

# Inhaltsverzeichnis

Editorial der TMF	VII
<b>1 Das TMF-Projekt zum Qualitätsmanagement von Hochdurchsatz-Genotypisierungsdaten</b>	<b>1</b>
<b>2 Affymetrix Genotypisierungs-Chips: Genotypbestimmung und Qualitätsfilter</b>	<b>9</b>
2.1 Technischer Hintergrund und Motivation	11
2.2 Systematische Literaturrecherche	12
2.3 Datenvorverarbeitung	14
2.3.1 Quantilnormalisierung	15
2.3.2 Robust Multi-Array Averaging (RMA)	16
2.3.3 Schätzung des logarithmierten Fehlers der Proben-Intensitäten (PLIER)	17
2.4 Beschreibung ausgewählter Genotypisierungs-Algorithmen	18
2.4.1 Dynamic Model-Algorithmus	18
2.4.2 Robustes lineares Modell mit Mahalanobis Distanz (RLMM) und Bayesianisches robustes lineares Modell mit Mahalanobis Distanz (BRLMM)	19
2.4.3 Korrigiertes robustes lineares Modell mit Mahalanobis Distanz (CRLMM)	21
2.4.4 CHIAMO	22
2.4.5 JAPL	23
2.4.6 BIRDSEED	23
2.5 Systematischer Vergleich ausgewählter Genotypisierungsalgorithmen	24
2.5.1 Vergleichsgrundlage HAPMAP II	24
2.5.2 Benutzerfreundlichkeit der Programme zur Genotypisierung	25
2.5.3 Ergebnisse der Genotypbestimmung	28
2.6 Standard-Qualitätskontrolle	30
2.7 Diskussion	32
2.7.1 Danksagung	36
Literaturverzeichnis	36
<b>3 Nicht-kanonische Fluoreszenz-Cluster und Begleitinformation für Genotypisierungsdaten</b>	<b>39</b>
3.1 Einführung	41
3.2 Motivation	42
3.2.1 Nicht-kanonische Fluoreszenzintensitätscluster	42
3.2.2 Begleitinformation für Hochdurchsatz-Genotypisierungsdaten	45

## Inhaltsverzeichnis

<b>3.3 Nicht-kanonische Fluoreszenz-Clusterbildung</b>	<b>46</b>
3.3.1 Datenbasis	46
3.3.2 Verteilung der call rates	47
3.3.3 Klassifikation der Clusterbildung	47
3.3.4 Übersicht über die Verteilung der Cluster-Klassen	48
3.3.5 Automatische Erkennung nicht-kanonischer Clusterbildung	50
3.3.6 Bedeutung und Interpretation nicht-kanonischer Fluoreszenzintensitäts-Cluster	50
<b>3.4 Begleitinformation für Genotypisierungsdaten</b>	<b>52</b>
<b>3.5 Zusammenfassung</b>	<b>53</b>
<b>3.6 Anhang: Begleitformular für Genotypisierungsdaten</b>	<b>54</b>
<b>4 Etablierung wissenschaftlich fundierter Plausibilitätskriterien für Genotypisierungsdaten</b>	<b>63</b>
4.1 Einleitung	65
4.2 Üblicherweise verwendete QM-Parameter	65
4.2.1 Minor Allele Frequency (MAF)	66
4.2.2 Per Person Call Rate (ppCR)	66
4.2.3 Per SNP Call Rate (psCR)	66
4.2.4 Hardy-Weinberg Equilibrium (HWE)	67
4.3 Empirische Bewertung von Schwellenwerten	67
4.3.1 Verwendete Stichproben	67
4.3.2 Analyse	67
4.4 Ergebnisse	68
4.5 Diskussion	70
Literaturverzeichnis	72
<b>5 Fehlererkennung und Fehlerkorrektur bei Genotypisierungsdaten</b>	<b>73</b>
5.1 Einleitung	75
5.2 Datengrundlage	75
5.3 Ergebnisse	76
5.3.1 Programmimplementation	76
5.3.2 Abschätzung der Fehlerraten	77
5.3.3 Systematische Untersuchung der high level-Qualitätsparameter	79
5.3.4 Systematische Untersuchung der medium level-Qualitätsparameter	81
5.3.5 Systematische Untersuchung der low level-Qualitätsparameter	82
Literaturverzeichnis	85

## Inhaltsverzeichnis

<b>6 Daten- und Qualitätsmanagement bei der inhaltlichen Replikation von Genotypisierungsexperimenten</b>	<b>87</b>
6.1 Einleitung	89
6.2 Motivation	89
6.3 Primäre Datentypen und Datenformate	90
6.4 Qualitätskontrolle mit Hilfe von REPLICHECKSNP	93
6.4.1 CHECKPOS	94
6.4.2 CHECKBLAT	95
6.4.3 CHECKHWE	96
6.5 Zusammenfassung	98
Literaturverzeichnis	98
<b>7 Datenstrukturen, Speicherung und Transfer von Genotypisierungsdaten</b>	<b>99</b>
7.1 Einleitung	101
7.2 Motivation	101
7.3 Primäre Datentypen und Datenformate	102
7.3.1 Vorsätze für Maßeinheiten	103
7.3.2 Primäre Datentypen	104
7.3.3 Datenformate	105
7.4 Datentransfer	108
7.4.1 Datentransport über feste Medien	108
7.4.2 Datentransfer über ein Netzwerk	109
7.4.3 Datenkomprimierung	111
7.5 Datenspeicherung	113
7.5.1 Relationale Datenbanken und binary large objects (BLOBS)	113
7.5.2 Datensicherung und Archivierung	119
7.6 Datenschutz	121
7.6.1 Mehrstufige Berechtigungskonzepte	122
7.6.2 Datentreuhändermodell	124
7.7 Testen von Software und Softwareentwicklung	127
7.7.1 Testen von Programmen und Dateiformaten	127
7.7.2 Programmentwicklung	130
7.8 Zusammenfassung	131
Literaturverzeichnis	132