Index

Numbers and Symbols

10X Genomics Chromium, 22

A

Actin, 8
Affymetrix, 72
Agent-based models
  cells and interactions, focus on, 10
  centroid models, 12
  continuous models, relationship between, 11
  future outlook for, 13
  lattice-based, 6
  lattice-free, 6
  systems biology approach, 11
Alzheimer's disease, 22, 46
Amino acids, 56, 57
Artificial intelligence, 38, 55, 59
Artificial neural networks (ANN), 38, 61–62
Autoencoder (AE), 41, 181

B

BASIC, 25
Basic Local Alignment Search Tool, 58–59
Bayes chains, 59
BaySeq, 87, 90, 91, 92
Biodiversity, 56
Biological fluids, analysis of, 124–125
Biological sequence analysis
  artificial neural networks (ANN) (*see* artificial neural networks (ANN))
  Basic Local Alignment Search Tool, 58–59
  heuristic local alignment, 58–59
  HHblits (*see* HHblits)
  next-generation sequencing (NGS) (*see* next-generation sequencing (NGS))
  overview, 55, 56
  pairwise alignment, 57
Biomarkers
  Alzheimer's disease, 46
  analyzing, 74, 77
  applications, 87
  cancer treatment, 45–46
  collecting, 73, 76
  data collection and analysis, 75, 76, 77
  detecting, 78, 145
  discoveries, 20, 41, 77, 121–122, 124
  exosomes as source of, 124
  identifying, 27, 87, 120–121
  measuring, 76
  meta-analysis, 24
  methylation studies, 108
  omics domains, interactions between-, 81
  prognostics, 27, 102
  redundancy analysis of, 79
  sepsis, for, 27
  subsets, 74, 75, 77
Bioprinting
  overview, 10–11
  quantitative imaging data, 12
  spatial mathematical modeling, 12
Bismark, 104, 105
Bulk RNA sequencing, 85, 86–87, 91, 93, 95, 96. *See also specific sequencing methods*

C

Cadherins, 8
Canonical correlation analysis (CCA), 71, 74–75, 76–77
Capillary electrophoresis (CE), 146
Cell biology, 8, 9
Cell lines, 122
CellSys, 6
Cellular Potts Model, 6, 12, 13
CEMiTool, 28
Central dogma, molecular biology, 78
Chaste, 6
Cheminformatics, 143
ChemSpider, 144
Chromatin regulation
  accessibility, assessing, 113
  chromatin immunoprecipitation (CnIP), 110–113
  function of, 109–110
  post-translational modifications (PTMs), 109–110
Chronic obstructive pulmonary disease (COPD), 26
Cistromics, 20
Clusters, cells, 9
Cologuard, 108
Complementary DNA (cDNA), 86, 171
CompuCell3D, 6
Computed tomography, 47
Continuous models, 6, 12
Convolutional neural networks (CNNs), 39, 41, 43, 44–46, 46–47, 59–60, 179
Cramer-von Mises test, 94
Crick, Francis, 56
Cryo-electron tomography (Cryo-ET), 175
  applications, 181, 183
  data analysis, 176
  deep learning, relationship between, 177, 178
  domain adaptation, 179
  model compression, 178–179
  overview, 176
  simultaneous classification and segmentation, 179–180
  structural patterns, 176
  subtomogram classification, 177
  template-free structural pattern mining, 180–181
Cuffdiff, 89, 92
Cuffdiff2, 89, 92
Cytoscape, 137, 155
Cytosine methylation, 105–106. *See also* methylation, DNA

D

Data mining, 143, 150
Data normalisation method, 163
Data summarization, 161
Decision trees, 153
Deep belief networks (DBNs), 40–41
Deep forward neural networks (DFFs), 39
Deep learning (DL) methods, 35, 46, 62, 177, 178
  algorithms, 36, 41
  complementary DNA (cDNA) (*see* complementary DNA (cDNA))
  epigenomics, 44
  genomics and sequence analysis, use in, 41, 43
  metabolomics, 44
  methods, 36—37
  omics data analysis, 41, 43
  overview, 36
  proteomics, 44–45

template-based structural mining, 176–177
template-free structural pattern mining, 180–181
transcriptomics, 43–44
DEREPLICATOR+, 150
DESeq, 90, 91, 92
DESeq2, 90, 92, 113
DEsingle, 93
Dictionary of Natural Products (DNP), 144
Diffbind, 113
Differential analysis, RNA, 87, 91–93, 94, 95–96
Discriminative learning, 92, 95
Disease mapping, 136. *See also* precision medicine
Diseases, human, biological approaches to
biomarkers (*see* biomarkers)
cell changes, 26
cell changesgenetic variability, 27
complexities of, 20
overview, 19, 20
DNA (deoxyribonucleic acid), 56
amplifying, 103
analysis of fragments, 62
binding proteins, relationship between, 102
chromatin regulation (*see* chromatin regulation)
decoding, 56
DNA methylation (*see* methylation, DNA)
exomes, 21
genomes, 21
hidden Markov methods for analysis, 60
modifications, 44, 102–103
next-generation sequencing (*see* next-generation sequencing (NGS))
NGS analysis of, 62, 65 (*see also* next-generation sequencing (NGS))
overview, 56
sequencing techniques, 41, 44, 45, 56, 58, 60, 101 (*see also* next-generation sequencing (NGS))
TF binding, 41
Drop-Seq, 22
Dynamic programming, 55, 57, 58

E

E-cadherin, 8
EBSeq, 90, 91, 92
EdgeR, 89–90, 91, 92
EMDomics, 92, 94
Ensemble algorithms, 153–154
Entropy
application, 162
concept of, 162
gene sets, selecting, 171–173
history of concept, 173
Shannon's entropy, 162, 163–171
Epigenetics, 22, 66, 102
Epigenome-wide association studies (EWAS), 107, 114
Epigenomics, 20, 44, 72, 102
Exosomes, 124
Expression quantitative trait loci (eQTLs), 27

F

FAIR guideline, 30
False discovery rate (FDR), 22–24
Focal adhesion kinase (FAK), 8

G

Gas chromatography (GC), 146
Gene expression, 102, 171
Gene ontology, 135, 137
Genetic regions (loci), 43

Genome Analysis Toolkit (GATK), 66
Genome-wide association
studies (GWAS), 43, 74, 107
Genomics
applications, 43
approaches, 2010s, 78
comparative, 59
data integration, 28
data repositories for, 22
data validation, 27
defining, 20
genome-wide association studies, 43
metabolomics, integration with, 154–155
overview, 20
research, 120
sequence analysis, 41–42
workflows, 137
GitHub, 30

H

Herpes zoster vaccine, 27
HHblits, 55, 59
Hidden Markov models, 55, 59, 60
Hierarchical clustering (HCA), 152, 171
High performance liquid chromatography
(HPLC), 146
High-throughput sequencing
methods, 72–74
Highly variable genes (HVG), 24
Human Metabolome
Database (HMDB), 149

I

Ilastik, 8
Illumina, 64, 65, 72
Imaging, medical, 46–47. *See also specific imaging types*
In vitro models, three-dimensional, 2
Indels, 65–66
InDrop, 22
Innate lymphoid cells (ILCs), 25
Integrative analysis, 28, 30
Integrative biology, 26–27, 30
Integrins, 8
Interactomes, 136
InterPro, 134
IsoScale, 132

K

KEGG pathway, 125, 154
Kernel method, 153
Kolmogorov– Smirnov test, 94

L

LC-MS/MS data, 137, 149–150
Light sheet-based fluorescence
microscopy (LSFM), 4, 10
Linnorm, 92, 95
Long-/short-term memory
(LTSM), 40, 43, 59, 62

M

Machine learning (ML), 38, 43, 45, 46,
109, 150–152, 153–154. *See also*
deep learning (DL) methods
biological data, 59–60
metabolomics, use in, 143
methylome-wide association studies
(MWAS), use in, 109
popularity of, 151
Madison Metabolomics Consortium
Database (MMCD), 149
Magnetic resonance imaging (MRI), 46
MarinLit, 144

Markov chains, 59
Markov clustering, 59–60
Mascot, 132
Mass spectronomy (MS), 44, 120, 121, 122, 132, 133–134
  defining, 146
  metabolomics, use in, 144, 146, 148, 149, 155
MAST, 93
MaxQuant platform, 125
MestReNova (Mnova) software, 150
MetaboAnalyst, 24, 155
MetaboAnalystR, 24
MetaboLights, 149
Metabolomics, 20, 22, 24, 30, 44, 72, 136
  analytical platforms, 155
  annotating and interpreting data, 154–155
  bottom-up, 28
  data acquisition and pre-processing, 146, 148–150, 153
  data analysis, 152
  data pre-processing, 148–149
  datasets, 150
  defining, 144
  DL methods, 45
  evolution of, 148, 155
  integration, 26, 154–155
  metabolite identification, 149–150
  nuclear magnetic resonance analytical platform, 148
  overview, 143, 144
  proteomics, relationship between, 44–45
  protocols, need for, 155
  random forests (RFs), use of (*see* random forests (RFs))
  software used in, 144–146
  unbiased, 26
Methylation, DNA, 102–103
  "partially methylated domains" (PMDs), 106
  analysis of data, 104–105
  arrays, 103
  assessments of cytosine modifications, 103–104
  computing, 106–107
  defining, 104
  disease research and risk assessment applications, 107–108
  fully methylated regions (FMRs), 105
  low-methylated regions (LMRs), 106
  patterns, 105–106
  unmethylated regions (UMRs), 106
Methylome-wide association studies (MWAS), 107, 108
Michaelis–Menten modeling of dropouts (M3Drop), 24
MicroRNA (mRNA), 26
  Microscopy, 4, 12, 176. *See also* light sheet-based fluorescence microscopy (LSFM)
Microtubules, 8
MiRNAomics, 20
MixOmics, 28–29
MMDiff, 113
Monocle2, 92, 94–95
Monte Carlo method, 12
MS-Deconv, 131
Multi-omics, 19
Multicellular organisms, characteristics of, 1
Multitask learning neural network model, 179–180
MxXCR, 25
MZmine2, 148, 149

N

N-cadherin, 8
Neeleman-Wunsch algorithm, 57
Next-generation sequencing (NGS), 21, 30, 55, 62–63, 65, 66, 144
  efficiency, 65
  evolution of, 62–63
  metabolomics, use in, 144
  steps in, 63–66
NMR spectroscopy, 148
NOIseq, 91, 92
Non-negative matrix factorization (NMF) method, 43
Non-small cell lung cancer (NSCLC), 10, 11
Nuclear magnetic resonance (NMR), 44, 45, 144, 148–149, 150, 152, 153, 155
Nucleic acids, 55, 56

O

Omics techniques, 72
  biomarkers, interactions between, 81
  future of, 81
  high-throughput techniques, 120–121
  history of, 73–74
  multivariate approaches, 74–75, 75–77, 78–79
  overview, 35
  univariate approaches, 74
Organ samples, 124
Orthogonal partial least squares discriminant analysis (OPLS-DA), 152

P

Partial least squares discriminant analysis (PLS-DA), 152
Partial least squares regression, 28, 152
Penalized multi-block partial least squares regression, 77
Penalized partial least squares regression, 75–76
Pfam, 134
PGMRA, 43
Phase field theory, 12
Precision medicine, 26, 45. *See also* biomarkers
  disease mapping, 136
  imaging, 46–47
Principal component analysis (PCA), 125, 152
PRINTS, 134
ProDom, 134
PROSITE, 134
Proteins, 55. *See also* chromatin regulation
  biological fluids, 124
  composition of, 56
  exosomes, 124
  expression, 120
  post-translational modifications (PTM), 120
Proteogenomics, 19, 27
Proteome Discoverer, 125, 131
ProteomeXchange Consortium, 125
Proteomics, 20, 44–45
  analysis, 120–121, 125
  applications, 134
  bottom-up data analysis, 125
  data acquisition, 125
  data analysis, 138
  data integration, 136
  data interpretation, 120–121
  data processing software, 132
  frameworks, integrated, 136–137
  gene ontology, 135, 137
  genomics, relationship between, 176

input, 30
integration with metabolomics, 26, 132–134, 136–137
integrative analysis, 28
interactions, 78
LC-MS/MS shotgun proteomics workflow, 137
mass spectronomy, relationship between, 176
metabolemics, relationship between, 44–45, 144
middle-down data analysis, 125, 132, 133
NMR technology, 45
organ samples, 124
overview, 44, 119, 120–121
pathway analysis, 135–136
statistical analysis, 134
tissue culture, 122
top-down, 131
top-down data analysis, 125, 131
visual proteomics, 176, 177, 180
workflow, 137
ProteoWizardmsConvert tool, 148

Q

QRT-PCR, 92
QuasR, 104, 105

R

Random forests (RFs), 153, 154
Recurrent neural networks (RNNs), 39, 40
Redundancy analysis (RDA), 71, 78–79
Regression algorithms, 109
Restricted Boltzmann machine (RBM), 40
RNA (ribonucleic acid), 56
bulk RNA sequencing (*see* bulk RNA sequencing)
differential analysis, RNA (*see* differential analysis, RNA)
sequencing data analysis, 85
single-cell RNA sequencing (*see* single-cell RNA sequencing)
RNA sequencing at single-cell level (scRNA-seq), 22, 24, 25

S

SAMseq, 90–91, 92
SAMtools, 66
Sanger sequencing methodology, 63
ScDD, 93
Selective or multiple reaction monitoring (SRM/MRM), 121
Sequest, 132
SigEMD, 94
SINCERA, 92, 94
Single-cell differential expression (SCDE), 93
Single-cell RNA sequencing, 85. *See also specific sequencing methods*
applications, 86–87
costs, 87
statistical methods for, 92–93, 94–95, 95–96
Single-layer high throughput data, 19, 21–22, 22–23
Single-nucleotide polymorphisms (SNPs), 26, 27, 65, 106
Single-nucleotide variants (SNV), 65, 66
Skyline software, 132
Sleuth, 91
Smith-Waterman algorithm, 57, 59
Sparse autoencoder (SAE), 41
Spheroids
applications, 9–10
Cellular Potts Model (*see* Cellular Potts Model)

compressive stress on, 9–10
experimental approaches to, 3
formation, 3, 8, 9
fusion, 3, 11, 12, 13
imaging techniques, 3–4
in-vitro representation, 1
mechanically perturbed, 11
modeling, 2, 6, 9, 10–11
quantitative imaging data, 10
radiation sensitivity, 10, 11
time-lapse images, 8
tumors, 11
Spike-in methods (51), 34
Stacked sparse autoencoder (SSAE), 41, 43
STRING database, 125
Subtomogram classification, 177
Support vector machines (SVMs), 153
Systems biology
agent-based models, 11
evolution of, 13
image-based, 1, 2, 3
modeling of data, 162
overview, 76, 162
spheroids, 8, 13
workflows, 1

T

Taylor's theorem, 91
Template-based structural mining, 176–177
Template-free structural pattern mining, 180–181
The Birmingham Metabolite Library Nuclear Magnetic Resonance database (BML-NMR), 148
The Cancer Genome Atlas (TCGA), 45
Tissue cultures, 122–123
Transcription factors (TFs), 102
binding sites, 113, 114
genome codes, 113
Transcriptomics, 20, 22, 26, 27, 28, 43–44, 72, 78, 137, 144
Trim Galore, 104
TRUST, 25
Type 2 diabetes mellitus, 20

U

Ultra performance liquid chromatography, 153
Ultra-high-pressure liquid chromatography (UHPLC), 148
Ultrasound imaging (US), 46

V

Vemurafenib, 146
Voom, 91, 92

W

Waddington, Conrad, 102
Watson, James, 56
Whole-genome bisulfite sequencing (WGBS), 103, 104, 105, 107

X

X-rays, 46
Xtract, 131

Y

YADA, 131

Z

ZINB model, 93

Doi: http://dx.doi.org/10.15586/computationalbiology.2019.ind