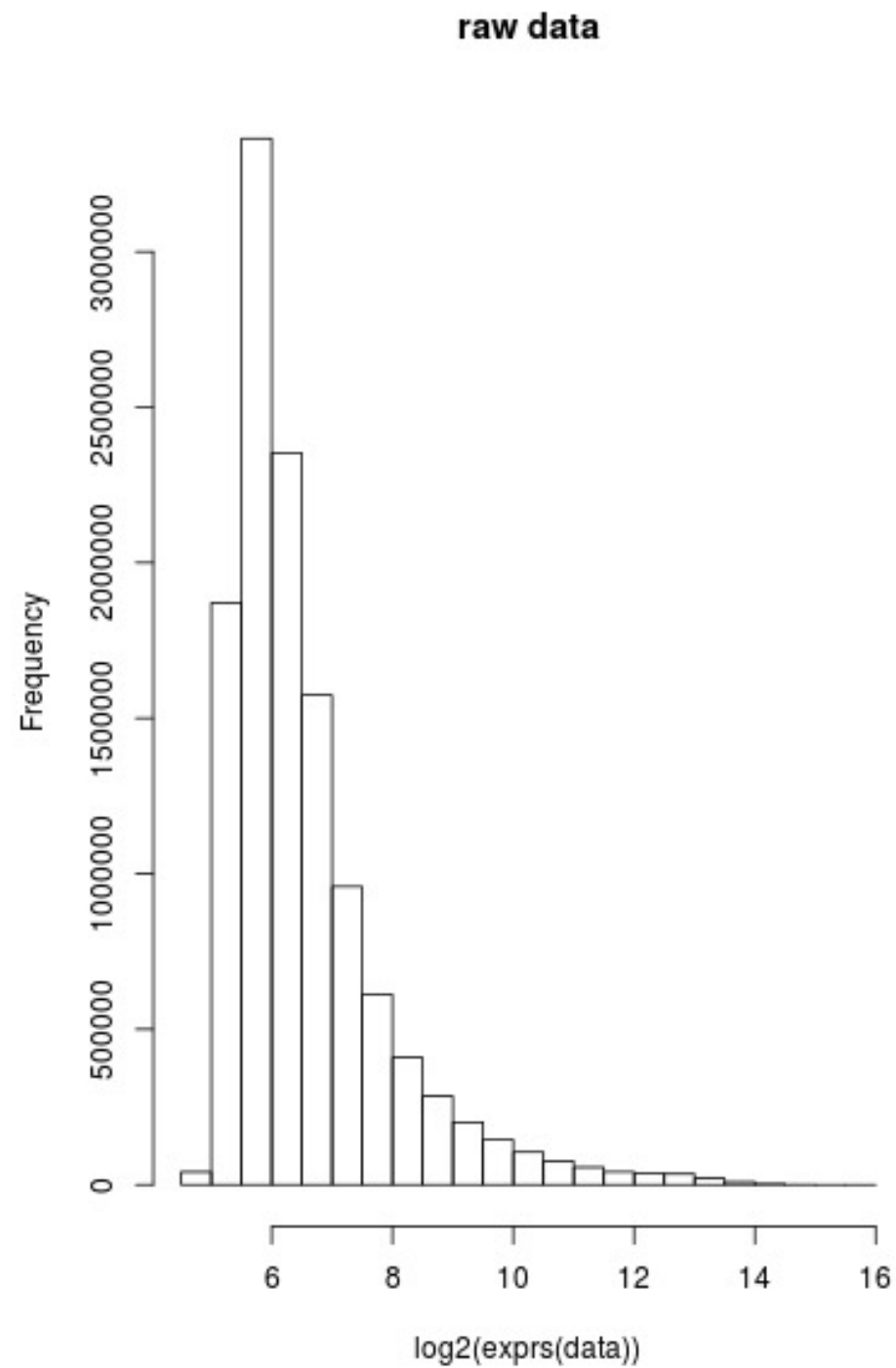


weitere Normalisierungsverfahren:

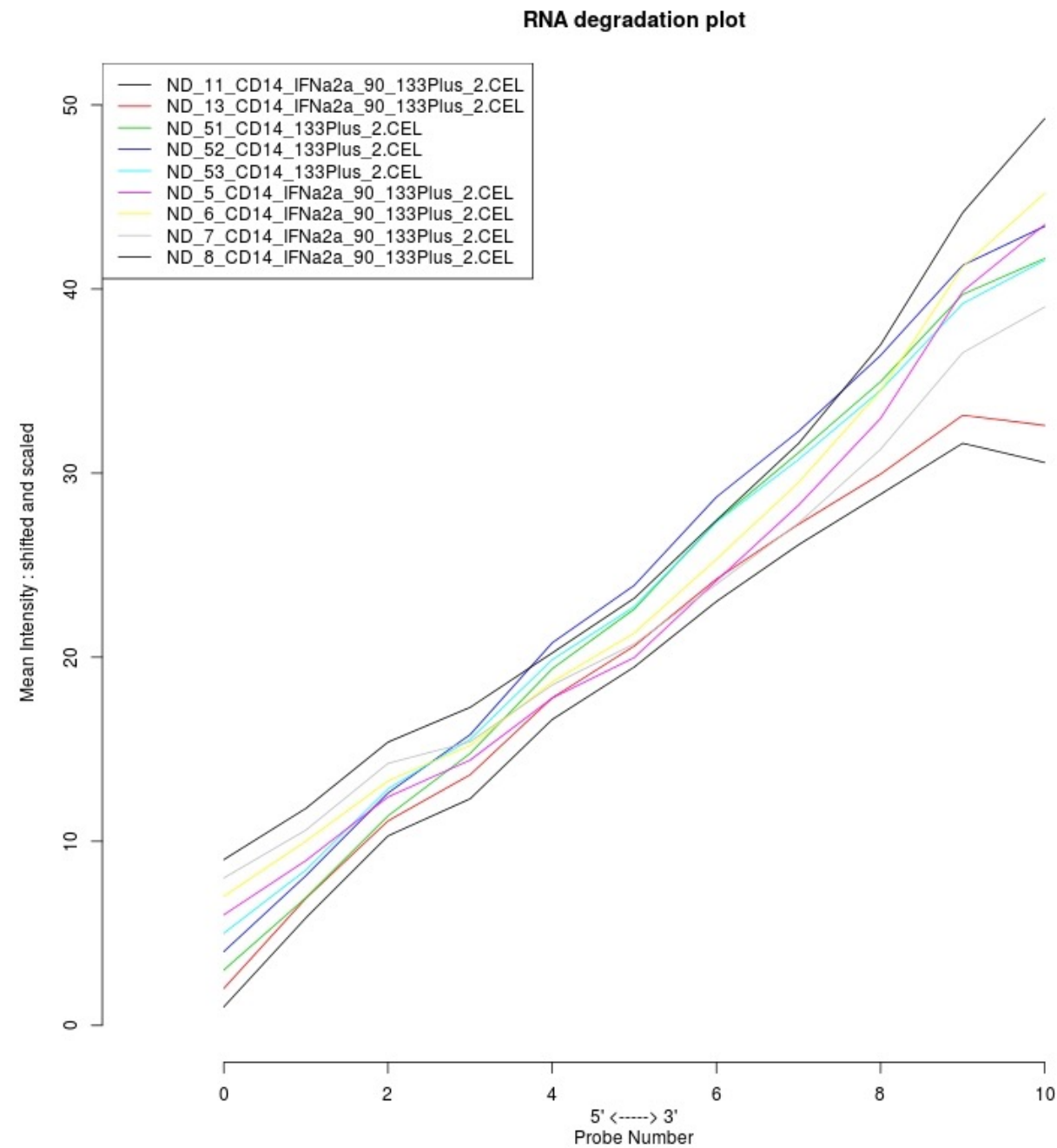
- GCRMA ähnelt RMA, nutzt aber MMs zur Hintergrundkorrektur
- MAS 5.0 basiert auf PM-MM
- dChip basiert auf PM-MM
- Cyclic loess nutzt Loess Normalisierung für Paare von Arrays
- VSN → Variance stabilizing normalization, kombiniert Hintergrundkorrektur und Normalisierung

Qualitätskontrolle

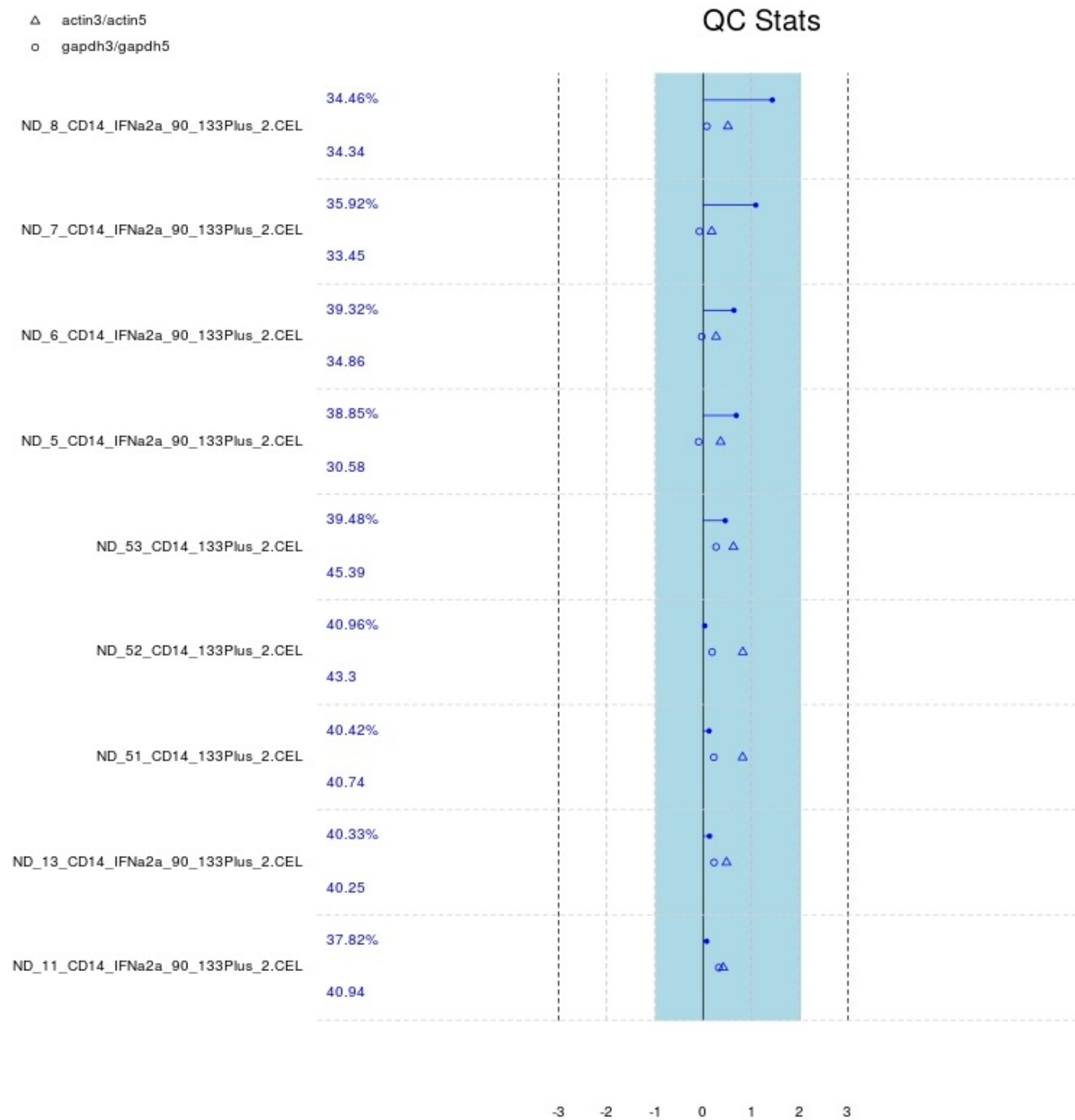
Histogramm



RNA-Degradation-Plot



Quality-Control-Plot



Quellen

- https://www.dkfz.de/gpcf/affymetrix_genechips.html
- <http://www.zhb.uni-luebeck.de/epubs/ediss817.pdf>
- <http://cswwww.essex.ac.uk/staff/langdon/genechip/>
- http://math.usu.edu/jrstevens/stat5570/1.4.Preprocess_4up.pdf
- http://www.ub.edu/stat/docencia/bioinformatica/microarrays/ADM/slides/2_PreprocessingMicroarrayData-2-Preprocessing%20and%20Normalization.pdf
- <https://genevestigator.com/userdocs/manual/qc.html>
- <http://arrayanalysis.org/main.html>
- <http://sites.stat.psu.edu/~naomi/talks/%20AffyIntro.pdf>
- <https://de.wikipedia.org/wiki/DNA-Chip-Technologie>
- http://repository.countway.harvard.edu/xmlui/bitstream/handle/10473/4713/hummel_affy_normalization.pdf?sequence=1