Index

a

acceleration of analysis 396
accuracy 424
adapter 314, 316
adapter fragment 9
adapter ligation 131f., 304, 306, 311, 374
adapter preparation 9
adapter sequence 265
adapter-tag formation 12
allele-specific expression 349f.
amplification 12, 19, 251, 253
amplified fragment length polymorphism (AFLP) 288
analysis 456ff.
analytical tool 467
annotation 455, 458, 471
annotation data 458f.
anoxic 171ff., 176ff., 181
antibody specifity 155
antibody validation 141
application programming interface (API) 396
authentication 465

b

background tag distribution 427
bacterial artificial chromosome (BAC) 287ff.
– BAC clone 287ff., 295ff.
– BAC library 291
– BAC pooling strategy 290f.
barcode 26, 59, 157, 265, 289, 422
bioinformatics 337f., 343f., 383ff., 393f., 404f., 449
biological data integration 407ff.
biotinylation 79
BLAST 4f., 15
Burros-Wheeler Aligner (BWA) 268f., 408
Burrows-Wheeler Transform (BWT) 269

c

cag pathogenicity island 114f.
CAGE 23ff., 47ff., 64ff.
CAGE library 27
CAGEscan 45, 47ff.
– applications 59ff.
– definition 48ff.
– key reagents 51f.
– library 53f.
– methods 51ff.
– perspectives 59ff.
– precautions 53
– protocols 51ff.
– RNA samples 53
cancer genome 283
cap analysis gene expression (CAGE) 23ff., 44f., 47f., 64, 353, 392, 410
– definition 24ff.
capping enzyme complex (CEC) 24, 48
Cariaco Basin 173, 175ff., 181
cDNA
– database 15
– library preparation 118, 120
– precipitation 231
– purification 55, 218
– synthesis 10, 75
cell lysis 248
ChIA-PET library 188
– construction 189f., 193, 200
– mapping 190
– quality control 207
– sequencing 190
chromatin 243ff, 250, 254
chromatin immunoprecipitation (ChIP) 139ff., 185, 188, 198, 245f., 412, 417, 423, 455
– buffers 142
– quality control 207
chromatin immunoprecipitation sequencing (ChIP-seq) 139, 327, 364, 368, 385, 393, 401ff., 413f., 417
– advantages 149
– applications 147
– challenges 149
– clinical application 149
– data analysis 146f.
– enrichment 430
– mapping of tags 402
– methods 141ff.
– negative control 426
– perspectives 150
– protocols 141ff.
– relation enrichment factor 421
– sequencing depth 421
– sequencing library preparation 144ff.
chromatin interaction analysis using paired-end tag sequencing (ChIA-PET) 185ff.
– anticipated results 207
– applications 187
– comparison with other methods 186
– development 186

Tag-based Next Generation Sequencing, First Edition. Edited by Matthias Harbers and Günter Kahl.
© 2012 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2012 by Wiley-VCH Verlag GmbH & Co. KGaA.

– equipment 194
– experimental design 187
– key reagents 192f.
– methods 192ff.
– oligonucleotide sequences 196
– overview 186
– perspectives 209
– protocols 192ff., 195
– quality control 207
– reagent set-up 195
– timeline 206
circular cDNA 75
CLC bio integrated platform 393ff., 397
– applications 396ff.
– features 394
– main components 394
– perspectives 404
CLC Genomics Server 395
CLC Genomics Workbench 394
cloning 278
cloud computing 449ff.
cloud-based infrastructure 461f.
cloud-based solution 385
cluster 408
collaboration 454
comparative sequence analysis 407
computational infrastructure 385ff.
– application 392
– computational consideration 388
– data dynamics 389
– perspectives 392
– postanalysis 390
– software 390
– staffing requirement 391f.
contaminant 320
contamination 431
contig 293
control 432
control library 191
copy number variation (CNV) 257f., 409
correlation measure 431
cross-linking 199, 202, 247
cross-linking and immunoprecipitation (CLIP) 153ff.
cycle number check 41
cytosine methylation profiling 299ff.

d

dA-tailing 305
data analysis 232ff., 236, 300, 385ff., 415, 452f.
data analysis algorithm 449, 459
data back-end 394
data collection 454f.
data distribution 455f.
data flow 394
data import 396f.
data management 391, 450
data processing 261f.
data quantity 450
data repository 449, 456f.
data retention requirement 390
data security 463
data size 453
data storage 385, 391
data visualization 449
database management system (DBMS) 445
de novo sequencing 394
deepCAGE 23ff., 418
– applications 43f.
– key steps 25
– methods 28ff.
– protocols 28ff.
deepCAGE library 25, 30
– preparation 32ff.
deep-RACE 65
– analysis 68
– applications 66
– choice of the sequencer 65
– limitations 66f.
– validation of promoter studies 65
deep-RACE library 67f.
deepSAGE 418ff., 423, 425f., 430
deepSuperSAGE 3ff., 6f., 17f.
– application 14ff.
– development 6f.
demultiplexing 397
deoxynucleotide triphosphate (dNTP) 328
differential RNA sequencing (dRNA-Seq) 109ff.
– advantages 113
– applications 119f.
– definition 111
– materials 115
– methods 115ff.
– perspectives 120f.
– precautions 116
– protocols 115ff.
– workflow 113
digital gene expression (DGE) 7, 349, 354
digital signature 465
dinoflagellate 176f.
– taxonomic groups 177
ditag-based deepSuperSAGE 6f.
ditag concatemer 280
ditag formation 12
ditag genome scanning (DGS) 277ff.
– application 283
– cloning-based protocol 278
– methods 278ff.
– non-cloning-based protocol 281
– protocol 278ff.
DNA fingerprinting 288
DNA fragment, blunting 251f.
DNA fragmentation 311, 370
DNA methylation 299, 308
DNA microarray 243f., 250, 337
DNA nick repair 85
DNA precipitation 249
DNA-protein complex 186
DNA purification 218, 231
DNA quality 319
DNA tag 393
double-stranded cDNA 39, 81, 110, 341
dRNA-Seq library 115
– preparation 116
– RNA samples 116
DSN digestion 375ff.

e

*Eco*P15I digestion 39, 85, 305
embryonic stem cells (ES) 148
encryption 465
end-repair reaction 305, 372
end trimming 398
exon 63ff., 379

exon count 348
exonuclease I 43
experimental control 425
expressed sequence tag (EST) 4, 15, 23, 64, 91f., 211, 339, 342, 344, 346, 348, 398, 458
expression analysis 403
expression profiling 123ff., 420

f
false-discovery rate (FDR) 402
FastaTrack 442
FASTQ format 262, 267, 451ff.
FASTX-toolkit 266
first-strand cDNA synthesis 54, 99, 110, 215, 217, 228, 306, 340, 371
formaldehyde-assisted isolation of regulatory elements (FAIRE) 243ff., 250ff.
– application 254
– equipment 246
– perspectives 254
– procedure 244f.
– procedure optimization 245
formalin-fixed paraffin-embedded sample (FFPS) 369
FPC program 293f.
FPC software 289
fragments per kilobase of exon per million fragments mapped (FPKM) 459f.
Framvaren Fjord 176ff.
FTP site 454
full genome resequencing 394
full-length cDNA 83
full-length transcript 337ff.
full-length transcriptome 338ff.
– generation 338ff.
– sequencing 341
functional annotation 407, 409
fusion transcript detection 347f.

g
gel electrophoresis 57
gel purification 166
gene annotation 380
gene discovery 342, 380
gene expression 23f., 26f., 44, 109f., 114f., 119, 123, 135f., 139, 148, 211ff., 232, 240f.
Gene Expression Omnibus (GEO) 455f.
gene ontology (GO) 16, 18
gene regulation 139, 149
gene transcript 342
general quality assessment 427
genetic disorder 284
genetic variation 287
genome analysis tool kit (GATK) 272, 274
genome analyzer (GA) 7, 73f.
genome annotation 407, 409ff.
genome browser 435ff., 441
– publishing 443
– visualization 89
genome enrichment 325, 331
genome equivalence (GE) 288
genome project 295
genome sequencer 6
genome sequencing 337
genome structural analysis 277ff., 283
Genome Wide Association Studies (GWAS) 460
genome-wide DNA methylation analysis 299
genomics 287
globin depletion 380
Google web toolkit (GWT) 437
*Gsu*I digestion 82
guideline 432

h
hairpin structure 472
header (HD) 270
Health Insurance Portability and Accountability Act (HIPAA) 464
healthcare data 464
Helicobacter pylori 114
Helicos DGE 354
– application 362ff.
– precautions 357
– primers 356
– reagent 355
– sample preparation 357
– thermocycler programs 355
Helicos single-molecule sequencing 353ff., 364
HELP-Tagging 299ff.
– adapter preparation 302
– application 308
– buffer preparation 302
– definition 300
– library preparation 302ff.
– material 301f.
– methods 301ff.
– perspectives 308
– precautions 303
– protocol 301ff.
high-frequency allele 260
high-speed sequencer (HSS) 465ff., 451ff.
high-throughput sequencing (HTS) 65, 110, 112, 123ff., 135ff., 185, 189f., 206, 252f., 393f., 396, 400, 405, 435f.
high-throughput SuperSAGE (HT-SuperSAGE) 5, 8, 10f., 13f.
– practical analysis 18f.
– protocol 18f.
high-throughput transcriptome sequencing 3ff.
human diseases 149
hybrid selection method 332

i
Illumina FASTQ 263
Illumina genome analyzer 211ff., 219, 223, 226, 245, 253, 261ff., 288, 290, 300f., 317, 318, 321, 325ff., 330ff., 367ff., 429
Illumina sequencing 292f., 307, 313
immunoprecipitation 153, 155ff., 159, 162
in silico cloning 394
in vitro tagmentation 311ff.
– materials 313
– methods 313f.
– protocols 313
in vitro transposition 321
incorporation 341
individual-nucleotide resolution CLIP (iCLIP) 153ff.
– methods 159ff.
– modifications 158
– protocol 155, 159ff.
– troubleshooting 158
input file 267

insertion/deletion (Indel) 257f., 269f., 273ff., 465
integrated analysis 393
interindividual analysis 149
interspecies analysis 149
Ion Torrent electronic sequencing 329f.
isoform 346
isogroup 342
isotig 346f., 350

l
laboratory information management (LIM) 388, 450, 452
lane effect 423
large-scale genome sequence 435
library 312, 314ff.
– preparation 311f., 314, 317, 321
– preparation strategy 259ff.
– quality control 279, 281
– statistics 235ff.
ligation product 37
lightweight directory access protocol (LDAP) 464
linker 189f., 265
linker ligation 40, 163, 201, 251
linker sequence 168f.
linker-tag fragment 12
Linux 438, 444, 461
local server 385
long serial analysis of gene expression (LongSAGE) 73, 211f., 235, 239f.

m
Mac OS X 438, 444
mapping 277f., 282ff., 287ff., 343, 345, 471
mapping to the genome (MG) 213, 235
mapping to the transcriptome (MT) 213, 235
marine environment 171ff., 181
massively parallel sequencing (MPS) 299ff., 308, 311
massively parallel tag sequencing 171ff., 181
MethylC-seq 299f., 308
micro-oxic 171, 173, 181
microRNA (miRNA) 123ff., 135ff., 378, 458, 469ff.
– discovery 469ff., 474
– expression profiles 474
– identification 469
– methods for identification 469
– primary transcript (pri-miRNA) 469
miRNA precursor (pre-miRNA) 469
miRNA profiling 430
miRNA signature 472f.
mirtron 124
model organism 344
molecular inversion probe (MIP) 331
molecular profiling 325
MOSAIK 267ff.
mRNA fragmentation 339f.
mRNA preparation 339

n
nanopore sequencing 328f.
National Center for Biotechnology Information (NCBI) genome browser 435, 447, 451, 455
natural antisense transcript-derived siRNA (natsiRNA) 126
negative control 426
network attached storage (NAS) 389
Nextera 311ff., 316f., 319ff.
next-generation manufacturer 387
next-generation sequencing (NGS) 3, 5f., 19, 73ff., 91f., 103, 126f., 153ff., 186, 211ff., 239, 243f., 257ff., 275, 277, 287ff., 311, 321, 325ff., 331ff., 337, 385ff., 407ff., 417ff., 449
– advantages 331
– application 412f.
– comparison of NGS technologies 334
– experimental design 417ff.
– quality control 417ff.
NGS analyzer 412
nick translation repair 86
*Nla*III digestion 10, 19
nucleotide frequency characteristics 428
nucleotide variation 257ff.
number of sequences 427

o
oligonucleotide 30
oligonucleotide design 156f.
operating system (OS) 463
operational taxonomic unit (OUT) 176, 178, 180
oxic 171, 173, 176ff.

p
paired-end data 269
paired-end sequencing 277ff., 282ff., 408, 419
paired-end tag (PET) 73ff., 185ff., 279
– clustering 89
– mapping 88
– PET sequencing with SOLiD 88
PCR amplification 86, 133, 137, 157, 167, 174, 205, 307, 331, 374, 377, 428
peak refinement 402
percentage duplicate reads 428
percentage in repeat regions 430
percentage of reads aligned 430
peroxisome proliferator-activated receptor-γ (PPAR-γ) 148, 414
phenol/chloroform extraction 249
phylogenetic annotation 407, 409f.
physical mapping 287ff., 295ff.
– history 288
pileup file 272
Piwi-interacting RNA (piRNA) 123, 125ff., 135f.
plasmid-safe treatment 85
plasmid vector 278
platform 418
– advantages 419
– disadvantages 419
platform-specific quality score 428
poly(A) mRNA 75, 78, 95, 312
poly(A) selection 378f.
poly(A) tail 359ff., 364, 380
polyacrylamide gel electrophoresis (PAGE) 8
polymerase chain reaction (PCR) 8, 49f., 63, 140, 174, 187f., 224, 243, 250, 278, 301, 375, 422
– cycling conditions 87
polymorphic repeats 335

polymorphism 333
polynucleotide kinase (PNK) 156
population genomics 259
post-transcriptional regulation 153
precaution 30
precursor microRNA (pre-miRNA) 123f., 135, 472
preparative PCR 42, 167
primary transcriptome 109ff.
primer 101, 314
primer design 67, 394
primer extension 82
primer ligation 166
primer sequence 168f.
proximity ligation 185f., 188ff., 202, 206
promoter 23f., 26f., 44f., 65
promoter-associated small RNAs (PASRs) 51
promotome 47ff.
protein-DNA interaction 139ff.
protein-RNA interaction 153ff.
protist community 171f., 178f.
protistan gene sequence 174
pyrosequencing 9f., 12, 14, 73, 171, 173ff., 184, 264

q
quality check 430
quality control 155
quality evaluation 265
quality score 429
quality trimming 266
quantitative polymerase chain reaction (qPCR) 3, 38, 244, 250

r
RACE kits 64
random primer (RT-primer) 49f., 120
randomization 422, 424
rapid amplification of cDNA ends (RACE) 3, 6, 63ff., 110
– applications 63ff.
reactive oxygen species (ROS) 16f.
read 408, 418
read-depth distribution 272
read group (RG) 270
read length 418, 427
reads per kilobase of exon model per million mapped reads (RPKM) 349, 399, 420
ReadTrack 442
real-time sequencing 327f.
reduced representation bisulfite sequencing (RRBS) 301
reduced representation library (RRL) 258, 260f., 266
reference genome 257f., 266ff., 274
regulatory element 243ff., 254
regulatory sequence 411
replicate 422ff.
restriction endonuclease 288
restriction enzyme 299f., 303
restriction fragment 277f., 283f.
retinoid X receptor (RXR) 148
reverse transcript primer 31f.
reverse transcription 132, 157, 165
reverse transcription-polymerase chain reaction (RT-PCR) 3, 38, 348
ribosomal RNA (rRNA) 127, 130
RNA 91ff.
RNA digestion 161, 358
RNA fragmentation 98
RNA immunoprecipitation (RIP) 153
RNA integrity number (RIN) 32, 54, 95, 357, 369
RNA interference (RNAi) 141
RNA isolation 164
RNA preparation 30
RNA quantitation 353ff., 364
RNA sequencing (RNA-seq) 91ff., 109ff., 113f., 120, 327, 353, 362ff., 393, 398ff., 404, 417, 430, 459
– application 93, 104f.
– bioinformatic considerations 103f.
– key equipment 94
– key reagents 94
– materials 94
– methods 93ff.
– perspectives 105f.
– protocol 93ff., 370
– RNA preparation 95
RNA-binding protein (RBP) 135, 153, 155f., 158, 162
RNA-dependent RNA polymerase (RdRP) 126
RNA-induced silencing complex (RISC) 124, 470, 474
RNA-PET 73ff.
– methods 75ff.
– oligonucleotide sequences 76
– protocols 75, 78ff.
RNA-seq library 368

s
salt-overly-sensitive (SO) pathway 16
SAM format 270f.
sample clean-up 358, 361
sample evaluation 465
sample size 424
SAMtools 271ff.
Sanger FASTQ 263
sanger sequencing 326
sans singletons and one-offs of highly expressed tags (SSOOHE) 213, 235
scale-up PCR 134
SCARF format 264f.
SDS-PAGE 155, 162f.
second-generation sequencing 311ff.
second-strand cDNA synthesis 38f., 56, 82, 215, 217, 340, 372
security 462f.
sequence dictionary (SQ) 270
sequence reads 333
sequence reads archive (SRA) 451
sequence tag analysis of genomic element (STAGE) 140
sequencer 65, 106
sequencing 26f., 30, 43ff., 68, 139ff., 144ff., 149f.
sequencing adapter 86
sequencing budget 295
sequencing-by-hybridization 386
sequencing-by synthesis (SBS) 127, 367ff., 386
sequencing depth 420f.
serial analysis of binding elements (SABE) 140
serial analysis of chromatin occupancy (SACO) 140

serial analysis of gene expression (SAGE) 3f., 23f., 26, 48, 110, 211, 353
shogun sequencing 91ff.
short reads 333
short reads alignment 335
short tail correction 361
shotgun library 91, 98
Simbiot 465f.
single instruction multiple data (SIMD) technology 396, 404
single nucleotide polymorphism (SNP) 91, 93, 104f., 257ff., 269ff., 296, 333, 337ff., 343ff., 349f., 353, 362, 409, 419, 447, 458f.
– SNP assay 259
– SNP calling 345
– SNP chip 259
– SNP discovery 259ff., 265
single-end sequencing 419
single-molecule sequencing 353ff., 357, 364
single-molecule template 327
single-nucleotide misincorporation 341
single-nucleotide resolution 153
single-strand linker ligation 37f.
single-tag-based deepSuperSAGE 8f.
small interfering RNA (siRNA) 123f., 126, 131, 136, 471
small nuclear RNA (snRNA) 127
small nucleolar RNA (snoRNA) 127
small regulatory RNA (sRNA) 111
small RNA 123ff., 469ff.
small RNA isolation 129f.
small RNA library preparation 128ff.
– application 134ff.
– library preparation 128
– methods 128ff.
– protocols 128ff.
small RNA sequencing 393, 399, 400f.
small RNA sequencing data 469ff.
small subunits (SSU) 172, 174f., 177ff.
SOLiD 73ff., 88, 94, 97, 105, 119, 127, 190, 264, 288, 327, 398, 429, 456f.
somatic rearrangement 283
sonication 248
spike-in 425
splice junction 379
SSU rDNA 172
SSU rRNA 174f., 177f., 180f.
statistical testing 422
storage area network (SAN) 389
structural variant (SV) 257
SuperSAGE 3ff.
– principle 5
SuperSAGE tag 5f.
suplex-specific nuclease (DSN) 369
systematic evolution of ligands by exponential enrichment (SELEX) 153

t

tag 353ff.
tag-based expression profiling 417
tag-based RNA quantitation 353ff.
tag-based transcriptome technology 392
tag extraction 10f.
tag sequencing (Tag-seq) 211ff., 393, 400
– annotate tag experiment tool 401
– applications 239f.
– create virtual tag list tool 401
– critical parameters 214
– extract and count tool 400
– materials 215f., 219, 223, 226f., 232
– methods 215ff.
– protocol 212f., 215ff., 219ff., 223ff., 226ff., 234
– timeline 213
– troubleshooting 214
tagmentation reaction 314, 317, 320
Tag-seq library 212, 241
– construction 227
technical variability 423
template 311f., 316, 319f., 331
template preparation 14, 325ff., 334
template-switching primer (TS primer) 49f.
terminator exonuclease (TEX) 111ff., 117f., 120
thermocycler 356
third-generation sequencing technology 325ff., 453
tobacco acid pyrophosphatase (TAP) 112, 117f.
total RNA isolation 129
total RNA-seq 367ff.
– application 380
– data collection 378ff.
– interpretion 378ff.
– methods 370ff.
trans-acting siRNA (tasiRNA) 126
transcript length coverage 344, 347
transcription factor (TF) 140, 410, 412, 414
transcription factor binding site (TFBS) 410ff.
– *de novo* definition 413
transcription start site (TSS) 23f., 26, 44, 47f., 50f., 59f., 64, 109ff., 119ff., 147f.
transcriptome 26, 47ff., 50, 63, 68, 106, 367ff.
transcriptome analysis 337f., 346
transcriptome annotation 407, 409
Transcriptome Assembler 342f.
transcriptome assembly 342, 345
transcriptome mapper 343
transcriptome profiling 335
transcriptome sequencing 326, 335, 337
transcripts per million (TPM) 420
transfer RNA (tRNA) 127, 130
transposase 312
transposome 311ff., 317ff.
transposon 213ff., 317, 320f.
trans-splicing 347f.

u

understanding 391
UniTag 16f.
University of California Santa Cruz (UCSC) genome browser 435f., 445, 471
University of Tokyo genome browser (UTGB) toolkit 435ff.
– add a new track 442
– applications 444
– configure track views 441
– database connection 446
– develop own tracks 444
– help messages 439f.
– installation 438
– keyword search 443
– manual installation 444

– methods 438ff.
– new UTGB project 440
– object-database mapping 446
– overview 436ff.
– perspectives 447
– portable database engine 445
– portable web server 441, 444f.
– prerequisites 438
– run the UTGB toolkit 439f.
– web action 446
– web application 444, 446
untranslated region (UTR) 66, 104, 114, 123, 338, 346
UV cross-linking 153, 155, 159, 164

v
variant calling 257, 261, 266, 269f., 273ff.
virus-induced gene-silencing (VIGS) 18

w
whole-genome profiling (WGP) 287ff.
– applications 294f.
– definition 289
– perspectives 296f.
– project funding 295
– WGP project 289f.
WigTrack 443
Windows 439, 444, 472

y
yeast artificial chromosome (YAC) 287f.

z
zero-mode waveguide (ZMW) 327
zinc acetate 98
zymo protocol 316, 318