

Index

Note: Page numbers followed by ‘f’ and ‘t’ indicate figures and tables.

- accelerated failure time (AFT) model, 227
accuracy quality control (AQCg; AQCp),
 MetaQC package, 41
aCGH (array comparative genomic
 hybridization), 333, 403
AD (Alzheimer’s disease), 102–103
adaptively weighted Fisher’s (AW-Fisher’s)
 method, 43
adjusted Rand index, 242
Affymetrix arrays, 400–401
AFT (accelerated failure time) model, 227
allele frequencies, comparison of, 14f
allele-specific binding (ASB), in ChIP-seq,
 109, 123–128
Alzheimer’s disease (AD), 102–103
Amandine, E., 202
AMD GWAS, 19–20
AQC (accuracy quality control), MetaQC
 package, 41
ARACNe algorithm, 297
Argonaute protein, 285–286, 292
array comparative genomic hybridization
 (aCGH), 333, 403
ArrayExpress, 397
ASB (allele-specific binding), in ChIP-seq,
 109, 123–128
autologistic regression, 206, 208–209
AW-Fisher’s (adaptively weighted Fisher’s)
 method, 43
background window, ChIP-chip peak calling,
 119
bacterial artificial chromosome (BAC) probes,
 aCGH, 403–404
Bayesian Consensus Clustering (BCC)
 method, 251–260
 application to TCGA data, 257–260
 defined, 238
Dirichlet mixture model, 252–253
estimation, 254–256
illustrative example, 256–257
multisource model, 253–254
overview, 251–252
Bayesian false discovery rate (BFDR), 362
Bayesian graphical models
 BayesGraph for TCGA integration,
 201–214
 graphical models, 203
 markov random fields, 204–205
 MCMC Simulations, 207–208
 overview, 201–203
 posterior inference using false discovery
 rates, 208
 probability model, 205–207
 simulation study, 208–211
 TCGA integrative analysis, 211–214
 data description, 226–227
 iBAG models, 220–226
 linear, 221–224
 non-linear extensions, 224–226
 overview, 220–221
 illustrations, 226
 overview, 4–5, 217–220
 results, 227–233
Bayesian information criterion (BIC)
 correlation motif model, 117–118
 directed acyclic graph, 273
BBID database, 335
BCC method. *See* Bayesian Consensus
 Clustering method
Benjamini-Hochberg (BH) procedure, 51
BFDR (Bayesian false discovery rate), 362
BH (Benjamini-Hochberg) procedure, 51
BIC. *See* Bayesian information criterion

- binding, miRNAs
 ComiR targeting method, 292–294
 effect of single and multiple targets, 290
 Fermi-Dirac combination of targets, 290–292
 thermodynamics of, 289–290
- binding, TF
 cause or consequence relationship between gene expression, 393–394
 ENCODE K562 and GM12878 data, 378–379
 framework for integrating with gene expression data, 376–377
 interplay between histone modification and other chromatin features, 391–392
 machine learning methods used in predictive models, 377–378
 ModENCODE Early Embryo data, 379
 Mouse ESC data, 379
 overview, 374–375
 performance evaluation of models, 378
 predicting differential gene expression, 388–389
 predicting expression levels for genes with HCP and LCP content, 391
 predicting expression of noncoding genes, 389–390
 predicting gene expression by combining with histone modifications, 385–388
 predicting gene expression from, 379–382
 regulatory signals in distal regions, 392–393
 Yeast and Fly data, 379
- BioCarta database, 335, 336f, 422
- biological pathways
 defined, 421
 iFad method, drug-pathway association analysis, 422
 iPad method, drug-pathway association analysis, 425
- BioPAX pathway format, 306
- burden tests, meta-analysis of GWAS
 that assume distribution of variant effect sizes, 25–26
 that assume variants have similar effect sizes for a simple burden test in study *k*, the impact of multiple rare variants, 24
- cancer (sub)type analysis, 350–352, 352f
 cancer genomics. *See also* latent variable approach, integrative clustering analysis; somatic mutations in cancer genomes
 active subnetwork search and discovery, 309–310
 joint NMF, 134–139
 network-regularized joint NMF Method, 139–143
 overview, 131–134, 304–305
 PARADIGM pathway method, 310–319
 applications of, 317, 319
 interaction parameters, 316–317
 interactions and probabilistic factors, 314–316
 matrix of, 318f
 overview, 310–312
 variables, 312–314
 PARADIGM-SHIFT pathway method, 319–322
 applications of, 320–322
 overview, 319–320
 pathway databases, 305–306
 pathway methods, 307–308
 pathway-based mutation assessment, 308–309
 sparse Multiple Block PLS method, 143–147
 TieDIE pathway method, 322–328
 cancer-related lncRNAs, 406, 407f, 408f
 CancerResource Database, 432
 canonical correlation analysis (CCA), 241
 causal inference, eQTLs, 270–271
 CCA (canonical correlation analysis), 241
 CD4:CD3 ratio, 172
 analysis results by applying MCP to each outcome separately, 191t–192t
 analysis results of gMCP, 185t, 193t–194t
 analysis results of gMCP with Laplacian penalty, 195t–198t
 analysis results of sparse gMCP, 186t–187t
 overlaps of different analysis methods, 184t
 CD4/CD8 ratio (T-Lymphocyte Helper/Suppressor Profile), 171–172
 analysis results by applying MCP to each outcome separately, 191t–192t
 analysis results of gMCP, 185t, 193t–194t
 analysis results of gMCP with Laplacian penalty, 195t–198t
 analysis results of sparse gMCP, 186t–187t
 overlaps of different analysis methods, 184t
 centered parametrization, 206
 change-point model, gene expression regulation, 358–360
 ChIP (chromatin immunoprecipitation), 108.
See also ChIP-X data
 ChIP-chip analysis, 118–123. *See also* ChIP-X data
 ChIP-seq (chromatin immunoprecipitation sequencing), 108, 378–380, 385, 388, 398. *See also* ChIP-X data
 ChIP-X data
 allele-specific binding in ChIP-seq, 123–128

- ChIP-chip analysis, 118–123
- ChIP-seq, 108, 378–380, 385, 388, 398
- correlation motif approach, 112–118
- general problem setting and motivations, 110–112
- overview, 108–110
- chromatin immunoprecipitation (ChIP), 108.
- See also* ChIP-X data
- chromatin immunoprecipitation sequencing (ChIP-seq), 108, 378–380, 385, 388, 398. *See also* ChIP-X data
- chromosome instability (CIN), 156, 164
- CIFA (common and individual feature analysis), 242
- CIMP (CpG island methylator phenotype), 164
- CIN (chromosome instability), 156, 164
- cis-eQTLs, 87, 94, 268–271, 269f
- clinical iBAG model, 222–224
- clipper method, analysis of gene expression, 307–308
- clustering. *See also* latent variable approach, integrative clustering analysis
- Bayesian Consensus Clustering method, 251–260
- differential clustering algorithm, 101
- exploratory methods for multisource data, 242–243
- cMCP (composite MCP), 176–177, 179–180
- CNAs (copy number aberrations), 333–334
- CNVs (copy number variations), 132, 143–147, 201–203, 207, 214
- Cochran-Mantel-Haenszel method, single-variant association test statistics, 24
- coexpression clusters, 76
- coexpression network, 58, 59f, 67, 68, 69f, 70, 100–102
- coherent FFLs, 295f
- collapsed Gibbs sampling algorithm, iFad method, 424–425
- colon cancer, 156
- colorectal carcinoma (CRC) study, 163–165, 164f, 165f
- combining effect sizes analysis, microarrays, 44
- combining *p*-values analysis, microarrays
- evidence aggregation methods, 42–44
- order statistics methods, 44
- combining ranks analysis, microarrays, 44–45
- ComiR targeting prediction algorithm, 287t, 288, 292–294, 293f
- common and individual feature analysis (CIFA), 242
- complete conditionals, 234–235
- complex correlation structures, 219
- complex diseases
- disease subtype discovery, 56–57
- MetaNetwork for differential network detection, 58
- MetaPath for pathway analysis, 52–55
- network integration of genetically regulated gene expression
- diabetes genes, 93–100
- differential connectivity in coexpression network, 100–102
- late-onset Alzheimer's disease brain study, 102–103
- LINKER approach, 92
- modeling genetic information flow, 88–91
- overview, 86–88
- PRINCE approach, 91–92
- prize collecting Steiner tree problem, 92–93
- random walk approach, 91
- network integration of genetically regulated gene expression to study, 86–104
- composite MCP (cMCP), 176–177, 179–180
- composite penalization, 180
- computational burden, exploratory methods for multisource data, 261
- computational cancer genomics. *See* cancer genomics
- computational methods. *See* integrative analysis; integrative quantitative models; latent variable approach, integrative clustering analysis
- conditional analyses
- meta-analysis of GWAS, 26–28
- results of conditional association analysis for LDL and variants in LDLR, 31t
- consensus clustering (ensemble clustering), 242–243, 260f
- consensus PCA, 241
- consistency quality control (CQCg; CQCp), MetaQC package, 41
- context score, TargetScan targeting prediction algorithm, 287, 294
- Conway, A. R. A., 156
- cooperative functional effects, mdmodules, 136–138
- copy number aberrations (CNAs), 333–334
- copy number data, *IGF1R* gene, 225f
- copy number variations (CNVs), 132, 143–147, 201–203, 207, 214
- core modules, 347, 350f
- correlation motif model
- data generative process, 113f
- integrative analysis of ChIP-X data, 112–118

- coupled transcription-splicing modules
 - mechanisms of, 81
 - methods and materials, 77–80
- CpG island methylator phenotype (CIMP), 164
- CQC (consistency quality control), MetaQC package, 41
- CRC (colorectal carcinoma) study, 163–165, 164f, 165f
- cross validation
 - data partitioning, 430f
 - generalized cross validation, 225
 - V-fold, 178
- Cytoscape biological network viewer, 306
- DAG. *See* directed acyclic graph
- DANCR lncRNA, 406
- data generative process, correlation motif model, 113f
- data partitioning, cross validation procedure, 430f
- data set membership vector, FCC, 79
- databases
 - BBID, 335
 - BioCarta, 335, 336f, 422
 - CancerResource Database, 432
 - hmChIP, 109
 - JASPAR, 81
 - KEGG, 335, 336f
 - KEGG MEDICUS, 431
 - pathway, 305–306
 - Pathway Commons, 306
 - Pathway Interaction, 306
- DAVID tool
 - pathways with lowest FDR, 336f
 - somatic mutations in cancer genomes, 335
- DCA (differential clustering algorithm), 101
- DCA (differential coexpression analysis), 281–282
- DCGs (directed cyclic graphs), 266, 276–277
- De Novo Driver Exclusivity
 - Dendrix algorithm, 343–346, 347–349
 - Multi-Dendrix algorithm, 343–344, 346–347, 349–351
- Dendrix algorithm, 309, 343–346, 345f, 347–349
- dependent clustering, 257, 258f
- diabetes genes, 93–100
- DiffCoEx method, differential coexpression, 102
- differential clustering algorithm (DCA), 101
- differential coexpression analysis (DCA), 281–282
- differential connectivity
 - in coexpression network, 100–102
 - modular differential connectivity, 103, 103f
- differential gene expression, 389f
- differential principal component analysis (dPCA), 128
- dimension reduction. *See also* joint and individual variation explained method
 - iCluster method, 3, 162f, 220, 243
 - MetaPCA, 3, 39t, 55–56
 - nonnegative matrix factorization, 3, 55, 133f, 134–139
 - overview, 3
 - partial least squares, 3, 241–242
 - sparse multi-block partial least squares regression, 3, 143, 144, 148–149
- directed acyclic graph (DAG)
 - Bayesian framework for inference, 278–279
 - hybrid methods, 274–276
 - Markov equivalence classes, 272f
 - method and software, 281t
 - overview, 271–273
 - PC algorithm, 274
 - search-and-score methods, 273–274
 - structure equation models, 279–280
- directed cyclic graphs (DCGs), 266, 276–277
- directed graphical models
 - Bayesian framework for DAG inference, 278–279
 - directed acyclic graph
 - hybrid methods, 274–276
 - method and software, 281t
 - overview, 271–273
 - PC algorithm, 274
 - search-and-score methods, 273–274
 - directed cyclic graphs, 266, 276–277
 - overview, 265–266, 276–277
 - QTL directed dependency graph, 277–278
- Dirichlet mixture model, BCC method, 252–253
- distant eQTL, 267
- DM. *See* DNA methylation
- DNA. *See also* Encyclopedia of DNA Elements
 - CNVs, 132, 143–147, 201–203, 207, 214
 - GWAS
 - AMD GWAS, 19–20
 - GWAS-tailored software, 29–32
 - imputation, 8–9
 - methods for single marker test, 9–19
 - number of publications by year, 8f
 - overview, 7–8
 - plasma lipid levels, 28–29
 - rare variant associations, 20–28
 - workflow of, 10t
 - targeted cancer treatment and, 218
- DNA methylation (DM), 201–202
 - mdmodules, 136
 - TCGA project, 132
- dPCA (differential principal component analysis), 128

- driver mutations, 308, 331, 400–401
Drosophila targets, 286, 293
 drug discovery, 419. *See also* drug-pathway association analysis
 drug sensitivity, 421–422. *See also* drug-pathway association analysis
 drug-pathway association analysis
 iFad method, 422–425
 iPad method, 425–430, 434t–435t
 iterative signature algorithm, 435
 NCI-60 project, 431–433
 overview, 419–422
- EGOT lncRNA, 406
 EM (expectation-maximization) algorithm, 115–117, 311
 embryonic stem cells (ESCs), 385
 EMT (epithelial-to-mesenchymal transition), 296
 Encyclopedia of DNA Elements (ENCODE), 71, 81, 87, 109, 375, 382
 CAGE data from, 389–390
 K562 and GM12878 data, 378–379, 389f
 transcriptome profiling in human cells from, 398
 ensemble clustering (consensus clustering), 242–243, 260f
 epigenomic analysis, 75–76
 epithelial-to-mesenchymal transition (EMT), 296
 EQC (external quality control), MetaQC package, 40–41
 eQTL mapping, 267–270, 281t
 eQTL meta-analysis, 43
 eQTLs. *See* expression quantitative trait loci
 ESCs (embryonic stem cells), 385
 E-step, EM algorithm, 116
 estimates
 analysis results by applying MCP to each outcome separately, 191t–192t
 analysis results of gMCP, 185t, 193t–194t
 analysis results of gMCP with Laplacian penalty, 195t–198t
 analysis results of sparse gMCP, 186t–187t
 BCC method, 254–256
 estimated graph, simulated data sets, 210f
 JIVE method, 245–246, 247f
 mass-action-based model for gene expression regulation, 360–363
 eukaryotic gene expression, 374–375
 evidence aggregation methods, combining *p*-values analysis, 42–44
 exon membership vector, FCC, 79
 exonic lncRNAs, 400f
 exons
 co-splicing mechanisms and, 76
 co-splicing networks, 78f
- expectation-maximization (EM) algorithm, 115–117, 311
 exploratory methods for multisource data. *See* multi-source data, exploratory methods for
 expression quantitative trait loci (eQTLs)
 causal inference and, 270–271
 cis-, 87, 94, 268–271, 269f
 diabetes genes, 94–97
 differential coexpression analysis, 281–282
 directed acyclic graph, 271–276, 281t
 hybrid methods, 274–276
 overview, 271–273
 PC algorithm, 274
 search-and-score methods, 273–274
 directed graphical model estimation using, 276–280
 Bayesian framework for DAG inference, 278–279
 overview, 276–277
 QTL directed dependency graph, 277–278
 structure equation models, 279–280
 eQTL mapping, 267–270, 281t
 gene transcripts, 95f
 identifying regulatory SNPs, 132
 local eQTL versus distant eQTL, 267
 modeling genetic information flow in network, 88–91
 overview, 265–266
 protein QTL data, 281
 trans-, 87, 94, 268–271
 external quality control (EQC), MetaQC package, 40–41
- factorization methods, 241–242
 false discovery rate (FDR), 202, 336, 336f, 337f
 LOAD brain study, 103, 103f
 posterior inference using, 208
 FCC (frequent coupled cluster), 77–80, 78f
 FDR. *See* false discovery rate
 feed-forward loop (FFL), 294–296, 295f
 FEM. *See* fixed effects model
 Fermi-Dirac combination of targets, 290–292, 291f
 FFL (feed-forward loop), 294–296, 295f
 Fisher's method, microarray meta-analysis, 42–43
 meta-analysis of GWAS, 16
 MetaDE package, 42–44, 46t, 47–48, 49t, 51
 fixed effects model (FEM)
 meta-analysis of GWAS, 16
 MetaDE package, 44
 formal framework, exploratory methods for multisource data, 240–241

- FOS* regulatory factor, 75–76
 fragments per kilobase of exon per million fragments mapped (FPKM), 74
 frequent coupled cluster (FCC), 77–80, 78f
- GA (genetic algorithm), 280
GABP regulatory factor, 75
 gain-of-function (GOF), PARADIGM-SHIFT pathway method, 319
 GAM (generalized additive models), 224–225, 225f
 GAS5 lncRNA, 406
 Gaussian graphical models (GGMs), 203
 Gaussian mixture model (GMM), 153
 GBM (Glioblastoma Multiforme), 219, 226, 406, 407, 409
 GCV (generalized cross validation), 225
 GE. *See* gene expression
 GENCODE, 398
 gene expression (GE). *See also* histone modifications; mass-action-based model for gene expression regulation; transcription factor binding
 allele-specific gene expression, 269–270
 eQTLs and, 268f
 JIVE method and, 248–251
 mdmodules, 136
 pathway methods for analysis of, 307–308
 TCGA project, 132
 Gene Expression Omnibus (GEO), 37, 109, 397, 406
 gene expression profiles
 drug-pathway association analysis, 421
 iFad method, drug-pathway association analysis, 422
 iPad method, drug-pathway association analysis, 425
 gene expression regulation
 analysis of osmotic shock in yeast, 366–367
 change-point model, 358–360
 characterizing link between regulatory processes, 368–371
 data integration to study, 5–6
 estimation and inference, 360–363
 overview, 356–358
 scoring protein-level regulation changes, 367–368
 simulation study, 363–366
 gene membership vector, FCC, 79
 gene regulation pathways, 421
 Gene Set Enrichment Analysis (GSEA), 307, 323
 gene sets with lowest FDR, 337f
 overview, 421–422
 somatic mutations in cancer genomes, 335–336
 gene set methods, 307
 gene sets, 337f
 generalized additive models (GAM), 224–225, 225f
 generalized cross validation (GCV), 225
 genetic algorithm (GA), 280
 genetic interaction, 201–203
 GenMiR++ algorithm, 297–298
 genome-wide association studies (GWAS)
 meta-analysis of, 7–33
 AMD GWAS, 19–20
 GWAS-tailored software, 29–32
 imputation, 8–9
 methods for single marker test, 9–19
 overview, 7–8
 plasma lipid levels, 28–29
 rare variant associations, 20–28
 workflow of, 10t
 number of publications by year, 8f
 genomics. *See also* cancer genomics; latent variable approach, integrative clustering analysis
 epigenomic analysis, 75–76
 iBAG models
 linear, 221–224
 non-linear extensions, 224–226
 overview, 220–221
 Roadmap Epigenomics project, 109
 Genotype of Tissue Expression (GTEx) project, 87
 GEO (Gene Expression Omnibus), 37, 109, 397, 406
 germline variants, somatic mutations in cancer genomes, 333
 GES (greedy equivalence search) algorithm, DAG, 273
 GGMs (Gaussian graphical models), 203
*GI*₅₀ value, drug sensitivity, 421, 422
 Gibbs sampling procedure, 254–255
 GISTIC2 algorithm, 334
 Glioblastoma Multiforme (GBM), 219, 226, 406, 407, 409
 Glymour, Clark, 274
 GM12878 data, 378–379
 gMCP. *See* group MCP
 GMM (Gaussian mixture model), 153
 GOF (gain-of-function), PARADIGM-SHIFT pathway method, 319
 graphical (regression) networks, miRNAs, 297
 graphical models. *See also* Bayesian graphical models; directed graphical models
 graphical model and network analysis, 4
 probabilistic, 310
 greedy equivalence search (GES) algorithm, DAG, 273
 group MCP (gMCP), 102, 180
 analysis results of, 185t, 193t–194t

- analysis results of, with Laplacian penalty, 195t–198t
 defined, 175
 marker selection under heterogeneity model, 175–176
- GSEA. *See* Gene Set Enrichment Analysis
- GTEx (Genotype of Tissue Expression) project, 87
- GWAS. *See* genome-wide association studies
- H3K27me3* transcription factor, 146
- H3K4me3* mark, 387–388
- Hammersley Clifford theorem, 204
- HCPs (high CpG promoters), 391
- heat kernel model, cancer genomics, 310
- HER2* tumor subtype, 161–162
- heterogeneity model
 meta-analysis of GWAS, 16–17
 MetaDE package and, 49
 penalized integrative analysis of
 high-dimensional omics data, 174–180
- heterogeneous stock mice, 171, 183–188
- high CpG promoters (HCPs), 391
- high-dimensional transcriptional and drug sensitivity profile. *See* drug-pathway association analysis
- high-dimensionality, 219
- high-order cooperativity, in transcription regulatory networks, 71–73
- histone modifications (HMs), 109
 cause or consequence relationship between gene expression, 393–394
 ENCODE K562 and GM12878 data, 378–379
 framework for integrating with gene expression data, 376–377
 interplay between TF binding and other chromatin features, 391–392
 machine learning methods used in predictive models, 377–378
 ModENCODE Early Embryo data, 379
 Mouse ESC data, 379
 overview, 374–375
 performance evaluation of models, 378
 predicting differential gene expression, 388–389
 predicting expression levels for genes with HCP and LCP content, 391
 predicting expression levels of human promoters, 383f
 predicting expression of noncoding genes, 389–390
 predicting gene expression by combining with TF binding, 385–388
 predicting gene expression from, 382–385
 regulatory mechanism of, 392f
- regulatory signals in distal regions, 392–393
- Yeast and Fly data, 379
- hmChIP database, 109
- HMs. *See* histone modifications
- homogeneity model
 Cochran’s homogeneity test, 49
 penalized integrative analysis of
 high-dimensional omics data, 174–175
- horizontal meta-analysis, 1, 2f
- hot spots, eQTL, 267
- HOTAIR lncRNA, 408–409
- HotNet algorithm, 310, 323, 326
 applying to mutation data, 340–341
 config file used for running on TCGA GBM data, 353
 diffusion time used for PPI networks, 343f
 overview, 336–340
 parameter selection, 341–343
- HOTTIP lncRNA, 406
- HOX* family genes, 145
- hybrid methods, DAG, 274–276
- hypothesis settings, MetaDE package, 45, 46t, 47–48
- iASeq model, ASB, 123–125, 127f
- iBAG. *See* integrative Bayesian analysis of genomics data models
- iCluster method, 220
 exploratory methods for multisource data, 243
 joint analysis with lasso iCluster method, 162f
- iFad method, drug-pathway association analysis, 422–425
- ILP (integer linear program), 347
- imputation, meta-analysis of GWAS, 8–9
- incoherent FFLs, 295f
- individual structures
 JIVE method, 244–248, 247f
 miRNA and gene expression, 248–251, 250f
- inference
 Bayesian framework for DAG inference, 278–279
 causal inference, eQTLs, 270–271
 iFad method, 424–425
 mass-action-based model for gene expression regulation, 360–363
 network inference algorithms, 298
 posterior inference using false discovery rates, 208
- information flow, modeling, 88–91
- Ingenuity PathwayAnalysis (IPA) system, 142–143, 146, 146f
- inner lasso penalty, SNP, 176
- inner MCP penalty, 176

- integer linear program (ILP), 347
- Integrated Druggable Genome Database Project, 431
- integrated pathway level (IPL), 316
- integrated subtypes of colorectal cancer, 164–165
- integration with biological pathway information, 3
- integrative analysis. *See also* Bayesian graphical models; latent variable approach, integrative clustering analysis; penalized integrative analysis of high-dimensional omics data
 - of ChIP-X data, 108–128
 - allele-specific binding in ChIP-seq, 123–128
 - ChIP-chip peak calling, 118–123
 - correlation motif approach, 112–118
 - general problem setting and motivations, 110–112
 - overview, 108–110
 - of gene regulation
 - coupled transcription-splicing modules, 77–81
 - overview, 66–68
 - splicing modules, 73–77
 - transcriptional modules, 68–73
- integrative Bayesian analysis of genomics data models (iBAG models)
 - linear, 221–224
 - non-linear extensions, 224–226
 - overview, 220–221
- integrative quantitative models. *See also* histone modifications; transcription factor binding
 - cause or consequence relationship between gene expression, 393–394
 - ENCODE K562 and GM12878 data, 378–379
 - framework for integrating with gene expression data, 376–377
 - interplay between histone modification and other chromatin features, 391–392
 - machine learning methods used in predictive models, 377–378
 - ModENCODE Early Embryo data, 379
 - Mouse ESC data, 379
 - overview, 374–375
 - performance evaluation of models, 378
 - predicting differential gene expression, 388–389
 - predicting expression levels for genes with high and low CpG content, 391
 - predicting expression of noncoding genes, 389–390
 - predicting gene expression by combining with histone modifications, 385–388
 - predicting gene expression from, 379–382
 - regulatory signals in distal regions, 392–393
 - Yeast and Fly data, 379
- interaction potential (IP), TIE score, 97
- intergenic lncRNAs, 400f
- internal quality control (IQC), MetaQC package, 40
- interpretation, 219
 - of meta-analysis results, 19
 - patient-specific, 304–328
- interventional Markov equivalence classes, 273–274
- intronic lncRNAs, 400f
- inverse Wishart (IW) prior, 204
- IP (interaction potential), TIE score, 97
- IPA (Ingenuity Pathway Analysis) system, 142–143, 146, 146f
- iPad method, drug-pathway association analysis, 425–430, 434t–435t
- IPL (integrated pathway level), 316
- IQC (internal quality control), MetaQC package, 40
- iterative signature algorithm (ISA), 435
- IW (inverse Wishart) prior, 204
- jActiveModules plugin, 309
- JAMIE method
 - correlation motif model, 118–123
 - joint peak calling by, 121f
- JASPAR database, 81
- Ji, H., 122
- JIVE method. *See* joint and individual variation explained method
 - joint analysis with lasso iCluster method, 162f
 - joint and individual variation explained (JIVE) method, 244–251
 - application to TCGA data, 248–251
 - defined, 238
 - estimation, 245–246, 247f
 - gene expression (GE) and, 250f
 - illustrative example, 246–248
 - MetaPCA package, 55, 56f
 - microRNA (miRNA) and, 250f
 - model, 244–245
 - joint clustering, 256, 258f
 - joint NMF, 133f, 134–139
 - joint structure
 - JIVE method, 244–248, 247f
 - miRNA and gene expression, 248–251, 250f
- Kaplan-Meier curve
 - lung squamous cell carcinoma (lung SCC), 410f
 - ovarian cancer (OvCa), 410f
- Kaplan-Meier survival analysis, 138f
- KEAP1 mutation, 321–322, 321f
- KEGG database, 335, 336f

- KEGG MEDICUS database, 431
- KEGG pathways, 422
- drug-pathway associations identified by iPad, 434t–435t
 - network-regularized joint NMF method and, 142
- K-means clustering method, 153–154
- Laplace prior, 223
- Laplacian penalty, 182–183, 195t–198t
- “large *d*, small *n*” data, 170
- lasso (least absolute shrinkage and selection operator)
- inner lasso penalty, SNP, 176
 - joint analysis with lasso iCluster method, 162f
 - lasso iCluster method, 162f
 - Lasso prior, 223
 - overview, 426
- lasso iCluster method, 162f
- Lasso prior, 223
- latent variable approach, integrative clustering analysis
- example, 161–162
 - Gaussian mixture model, 153
 - integrated subtypes of colorectal cancer, 164–165
 - K-means clustering method, 153–154
 - latent variable models, 156–159
 - model selection, 160–161, 163–164
 - overview, 151–153
 - principal component analysis, 154–155
 - subtype analysis and, 156
 - TCGA colorectal cancer data set, 163
- late-onset Alzheimer’s disease (LOAD) brain study, 102–103, 103f
- LCPs (low CpG promoters), 391
- LD (linkage disequilibrium), 10t, 171
- least absolute shrinkage and selection operator. *See* lasso
- ligand-based approach, drug-pathway association analysis, 420
- linear iBAG model
- clinical model, 222–224
 - mechanistic model, 222
 - overview, 221–224
 - posterior probabilities, 228f–230f
 - prognostic markers, 231–232
- linkage disequilibrium (LD), 10t, 171
- LINKER approach, network integration of genetically regulated gene expression, 92
- linker genes, 322–323, 324f
- liver tissue
- gene transcripts in tissue containing eQTLs that overlap with insulin QTLs, 95f
 - top five genes ranked by TIE scores, 99t
 - lncRNAs. *See* long noncoding RNAs
- LOAD (late-onset Alzheimer’s disease) brain study, 102–103, 103f
- local eQTL, 267
- Lock, E. F., 152, 245, 248
- LOF (loss-of-function), PARADIGM-SHIFT pathway method, 319
- log-ratios of copy number, lung cancer sample, 159f
- Logsdon, Benjamin, 279
- long noncoding RNAs (lncRNAs)
- identifying, 405f
 - integrating lncRNA expression, 402–405
 - integrative analyses of in four cancer types, 406–413
 - overview, 398–399
 - repurposing microarray data to interrogate lncRNA expression, 399–402
- long tail phenomenon, 332f
- loss-of-function (LOF), PARADIGM-SHIFT pathway method, 319
- low CpG promoters (LCPs), 391
- lung squamous cell carcinoma (lung SCC), 406, 407, 409
- macular degeneration, 20f
- MAPE (meta-analysis pathway enrichment) methods, 52–55, 53f
- markers, integrative analysis, 170
- Markov chain Monte Carlo (MCMC), 204, 207–208, 255, 345, 424
- Markov equivalence classes (MECs), 272f
- Markov property, MRF, 204–205
- Markov random fields (MRF), BayesGraph for TCGA integration, 204–205
- MASS software package, 32
- mass spectrometry (MS), 357, 366–367
- mass-action-based model for gene expression regulation
- analysis of osmotic shock in yeast, 366–367
 - change-point model, 358–360
 - characterizing link between regulatory processes, 368–371
 - estimation and inference, 360–363
 - overview, 356–358
 - scoring protein-level regulation changes, 367–368
 - simulation study, 363–366
- MAT peak calling method, 122
- matrix decomposition, 422, 425, 435
- Matrix eQTL software, 270
- maximum p-value (maxP) statistic method, MetaDE package, 44, 46t, 47–48, 49t, 51
- max-min hill-climbing (MMHC) algorithm, 274–275
- MCC (multiclass correlation) method, 51
- mCCA (multiple canonical correlation analysis), 242

- MCMC (Markov chain Monte Carlo), 204, 207–208, 255, 345, 424
- MCP. *See* minimax concave penalty
- MDC (modular differential connectivity), LOAD brain study, 103, 103f
- MDI (multiple data set integration), 243
- MDL (minimum description length), DAG, 273
- mdmodules (multi-dimensional modules)
 - biological relevance of, 136–138
 - clinical associations of, 138–139
- MDRM. *See* multidimensional regulatory module
- MDS (multidimension scaling), 55
- ME (microRNA expression), 132, 136, 379, 389f, 390
- mean decreased Gini, TF binding, 380
- mean squared errors (MSE), 430
- mechanistic iBAG model, 222
- MECs (Markov equivalence classes), 272f
- MEMo method, 309
- Memorial Sloan-Kettering Cancer Center (MSKCC) Prostate Oncogenome Project, 405–406
- Mendelian randomization, 271
- Menezes, R. X., 202
- messenger RNA (mRNA)
 - concentration, 369f
 - gene expression, 201–203
 - iBAG models, 220–226
 - targeted cancer treatment and, 218
- meta-analysis methods, 3–4. *See also*
 - Bayesian Consensus Clustering method; meta-analysis of GWAS; principal components analysis
- meta-analysis of GWAS, 7–33
 - age-related macular degeneration, 20f
 - AMD GWAS, 19–20
 - GWAS-tailored software, 29–32
 - imputation, 8–9
 - methods for single marker test, 9–19
 - overview, 7–8
 - plasma lipid levels, 28–29
 - rare variant associations, 20–28
 - approaches, 23
 - burden tests that assume a distribution of variant effect sizes, 25–26
 - burden tests that assume variants have similar effect sizes for a simple burden test in study k, the impact of multiple rare variants, 24
 - conditional analyses, 26–28
 - meta-analysis of single-variant association test statistics, 24
 - Monte Carlo method for empirical assessment of significance, 26
 - overview, 20, 22–23
 - sharing summary statistics, 23–24
 - summary of loci, 21t
 - variable threshold tests with an adaptive frequency threshold, 25
 - workflow of, 10t
- meta-analysis pathway enrichment (MAPE) methods, 52–55, 53f
- meta-analytic framework for the liquid association (MetaLA) method, 58
- metabolic pathways, 421
- MetaClust package, 39t, 56–58
- MetaDE package, 39t, 42–52, 59–61
- MetaDiffNet network, 58
- MetaGeneModule approaches, MetaClust package, 57
- MetaLA (meta-analytic framework for the liquid association) method, 58
- MetaNetwork package, 39t, 58
- MetaOmics software
 - MetaClust package, 39t, 56–58
 - MetaDE package, 42–52, 59–61
 - MetaNetwork package, 39t, 58
 - MetaPath package, 39t, 52–55, 59–61
 - MetaPCA package, 3, 39t, 55–56
 - MetaPredict package, 39t, 58–59
 - MetaQC package, 38–42, 59–61
 - overview, 37–38
- MetaPath package, 39t, 52–55, 59–61
- MetaPCA package, 3, 39t, 55–56
- MetaPredict package, 39t, 58–59
- MetaQC package, 38–42, 39t, 59–61
- MetaSKAT software package, 32
- MetaSparseKmeans method, 56–57, 57f
- methods and materials. *See also names of specific methods*
 - BayesGraph for TCGA integration, 205–208
 - coupled transcription-splicing modules, 77–80
 - splicing modules, 73–74
 - transcriptional modules, 68–70
- methylation data
 - DNA methylation, 132, 136, 201–202
 - IGF1R* gene, 225f
- Metropolis-Hastings ratio, 279, 360–362
- MIAT lncRNA, 408
- microarrays. *See also* clustering; latent variable approach, integrative clustering analysis
 - combining effect sizes analysis, 44
 - combining *p*-values analysis, 42–44
 - evidence aggregation methods, 42–44
 - order statistics methods, 44
 - combining ranks analysis, 44–45
 - conventions, 2
 - for detecting differentially expressed genes, 38, 40–42

- modeling data sets, 69f
- repurposing microarray data to interrogate
 - lncRNA expression, 399–402
 - sequencing cancer genomes, 333–334
- microRNA expression (ME), 132, 379, 389f, 390
- microRNAs (miRNAs)
 - binding, 289–292
 - ComiR targeting method, 292–294
 - effect of single and multiple targets, 290
 - Fermi-Dirac combination of targets, 290–292
 - thermodynamics of, 289–290
 - cooperation between genes and, 148
 - JIVE method and, 248–251
 - network inference algorithms, 298
 - as network regulators, 294–298
 - network-regularized joint NMF method and, 141–143
 - overexpressing genes, 298–299
 - overview, 285–286
 - sponge effect, 298
 - target prediction algorithms, 286–289, 294
- microsatellite instability (MIN), 156, 164
- minimax concave penalty (MCP)
 - analysis results by applying MCP to each outcome separately, 191t–192t
 - defined, 174
 - marker selection under heterogeneity model, 175–176
 - mismatched penalties, 180
 - sparse group MCP (gMCP), 175–176
- minimum description length (MDL), DAG, 273
- minimum p-value (minP) statistic method, MetaDE package, 44, 46t, 47–48, 49t, 51
- miRanda targeting prediction algorithm, 287, 287t, 291f, 293f
- mirConnX algorithm, 297
- miRNA expression (ME), mdmodules, 136
- miRNAs. *See* microRNAs
- mirSVR targeting prediction algorithm, 287t, 289, 293f
- mirWIP targeting prediction algorithm, 287t, 289
- MMHC (max-min hill-climbing) algorithm, 274–275
- Mo, Qianxing, 152, 157
- model-based approach, integrative clustering. *See also* latent variable approach, integrative clustering analysis
 - integrative clustering analysis, 160–161, 163–164
- modeling genetic information flow, 88–91
- modENCODE project, 109, 379, 389f
- modular differential connectivity (MDC), LOAD brain study, 103, 103f
- molecular interaction network, 146f
- Monte Carlo method, meta-analysis of GWAS, 26
- Mouse ESC data, 379
- MRF (Markov random fields), BayesGraph for TCGA integration, 204–205
- mRNA. *See* messenger RNA
- MS (mass spectrometry), 357, 366–367
- MSE (mean squared errors), 430
- MSigDB gene set collection, 335
- MSKCC (Memorial Sloan-Kettering Cancer Center) Prostate Oncogenome Project, 405–406
- M-step, EM algorithm, 116–117
- multi-cancer markers, 171
- multiclass correlation (MCC) method, 51
- Multi-Dendrix algorithm, 343–344, 346–347, 346f, 349–351
- multidimension scaling (MDS), 55
- multi-dimensional modules (mdmodules)
 - biological relevance of, 136–138
 - clinical associations of, 138–139
- multidimensional regulatory module (MDRM)
 - regulatory analysis and, 146–147
 - synergistic functions across multiple dimensions, 145–146
- multimodality, TCGA, 201
- multi-platform datasets, schematic representation of, 218f
- multiple canonical correlation analysis (mCCA), 242
- multiple data set integration (MDI), 243
- multiple data sets, 172, 189–190. *See also* integrative analysis
- multi-source data, exploratory methods for
 - BCC method, 251–260
 - application to TCGA data, 257–260
 - Dirichlet mixture model, 252–253
 - estimation, 254–256
 - illustrative example, 256–257
 - multisource model, 253–254
 - overview, 251–252
 - clustering methods, 242–243
 - computational burden, 261
 - factorization methods, 241–242
 - formal framework, 240–241
 - JIVE method, 244–251
 - application to TCGA data, 248–251
 - estimation, 245–246
 - illustrative example, 246–248
 - model, 244–245
 - overview, 238–240
- MutationAssessor, 308
- mutations. *See also* somatic mutations in cancer genomes

- MutSig method, 308, 327f
MutSigCV algorithm, 334
- National Human Genome Research Institute (NHGRI), 421
NCI-60 project, 132, 431–433
NEAT1 lncRNA, 406
negative markers, iBAG models, 227–232, 232t
Neto, Elias Chaibub, 277, 278
network analysis, SNP data, 182–183
network inference algorithms, miRNAs, 298
network integration of genetically regulated gene expression
 diabetes genes, 93–100
 differential connectivity in coexpression network, 100–102
 LINKER approach, 92
 LOAD brain study, 102–103
 modeling genetic information flow, 88–91
 overview, 86–88
 PCST problem, 92–93
 PRINCE approach, 91–92
 random walk approach, 91
network regulators, miRNAs as, 294–298
network-regularized joint NMF method, 133f
 IPA system, 142–143
 KEGG pathways, 142
 miRNAs, 141–143
 overview, 139–141
 sparse network-regularized NMF algorithm, 140–141
network-regularized multiple NMF (NRNMF) framework, 140
Newton, Michael A., 208
next-generation sequencing (NGS), 37, 397, 402, 404
NFE2L2 (Nrf2) oncogene, 320–322, 321f
NFYB regulatory factor, 75–76
NGS (next-generation sequencing), 37, 397, 402, 404
NHGRI (National Human Genome Research Institute), 421
NMF. *See* nonnegative matrix factorization
noncoding RNA studies, data integration on application, 405–413
 clinical information, 405
 integrating lncRNA expression, 402–405
 overview, 397–399
 repurposing microarray data to interrogate lncRNA expression, 399–402
 somatic copy number alteration data, 403–405
non-linear iBAG model
 posterior probabilities, 228f–230f
 prognostic markers, 231–232
nonnegative matrix factorization (NMF), 55
 defined, 3
 joint NMF, 133f, 134–139
Normal-Exponential prior, 223
Normal-Gamma prior, 223, 225–226
not allele specific (NS) state, ASB, 124
Nrf2 (*NFE2L2*) oncogene, 320–322, 321f
NRNMF (network-regularized multiple NMF) framework, 140
NS (not allele specific) state, ASB, 124
nucleosome positioning, 109
observed occurrence index (OOI)
 analysis results by applying MCP to each outcome separately, 191t–192t
 analysis results of gMCP, 185t, 193t–194t
 analysis results of gMCP with Laplacian penalty, 195t–198t
 analysis results of sparse gMCP, 186t–187t
 defined, 188
oligonucleotide aCGH, 403–404
Oncodrive FM method, 308–309
“one drug – one target” approach, 419–420
OOI. *See* observed occurrence index
optimization algorithm, iPad method, 426–430
order statistics methods, combining *p*-values analysis, 44
outer MCP penalty, 176
ovarian cancer (OvCa), 406, 407, 409
overexpressing genes, 298–299
overlapping lncRNAs, 400f
overlapping subjects, 17
OWL (Web Ontology Language), 306
p53 protein, 313–314
PageRank teleporting random walk, 92
pairwise correlation matrices (PCMs), 101
pancreatic islets, 95f
PARADIGM pathway method, 298
 applications of, 317, 319
 interaction parameters, 316–317
 interactions and probabilistic factors, 314–316
 matrix of activities, 318f
 modeling components, 312f
 overview, 310–312
 variables, 312–314
PARADIGM-SHIFT pathway method
 analysis of *NFE2L2* and *KEAP1* mutations, 321f
 applications of, 320–322
 calculating shift scores, 319f
 overview, 319–320
parameter tuning, iPad method, 430
partial correlation, 211
partial least squares (PLS), 3, 241–242
partitioning explained variation, 235

- passenger mutations, 308–309, 331
 Pathifier method, 307
 PathOlogist method, 307
 PathScan approach, 334
 pathway analysis, 181–182, 307–308, 336f.
 See also drug-pathway association analysis
 Pathway Commons database, 306
 pathway databases, cancer genomics, 305–306
 Pathway Interaction Database, 306
 pathway-based drug discovery (polypharmacology), 420. *See also* drug-pathway association analysis
 pathway-based mutation assessment, 308–309
 PBR (potential binding regions), ChIP-chip peak calling, 120, 121f
 PC algorithm, DAG, 274–275, 281t
 PCA. *See* principal components analysis
 PCAN-R1 lncRNA, 411, 412f, 413
 PCAN-R2 lncRNA, 411, 412f, 413
 PCGs (protein encoding genes), 398
 PCMs (pairwise correlation matrices), 101
 PCST (prize-collecting Steiner tree) method, 92–93, 309
 PDIs (protein-DNA interactions) genome, 108
 peak calling, ChIP-chip, 118–123
 PECA (Protein Expression Control Analysis), 357–358
 penalization
 composite, 180
 methods, 5
 penalized integrative analysis of
 high-dimensional omics data
 data quality control and processing, 189–190
 examples, 170–173
 heterogeneity model
 marker selection, 175–180
 overview, 174
 heterogeneous stock mice, WTCCC, 183–188
 homogeneity model
 marker selection, 175
 overview, 174
 interplay among SNPs, 181–183
 network analysis, 182–183
 pathway analysis, 181–182
 overview, 170
 PenPC algorithm, 275
 phenotype-based approach, drug-pathway association analysis
 defined, 420
 iFad method, 422–425
 iPad method, 425–430
 phyloCSF method, 411
 PicTar targeting prediction algorithm, 287t, 288
 Ping-Pong algorithm, 132
 PITA targeting prediction algorithm, 287t, 288, 291f, 293f
 plasma lipid levels, meta-analysis of GWAS, 28–29
 PLS (partial least squares), 3, 241–242
 polypharmacology (pathway-based drug discovery), 420. *See also* drug-pathway association analysis
 positive markers, iBAG models, 227–232, 233t
 posterior inference
 for genes, 212t
 using false discovery rates, 208
 posterior probabilities
 linear iBAG model, 228f–230f
 non-linear iBAG model, 228f–230f
 potential binding regions (PBR), ChIP-chip peak calling, 120, 121f
 PPI (protein-protein interaction) networks, 81, 87, 265–266, 336–338, 338f, 343f
 pQTL (protein QTL) data, 281
 PR (Product of ranks) method, MetaDE package, 45
 predicting gene expression
 by combining with TF binding and histone modifications, 385–388
 differential gene expression, 388–389
 with high and low CpG content, 391
 from histone modifications, 382–385
 of noncoding genes, 389–390
 from TF binding, 379–382
 PRINCE approach, 91–92
 principal components analysis (PCA), 55
 consensus clustering, 260f
 exploratory methods for multisource data, 241, 247f
 integrative clustering analysis, 154–155
 mechanistic iBAG model, 222
 MetaQC package, 41, 60f
 PRINS lncRNA, 406
 prize-collecting Steiner tree (PCST) method, 92–93, 309
 probabilistic factors, PARADIGM pathway method, 314–316
 probabilistic graphical models, PARADIGM pathway method, 310
 probability model, BayesGraph for TCGA integration, 205–207
 Product of ranks (PR) method, MetaDE package, 45
 prognostic markers, iBAG models, 227–232
 protein concentration, 369f
 protein encoding genes (PCGs), 398
 Protein Expression Control Analysis (PECA), 357–358
 protein QTL (pQTL) data, 281

- protein synthesis (translation), 356–357, 364f, 369f
- protein-DNA interactions (PDIs) genome, 108
- protein-level regulation changes, scoring, 367–368
- protein-protein interaction (PPI) networks, 81, 87, 265–266, 336–338, 338f, 343f
- proteomics, 357, 359. *See also specific protein entries*
- QC measures. *See* quality control measures
- QDG (QTL directed dependency graph), 277–278
- QTL directed dependency graph (QDG), 277–278
- QTLnet method, 278
- quality control (QC) measures
MetaQC package, 40–41
Single Nucleotide Polymorphisms, 15t
- RACE (rapid amplification of cDNA ends), 411, 412f, 413
- random effects model (REM)
meta-analysis of GWAS, 16–17
MetaDE package, 44
- random forest (RF) method, 377, 380, 381f, 388, 393
- random walk approach, 89f, 91
- RankProd (RP) method, MetaDE package, 44
- rapid amplification of cDNA ends (RACE), 411, 412f, 413
- rare variant associations, meta-analysis of GWAS
approaches, 23
burden tests that assume distribution of variant effect sizes, 25–26
burden tests that assume variants have similar effect sizes for a simple burden test in study k, the impact of multiple rare variants, 24
conditional analyses, 26–28
meta-analysis of single-variant association test statistics, 24
Monte Carlo method for empirical assessment of significance, 26
overview, 20, 22–23
results for meta-analysis of gene-level rare variant association test, 30t
sharing summary statistics, 23–24
summary of loci, 21t
variable threshold tests with an adaptive frequency threshold, 25
- RAREMETAL software package, 32
- RDF (Resource Description Framework), 306
- Reactome pathways, 422
- recurrent heavy subgraphs (RHSs), 68, 69–70, 69f
- regression (graphical) networks, miRNAs, 297
- regularization methods, 5
- regularization parameters, 177
- regulatory processes, gene expression, 368–371
- relevance networks, miRNAs, 296
- REM. *See* random effects model
- Resource Description Framework (RDF), 306
- reversible edge (REV) proposal, 279
- reversible-jump Markov chain Monte Carlo (MCMC), 360, 367
- RF (random forest) method, 377, 380, 381f, 388, 393
- RHSs (recurrent heavy subgraphs), 68, 69–70, 69f
- RMST lncRNA, 408
- RNA. *See also* microRNAs
lncRNAs
identifying, 405f
integrating lncRNA expression, 402–405
integrative analyses of in four cancer types, 406–413
overview, 398–399
repurposing microarray data to interrogate lncRNA expression, 399–402
- mRNA
concentration, 369f
gene expression, 201–203
iBAG models, 220–226
targeted cancer treatment and, 218
- siRNAs, 412f, 413
- sncRNAs, 398
- RNA sequencing (RNA-seq), 87, 398–402, 406
- RNA synthesis (transcription), 356–357, 369f
- rna22 targeting prediction algorithm, 287t, 288
- RNA-seq (RNA sequencing), 87, 398–402, 406
- Roadmap Epigenomics project, 109
- RP (RankProd) method, MetaDE package, 44
- rth ordered p-value (rOP) statistic method, MetaDE package, 44, 46t, 47–48, 49t, 51
- S. cerevisiae* data with osmotic stress, 370f
- sample statistical analysis plan, 12t
- SCNA (somatic copy number alteration) data, 399, 403–406, 409
- search-and-score methods, DAG, 273–274
- SEMs (structure equation models), 277, 279–280
- separate clustering, 256, 258f
- sequence kernel association tests (SKAT), 22
- Sequence Read Archive (SRA), 37
- sequencing cancer genome/exome, 333–334
- short interfering RNAs (siRNAs), 412f, 413

- SIF (Simple Interchange Format), 306
 SIFT, 308
 signal transduction pathways, 421
 Signaling Pathway Impact Analysis (SPIA), 307
 significance test, iPad method, 430
 significantly mutated subnetworks, 336–343
 Simple Interchange Format (SIF), 306
 simulated data sets, 210f
 simulation study
 BayesGraph for TCGA integration, 208–211
 mass-action-based model for gene expression regulation, 363–366, 365f
 single marker test, meta-analysis of GWAS, 9–19
 single nucleotide polymorphisms (SNPs), 9, 333
 analysis results by applying MCP to each outcome separately, 191t–192t
 analysis results of gMCP, 185t, 193t–194t
 analysis results of gMCP with Laplacian penalty, 195t–198t
 analysis results of sparse gMCP, 186t–187t
 arrays, 404
 eQTLs and, 267, 268f
 meta-analysis of GWAS, 18t
 MetaDE for marker gene detection, 42
 penalized integrative analysis of
 high-dimensional omics data, 181–183
 network analysis, 182–183
 pathway analysis, 181–182
 quality control, 15t
 TCGA project, 132
 single-data-set analysis, 171
 single-nucleotide variants (SNVs), 333–334
 single-variant association test statistics
 Cochran-Mantel-Haenszel method, 24
 meta-analysis of GWAS, 24
 singular value decomposition (SVD), 246
 siRNAs (short interfering RNAs), 412f, 413
 SKAT (sequence kernel association tests), 22
 skeleton, DAG, 272
 skewed to the nonreference allele (SN) state, ASB, 124
 small noncoding RNAs (sncRNAs), 398
 SMBPLS. *See* sparse multi-block partial least squares regression
 SMR (standardized mean ranks), MetaQC package, 41
 SN (skewed to the nonreference allele) state, ASB, 124
 sncRNAs (small noncoding RNAs), 398
 SNMRMF (sparse network-regularized NMF) algorithm, 140–141
 SNPs. *See* single nucleotide polymorphisms
 SNVs (single-nucleotide variants), 333–334
 software, 29–32. *See also* MetaOmics software
 somatic copy number alteration (SCNA) data, 399, 403–406, 409
 somatic mutations in cancer genomes
 cancer (sub)type analysis, 350–352
 DAVID tool, 335
 Dendrix algorithm, 343–346, 347–349
 GSEA algorithm, 335–336
 Multi-Dendrix algorithm, 343–344, 346–347, 349–351
 overview, 331–333
 sequencing, 333–334
 significantly mutated subnetworks, 336–343
 sparse group MCP (gMCP), 178–180, 186t–187t
 sparse multi-block partial least squares regression (sMBPLS), 3, 133f, 143, 144, 148–149
 multidimensional regulatory module, 145–147
 overview, 143–144
 sparse network-regularized NMF (SNMRMF) algorithm, 140–141
 SPIA (Signaling Pathway Impact Analysis), 307
 splicing modules
 exons, 76–77
 identifying novel functions associated with co-splicing but not coexpression, 76
 methods and materials, 73–74
 transcriptional and epigenomic analysis, 75–76
 sponge effect, miRNAs, 298
 squared Euclidean error function, NMF, 134–135
 SR (Sum of ranks) method, MetaDE package, 45
 SRA (Sequence Read Archive), 37
 SRF transcription factor, 145
 SSC (sum of squared cosines), MetaPCA package, 55
 standardized mean ranks (SMR), MetaQC package, 41
 STAT1 transcription factor, 145
 statistical methods. *See* integrative analysis; latent variable approach, integrative clustering analysis; MetaOmics software
 Stouffer's method
 MetaDE package, 43, 46t, 47–48, 49t, 51
 microarray meta-analysis, 43
 structural variants (SVs), 333–334
 structure equation models (SEMs), 277, 279–280
 structured pathway methods, analysis of gene expression, 307

- subtype analysis, integrative clustering analysis, 156
- Sum of ranks (SR) method, MetaDE package, 45
- sum of squared cosines (SSC), MetaPCA package, 55
- sum of variance (SV), MetaPCA package, 55
- support vector machine (SVM), 293, 377, 388, 393
- SV (sum of variance), MetaPCA package, 55
- SVD (singular value decomposition), 246
- SVM (support vector machine), 293, 377, 388, 393
- SVs (structural variants), 333–334
- TAF8* regulatory factor, 75
- target exclusivity, miRNAs, 289
- target prediction algorithms, miRNAs, 294
 - ComiR, 287t, 288
 - miRanda, 287, 287t
 - mirSVR, 287t, 289
 - mirWIP, 287t, 289
 - overview, 286–287
 - PicTar, 287t, 288
 - PITA, 287t, 288
 - rna22, 287t, 288
 - TargetScan, 287–288, 287t
- target-based approach
 - differential coexpression, 101
 - drug-pathway association analysis, 420
- targets, miRNAs, 290
- TargetScan targeting prediction algorithm, 287–288, 287t
- TCGA project. *See* The Cancer Genome Atlas project
- TF binding. *See* transcription factor binding
- TF+HM model, 386–387, 386f
- TFBSs (transcription factor binding sites), 108
- TFs. *See* transcription factors
- The Cancer Genome Atlas (TCGA) project, 132, 152, 163, 165f. *See also* Bayesian graphical models; somatic mutations in cancer genomes
 - application of BCC method to data, 257–260
 - application of JIVE method to data, 248–251
 - BayesGraph for TCGA integration, 211–214
 - cancer types and samples for integrative analysis, 212f
 - integrative clustering analysis, 163
 - TCGA GBM study, 334
- third-order tensor, transcriptional regulatory modules, 69f
- TIE score, 97–98, 99t
- TieDIE pathway method, 322–328, 327f
- TileMap peak calling method, 122
- time course experiments
 - analysis of osmotic shock in yeast, 366–367
 - change-point model, 358–360
 - characterizing link between regulatory processes, 368–371
 - estimation and inference, 360–363
 - overview, 356–358
 - scoring protein-level regulation changes, 367–368
 - simulation study, 363–366
- TIPC (trait-IP correlation), 97–98
- T-Lymphocyte Helper/Suppressor Profile. *See* CD4/CD8 ratio
- top scoring pair (TSP) algorithm, prediction analysis, 58–59
- Tpil* gene, 99t
- trait-IP correlation (TIPC), 97–98
- transcription (RNA synthesis), 356–357, 369f
- transcription, defined, 201
- transcription factor binding sites (TFBSs), 108
- transcription factor (TF) binding
 - cause or consequence relationship between gene expression, 393–394
 - ENCODE K562 and GM12878 data, 378–379
 - framework for integrating with gene expression data, 376–377
 - interplay between histone modification and other chromatin features, 391–392
 - machine learning methods used in predictive models, 377–378
 - ModENCODE Early Embryo data, 379
 - Mouse ESC data, 379
 - overview, 374–375
 - performance evaluation of models, 378
 - predicting differential gene expression, 388–389
 - predicting expression levels for genes with HCP and LCP content, 391
 - predicting expression of noncoding genes, 389–390
 - predicting gene expression by combining with histone modifications, 385–388
 - predicting gene expression from, 379–382
 - regulatory signals in distal regions, 392–393
 - Yeast and Fly data, 379
- transcription factors (TFs), 71–72, 72f
 - gene expression, 87, 89f
 - GLI1, 110, 111f, 112
 - GLI3, 110, 111f, 112
 - H3K27me3*, 146
 - predicting expression levels of human promoters, 381f
 - regulatory mechanism of, 392f
 - SRF*, 145
 - STAT1*, 145

- transcriptional analysis, splicing modules, 75–76
- transcriptional modules
 - high-order cooperativity and regulation in transcription regulatory networks, 71–73
 - methods and materials, 68–70
- transcriptional regulation, 72f, 379, 387–388, 391–392
- transcriptomics meta-analysis. *See also* mass-action-based model for gene expression regulation
 - for differential network detection, 58
 - MetaClust package, 56–58
 - MetaDE package, 42–52, 59–61
 - MetaNetwork, 58
 - MetaPath package, 52–55, 59–61
 - MetaPCA, 55–56
 - MetaPredict, 58–59
 - MetaQC package, 38–42, 59–61
 - overview, 37–38
- trans-eQTLs, 87, 268–271, 269f
- translation (protein synthesis), 356–357, 364f, 369f
- trans-regulated gene expression, 87
- TRe-CASE model, 270
- TSP (top scoring pair) algorithm, prediction analysis, 58–59
- txCdsPredict method, 411
- undirected networks, 190
- uniform design (UD), sampling method, 160
- unsupervised analysis
 - Bayesian consensus clustering (BCC), 3
 - cluster analysis, 3
 - iCluster method, 3
 - MetaSparseKmeans method, 3
 - overview, 3
- untargeted approach, differential coexpression, 101
- variable selection, 170, 174
- variables
 - PARADIGM pathway method, 312–314
 - variable threshold tests with an adaptive frequency threshold, 25
- Venn diagram
 - enriched pathways identified by MAPE, 60f
 - lncRNA located in SCNA regions of cancer, 410f
 - subtype-specific lncRNA in cancers, 408f
- vertical multi-omics analysis, 1, 2f
- V-fold cross-validation, 178
- v-structures, DAG, 272–273
- walking in gene network, 89f
- Web Ontology Language (OWL), 306
- Wellcome Trust Case Control Consortium (WTCCC), 171, 183–188
- WGCNA package, 102
- white adipose tissue
 - gene transcripts in tissue containing eQTLs that overlap with insulin QTLs, 95f
 - top five genes ranked by TIE scores, 99t
- whole-cell pathway model
 - active subnetwork search and discovery, 309–310
 - PARADIGM pathway method, 310–319
 - PARADIGM-SHIFT pathway method, 319–322
 - pathway databases, 305–306
 - pathway methods, 307–308
 - pathway-based mutation assessment, 308–309
 - TieDIE pathway method, 322–328
- whole-exome sequencing, 333
- whole-genome sequencing, 333
- WTCCC (Wellcome Trust Case Control Consortium), 171, 183–188
- Yeast and Fly data, 379