INDEX

adaptive shape segmentation, 68
melting temperature, 55
probe selection for, 49, 51–54
Affymetrix technology, 6, 62
data and, 74–75
experimental design and, 214–21
labelling in, 10
negative intensities and, 75
spot quality
age, defined, 253
Agilent 6, 8
analysis of variance (ANOVA), 112, 127, 220
multifactor, 135
one-way, 134–36
ANOVA. See analysis of variance
arrangement, of samples, 214–21
attachment chemistry, 4
average intensity, 81–90
average linkage, 159, 166
background signal, 68–69
subtraction, 74
balanced design, 213
classification methods, 94, 97, 184–85, 189
advantages and disadvantages, 188
centroid classification and, 188, 191, 193
dimensionality reduction, 198–208
diseases and, 183
individual gene selection, 203
K-nearest-neighbours, 191–92
neural networks and, 193
noise and, 198
pairwise gene selection, 204
validation and, 195–98
voting algorithms, 205
See also specific methods, parameters
clustering, 94, 97
bootstrapping and, 169–71
chaining, 183
distance measures and, 167
ESTs, 28, 33
Euclidean distance and, 167
hierarchical clustering, 158, 174
k-means algorithm, 174–78
machine-learning methods for, 174–80
Pearson correlation and, 167
reliability of methods, 168–74
Spearman correlation and, 167
CMR. See Comprehensive Microbial Resource
colour swaps, 217–21
complete linkage, 166
composite sequence information, 254
Comprehensive Microbial Resource (CMR), 38
computer design, of probes, 43–61
confidence levels, 222–23
confounded variables, 106
bias and, 214
blocking and, 213
colour swaps and, 217–18
time series and, 220
consensus trees, 171–74
correlation coefficients, defined, 142–44, 167
coupling efficiencies, 7
covalent attachment, 4
cross-hybridisation, 44, 45, 47–51
cross-validation, 197, 198
data analysis, 110–210
cleaning data, 73–79
methods for. See specific methods software for. See specific programs standards for, 231
storage methods, 231
three types of data, 235
transformation of data, 73–79
databases. See specific systems DDBJ. See DNA Data Bank of Japan degrees of freedom, 52, 115, 117
dendograms, 161–65, 167
deprotection, methods for, 6
diagnostic tools, 199
differential expression assumption, 80
differently expressed genes, identification of, 110–38
dimensionality reduction, 151–58
classification and, 198–208
methods for, 199
directed acryclic graph, 248
distance measures, 141, 142, 167
distribution normalisation, 96–97
DNA Data Bank of Japan (DDBJ), 23–27
duplexes, 45, 52–56
EBI. See European Bioinformatics Institute EMBL. See European Molecular Biology Laboratory
Ensembl project, 37–38
entrophy, 52
EST. See Expressed Sequence Tag resources Euclidean distance, 165
classification and, 167
disadvantages of, 152
scale invariance and, 148–51
 Spearman correlation and, 151
European Bioinformatics Institute (EBI), 37
European Molecular Biology Laboratory (EMBL), 23–27
experiment design, 254
Expressed Sequence Tag (EST) resources, 23
false negatives, 222
false positives, 131–32, 222
Fasta algorithm, 48
feature extraction, 62
image processing, 62–72
irregular features, 68
software for, 233–35
filtering, low-complexity and, 45–47
fixed circle segmentation, 66–67
flagged features, 71–74
fold-ratio of data, 76, 197, 222, 226
freedom, degrees of, 52, 115, 117
full-factorial design, 217
GeneBank database, 23–35, 38
gene expression data tables, 235
Gene Indices, TIGR, 33–35
Gene Ontology (GO) Consortium, 245–48
general linear models, 136
genetic algorithms, 199, 206–8
Genomic Database Resources, 37–41
GO. See Gene Ontology Consortium
Golden Path sequence, 37
graph structures, 248
graphical data analysis, 235
grids, 64–66
hardness, and complexity, 199
hepatocellular carcinomas, 214–16
hierarchical clustering, 158–68
reliability and, 168–74
histogram segmentation, 67, 79
homology search algorithms, 44, 48
hotel, of plates, 4
hybridisation, 11–13, 53
defined, 255
pixel analysis, 62
termology for, 3n2
variability in, 106
hypothesis generation, 199
hypothesis testing, 113–14
IMAGE. See Integrated Molecular Analysis of Gene Expression
image processing, 62–72, 235
in-situ synthesis, 6, 62
independence, statistical, 114–15
individual gene selection, 199, 202
individual number, defined, 255
inkjet methods, 7
INDEX

Institute for Genetics Research (TIGR), 33–35
INDEX

oligonucleotides

cDNAs and, 4
coupling efficiencies and, 7
in-situ synthesized, 5–6
microarrays and, 5
probe design, 43–59
specificity of, 43, 48
spotting and, 3
one-sample t-test. See t-test
one-way ANOVA, 134–36
ontologies, 239, 244–51
outliers, and variability measures, 123
p-values, 113–14
false positives and, 132
hypothesis testing and, 114
multiplicity of, 131–32
paired t-tests, 115, 116
pairwise gene selection, 199, 204–5
palindromic sequences, 56, 57
parametric bootstrapping, 169–71
PCA. See principal component analysis
Pearson correlation, 163, 164
linkage methods, 163
negative correlations and, 167
Spearman correlation and, 147
strengths and weaknesses of, 152
photodeprotection, 6
photolithography, 6
photomultiplier tubes (PMT), 14
pilot studies, 101–2
pin-spotted arrays, 62–72
pixels, 15–16, 62, 68–69
PMT. See photomultiplier tubes
polymerase chain reaction (PCR), 3, 183
power analysis, 101–2, 124, 222–28
principal component analysis (PCA), 152–55
advantages and disadvantages of, 199
classification and, 200–4
multidimensional scaling and, 158
public sequence databases, 23–27
random effects, 135
randomisation, 231–14
redundancy, 25
reference sequence project, 34–37
reliability, 168–74, 244
repetitive sequences, 44, 45–47
replication, 25, 101, 222–24
biological, 101, 113, 223
technical, 101, 223
reporter sequence information, 256
robustness, 168–74, 244
Rosetta, 6, 8
S. cerevisiae Genome Database (SGD), 41
S. pombe Genome Database (SGD), 42
S. pombe. See Saccharomyces Genome Database
SAM. See significance analysis of microarrays
samples, 10–11, 113n4
scale invariance, 148–51
scaling, 96
scanners, 14, 18
seed pixels, 68
segmentation methods, 66–68
self-hybridisation, 45, 56
self-organised maps, 177–80
self-self-hybridisation, 106
separability, 184–87
sequence databases, 23–27, 233
SGD. See Saccharomyces Genome Database
sign-rank test, 124
significance analysis of microarrays (SAM), 128n7, 130
similarity measures, 140–51
single colour systems, 220
single linkage, 166
Smith-Waterman alignments, 48
software standards, 232–39
spatial bias, 87, 90
Spearman correlation, 145–48, 163, 165
clustering and, 167
Euclidean distance and, 151
negative correlations and, 167
Pearson correlation and, 147
strengths and weaknesses of, 152
splice variants, 21–23, 33
spotting, 3–5, 9, 64–66
standard deviation, 96, 103, 105, 106, 107, 115, 117, 224–25, 226
standards, for software, 233–39
statistical inference, 112–13
stem-loop structure, 56
string-matching algorithms, 57n3
supervised learning, 184
support vector machines (SVMs), 188, 193–95
SVMs. See support vector machines
tagged image file format (TIFF), 15
taxonomic ontologies, 245
thermodynamics, 44, 45, 52–56
calculations in, 54–55, 57n3
of nucleic acid duplexes, 44, 45, 52–56
of palindromes, 57
INDEX

TIFF files, 15–16, 235
TIGR. See Institute for Genetics Research
time-course experiments, 219–21
time series correlations, 145
tissue-specific microarrays, 29
tissues, classification of, 183–210
training sets, 195, 197
triangle inequality, 141
t-test, 115–23, 127, 202, 217
two-colour array system, 214
two-sample t-tests. See unpaired t-tests
type I errors, 222
type II errors, 222
UniGene database, 28–33
unpaired t-tests, 117–23
up-or-down-regulated genes, 110
validation, classification and, 195–98
variability
of arrays, 100–9
coefficient of, 102, 103, 104, 107–8, 216
distributions of, 105
dot plots for, 104
estimation of, 216–17
experimental conditions and, 93
hybridisations and, 106
between individuals, 108
log-normal model, 102
measuring, 100–8
normal distribution, 105
outliers and, 123
quantifying, 100–8
sources of, 101
standard deviation and, 102–3
variable circle segmentation, 67
variance-covariance matrix, 153
visualisation of data, 235
voting algorithms, 206–7
washing, of slides, 13
Watson-Crick formation, 1
weightings, 193
Wellcome Trust Sanger Institute, 37
Wilcoxon test, 124
within-array normalisation, 80–90
World Wide Web, 58
yeast, 40–41, 139, 154–55, 156–57, 219–21, 245