

Inhaltsverzeichnis

Editorial der TMF _____	VII
1 Das TMF-Projekt zum Qualitätsmanagement von Hochdurchsatz-Genotypisierungsdaten _____	1
2 Affymetrix Genotypisierungs-Chips: Genotypbestimmung und Qualitätsfilter _____	9
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
3 Nicht-kanonische Fluoreszenz-Cluster und Begleitinformation für Genotypisierungsdaten _____	39
3.1 Einführung _____	41
3.2 Motivation _____	42
3.2.1 Nicht-kanonische Fluoreszenzintensitätscluster _____	42
3.2.2 Begleitinformation für Hochdurchsatz-Genotypisierungsdaten _____	45

3.3 Nicht-kanonische Fluoreszenz-Clusterbildung	46
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
3.4 Begleitinformation für Genotypisierungsdaten	52
3.5 Zusammenfassung	53
3.6 Anhang: Begleitformular für Genotypisierungsdaten	54
4 Etablierung wissenschaftlich fundierter Plausibilitätskriterien für Genotypisierungsdaten	63
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
5 Fehlererkennung und Fehlerkorrektur bei Genotypisierungsdaten	73
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

6	Daten- und Qualitätsmanagement bei der inhaltlichen Replikation von Genotypisierungsexperimenten	87
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
7	Datenstrukturen, Speicherung und Transfer von Genotypisierungsdaten	99
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