

Cambridge University Press

978-1-107-02752-7 - Advances in Statistical Bioinformatics Models and Integrative Inference for High-Throughput Data

Edited by Kim-Anh Do, Zhaohui Steve Qin and Marina Vannucci

Index

[More information](#)

Index

- 1000 Genomes Project, 23, 470
 3C, 147
 4C, 147
 5C, 147
- Affymetrix, 1, 11, 59, 72, 390, 401, 406
 MIPS array, 244, 247, 265
- alignment, 141
- allele, 54, 78
 alternative, 59, 64, 66, 70
 B-allele frequency, 262
 heterozygous, 249–250
 ratio, 247–251, 256, 262, 263, 267
 reference, 59, 61, 64
- allelic imbalance, 99–100
- alternative splicing, 84
- amplification, PCR, 141, 144
- annotation, 163, 359, 398, 401, 416, 419
- antibody, 141, 142, 161
- array-based resequencing, 59
- Array CGH, 11–13, 331, 332–333, 334, 335, 345
- association, 224, 226, 228, 402, 404, 418, 419
 marginal, 234
- base call, 60, 64, 67, 69, 72, 108, 127, 128–129, 130
- Bayes factor, 215–216
- Bayesian hierarchical model *see* hierarchical model, Bayesian
- Bayesian model averaging, 200, 208, 272
- Berhu, 382, 385ff
- beta-binomial, 87, 95
- bias, sequencing, 91, 107, 144
- binding, 138, 140, 148, 154, 156, 163, 165
 functional, 156
 site validation, 156
- biomarker selection, 280
- bootstrapping, 465
- branch lengths, 465, 468
- BS-Seq, 20
- call rate, 61, 69
- cancer, 244
 blood, 263, 269
 breast, 20, 47, 138, 251, 256, 267, 277, 334–335, 341, 399, 401, 405, 418
 lung, 6, 32, 47, 218–219, 390
 ovarian, 32, 369, 372, 375
 prostate, 120, 424–425, 435, 437, 440
- Cancer Genome Atlas, The, 31ff, 244, 292, 369, 371, 374–375, 424
- Cancer Genomics Browser, 51
- cancer subtyping, 367
- case-control study, 208, 210, 212, 218
- ChIA-PET, 147
- ChIP *see* Chromatin Immunoprecipitation
- ChIP-chip, 139, 156, 160
- ChIP-exo, 165
- ChIP-PET, 142
- ChIP-qPCR, 156
- ChIP-Seq, 17, 138ff, 172
 analysis protocol, 163
 high-throughput, 140, 166
 mathematical model, 148–149
- Chromatin Immunoprecipitation (ChIP), 139
 library, 159
- Chromatrap, 165
- circular binary segmentation (CBS), 246
- classical factor analysis models, 350–351
- classification, molecular, 399, 419
- cloud computing, 470
- coalescence process, 467
- Comparative genomic hybridization, 10
- conditional Gaussian graphical model, 319
- constrained optimization, 320
- convergence, 465, 467

- Copy number, 10, 244ff, 332
 algorithms for estimating, 261
 allele-specific, 249, 263
 parent-specific, 264
 Predict integral copy numbers in cancer (PICNIC), 263
- correlation, 399
 integrative, 398, 403–405
- Cox proportional hazard model, 392, 402–403
- cross-linking, 140
 reverse, 141
- cross-study comparison, 398, 413
- cross-study validation, 399, 418
- cross talk, 127, 128, 129
- data integration, 369
- Data storage, 22, 35
- Databases
 Biocarta, 433
 CGHub, 43
 dbSNP, 23, 66
 GEO, 23
 GO, 433
 KEGG, 282, 423
 Mbatch, 46
 OMIM, 23
 SAGEmap, 23
- differential co-expression, 315
 gene-set based, 317
- differential expression, 399, 426
 analysis, Bayesian, 97
 analysis, frequentist, 96
 nonparametric approach, 99
- directed acyclic graph (DAG), 284–285
- discriminant analysis, pathways, 281–283
- disease genes, 69
- disease heterogeneity, 368
- DNA methylation, 3, 4–6
- drug target pathway inference, 360, 361
- Elastic Net, 382, 385–386, 389–392, 395, 396
- enrichment tests, 425
- eQTL, 312–313, 314
 multi-tissue, 323
- estrogen receptor (ER), 138
- evolution, 460–464, 467–468, 470
- expression quantification, 82
 Poisson models for, 84
- exome, 3
- Fading, 128, 129, 130, 133
- false discovery rate (FDR), 84–87, 216, 405, 409, 412
- false negative rate, 55
- false positive rate, 55
- folate metabolism, 218, 221
- FPKM, 88
- fragment length estimation, 145, 147, 150–152
- fragment size distribution, 91, 165
- fragmentation, 140
- Gaussian graphical model (GGM), 285, 297–299
- Gaussian mixture model (GMM), 66, 127
 bivariate, 64
- GC-content, 119
- gene
 annotation, 398
 clustering, 447–448, 454
 isoforms, 77
 silencing, 4, 6, 7
- gene expression, 2, 4, 17, 106, 293, 295–296, 301, 398ff, 434
 classifier, 415, 419
 differential gene expression *see* differential expression
 microarray, 332, 347
- Gene Set Enrichment Analysis (GSEA) test, 424, 426, 440
- genetic analysis, 358–362
- genetic association studies, 208ff, 272
 statistical analysis for, 209–212
 SSVS for, 217, 221
- genome, 3
 complete, 464, 468
 sequences, 106
- Genome Analyzer, Illumina, 108, 127
 GA-I, 127ff
 GA-II, 131–135
- genome-wide association study, 12, 13, 188ff, 224, 229–231, 234, 239
- genomic aberrations, 368, 369, 371, 372
 Lasso approaches for, 195–197
 machine learning approaches for, 239
 Bayesian approaches for, 198–204
- genotype, 66–67, 226, 229, 238, 241
- Gibbs sampling, 87, 153
- graphics processing units (GPU), 469
- GWAS *see* genome-wide association study
- H1N1, 464
- H3N2, 464
- haplo-isoforms *see* isoform, haplotype-specific
- haplotypes, 78, 82
- heat map, next-generation clustered, 48
- heritability, 226, 228, 231
 missing, 225
- heterogeneity, 368, 467, 468
- heterozygous variant, 59, 64
- HiC, 147
- Hidden Markov Model (HMM), 153, 263
- Hidden Ising Model, 154

- hierarchical models, Bayesian, 128–130, 172, 468–469
 Poisson-Gamma, 326
 Poisson-Gamma mixture, 95
- histone, 145, 155
 modification of, 145, 155, 179
 HKY85, 467
- homozygous variant, 59, 64
- hyper-inverse Wishart (HIW) g-prior, 298–299
- ImmunoPrecipitation, 140
- importance sampling, 468–469
- imprinting, 100
- influenza A, 462–464
- Integration, 25–26, 331, 332–334
- integrative
 association, 399, 406, 409
 correlation, 398, 403–405
 model, 273, 283, 292, 294
- Integrative Genome Viewer, 50
- intrinsic gene classification, 413–414, 419
- intrinsic gene panel, 399, 400, 403, 406
- Irreproducible Discovery Rate (IDR), 159
- isoform, 77–86, 99
 haplotype-specific, 78–80, 85, 99
- kernel smoothing, 152
- Kimura two-parameter model, 466
- knock-down, 142
- Lasso, 195–197, 235–236, 382–383, 386–387, 390–395
- latent Dirichlet allocation (LDA), 367–369
 survival-supervised (survLDA), 369–371
- Latent probit score, 335
- library, ChIP, 139
- Linear models, 273–279
 multivariate, 192–195
- mapping, short reads, 109
- Markov Chain Monte Carlo (MCMC), 98, 154, 214–216, 219–220, 232, 234, 238, 276, 282, 338, 345, 433, 467
 dynamic iteratively reweighting algorithm, 469
- Markov random field priors, 275
- matrix normal graphical model, 323
- MedIP-seq, 147
- meta-analysis, 398, 404
- Microarray, 2, 9, 59, 63, 138, 139, 156, 277, 332, 390, 401
 DNA resequencing, 59
- microRNA targets, 285, 294, 297, 302, 308
- minimum spanning tree, 448
- miRNA, 8, 285–286, 301
- mitochondrial disorder, 68–69, 72
- Mixture model, 172, 176, 180, 184, 280, 283, 334–335, 468
- mixture priors, 274, 280, 286
- mRNA, 273, 301
- mseq, 117–118
- model selection, 297
- molecular clock, 465
- Molecular inversion probes (MIP), 9–11, 57, 244
- molecular profiling, 34
- molecular taxonomy, 399, 400
- motif, 157
- multiple-instruction multiple-data (MIMD), 469
- multi-SNP, 225
- negative binomial, 87, 95
- Negative Binomial distribution, 150, 153
- Nextera, 164
- Normal contamination, 264, 265
- normalization, 87–93, 148, 164
 count, 164
 global-scaling, 119–120
 quantile, 92, 119
 nonparametric, 92
 regression, 119
- nucleosome, 145–146
- Nucleosome positioning, 146, 170, 179–183
- nucleotide, 59
- objective Bayesian prior, 297
- over-dispersion, 93–94, 110
- padlock probes, 56–57
- pandemics, 460, 462–464
- parallel computing, 469
- pathway analysis, 273, 342, 351, 356–358, 424, 426, 429, 432
- patient subgroups, 368, 371, 379
- patient-specific inference, 369, 371, 374–375, 378
- personalized medicine, 366, 377
- Phasing, 128, 129, 130, 133, 134
- phenotype, 399, 403, 412, 417, 419
 binary, 238, 402
- phenotypic variance explained, 225, 229, 232, 235, 236
 for binary data, 238
- phylogenetics, 461, 463, 470
- phylogenetics, 460ff
- poisson
 model for expression quantification, 84–94, 112
 distribution, 86, 90, 95, 112, 150
 mixed-effect model, 114–117

- poisson (*cont.*)
 noise, 85, 93
 rate, 89, 91, 93
 regression, 117
- polygenic model, 189
- posterior probability, 66, 212, 215, 276
- power, 156, 159
- precision matrix, 313, 319, 320, 322
- prediction, 399
- predictor, 399, 409, 413
- prior probability, 64, 99
- prognosis, 415, 419
- proportion of variance explained, 235
- protein, 138–141
- PVE *see* phenotypic variance explained
- quality control, 61–64
- quality score, 67–68
- rare variants, 12, 54
- read count models, 85–87, 99–100, 152–153
- Read count data, 84–86, 89, 93, 96, 102, 110, 117
- reChIP, 146
- reference base, 73
- regression
 linear, 226
 logistic, 210, 211, 217, 218, 402
 model, 401, 412
 penalized, 226, 235
 probit, 335
 ridge, 385
- regularization, 315, 316, 382–386, 390
- relapse-free survival (RFS), 399, 402, 406, 412, 417
- replicate
 biological, 160–161
 technical, 158
- replication, 157–161
- reproducibility, 398–400, 409–415, 418–419
- resampling, 431
- Reverse-phase protein array, 13
- RNA biology, 77
- RNA interference, 4, 6
- RNA Technology, 78
- RNA virus, 460, 461, 465
- RNA-Seq, 17, 79, 101, 102, 105ff, 324
 messenger, 19
 non-coding, 19
 miRNA-Seq, 19
- rooted trees, 465–466
- RPKM, 88, 107, 114, 117, 119
- sample progression discovery, 446ff
- sample size, 374, 375, 376, 378, 380
- Segmentation
 copy number, 246, 262, 264
 hypersegmentation, 256, 267
- sequencing, 141, 461, 464, 465, 468
 depth of, 144
 high-throughput, 2–3, 9–14, 59, 63, 79, 85
 high-throughput sequencing of RNA, *see* RNA-Seq
 next generation, 2, 14–25
 paired-end, 141, 142, 147
 whole-exome sequencing, 15
- shrinkage, 233, 236
- simulation, 374
- single-cell, 165
- single-instruction multiple-data (SIMD), 469
- single-nucleotide polymorphism (SNP), 9, 32, 35, 208ff
- siRNA, 7
- SNP *see* single-nucleotide polymorphism
- SNV, 55
- Software
 ABSOLUTE, 267
 ADMIXTURE, 358
 ASCAT, 263
 BayesPeak, 153, 174–177
 BEAST, 466
 BLAST, 24, 401, 406, 409
 BLAT, 24
 BRDU-seq, 147
 CCAT, 155
 ChIP-PaM, 153, 155, 157
 ChIPpeakAnno, 157
 CNAnorm, 264
 CSAR, 152
 Cufflinks, 118–119
 DAVID, 433
 DEXSeq, 325
 DiffBind, 165
 FREEC, 264
 F-seq, 153
 GeneGo, 342
 GPseq, 112–114
 HPeak, 153
 iFad, 360
 iSeq, 154, 174
 MACS, 152, 174–177
 MatchMiner, 406
 miRanda, 296, 308
 mirSVR, 296, 308
 MrBayes, 466
 mseq (R package), 118
 OncoSNP, 262
 OntoTools, 433
 PathWave, 433
 PCSN, 264
 PhyloBayes, 468
 PICNIC, 263
 PICS, 155, 172–179
 PicTar, 296

Cambridge University Press

978-1-107-02752-7 - Advances in Statistical Bioinformatics Models and Integrative Inference for High-Throughput Data

Edited by Kim-Anh Do, Zhaohui Steve Qin and Marina Vannucci

Index

[More information](#)

Index

481

- PING, 180–183
 PITA, 296
 pLINK, 209, 224
 PMseq, 114–117
 POHE
 QuEST, 153
 Rcade, 166
 Regulome Explorer, 48
 Repitools, 164
 RSEG, 155
 SICER, 155
 SOURCE, 406
 TAPS, 264
 TargetScan, 296
 Tightrope, 265
 T-PIC, 155
 XSET, 152
 sonication, 140
 source-sink hypothesis, 468
 Spanish flu, 460, 463, 464
 sparse factor models, Bayesian, 351–354
 spatial gene expression patterns, 359
 Spike-and-slab prior, 213, 357
 spike-in, 161
 stochastic search, 201
 stochastic search variable selection (SSVS),
 201, 212–218, 221, 272
 substitution, 466–467, 468
 subtype, 463–464
 supermatrix, 468
 supertree, 468
 support vector machine (SVM), 239–241
 survival analysis, 399, 402–403
 swine flu, 463
 synergistic effect, 72

 Tamoxifen, 139
 target capture, 72
 target density, 467–468

 TCGA *see* Cancer Genome Atlas
 test of equality of covariance matrices, 313,
 318
 text mining, 367, 380
 topic models, 367, 376–377
 topology, 465, 468
 Transcript expression level, 106
 Transcript length, 89
 Transcription Factor (TF), 4, 138, 342, 351
 Transcription Factor Binding Site (TFBS),
 138, 170
 transcriptional regulatory network, 350,
 351
 Transcriptome, 3, 105–107
 assembly, *de novo*, 80
 assembly, genome guided, 81
 reference, 83
 transitions, 466–467
 transversions, 466–467
 tree space, 466–467
 Tumor
 fraction, 268
 DNA fraction, 253, 256, 261, 262, 263, 264,
 265, 267, 268
 ploidy, 251–252, 263

 validation, 155–161, 390
 variable selection, 382, 396
 Bayesian, 211, 229
 variational inference, 373
 virulence, 464

 wavelet, 152, 155
 whole-genome, 224
 multi-snp analysis, 225

 Yule prior, 467

 Zellner's *g*-prior, 275